# Combinatorics

an upper-level introductory course
in enumeration, graph theory, and design theory

by

Joy Morris

*University of Lethbridge*

Version 2.1.1 of March 2023

Please send comments and corrections to:
Joy.Morris@uleth.ca

# Contents

# Part III. Design Theory

Chapter 1

# What is Combinatorics?

Combinatorics is a subfield of "discrete mathematics," so we should begin by asking what discrete mathematics means. The differences are to some extent a matter of opinion, and various mathematicians might classify specific topics differently.

"Discrete" should not be confused with "discreet," which is a much more commonly-used word. They share the same Latin root, "discretio," which has to do with wise discernment or separation. In the mathematical "discrete," the emphasis is on separateness, so "discrete" is the opposite of "continuous." If we are studying objects that can be separated and treated as a (generally countable) collection of units rather than a continuous structure, then this study falls into discrete mathematics.

In calculus, we deal with continuous functions, so calculus is not discrete mathematics. In linear algebra, our matrices often have real entries, so linear algebra also does not fall into discrete mathematics.

Text books on discrete mathematics often include some logic, as discrete mathematics is often used as a gateway course for upper-level math. Elementary number theory and set theory are also sometimes covered. Algorithms are a common topic, as algorithmic techniques tend to work very well on the sorts of structures that we study in discrete mathematics.

In Combinatorics, we focus on combinations and arrangements of discrete structures. There are five major branches of combinatorics that we will touch on in this course: enumeration, graph theory, Ramsey Theory, design theory, and coding theory. (The related topic of cryptography can also be studied in combinatorics, but we will not touch on it in this course.) We will focus on enumeration, graph theory, and design theory, but will briefly introduce the other two topics.

## 1.1. Enumeration

Enumeration is a big fancy word for counting. If you've taken a course in probability and statistics, you've already covered some of the techniques and problems we'll be covering in this course. When a statistician (or other mathematician) is calculating the "probability" of a particular outcome in circumstances where all outcomes are equally likely, what they usually do is enumerate all possible outcomes, and then figure out how many of these include the outcome they are looking for.

**EXAMPLE 1.1.1.** What is the probability of rolling a 3 on a 6-sided die?

**SOLUTION.** To figure this out, a mathematician would count the sides of the die (there are six) and count how many of those sides display a three (one of them). They would conclude that the probability of rolling a 3 on a 6-sided die is 1/6 (one in six). □

That was an example that you could probably figure out without having studied enumeration or probability, but it nonetheless illustrates the basic principle behind many calculations of probability. The object of enumeration is to enable us to count outcomes in much more complicated situations. This sometimes has natural applications to questions of probability, but our focus will be on the counting, not on the probability.

After studying enumeration in this course, you should be able to solve problems such as:

- If you are playing Texas Hold'em poker against a single opponent, and the two cards in your hand are a pair, what is the probability that your opponent has a higher pair?
- How many distinct Shidokus (4-by-4 Sudokus) are there?
- How many different orders of five items can be made from a bakery that makes three kinds of cookies?
- Male honeybees come from a queen bee's unfertilised eggs, so have only one parent (a female). Female honeybees have two parents (one male, one female). Assuming all ancestors were distinct, how many ancestors does a male honeybee have from 10 generations ago?

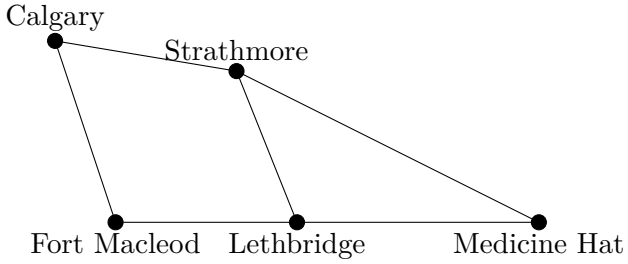
Although all of these questions (and many others that arise naturally) may be of interest to you, the reason we begin our study with enumeration is because we'll be able to use many of the techniques we learn, to count the other structures we'll be studying.

## 1.2. Graph Theory

When a mathematician talks about graph theory, they are not referring to the "graphs" that you learn about in school, that can be produced by a spreadsheet or a graphing calculator. The "graphs" that are studied in graph theory are models of networks.

Any network can be modeled by using dots to represent the nodes of the network (the cities, computers, cell phones, or whatever is being connected) together with lines to represent the connections between those nodes (the roads, wires, wireless connections, etc.). This model is called a graph. It is important to be aware that only at a node may information, traffic, etc. pass from one edge of a graph to another edge. If we want to model a highway network using a graph, and two highways intersect in the middle of nowhere, we must still place a node at that intersection. If we draw a graph in which edges cross over each other but there is no node at that point, you should think of it as if there is an overpass there with no exits from one highway to the other: the roads happen to cross, but they are not connecting in any meaningful sense.

**EXAMPLE 1.2.1.** The following diagram:




is a graph.

Many questions that have important real-world applications can be modeled with graphs. These are not always limited to questions that seem to apply to networks. Some questions can be modeled as graphs by using lines to represent constraints or some other relationship between them (e.g. the nodes might represent classes, with a line between them if they cannot

be scheduled at the same time, or some nodes might represent students and others classes, with a line between a student and each of the classes they are taking).

After studying graph theory in this course, you should be able to solve problems such as:

- How can we find a good route for garbage trucks to take through a particular city?
- Is there a delivery route that visits every city in a particular area, without repetition?
- Given a collection of project topics and a group of students each of whom has expressed interest in some of the topics, is it possible to assign each student a unique topic that interests them?
- A city has bus routes all of which begin and end at the bus terminal, but with different schedules, some of which overlap. What is the least number of buses (and drivers) required in order to be able to complete all of the routes according to the schedule?
- Create a schedule for a round-robin tournament that uses as few time slots as possible.

Some of these questions you may only be able to answer for particular kinds of graphs.

Graph theory is a relatively "young" branch of mathematics. Although some of the problems and ideas that we will study date back a few hundred years, it was not until the 1930s that these individual problems were gathered together and a unified study of the theory of graphs began to develop.

### 1.3. Ramsey Theory

Ramsey theory takes its name from Frank Plumpton Ramsey (1903—1930), a British mathematician who died at the tragically young age of 26.

Ramsey was a logician. A result that he considered a minor lemma in one of his logic papers now bears the name "Ramsey's Theorem" and was the basis for this branch of mathematics. Its statement requires a bit of graph theory: given $c$ colours and fixed sizes $n_1, \ldots, n_c$, there is an integer $r = R(n_1, \ldots, n_c)$ such that for any colouring the edges of of a complete graph on $r$ vertices, there must be some $i$ between 1 and $c$ such that there is a complete subgraph on $n_i$ vertices, all of whose edges are coloured with colour $i$.

In addition to requiring some graph theory, that statement was a bit technical. In much less precise terms that don't require so much background knowledge (but could be misleading in specific situations), Ramsey theory asserts that if a structure is big enough and contains a property we are interested in, then no matter how we cut it into (a limited number of) pieces, at least one of the pieces should also have that property. One major theorem in Ramsey theory is van der Waerden's Theorem, which states that for any two constants $c$ and $n$, there is a constant $V(c, n)$ such that if we take $V(c, n)$ consecutive numbers and colour them with $c$ colours, there must be an arithmetic progression of length $n$ all of whose members have been coloured with the same colour.

**EXAMPLE 1.3.1.** Here is a small example of van der Waerden's Theorem. With two colours and a desired length of 3 for the arithmetic progression, we can show that $V(2, 3) > 8$ using the following colouring:

**3 4** 5 6 **7 8** 9 10

(In case it is difficult to see, we point out that 3, 4, 7, and 8 are black, while 5, 6, 9, and 10 are grey, a different colour.) Notice that with eight consecutive integers, the difference in a three-term arithmetic progression cannot be larger than three. For every three-term arithmetic progression with difference of one, two, or three, it is straightforward to check that the numbers have not all received the same colour.

In fact, $V(2, 3) = 9$, but proving this requires exhaustive testing.

We will touch lightly on Ramsey theory in this course, specifically on Ramsey's Theorem itself, in the context of graph theory.

## 1.4. Design Theory

Like graph theory, design theory is probably not what any non-mathematician might expect from its name.

When researchers conduct an experiment, errors can be introduced by many factors (including, for example, the timing or the subject of the experiment). It is therefore important to replicate the experiment a number of times, to ensure that these unintended variations do not account for the success of a particular treatment. If a number of different treatments are being tested, replicating all of them numerous times becomes costly and potentially infeasible. One way to reduce the total number of trials while still maintaining the accuracy, is to test multiple treatments on each subject, in different combinations.

One of the major early motivations for design theory was this context: given a fixed number of total treatments, and a fixed number of treatments we are willing to give to any subject, can we find combinations of the possible treatments so that each treatment is given to some fixed number of subjects, and any pair of treatments is given together some fixed number of times (often just once). This is the basic structure of a block design.

**EXAMPLE 1.4.1.** Suppose that we have seven different fertilisers and seven garden plots on which to try them. We can organise them so that each fertiliser is applied to three of the plots, each garden plot receives three fertilisers, and any pair of fertilisers is used together on precisely one of the plots. If the different fertilisers are numbered one through seven, then a method for arranging them is to place fertilisers 1, 2, and 3 on the first plot; 1, 4, and 5 on the second; 1, 6, and 7 on the third; 2, 4, and 6 on the fourth; 2, 5, and 7 on the fifth; 3, 4, and 7 on the sixth; and3, 5, and 6 on the last. Thus,

| | | |
|---|---|---|
| 123 | 145 | 167 |
| 246 | 257 | 347 |
| | 356 | |

is a design.

This basic idea can be generalised in many ways, and the study of structures like these is the basis of design theory. In this course, we will learn some facts about when designs exist, and how to construct them.

After studying design theory in this course, you should be able to solve problems such as:

- Is it possible for a design to exist with a particular set of parameters?
- What methods might we use in trying to construct a design?

## 1.5. Coding Theory

In many people's minds "codes" and "cryptography" are inextricably linked, and they might be hard-pressed to tell you the difference. Nonetheless, the two topics are vastly different, as is the mathematics involved in them.

Coding theory is the study of encoding information into different symbols. When someone uses a code in an attempt to make a message that only certain other people can read, this becomes cryptography. Cryptographers study strategies for ensuring that a code is difficult to "break" for those who don't have some additional information. In coding theory, we ignore the question of who has access to the code and how secret it may be. Instead, one of the primary concerns becomes our ability to detect and correct errors in the code.

Codes are used for many purposes in which the information is not intended to be secret. For example, computer programs are transformed into long strings of binary data, that a computer reinterprets as instructions. When you text a photo to a friend, the pixel and colour information are converted into binary data to be transmitted through radio waves. When you listen to an audio file, the sound frequencies of the music have been converted into binary data that your computer decodes back into sound frequencies.

Electronic encoding is always subject to interference, which can cause errors. Even when a coded message is physically etched onto a device (such as a dvd), scratches can make some parts of the code illegible. When you are connected to the internet (for example, if you are streaming), wifi signals can fail, suffer interference from other nearby networks, or become overloaded. Any of these problems can cause your connection to lose or misread part of the data that is being transmitted. People don't like it when their movies, music, or apps freeze, crash, or skip over something. This can become even more problematic if it happens during an important online meeting or class, or while you are making an electronic financial transaction. To avoid this problem, engineers use codes that allow our devices to automatically detect, and correct, minor errors that may be introduced.

In coding theory, we learn how to create codes that allow for error detection and correction, without requiring excessive memory or storage capacity. Although coding theory is not a focus of this course, designs can be used to create good codes. We will learn how to make codes from designs, and what makes these codes "good."

**EXAMPLE 1.5.1.** Suppose we have a string of binary information, and we want the computer to store it so we can detect if an error has arisen. There are two symbols we need to encode: 0 and 1. If we just use 0 as the code for indicating 0 and 1 as the code for 1, we'll never know if a bit has been flipped (from 0 to 1 or vice versa). If we use 00 for 0 and 01 for 1, then if the first bit gets flipped we'll know there was an error (because the first bit should never be 1), but we won't notice if the second was flipped. If we use 00 for 0 and 11 for 1, then we will be able to detect an error, as long as at most one bit gets flipped, because flipping one bit of either code word will result in either 01 or 10, neither of which is a valid code word. Thus, this code allows us to detect an error. It does not allow us to correct an error, because even knowing that a single bit has been flipped, there is no way of knowing whether a 10 arose from a 00 with the first bit flipped, or from a 11 with the second bit flipped. We would need a longer code to be able to correct errors.

After our study of coding theory, you should be able to solve problems such as:

- Given a code, how many errors can be detected?
- Given a code, how many errors can be corrected?
- Construct some small codes that allow detection and correction of small numbers of errors.

**EXERCISE 1.5.2.** Can you come up with an interesting counting problem that you wouldn't know how to solve?

**SUMMARY:**

- enumeration
- graph theory
- Ramsey theory
- design theory
- coding theory

# Part I

# Enumeration

# Chapter 2

# Basic Counting Techniques

When we are trying to count the number of ways in which something can happen, sometimes the answer is very obvious. For example, if a doughnut shop has five different kinds of doughnuts for sale and you are planning to buy one doughnut, then you have five choices.

There are some ways in which the situation can become slightly more complicated. Perhaps you haven't decided whether you'll buy a doughnut or a bagel, and the store also sells three kinds of bagels. Or perhaps you want a cup of coffee to go with your doughnut, and there are four different kinds of coffee, each of which comes in three different sizes.

These particular examples are fairly small and straightforward, and you could list all of the possible options if you wished. The goal of this chapter is to use simple examples like these to demonstrate two rules that allow us to count the outcomes not only in these situations, but in much more complicated circumstances. These rules are the *product rule*, and the *sum rule*.

## 2.1. The product rule

The product rule is a rule that applies when we there is more than one variable (i.e. thing that can change) involved in determining the final outcome.

**EXAMPLE 2.1.1.** Consider the example of buying coffee at a coffee shop that sells four varieties and three sizes. When you are choosing your coffee, you need to choose both variety and size. One way of figuring out how many choices you have in total, would be to make a table. You could label the columns with the sizes, and the rows with the varieties (or vice versa, it doesn't matter).

| | **Small** | **Medium** | **Large** |
|---|---|---|---|
| **Latte** | small latte | medium latte | large latte |
| **Mocha** | small mocha | medium mocha | large mocha |
| **Espresso** | small espresso | medium espresso | large espresso |
| **Cappuccino** | small cappuccino | medium cappuccino | large cappuccino |

As you can see, a different combination of variety and size appears in each entry of the table, and every possible combination of variety and size appears somewhere. Thus the total number of possible choices is the number of entries in this table. Although in a small example like this we could simply count all of the entries and see that there are twelve, it will be more useful to notice that elementary arithmetic tells us that the number of entries in the table will be the number of rows times the number of columns, which is four times three.

In other words, to determine the total number of choices you have, we multiply the number of choices of variety (that is, the number of rows in our table) by the number of choices of size (that is, the number of columns in our table). This is an example of what we'll call the *product rule*.

We're now ready to state the product rule in its full generality.

**THEOREM 2.1.2 (Product Rule).** *Suppose that when you are determining the total number of outcomes, you can identify two* different *aspects that can vary. If there are $n_1$ possible outcomes for the first aspect, and for each of those possible outcomes, there are $n_2$ possible outcomes for the second aspect, then the total number of possible outcomes will be $n_1 n_2$.*

In the above example, we can think of the aspects that can change as being the variety of coffee, and the size. There are four outcomes (choices) for the first aspect, and three outcomes (choices) for the second aspect, so the total number of possible outcomes is $4 \cdot 3 = 12$.

Sometimes it seems clear that there are more than two aspects that are varying. If this happens, we can apply the product rule more than once to determine the answer, by first identifying two aspects (one of which may be "all the rest"), and then subdividing one or both of those aspects. An example of this is the problem posed earlier of buying a doughnut to go with your coffee.

**EXAMPLE 2.1.3.** Kyle wants to buy coffee and a doughnut. The local doughnut shop has five kinds of doughnuts for sale, and sells four varieties of coffee in three sizes (as in Example 2.1.1). How many different orders could Kyle make?

**SOLUTION.** A natural way to divide Kyle's options into two aspects that can vary, is to consider separately their choice of doughnut, and their choice of coffee. There are five choices for the kind of doughnut they order, so $n_1 = 5$. For choosing the coffee, we have already used the product rule in Example 2.1.1 to determine that the number of coffee options is $n_2 = 12$.

Thus the total number of different orders Kyle could make is $n_1 n_2 = 5 \cdot 12 = 60$. □

Let's go through an example that more clearly involves repeated applications of the product rule.

**EXAMPLE 2.1.4.** Chloë wants to manufacture children's t-shirts. There are generally three sizes of t-shirts for children: small, medium, and large. She wants to offer the t-shirts in eight different colours (blue, yellow, pink, green, purple, orange, white, and black). The shirts can have an image on the front, and a slogan on the back. She has come up with three images, and five slogans.

To stock her show room, Chloë wants to produce a single sample of each kind of shirt that she will be offering for sale. The shirts cost her $4 each to produce. How much are the samples going to cost her (in total)?

**SOLUTION.** To solve this problem, observe that to determine how many sample shirts Chloë will produce, we can consider the size as one aspect, and the style (including colour, image, and slogan) as the other. There are $n_1 = 3$ sizes. So the number of samples will be three times $n_2$, where $n_2$ is the number of possible styles.

We now break $n_2$ down further: to determine how many possible styles are available, you can divide this into two aspects: the colour, and the decoration (image and slogan). There are $n_{2,1} = 8$ colours. So the number of styles will be eight times $n_{2,2}$, where $n_{2,2}$ is the number of possible decorations (combinations of image and slogan) that are available.

We can break $n_{2,2}$ down further: to determine how many possible decorations are available, you divide this into the two aspects of image and slogan. There are $n_{2,2,1} = 3$ possible images,

and $n_{2,2,2} = 5$ possible slogans, so the product rule tells us that there are $n_{2,2} = 3 \cdot 5 = 15$ possible combinations of image and slogan (decorations).

Putting all of this together, we see that Chloë will have to create $3(8(3 \cdot 5)) = 360$ sample t-shirts. As each one costs \$4, her total cost will be \$1440. □

Notice that finding exactly two aspects that vary can be quite artificial. Example 2.1.4 serves as a good demonstration for a generalisation of the product rule as we stated it above. In that example, it would have been more natural to have considered from the start that there were four apparent aspects to the t-shirts that can vary: size; colour; image; and slogan. The total number of t-shirts she needed to produce was the product of the number of possible outcomes of each of these aspects: $3 \cdot 8 \cdot 3 \cdot 5 = 360$.

**THEOREM 2.1.5 (Product Rule for many aspects).** *Suppose that when you are determining the total number of outcomes, you can identify $k$* different *aspects that can vary. If for each $i$ between 1 and $k$ there are $n_i$ possible outcomes for the $i$th aspect, then the total number of possible outcomes will be $\prod_{i=1}^{k} n_i$ (that is, the product as $i$ goes from 1 to $k$ of the $n_i$).*

Now let's look at an example where we are trying to evaluate a probability. Since this course is about counting rather than probability, we'll restrict our attention to examples where all outcomes are equally likely. Under this assumption, in order to determine a probability, we can count the number of outcomes that have the property we're looking for, and divide by the total number of outcomes.

**EXAMPLE 2.1.6.** Iswar has flipped a coin twice, and come up with heads both times. What is the probability that he will flip heads each of his next two tries?

**SOLUTION.** To answer this, we consider each coin flip as a different aspect. There are two possible results for the third flip: heads or tails. For each of these, there are two possible results for the fourth flip: heads or tails. So in total, the product rule tells us that there are $2 \cdot 2 = 4$ possible combinations for the results of the third and fourth flips. This will be the denominator of the probability.

To determine the numerator (that is, the number of ways in which both flips can result in heads), we again consider each flip as a different aspect. There is only one possible way for the third flip to be heads, and then there is only one possible way for the fourth flip to be heads. So in total, only one of the four possible combinations of outcomes involves both coins landing as heads.

The probability that Iswar's next two flips will also result in heads is 1/4. □

Notice that in this example, the fact that Iswar's first two flips were heads was irrelevant to our calculations, because it was already a known outcome, over and done with, so is true no matter what may happen with his later flips. If Iswar hadn't yet flipped the coin and we asked for the probability that his first four flips will all be heads, then our calculations would have to include both possible options for the outcome of each of his first two flips. In this case, the final probability would be 1/16 (there are 16 possible combinations for the outcomes of four coin flips, only one of which involves all four being heads).

**EXERCISES 2.1.7.** Use only the product rule to answer the following questions:

1) The car Ace wants to buy comes in four colours; with or without air conditioning; with five different options for stereo systems; and a choice of none, two, or four floor mats. If the dealership they visit has three cars in the lot, each with different options, what is the probability that one of the cars the dealership has in stock has exactly the options Ace wants?

2) Candyce is writing a "Choose your own adventure" book in which she wants every possible choice to result in a different ending. If there are four points at which choices must be made in every storyline, and there are three choices the first time but two every time after that, how many endings does Candyce need to write?
3) William is buying five books. For each book he has a choice of version: hardcover, paperback, or electronic. In how many different ways can he make his selection?

## 2.2. The sum rule

The sum rule is a rule that can be applied to determine the number of possible outcomes when there are two different things that you might choose to do (and various ways in which you can do each of them), and you cannot do both of them. Often, it is applied when there is a natural way of breaking the outcomes down into cases.

**EXAMPLE 2.2.1.** Recall the example of buying a bagel *or* a doughnut at a doughnut shop that sells five kinds of doughnuts and three kinds of bagels. You are only choosing one or the other, so one way to determine how many choices you have in total, would be to write down all of the possible kinds of doughnut in one list, and all of the possible kinds of bagel in another list:

| **Doughnuts** | **Bagels** |
|---|---|
| chocolate glazed | blueberry |
| chocolate iced | cinnamon raisin |
| honey cruller | plain |
| custard filled | |
| original glazed | |

The total number of possible choices is the number of entries that appear in the two lists combined, which is five plus three.

In other words, to determine the number of choices you have, we add the number of choices of doughnut (that is, the number of entries in the first list) and the number of choices of bagel (that is, the number of entries in the second list). This is an example of the *sum rule*.

We're now ready to state the sum rule in its full generality.

**THEOREM 2.2.2 (Sum Rule).** *Suppose that when you are determining the total number of outcomes, you can identify two distinct cases with the property that every possible outcome lies in exactly one of the cases. If there are $n_1$ possible outcomes in the first case, and $n_2$ possible outcomes in the second case, then the total number of possible outcomes will be $n_1 + n_2$.*

It's hard to do much with the sum rule by itself, but we'll cover a couple of additional examples and then in the next section, we'll get into some more challenging examples where we combine the two rules.

Sometimes the problem naturally splits into more than two cases, with every possible outcome lying in exactly one of the cases. If this happens, we can apply the sum rule more than once to determine the answer. First we identify two cases (one of which may be "everything else"), and then subdivide one or both of the cases. Let's look at an example of this.

**EXAMPLE 2.2.3.** Iswar is planning to flip his coin up to three times. What are the possible combinations of heads and tails he might end up with, if we aren't keeping track of the order of the outcomes? (By not keeping track of the order of the outcomes, I mean that we'll consider flipping two heads followed by one tails as being the same as flipping two heads and one tails in any other order.)

**SOLUTION.** To answer this question, we'll break the problem into cases. First we'll divide the problem into two possibilities: Iswar flips the coin no times; or he flips it at least once. If Iswar does not flip the coin, there is only one possible outcome (no heads and no tails). If Iswar flips the coin at least once, then he flips it between one and three times. We'll have to break this down further to find how many outcomes are involved.

We break the case where Iswar flips the coin between one and three times down into two cases: he might flip it exactly once, or he might flip it more than once. If he flips it exactly once, the outcome of that might be heads or tails, so there are two possible outcomes. If he flips it more than once, again we'll need to further subdivide this case.

The case where Iswar flips the coin either two or three times naturally breaks down into two cases: he might flip it twice, or he might flip it three times. If he flips it twice, the number of heads he gets might be zero, one, or two, so there are three possible outcomes (the remaining results, if any, must all be tails). If he flips it three times, the number of heads he gets might be zero, one, two, or three, so there are four possible outcomes (again, any remaining results must be tails).

Now we put all of these outcomes together with the sum rule. We conclude that in total, there are $1 + (2 + (3 + 4)) = 10$ different combinations of heads and tails that Iswar might end up with. □

Notice that it was artificial to repeatedly break this example up into two cases at a time. Thus, Example 2.2.3 serves as a good demonstration for a generalisation of the sum rule as we stated it above. It would have been more natural to have broken the problem of Iswar's coin flips up into four cases from the beginning, depending on whether he flips the coin zero, one, two, or three times. The total number of combinations of heads and tails that Iswar might end up with, is the sum of the combinations he can end up with in each of these cases; that is, $1 + 2 + 3 + 4 = 10$.

**THEOREM 2.2.4 (Sum Rule for many cases).** *Suppose that when you are determining the total number of outcomes, you can identify $k$ distinct cases with the property that every possible outcome lies in exactly one of the cases. If for each $i$ between 1 and $k$ there are $n_i$ possible outcomes in the $i$th case, then the total number of possible outcomes will be $\sum_{i=1}^{k} n_i$ (that is, the sum as $i$ goes from 1 to $k$ of the $n_i$).*

There is one other important way to use the sum rule. This application is a bit more subtle. Suppose you know the total number of outcomes, and you want to know the number of outcomes that *don't* include a particular event. The sum rule tells us that the total number of outcomes is comprised of the outcomes that *do* include that event, together with the ones that don't. So if it's easy to figure out how many outcomes include the event that interests you, then you can subtract that from the total number of outcomes to determine how many outcomes *exclude* that event. Here's an example.

**EXAMPLE 2.2.5.** There are 216 different possible outcomes from rolling a white die, a red die, and a yellow die. (You can work this out using the product rule.) How many of these outcomes involve rolling a one on two or fewer of the dice?

**SOLUTION.** Tackling this problem directly, you might be inclined to split it into three cases: outcomes that involve rolling no ones, those that involve rolling exactly one one, and those that involve rolling exactly two ones. If you try this, the analysis will be long and fairly involved, and will include both the product rule and the sum rule. If you are careful, you will be able to find the correct answer this way.

We'll use a different approach, by first counting the outcomes that we *don't* want: those that involve getting a one on all three of the dice. There is only one way for this to happen: all

three of the dice have to roll ones! So the number of outcomes that involve rolling ones on two or fewer of the dice, will be $216 - 1 = 215$. □

**EXERCISES 2.2.6.** Use only the sum rule to answer the following questions:

1) I have four markers on my desk: one blue and three black. Every day on my way to class, I grab three of the markers without looking. There are four different markers that could be left behind, so there are four combinations of markers that I could take with me. What is the probability that I take the blue marker?
2) Ocean is thinking of either a letter, or a digit. How many different things could they be thinking of?
3) How many of the 16 four-bit binary numbers have at most one 1 in them?

## 2.3. Putting them together

When we combine the product rule and the sum rule, we can explore more challenging questions.

**EXAMPLE 2.3.1.** Grace is staying at a bed-and-breakfast. In the evening, she is offered a choice of menu items for breakfast in bed, to be delivered the next morning. There are three kinds of items: main dishes, side dishes, and beverages. She is allowed to choose up to one of each, but some of them come with optional extras. From the menu below, how many different breakfasts could she order?

***Menu***

| Mains | Sides | Beverages |
|---|---|---|
| pancakes | fruit cup | coffee |
| oatmeal | toast | tea |
| omelette | | orange juice |
| waffles | | |

Pancakes, waffles, and toast come with butter.
Coffee and tea come with milk and sugar.

**Optional extras:**
marmalade, lemon curd, or blackberry jam for toast;
maple syrup for pancakes or waffles.

**SOLUTION.** We see that the number of choices Grace has available depends partly on whether or not she orders an item or items that include optional extras. We will therefore divide our consideration into four cases:

1) Grace does not order any pancakes, waffles, or toast.
2) Grace orders pancakes or waffles, but does not order toast.
3) Grace does not order pancakes or waffles, but does order toast.
4) Grace orders toast, and also orders either pancakes or waffles.

In the first case, Grace has three possible choices for her main dish (oatmeal, omelette, or nothing). For each of these, she has two choices for her side dish (fruit cup, or nothing). For each of these, she has four choices for her beverage (coffee, tea, orange juice, or nothing). Using the product rule, we conclude that Grace could order $3 \cdot 2 \cdot 4 = 24$ different breakfasts that do not include pancakes, waffles, or toast.

In the second case, Grace has two possible choices for her main dish (pancakes, or waffles). For each of these, she has two choices for her side dish (fruit cup, or nothing). For each of these, she has four choices for her beverage. In addition, for each of her choices of pancakes or waffles, she can choose to have maple syrup, or not (two choices). Using the product rule, we conclude that Grace could order $2 \cdot 2 \cdot 4 \cdot 2 = 32$ different breakfasts that include pancakes or waffles, but not toast.

In the third case, Grace has three possible choices for her main dish (oatmeal, omelette, or nothing). For each of these, she has only one possible side dish (toast), but she has four choices for what to put on her toast (marmalade, lemon curd, blackberry jam, or nothing). For each of these choices, she has four choices of beverage. Using the product rule, we conclude that Grace could order $3 \cdot 4 \cdot 4 = 48$ different breakfasts that include toast, but do not include pancakes or waffles.

In the final case, Grace has two possible choices for her main dish (pancakes, or waffles). She has two choices for what to put on her main dish (maple syrup, or only butter). She is having toast, but has four choices for what to put on her toast. Finally, she again has four choices of beverage. Using the product rule, we conclude that Grace could order $2 \cdot 2 \cdot 4 \cdot 4 = 64$ different breakfasts that include toast as well as either pancakes or waffles.

Using the sum rule, we see that the total number of different breakfasts Grace could order is $24 + 32 + 48 + 64 = 168$. □

Here's another example of combining the two rules.

**EXAMPLE 2.3.2.** The types of license plates in Alberta that are available to individuals (not corporations or farms) for their cars or motorcycles, fall into one of the following categories:

- vanity plates;
- regular car plates;
- veteran plates; or
- motorcycle plates.

None of these license plates use the letters I or O.

Regular car plates have one of two formats: three letters followed by three digits; or three letters followed by four digits (in the latter case, none of the letters A, E, I, O, U, or Y is used).

Veteran plates begin with the letter V, followed by two other letters and two digits. Motorcycle plates have two letters followed by three digits.

Setting aside vanity plates and ignoring the fact that some three-letter words are avoided, how many license plates are available to individuals in Alberta for their cars or motorcycles?

**SOLUTION.** To answer this question, there is a natural division into four cases: regular car plates with three digits; regular car plates with four digits; veteran plates; and motorcycle plates.

For a regular car plate with three digits, there are 24 choices for the first letter, followed by 24 choices for the second letter, and 24 choices for the third letter. There are 10 choices for the first digit, 10 choices for the second digit, and 10 choices for the third digit. Using the product rule, the total number of license plates in this category is $24^3 \cdot 10^3 = 13,824,000$.

For a regular car plate with four digits, there are 20 choices for the first letter, followed by 20 choices for the second letter, and 20 choices for the third letter. There are 10 choices for the first digit, 10 choices for the second digit, 10 choices for the third digit, and 10 choices for the fourth digit. Using the product rule, the total number of license plates in this category is $20^3 \cdot 10^4 = 80,000,000$.

For a veteran plate, there are 24 choices for the first letter, followed by 24 choices for the second letter. There are 10 choices for the first digit, and 10 choices for the second digit. Using the product rule, the total number of license plates in this category is $24^2 \cdot 10^2 = 57,600$.

Finally, for a motorcycle plate, there are 24 choices for the first letter, followed by 24 choices for the second letter. There are 10 choices for the first digit, 10 choices for the second digit, and 10 choices for the third digit. Using the product rule, the total number of license plates in this category is $24^2 \cdot 10^3 = 576,000$.

Using the sum rule, we see that the total number of license plates is

$$13,824,000 + 80,000,000 + 57,600 + 576,000$$

which is $94,457,600$. □

It doesn't always happen that the sum rule is applied first to break the problem down into cases, followed by the product rule within each case. In some problems, these might occur in the other order. Sometimes there may seem to be one "obvious" way to look at the problem, but often there is more than one equally effective analysis, and different analyses might begin with different rules.

In Example 2.3.1, we could have begun by noticing that no matter what else she may choose, Grace has four possible options for her beverage. Thus, the total number of possible breakfast orders will be four times the number of possible orders of main and side (with optional extras). Then we could have proceeded to analyse the number of possible choices for her main dish and her side dish (together with the extras). Breaking down the choices for her main and side dishes into the same cases as before, we could see that there are $3 \cdot 2 = 6$ choices in the first case; $2 \cdot 2 \cdot 2 = 8$ choices in the second case; $3 \cdot 4 = 12$ choices in the third case; and $2 \cdot 2 \cdot 4 = 16$ choices in the fourth case. Thus she has a total of $6 + 8 + 12 + 16 = 42$ choices for her main and side dishes. The product rule now tells us that she has $4 \cdot 42 = 168$ possible orders for her breakfast.

Let's run through one more (simpler) example of using both the sum and product rules, and work out the answer in two different ways.

**EXAMPLE 2.3.3.** Ming plans to buy her Dad a shirt for his birthday. The store she goes to has three different colours of short-sleeved shirts, and six different colours of long-sleeved shirts. They will gift-wrap in her choice of two wrapping papers. Assuming that she wants the shirt gift-wrapped, how many different options does she have for her gift?

**SOLUTION.** Let's start by applying the product rule first. There are two aspects that Ming can vary: the shirt, and the wrapping. She has two choices for the wrapping, so her total number of options will be twice the number of shirt choices that she has. For the shirt, we break her choices down into two cases: if she opts for a short-sleeved shirt then she has three choices (of colour), while if she opts for a long-sleeved shirt then she has six choices. In total she has $3 + 6 = 9$ choices for the shirt. Using the product rule, we see that she has $2 \cdot 9 = 18$ options for her gift.

Alternatively, we could apply the sum rule first. We will consider the two cases: that Ming buys a short-sleeved shirt; or a long-sleeved shirt. If she buys a short-sleeved shirt, then she has three options for the shirt, and for each of these she has two options for the wrapping, making (by the product rule) $3 \cdot 2 = 6$ options of short-sleeved shirts. If she buys a long-sleeved shirt, then she has six options for the shirt, and for each of these she has two options for the wrapping, making (by the product rule) $6 \cdot 2 = 12$ options of long-sleeved shirts. Using the sum rule, we see that she has $6 + 12 = 18$ options for her gift. □

**EXERCISES 2.3.4.** How many passwords can be created with the following constraints:

1) The password is three characters long and contains two lowercase letters and one digit, in some order.

2) The password is eight or nine characters long and consists entirely of digits.

3) The password is five characters long and consists of lowercase letters and digits. All of the letters must come before all of the digits in the password, but there can be any number of letters (from zero through five).

4) The password is four characters long and consists of two characters that can be either digits or one of 16 special characters, and two lowercase letters. The two letters can be in any two of the four positions.

5) The password is eight characters long and must include at least one letter and at least one digit.

6) The password is eight characters long and cannot include any character more than once.

**EXERCISES 2.3.5.**

1) There are 8 buses a day from Toronto to Ottawa, 20 from Ottawa to Montreal, and 9 buses directly from Toronto to Montreal. Assuming that you do not have to complete the trip in one day (so the departure and arrival times of the buses is not an issue), how many different schedules could you use in travelling by bus from Toronto to Montreal?

2) How many 7-bit ternary strings (that is, strings whose only entries are 0, 1, or 2) begin with either 1 or 01?

## 2.4. Summing up

Very likely you've used the sum rule or the product rule when counting simple things, without even stopping to think about what you were doing. The reason we're going through each of them very slowly and carefully, is because when we start looking at more complicated problems, our uses of the sum and product rules will become more subtle. If we don't have a very clear understanding in very simple situations of what we are doing and why, we'll be completely lost when we get to more difficult examples.

It's dangerous to try to come up with a simple guideline for when to use the product rule and when to use the sum rule, because such a guideline will often go wrong in complicated situations. Nonetheless, a good question to ask yourself when you are trying to decide which rule to use is, "Would I describe this with the word 'and,' or the word 'or'?" The word "and" is generally used in situations where it's appropriate to use the product rule, while "or" tends to go along with the sum rule.

Let's see how this applies to each of the examples we've looked at in this chapter.

In Example 2.1.1, you needed to choose the size *and* the variety for your coffee. In Example 2.1.3, Kyle wanted to choose a doughnut *and* coffee. In Example 2.1.4, Chloë needed to determine the size *and* the colour *and* the image *and* the slogan for each t-shirt. In Example 2.1.6, we wanted to know the results of Iswar's third *and* fourth coin flips. So in each of these examples, we used the product rule.

In Example 2.2.1, you needed to choose a bagel *or* a doughnut. In Example 2.2.3, Iswar could have flipped the coin zero *or* one *or* two *or* three times. So in each of these examples, we used the sum rule.

You definitely have to be careful in applying this guideline, as problems can be phrased in a misleading way. We could have said that in Example 2.2.1, we want to know how many different kinds of doughnuts *and* of bagels there are, altogether. The important point is that you aren't choosing both of these things, though; you are choosing just one thing, and it will be either a doughnut, *or* a bagel.

In Example 2.3.1, Grace was choosing a main dish *and* a side dish *and* a beverage, so we used the product rule to put these aspects together. Whether or not she had extra options available for her main dish depended on whether she chose pancakes *or* waffles *or* oatmeal *or* omelette *or* nothing, so the sum rule applied here. (Note that we didn't actually consider each of these four things separately, since they naturally fell into two categories. However, we would have reached the same answer if we had considered each of them separately.) Similarly, whether or not she had extra options available for her side dish depended on whether she chose toast *or* not, so again the sum rule applied.

In Example 2.3.2, the plates can be regular (in either of two ways) *or* veteran *or* motorcycle plates, so the sum rule was used. In each of these categories, we had to consider the options for the first character *and* the second character (and so on), so the product rule applied.

Finally, in Example 2.3.3, the shirt Ming chooses can be short-sleeved *or* long-sleeved, so the sum rule applies to that distinction. Since she wants to choose a shirt *and* gift wrap, the product rule applies to that combination.

**EXERCISES 2.4.1.** For each of the following problems, do you need to use the sum rule, the product rule, or both?

1) Count all of the numbers that have exactly two digits, and the numbers that have exactly four digits.
2) How many possible outcomes are there from rolling a red die and a yellow die?
3) How many possible outcomes are there from rolling three dice, if you only count the outcomes that involve at most one of the dice coming up as a one?

---

## SUMMARY:

- Product rule
- Sum rule
- Combining the product and sum rules

---

# Chapter 3

# Permutations, Combinations, and the Binomial Theorem

The examples we looked at in Chapter 2 involved drawing things from an effectively infinite population — they couldn't run out. When you are making up a password, there is no way you're going to "use up" the letter b by including it several times in your password. In Example 2.1.4, Chloë's suppliers weren't going to run out of blue t-shirts after printing some of her order, and be unable to complete the remaining blue t-shirts she'd requested. The fact that someone has already flipped a coin once with a result of heads doesn't mean they've used up that side of the coin so won't be able to flip heads again.

In this chapter, we'll look at situations where we are choosing more than one item from a finite population in which every item is uniquely identified — for example, choosing people from a family, or cards from a deck.

The study of permutations and combinations has a very long history, particularly in India and China. Much of the older historical context that will be provided in this book comes from handouts that were developed by Prof. Randy Schwartz of Schoolcraft College in the U.S. He has made a study of the history of mathematics in a non-eurocentric context, for incorporation into his own classes, and his handouts on "Combinations and Their Sums" and "Binomial Coefficients and Subsets" are particularly relevant to this course.

## 3.1. Permutations

We begin by looking at permutations, because these are a straightforward application of the product rule. The word "permutation" means a rearrangement, and this is exactly what a permutation is: an ordering of a number of distinct items in a line. Sometimes even though we have a large number of distinct items, we want to single out a smaller number and arrange those into a line; this is also a sort of permutation.

**DEFINITION 3.1.1.** A **permutation** of $n$ distinct objects is an arrangement of those objects into an ordered line. If $1 \leq r \leq n$ (and $r$ is a natural number) then an **$r$-permutation** of $n$ objects is an arrangement of $r$ of the $n$ objects into an ordered line.

So a permutation involves choosing items from a finite population in which every item is uniquely identified, and keeping track of the order in which the items were chosen.

Since we are studying enumeration, it shouldn't surprise you that what we'll be asking in this situation is *how many* permutations there are, in a variety of circumstances. Let's begin with an example in which we'll calculate the number of 3-permutations of ten objects (or in this case, people).

**EXAMPLE 3.1.2.** Ten athletes are competing for Olympic medals in women's speed skating (1000 metres). In how many ways might the medals end up being awarded?

**SOLUTION.** There are three medals: gold, silver, and bronze, so this question amounts to finding the number of 3-permutations of the ten athletes (the first person in the 3-permutation is the one who gets the gold medal, the second gets the silver, and the third gets the bronze).

To solve this question, we'll apply the product rule, where the aspects that can vary are the winners of the gold, silver, and bronze medals. We begin by considering how many different athletes might get the gold medal. The answer is that any of the ten athletes might get that medal. No matter which of the athletes gets the gold medal, once that is decided we move our consideration to the silver medal. Since one of the athletes has already been awarded the gold medal, only nine of them remain in contention for the silver medal, so for any choice of athlete who wins gold, the number of choices for who gets the silver medal is nine. Finally, with the gold and silver medalists out of contention for the bronze, there remain eight choices for who might win that medal. Thus, the total number of ways in which the medals might be awarded is $10 \cdot 9 \cdot 8 = 720$. □

We can use the same reasoning to determine a general formula for the number of $r$-permutations of $n$ objects:

**THEOREM 3.1.3.** *The number of $r$-permutations of $n$ objects is $n(n-1)\dots(n-r+1)$.*

**PROOF.** There are $n$ ways in which the first object can be chosen (any of the $n$ objects). For each of these possible choices, there remain $n-1$ objects to choose for the second object, etc.□

It will be very handy to have a short form for the number of permutations of $n$ objects.

**NOTATION 3.1.4.** We use $n!$ to denote the number of permutations of $n$ objects, so

$$n! = n(n-1)\dots 1.$$

By convention, we define $0! = 1$.

**DEFINITION 3.1.5.** We read $n!$ as "***n* factorial**," so $n$ factorial is $n(n-1)\dots 1$.

Thus, the number of $r$-permutations of $n$ objects can be re-written as $n!/(n-r)!$. When $n = r$ this gives $n!/0! = n!$, making sense of our definition that $0! = 1$.

**EXAMPLE 3.1.6.** In 1150CE, the Indian mathematician Bhāskara II (1114—1185) wrote a four-part book about mathematics and astronomy, entitled *Siddhānta Siromani* ("Crown of treatises"). One of the problems he considered was to find the sum of all numbers obtained by permuting the digits of 23456789.

**SOLUTION.** By the arguments we have discussed above, there are $8! = 40320$ summands. If we consider any fixed position, one-eighth of these summands have each of the 8 given digits in that position; that is, 5040 do. So the sum of the digits in this position is

$$5040(2+3+4+5+6+7+8+9) = 221760.$$

Therefore, the sum of all of these numbers is

$$221760(1 + 10 + 10^2 + 10^3 + 10^4 + 10^5 + 10^6 + 10^7) = 221760(111111111) = 2463999975360.$$

□

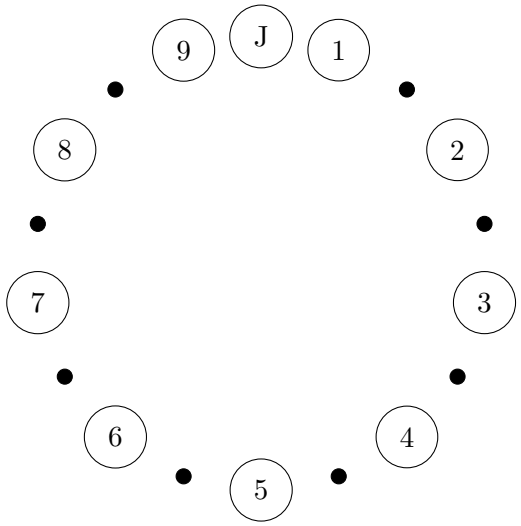
**EXAMPLE 3.1.7.** There are 36 people at a workshop. They are seated at six round tables of six people each for lunch. The Morris family (of three) has asked to be seated together (side-by-side). How many different seating arrangements are possible at the Morris family's table?

**SOLUTION.** First, there are $3! = 6$ ways of arranging the order in which the three members of the Morris family sit at the table. Since the tables are round, it doesn't matter which specific seats they take; only the order in which they sit matters. Once the Morris family is seated, the three remaining chairs are uniquely determined by their positions relative to the Morris family (one to their right, one to their left, and one across from them). There are 33 other people at the conference; we need to choose three of these people and place them in order into the three vacant chairs. There are $33!/(33 - 3)! = 33!/30!$ ways of doing this. In total, there are $6(33!/30!) = 196,416$ different seating arrangements possible at the Morris family's table. □

By adjusting the details of the preceding example, it can require some quite different thought processes to find the answer.

**EXAMPLE 3.1.8.** At the same workshop, there are three round dinner tables, seating twelve people each. The Morris family members (Joy, Dave, and Harmony) still want to sit at the same table, but they have decided to spread out (so no two of them should be side-by-side) to meet more people. How many different seating arrangements are possible at the Morris family's table now?

**SOLUTION.** Let's begin by arbitrarily placing Joy somewhere at the table, and seating everyone else relative to her. This effectively distinguishes the other eleven seats. Next, we'll consider the nine people who aren't in Joy's family, and place them (standing) in an order clockwise around the table from her. There are $33!/(33 - 9)!$ ways to do this. Before we actually assign seats to these nine people, we decide where to slot in Dave and Harmony amongst them.




(In the above diagram, the digits 1 through 9 represent the nine other people who are sitting at the Morris family's table, and the J represents Joy's position.) Dave can sit between any pair of non-Morrises who are standing beside each other; that is, in any of the spots marked by small black dots in the diagram above. Thus, there are eight possible choices for where Dave

will sit. Now Harmony can go into any of the remaining seven spots marked by black dots. Once Dave and Harmony are in place, everyone shifts to even out the circle (so the remaining black dots disappear), and takes their seats in the order determined.

We have shown that there are $(33!/24!)8 \cdot 7$ possible seating arrangements at the Morris table. That's a really big number, and it's quite acceptable to leave it in this format. However, in case you find another way to work out the problem and want to check your answer, the total number is $783,732,837,888,000$. □

**EXERCISES 3.1.9.** Use what you have learned about permutations to work out the following problems. The sum and/or product rule may also be required.

1) Six people, all of whom can play both bass and guitar, are auditioning for a band. There are two spots available: lead guitar, and bass player. In how many ways can the band be completed?
2) Your friend Garth tries out for a play. After the auditions, they text you that they got one of the parts they wanted, and that (including them) nine people tried out for the five roles. You know that there were two parts that interested them. In how many ways might the cast be completed (who gets which role matters)?
3) You are creating an 8-character password. You are allowed to use any of the 26 lowercase characters, and you must use exactly one digit (from 0 through 9) somewhere in the password. You are not allowed to use any character more than once. How many different passwords can you create?
4) How many 3-letter "words" (strings of characters, they don't actually have to be words) can you form from the letters of the word STRONG? How many of those words contain an s? (You may not use a letter more than once.)
5) How many permutations of $\{0, 1, 2, 3, 4, 5, 6\}$ have no adjacent even digits? For example, a permutation like 5034216 is not allowed because 4 and 2 are adjacent.

## 3.2. Combinations

Sometimes the order in which individuals are chosen doesn't matter; all that matters is whether or not they were chosen. An example of this is choosing a set of problems for an exam. Although the order in which the questions are arranged may make the exam more or less intimidating, what really matters is which questions are on the exam, and which are not. Another example would be choosing shirts to pack for a trip (assuming all of your shirts are distinguishable from each other). We call a choice like this a "combination," to indicate that it is the collection of things chosen that matters, and not the order.

**DEFINITION 3.2.1.** Let $n$ be a positive natural number, and $0 \leq r \leq n$. Assume that we have $n$ distinct objects. An ***r*-combination** of the $n$ objects is a subset consisting of $r$ of the objects.

So a combination involves choosing items from a finite population in which every item is uniquely identified, but the order in which the choices are made is unimportant.

Again, you should not be surprised to learn (since we are studying enumeration) that what we'll be asking is *how many* combinations there are, in a variety of circumstances. One significant difference from permutations is that it's not interesting to ask how many $n$-combinations there are of $n$ objects; there is only one, as we must choose all of the objects.

Let's begin with an example in which we'll calculate the number of 3-combinations of ten objects (or in this case, people).

**EXAMPLE 3.2.2.** Of the ten athletes competing for Olympic medals in women's speed skating (1000 metres), three are to be chosen to form a committee to review the rules for future competitions. How many different committees could be formed?

**SOLUTION.** We determined in Example 3.1.2 that there are $10!/7!$ ways in which the medals can be assigned. One easy way to choose the committee would be to make it consist of the three medal-winners. However, notice that if (for example) Wong wins gold, Sajna wins silver, and Andersen wins bronze, we will end up with the same committee as if Sajna wins gold, Andersen wins silver, and Wong wins bronze. In fact, what we've learned about permutations tells us that there are $3!$ different medal outcomes that would each result in the committee being formed of Wong, Sajna, and Andersen.

In fact, there's nothing special about Wong, Sajna, and Andersen — for any choice of three people to be on the committee, there are $3! = 6$ ways in which those individuals could have been awarded the medals. Therefore, when we counted the number of ways in which the medals could be assigned, we counted each possible 3-member committee exactly $3! = 6$ times. So the number of different committees is $10!/(7!3!) = 10 \cdot 9 \cdot 8/6 = 120$. □

We can use the same reasoning to determine a general formula for the number of $r$-combinations of $n$ objects:

**THEOREM 3.2.3.** *The number of $r$-combinations of $n$ objects is*

$$\frac{n!}{r!(n-r)!}.$$

**PROOF.** By Theorem 3.1.3, the number of $r$-permutations of $n$ object is $n!/(n-r)!$ . Suppose that we knew there are $k$ unordered $r$-subsets of $n$ objects (i.e. $r$-combinations). For each of these $k$ unordered subsets, there are $r!$ ways in which we could order the elements. This tells us that $k \cdot r! = n!/(n-r)!$. Rearranging the equation, we obtain $k = n!/(r!(n-r)!)$. □

It will also prove extremely useful to have a short form for the number of $r$-combinations of $n$ objects.

**NOTATION 3.2.4.** We use $\binom{n}{r}$ to denote the number of $r$-combinations of $n$ objects, so

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

**DEFINITION 3.2.5.** We read $\binom{n}{r}$ as "***n* choose *r***," so $n$ choose $r$ is $n!/[r!(n-r)!]$.

Notice that when $r = n$, we have

$$\binom{n}{r} = \frac{n!}{n!(n-n)!} = \frac{n!}{n!0!} = \frac{n!}{n!} = 1,$$

coinciding with our earlier observation that there is only one way in which all of the $n$ objects can be chosen. Similarly,

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = 1;$$

there is exactly one way of choosing none of the $n$ objects.

One of the earliest known historical examples of calculating combinations comes from India in the 500s (CE).

**EXAMPLE 3.2.6.** Varāhamihira (499—587), an astronomer and mathematician from the region of Ujjain, wrote the *Brhatsambitā*, an extensive treatise on divination. At one point in this book, he considers 16 possible fragrances that can be included in perfumes, and counts the number of distinct perfumes that could be made using any four of these fragrances. What is the correct answer to this?

**SOLUTION.** Now that we understand how to calculate combinations, the answer to this is very straightforward:

$$\binom{16}{4} = \frac{16!}{4!(16-4)!} = \frac{16 \cdot 15 \cdot 14 \cdot 13}{4 \cdot 3 \cdot 2} = 1820.$$

□

Let's go over another example that involves combinations.

**EXAMPLE 3.2.7.** Jasmine is holding three cards from a regular deck of playing cards. She tells you that they are all hearts, and that she is holding at least one of the two highest cards in the suit (Ace and King). If you wanted to list all of the possible sets of cards she might be holding, how long would your list be?

**SOLUTION.** We'll consider three cases: that Jasmine is holding the Ace (but not the King); that she is holding the King (but not the Ace), or that she is holding both the Ace and the King.

If Jasmine is holding the Ace but not the King, of the eleven other cards in the suit of hearts she must be holding two. There are $\binom{11}{2}$ possible choices for the cards she is holding in this case.

Similarly, if Jasmine is holding the King but not the Ace, of the eleven other cards in the suit of hearts she must be holding two. Again, there are $\binom{11}{2}$ possible choices for the cards she is holding in this case.

Finally, if Jasmine is holding the Ace and the King, then she is holding one of the other eleven cards in the suit of hearts. There are $\binom{11}{1}$ possible choices for the cards she is holding in this case.

In total, you would have to list

$$\binom{11}{2} + \binom{11}{2} + \binom{11}{1} = \frac{11!}{2!9!} + \frac{11!}{2!9!} + \frac{11!}{1!10!} = \frac{11 \cdot 10}{2} + \frac{11 \cdot 10}{2} + 11 = 55 + 55 + 11 = 121$$

possible sets of cards.

Here is another analysis that also works: Jasmine has at least one of the Ace and the King, so let's divide the problem into two cases: she might be holding the Ace, or she might be holding the King but not the Ace. If she is holding the Ace, then of the twelve other hearts, she is holding two; these can be chosen in $\binom{12}{2} = 66$ ways. If she is holding the King but not the Ace, then as before, her other two cards can be chosen in $\binom{11}{2} = 55$ ways, for a total (again) of 121.□

A common mistake in an example like this, is to divide the problem into the cases that Jasmine is holding the Ace, or that she is holding the King, and to determine that each of these cases includes $\binom{12}{2} = 66$ possible combinations of cards, for a total of 132. The problem with this analysis is that we've counted the combinations that include both the Ace and the King twice: once as a combination that includes the Ace, and once as a combination that includes the King. If you do this, you need to compensate by subtracting at the end the number of combinations that have been counted twice: that is, those that include the Ace and the King. As we worked out in the example, there are $\binom{11}{1} = 11$ of these, making a total of $132 - 11 = 121$ combinations.

We return to some simple examples of how combinations have been studied and used historically.

**EXAMPLE 3.2.8.** Abraham ibn Ezra (c. 1093—1167) lived in Spain while it was under Moorish rule. In his book *Se'fer Ha'Olam* ("The Book of the World"), he considered the number of "planetary conjunctions" that could occur involving any given number of the seven "planets". At the time, all of the "wandering" celestial bodies were thought to orbit the earth, and the seven such bodies that were known were called the planets. These seven were the moon, the sun, Mercury, Venus, Mars, Jupiter, and Saturn. When at least two of them appear to meet in the sky, it is called a conjunction. For each $k$ from 2 through 7, how many different ways can exactly $k$ of these be involved in a conjunction?

**SOLUTION.** For $k = 2$, the answer is

$$\binom{7}{2} = \frac{7!}{2!(7-2)!} = \frac{7 \cdot 6}{2} = 21.$$

For $k = 3$, the answer is

$$\binom{7}{3} = \frac{7!}{3!(7-3)!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2} = 35.$$

For $k = 4$, the answer is

$$\binom{7}{4} = \frac{7!}{4!(7-4)!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2} = 35.$$

For $k = 5$, the answer is

$$\binom{7}{5} = \frac{7!}{5!(7-5)!} = \frac{7 \cdot 6}{2} = 21.$$

For $k = 6$, the answer is

$$\binom{7}{6} = \frac{7!}{6!(7-6)!} = 7.$$

For $k = 7$, the answer is

$$\binom{7}{7} = \frac{7!}{7!(7-7)!} = 1$$

(recall that $0! = 1$). □

**EXAMPLE 3.2.9.** In 1144 at the age of 19, Al-Samaw'al ben Yahyā al-Maghribī (c.1130—c.1180) of Baghdad wrote *Al-Bāhir fi'l-jabr* (the "Brilliant in Algebra"). Among other things, he considered equations in multiple variables and counted the number of ways in which it is possible to take a sum of exactly 6 out of 10 possible variables. What is the correct answer to this?

**SOLUTION.** The answer to this is the number of ways to choose 6 of the 10 variables to sum up: that is,

$$\binom{10}{6} = \frac{10!}{6!(10-6)!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2} = 210.$$

□

**EXERCISES 3.2.10.** Use what you have learned about combinations to work out the following problems. Permutations and other counting rules we've covered may also be required.

1) For a magic trick, you ask a friend to draw three cards from a standard deck of 52 cards. How many possible sets of cards might they have chosen?
2) For the same trick, you insist that your friend keep replacing their first draw until they draw a card that isn't a spade. They can choose any cards for their other two cards. How many possible sets of cards might they end up with? (Caution: choosing $5\clubsuit, 6\diamondsuit, 3\spadesuit$ in that order, is *not* different from choosing $6\diamondsuit, 5\clubsuit, 3\spadesuit$ in that order. You do *not* need to take into account that some sets will be more likely to occur than others.)
3) How many 5-digit numbers contain exactly two zeroes? (We insist that the number contain exactly 5 digits.)
4) Sandeep, Hee, Sara, and Mohammad play euchre with a standard deck consisting of 24 cards (A, K, Q, J, 10, and 9 from each of the four suits of a regular deck of playing cards). In how many ways can the deck be dealt so that each player receives 5 cards, with 4 cards left in the middle, one of which is turned face-up? The order of the 3 cards that are left face-down in the middle does not matter, but who receives a particular set of 5 cards (for example, Sara or Sandeep) does matter.
5) An ice cream shop has 10 flavours of ice cream and 7 toppings. Their *megasundae* consists of your choice of any 3 flavours of ice cream and any 4 toppings. (A customer must choose *exactly* three different flavours of ice cream and four different toppings.) How many different megasundaes are there?

## 3.3. The Binomial Theorem

Here is an algebraic example in which "$n$ choose $r$" arises naturally.

**EXAMPLE 3.3.1.** Consider

$$(a+b)^4 = (a+b)(a+b)(a+b)(a+b).$$

If you try to multiply this out, you must systematically choose the $a$ or the $b$ from each of the four factors, and make sure that you make every possible combination of choices sooner or later.

One way of breaking this task down into smaller pieces, is to separate it into five parts, depending on how many of the factors you choose $a$s from (4, 3, 2, 1, or 0). Each time you choose 4 of the $a$s, you will obtain a single contribution to the coefficient of the term $a^4$; each time you choose 3 of the $a$s, you will obtain a single contribution to the term $a^3b$; each time you choose 2 of the $a$s, you will obtain a single contribution to the term $a^2b^2$; each time you choose 1 of the $a$s, you will obtain a single contribution to the term $ab^3$; and each time you choose 0 of the $a$s, you will obtain a single contribution to the term $b^4$. In other words, the coefficient of a particular term $a^ib^{4-i}$ will be the number of ways in which you can choose $i$ of the factors from which to take an $a$, taking a $b$ from the other $4-i$ factors (where $0 \leq i \leq 4$).

Let's go through each of these cases separately. By Theorem 3.2.3, there is $\binom{4}{4} = 1$ way to choose four factors from which to take $a$s. (Clearly, you must choose an $a$ from every one of the four factors.) Thus, the coefficient of $a^4$ will be 1.

If you want to take $a$s from three of the four factors, Theorem 3.2.3 tells us that there are $\binom{4}{3} = 4$ ways in which to choose the factors from which you take the $a$s. (Specifically, these four ways consist of taking the $b$ from any one of the four factors, and the $a$s from the other three factors). Thus, the coefficient of $a^3b$ will be 4.

If you want to take $a$s from two of the four factors, and $b$s from the other two, Theorem 3.2.3 tells us that there are $\binom{4}{2} = 6$ ways in which to choose the factors from which you take the $a$s (then take $b$s from the other two factors). This is a small enough example that you could easily work out all six ways by hand if you wish. Thus, the coefficient of $a^2b^2$ will be 6.

If you want to take $a$s from one of the four factors, Theorem 3.2.3 tells us that there are $\binom{4}{1} = 4$ ways in which to choose the factors from which you take the $a$s. (Specifically, these four ways consist of taking the $a$ from any one of the four factors, and the $b$s from the other three factors). Thus, the coefficient of $ab^3$ will be 4.

Finally, by Theorem 3.2.3, there is $\binom{4}{0} = 1$ way to choose zero factors from which to take $a$s. (Clearly, you must choose a $b$ from every one of the four factors.) Thus, the coefficient of $b^4$ will be 1.

Putting all of this together, we see that

$$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4.$$

In fact, if we leave the coefficients in the original form in which we worked them out, we see that

$$(a+b)^4 = \binom{4}{4}a^4 + \binom{4}{3}a^3b + \binom{4}{2}a^2b^2 + \binom{4}{1}ab^3 + \binom{4}{0}b^4.$$

This example generalises into a significant theorem of mathematics:

**THEOREM 3.3.2 (Binomial Theorem).** *For any $a$ and $b$, and any natural number $n$, we have*

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}.$$

*One special case of this is that*

$$(1+x)^n = \sum_{r=0}^{n} \binom{n}{r} x^r.$$

**PROOF.** As in Example 3.3.1, we see that the coefficient of $a^r b^{n-r}$ in $(a+b)^n$ will be the number of ways of choosing $r$ of the $n$ factors from which we'll take the $a$ (taking the $b$ from the other $n-r$ factors). By Theorem 3.2.3, there are $\binom{n}{r}$ ways of making this choice.

For the special case, begin by observing that $(1+x)^n = (x+1)^n$; then take $a = x$ and $b = 1$ in the general formula. Use the fact that $1^{n-r} = 1$ for any integers $n$ and $r$. □

The Binomial Theorem has been known for a long time. In China it was often used as a way to work out numerical estimations for high powers of mixed numbers, for example by taking the first few terms from the expansion of something like $(5 + .11)^5$.

From the theorem above, we see that the values $\binom{n}{r}$ are the coefficients of the terms in the Binomial Theorem.

**DEFINITION 3.3.3.** Expressions of the form $\binom{n}{r}$ are referred to as **binomial coefficients**.

There are some nice, simple consequences of the Binomial Theorem.

**COROLLARY 3.3.4.** *For any natural number $n$, we have*

$$\sum_{r=0}^{n} \binom{n}{r} = 2^n.$$

**PROOF.** This is an immediate consequence of substituting $a = b = 1$ into the Binomial Theorem. □

**COROLLARY 3.3.5.** *For any natural number $n$, we have*

$$\sum_{r=0}^{n} r\binom{n}{r}(-1)^{r-1} = 0.$$

**PROOF.** From the special case of the Binomial Theorem, we have

$$(1+x)^n = \sum_{r=0}^{n} \binom{n}{r} x^r.$$

If we differentiate both sides, we obtain

$$n(1+x)^{n-1} = \sum_{r=0}^{n} r\binom{n}{r} x^{r-1}.$$

Substituting $x = -1$ gives the result (the left-hand side is zero). □

*Remark 3.3.6.* We've encountered a number of "theorems" earlier in this book, which is probably a term you've seen previously, also. This is the first time we've stated a result that we haven't called a "theorem", so it's worth spending a few moments reviewing the various terms we'll use for results, and the circumstances under which each is appropriate.

The term *theorem* is generally reserved for significant results. A result that we might want to refer to from other courses or in other contexts should receive this term: an example of this is the *Fundamental Theorem of Calculus*, or from this course, the Binomial Theorem. In higher mathematics we don't generally call something a theorem if its proof is too easy, but within a course like this we tend to use the term more broadly, for results that are important within the context of the course.

A *corollary* such as the ones we see here, means a result that follows as an easy or direct consequence of another result. Typically the proof of a corollary should be very short. It may in large measure involve looking at the previous result (often a theorem) from a slightly different perspective, or taking a special case of it.

A *lemma* is a self-contained result that provides a stepping stone to one or more results of greater interest. Lemmas are used in a variety of situations, including:

- if a particular fact is needed more than once in a proof, it is often broken out into a lemma in order to avoid repeating the argument;
- if a proof is long and complex but breaks down into a series of steps that are reasonably self-contained, these may be separated into lemmas to make the arguments easier to follow;
- if a mathematician thinks that one piece of their argument may be of use in other contexts or in its own right, they may separate it out into a lemma.

A *proposition* is a self-contained result that is not a step along the way to another result, but is not sufficiently significant to deserve to be called a theorem.

These terms aren't always used with precision. When we encounter Euler's handshaking lemma later in the course, you might reasonably think it deserves to be called a theorem. However, perhaps because it is usually used to show other interesting results rather than on its own, the term "lemma" has become part of its title.

**EXERCISES 3.3.7.** Use the Binomial Theorem to evaluate the following:

1) $\sum_{i=1}^{n} \binom{n}{i} 2^i$.

2) the coefficient of $a^2b^3c^2d^4$ in $(a+b)^5(c+d)^6$.

3) the coefficient of $a^2b^6c^3$ in $(a+b)^5(b+c)^6$.

4) the coefficient of $a^3b^2$ in $(a+b)^5 + (a+b^2)^4$.

---

## SUMMARY:

- The number of $r$-permutations of $n$ objects is $n!/(n-r)!$.
- The number of $r$-combinations of $n$ objects is $\binom{n}{r} = \frac{n!}{r!(n-r)!}$.
- The Binomial Theorem
- Important definitions:
  - permutation, $r$-permutation
  - $n$ factorial
  - $r$-combination
  - $n$ choose $r$
  - binomial coefficients
- Notation:
  - $n!$
  - $\binom{n}{r}$

Chapter 4

# Bijections and Combinatorial Proofs

You may have learned previously that two sets have the same cardinality if there is a bijection between them. (A bijection is a one-to-one, onto function.) This leads us to another important method for counting a set: we come up with a bijection between the elements of the unknown set, and the elements of a set that we do know how to count. This idea is very closely related to the concept of a combinatorial proof, which we will explore in the second half of this chapter.

*Aside:* We won't explore the concepts of $P$ and $NP$ directly at all in this course (though they will be mentioned a few times). if you have studied these ideas in computer science courses, you may be interested to learn that most of the techniques used to prove that a particular problem is in $P$ or in $NP$ are related to the techniques discussed in this chapter. Usually a problem $X$ is proven to be in $NP$ (for example) by starting with a problem $Y$ that is already known to be in $NP$. Then the researcher uses some clever ideas to show that problem $Y$ can be related to problem $X$ in such a way that if problem $X$ could be solved in polynomial time, that solution would produce a solution to problem $Y$, still in polynomial time. Thus the fact that $Y$ is in $NP$ forces $X$ to be in $NP$ also. The same ideas may sometimes relate the number of solutions of problem $X$ to the number of solutions of problem $Y$.

## 4.1. Counting via bijections

It can be hard to figure out how to count the possible outcomes of a particular experiment (like shuffling cards, or flipping a number of coins). Sometimes it will be possible to find a different problem, and to prove that the two problems have the same number of outcomes (by finding a bijection between their outcomes). If we can work out how to count the outcomes for the second problem, then we've also solved the first problem! This may seem blatantly obvious intuitively, but this technique can provide simple solutions to problems that at first glance seem very difficult.

This technique of counting a set (or the number of outcomes to some problem) indirectly, via a different set or problem, is the bijective technique for counting. We begin with a classic example of this technique.

**EXAMPLE 4.1.1.** How many possible subsets are there, from a set of $n$ elements?

**SOLUTION.** One approach would be to figure out how many 0-element subsets there are, how many 1-element subsets, etc., and add up all of the values we find. This works, but there are so many pieces involved that it is prone to error. Also, the value will not be easy to calculate once $n$ gets reasonably large.

Instead, we imagine creating a table. The columns are indexed by the elements of the set, so there are $n$ columns. We index the rows by the subsets of our set (one per row). In each entry of the table, we place a 1 if the subset that corresponds to this row contains the element that corresponds to this column. Here's an example of such a table when the set is $\{x, y, z\}$:

| | $x$ | $y$ | $z$ |
|---|---|---|---|
| $\emptyset$ | 0 | 0 | 0 |
| $\{x\}$ | 1 | 0 | 0 |
| $\{y\}$ | 0 | 1 | 0 |
| $\{z\}$ | 0 | 0 | 1 |
| $\{x, y\}$ | 1 | 1 | 0 |
| $\{x, z\}$ | 1 | 0 | 1 |
| $\{y, z\}$ | 0 | 1 | 1 |
| $\{x, y, z\}$ | 1 | 1 | 1 |

As you can see, the pattern of 1s and 0s is different in each row of the table, since the elements of each subset are different. Furthermore, any pattern of 1s and 0s that has length 3, appears in some row of this table.

This is not a coincidence. In general, we can define a bijection between the binary strings of length $n$, and the subsets of a set of $n$ elements, as follows. We already know by the definition of cardinality, that there is a bijection between any set of $n$ elements, and the set $\{1, \ldots, n\}$, so we'll actually define a bijection between the subsets of $\{1, \ldots, n\}$ and the binary strings of length $n$. Since the composition of two bijections is a bijection, this will indirectly define a bijection between our original set, and the binary strings of length $n$.

Given a subset of $\{1, \ldots, n\}$, the binary string that corresponds to this subset will be the binary string that has 1s in the $i$th position if and only if $i$ is in the subset. This tells us how to determine the binary string from the subset. We can also reverse (invert) the process. Given a binary string of length $n$, the corresponding subset of $\{1, \ldots, n\}$ will be the subset whose elements are the positions of the 1s in the binary string.

Although we haven't directly proven that this map from subsets to binary strings is both one-to-one and onto, an invertible function must be a bijection, so the fact that we were able to find an inverse function does prove that this map is a bijection. (You should check that you agree that the function we've claimed as an inverse really does invert the original function.)

Now, our imaginary table wouldn't be much use if we actually had to write it out. In order to write it out, we would need to know all of the subsets of our set already; and if we knew them all, we could certainly count them! Fortunately, we do not need to write it out. Instead, we use the bijection we have just defined. Rather than count the number of subsets of an $n$-set, we count the number of binary strings of length $n$. We can do this using just the multiplication rule! In each position there is either a zero or a one, so there are 2 choices for each of the $n$ positions. Hence, there are $2^n$ binary strings of length $n$.

We conclude that $2^n$ subsets can be formed from a set of cardinality $n$. □

This has been known for a very long time. Roughly two millennia ago, the Indian doctor Sushruta (c.800BCE—c.700BCE) computed the number of flavour combinations that can be obtained by using at least one of the six known flavours (astringent, bitter, hot, salty, sour, or sweet) to be $2^6 - 1 = 63$ (removing the empty subset).

In some ways, we've actually been using this idea pretty much every time we've come up with more than one way to solve a problem. Implicitly, finding a different way of thinking of the problem is equivalent to finding a bijection between the solutions to these different approaches.

**EXAMPLE 4.1.2.** How many ways are there to choose ten people from a group of 30 men and 30 women, if the group must include at least one woman?

**SOLUTION.** Attacking this problem directly will get ugly. We would have to consider separately the cases of including one woman, two women, etc., all the way up to ten women, in our group, and add all of the resulting terms together. Instead, we note that there is an obvious bijection (the identity map) between groups that *do* include at least one woman, and groups that *do not* include exactly zero women. (We have previously seen this idea used intuitively in applications of the sum rule, for example when we worked out the number of ways of obtaining at least one 1 when rolling three six-sided dice, by actually working out the number of ways of *not* obtaining zero 1s.)

The number of groups that do not include zero women is relatively easy to figure out: there are $\binom{60}{10}$ possible groups of ten people that could be chosen from the 60 people. Of these, there are $\binom{30}{10}$ groups that do include zero women (since the members of any such group must be chosen entirely from the 30 men). Therefore, the number of groups that do not include exactly zero women, is $\binom{60}{10} - \binom{30}{10}$.

Thanks to our bijection, we conclude that the number of groups that can be chosen, that will include at least one woman, is also $\binom{60}{10} - \binom{30}{10}$. □

**EXERCISES 4.1.3.** The following problems should help you in working with the bijective technique for counting.

1) We define a structure that is like a subset, except that any element of the original set may appear 0, 1, or 2 times in the structure. How many of these structures can we form from the set $\{1, \ldots, n\}$?
2) Find a bijection between the coefficient of $x^r$ in $(1 + x)^n$, and the number of $r$-combinations of an $n$-set.
3) Find a bijection between the number of ways in which three different dolls can be put into ten numbered cribs, and the number of ways in which ten Olympic contenders can win the medals in their event.

## 4.2. Combinatorial proofs

As we said in the previous section, thinking about a problem in two different ways implicitly creates a bijection, telling us that the number of solutions we obtain will be the same either way. When we looked at bijections, we were using this idea to find an easier way to count something that seemed difficult. But if we actually can find a (possibly messy) formula that counts the answer to our problem correctly in some "difficult" way, and we can also find a different formula that counts the answer to the same problem correctly by looking at it in a different way, then we know that the values of the two formulas must be equal, no matter how different they may look.

This is the idea of a "combinatorial proof."

**THEOREM 4.2.1 (Combinatorial Proofs).** *If $f(n)$ and $g(n)$ are functions that count the number of solutions to some problem involving $n$ objects, then $f(n) = g(n)$ for every $n$.*

**DEFINITION 4.2.2.** Suppose that we count the solutions to a problem about $n$ objects in one way and obtain the answer $f(n)$ for some function $f$; and then we count the solutions to the same problem in a different way and obtain the answer $g(n)$ for some function $g$. This is a **combinatorial proof** of the identity $f(n) = g(n)$.

The equation $f(n) = g(n)$ is referred to as a **combinatorial identity**.

In the statement of this theorem and definition, we've made $f$ and $g$ functions of a single variable, $n$, but the same ideas hold if $f$ and $g$ are functions of more than one variable. Some of our examples will demonstrate this.

These ideas give us a methodical way to approach the proof of a combinatorial identity.

**Combinatorial Proof outline**

1) *Identify the problem you are counting.* The first step is to identify a problem for which we will be counting the solutions. Sometimes the problem may be provided directly in the question. If so, this first step is easy. If that didn't happen, examine one or the other of the formulas you have been asked to show are equal, and come up with any problem whose solutions can be counted by that formula. Explain this problem, and state that you will be counting the solutions in two ways.
2) *Counting method 1.* Explain why the left-hand side of the combinatorial identity counts the solutions to the problem you have identified. This is often easy, and perhaps even follows directly from definitions, especially if you came up with the problem yourself from that formula.
3) *Counting method 2.* Explain why the right-hand side of the combinatorial identity counts the solutions to the problem you have identified. Unless it was challenging to identify a problem in the first place (which can certainly happen), this is likely to be the most challenging part of the proof.
4) *Conclusion.* Note that since your methods both count the number of solutions to the same problem, the results must be equal, and state the identity that you have proven.

It is not always necessary to explicitly list and label each of these steps, but you may find it helpful as a method of clearly laying out your proof, and ensuring that it is correct and complete. Sometimes it may make sense to start with the right-hand side of the identity and then look at the left-hand side; this is perfectly acceptable.

**EXAMPLE 4.2.3.** Prove that for every natural number $n$ and every integer $r$ between 0 and $n$, we have

$$\binom{n}{r} = \binom{n}{n-r}.$$

**COMBINATORIAL PROOF..** *The problem.* We will count the number of ways of choosing $r$ objects from a set of $n$ distinct objects in two ways.

*Counting method 1:* By the definition of $\binom{n}{r}$, this is the number of ways of choosing $r$ objects from a set of $n$ distinct objects.

*Counting method 2:* Any time we choose $r$ objects from a set of $n$ objects, the other $n - r$ objects are being left out of the set we are choosing. So equivalently, instead of choosing the $r$ objects to include in our set, we could choose the $n - r$ objects to leave out of our set. By the definition of the binomial coefficients, there are $\binom{n}{n-r}$ ways of making this choice.

*Conclusion.* Since both of these methods count the number of solutions to the same problem, it must be the case that for every natural number $n$ and every integer $r$ between 0 and $n$, we

have

$$\binom{n}{r} = \binom{n}{n-r},$$

as desired. □

Of course, this particular identity is also quite easy to prove directly, using the formula for $\binom{n}{r}$, since

$$\binom{n}{n-r} = \frac{n!}{(n-r)!(n-(n-r))!} = \frac{n!}{(n-r)!r!} = \binom{n}{r}.$$

Many identities that can be proven using a combinatorial proof can also be proven directly, or using a proof by induction. The nice thing about a combinatorial proof is it usually gives us rather more insight into *why* the two formulas should be equal, than we get from many other proof techniques.

When you are asked to prove a result using a combinatorial proof, it is important that you use this technique (including all of the steps explained above). In some cases, there are other proofs that you may find much easier (such as taking a special case of the Binomial Theorem). However, the goal of these exercises is not for you to show that you can come up with a correct proof, but to give you practice and experience in using a proof technique that you haven't seen before. Using a different proof technique completely defeats the purpose.

In Example 4.1.1, we noted that one way to figure out the number of subsets of an $n$-element set would be to count the number of subsets of each possible size, and add them all up. We then followed a bijective approach to prove that the answer is in fact $2^n$. If we actually carry through on the first idea, this leads to another combinatorial identity (one that we already observed via the Binomial Theorem):

**EXAMPLE 4.2.4.** Prove that for every natural number $n$,

$$\sum_{r=0}^{n} \binom{n}{r} = 2^n.$$

**COMBINATORIAL PROOF..** *The problem.* We will count the number of subsets of an $n$-element set in two ways.

*Counting method 1:* For this example, we will start with the right-hand side of the identity. We have seen in Example 4.1.1 that the number of subsets of a set of $n$ elements is $2^n$.

*Counting method 2:* To determine the number of subsets of a set of $n$ elements, we break the problem down into $n+1$ cases, and use the sum rule. The cases into which we will divide the problem are the different possible cardinalities for the subsets: everything from 0 through $n$. There are $\binom{n}{r}$ ways to choose a subset of $r$ elements from the set of $n$ elements, so the number of subsets that contain $r$ elements is $\binom{n}{r}$. Thus, the total number of subsets of our original set must be $\sum_{r=0}^{n} \binom{n}{r}$.

*Conclusion.* Since we have counted the same problem in two different ways and obtained different formulas, Theorem 4.2.1 tells us that the two formulas must be equal; that is,

$$\sum_{r=0}^{n} \binom{n}{r} = 2^n,$$

as desired. □

This was in fact the idea that was used by Sushruta (c.800BCE—c.700BCE) in his flavour calculations. It has also been used in a variety of other contexts over the years. The Indian poet Halayudha (10th century CE) considered the number of kinds of poetic meter that were possible in a line with a fixed number of syllables, where the meter determined which syllables were short and which were long. If the line has $k$ syllables, then the number of choices is $2^k$, and he used combinations in his calculations. Bhāskara II (1114—1185), who we earlier saw also used permutations, asked in the same book about the number of ways of choosing some collection of doors to be open in a palace with 8 doors (requiring that at least one be open), and arrived at the answer $2^8 - 1 = 255$. Although it is an application of this particular combinatorial proof rather than an example of a combinatorial proof, we provide one other historical example in slightly more detail.

**EXAMPLE 4.2.5.** Abū-l'Abbās Ahmad ibn al-Bannā' (1256—1321) of Morocco considered a triangle to have 5 key associated values that might be of interest: the lengths of the three sides; the height; and the area. He asked how many geometrical problems associated with triangles could be posed, in which some number (at least one, but not all) of these quantities are provided and the rest are asked for. (Whether or not the problem is solvable from the given information is not relevant.) He also asked similar questions about circles (to which he attributed three associated values: diameter, perimeter, and area), and other shapes, but we'll only do the calculation for triangles.

**SOLUTION.** Any subset of the five quantities can be provided, except at least one value must be provided (so the subset cannot be empty) and the problem cannot provide all of the values (so the subset cannot be the entire set). The answer is therefore $2^5 - \binom{5}{0} - \binom{5}{5} = 32 - 2 = 30$.□

We can also produce an interesting combinatorial identity from a generalisation of the problem studied in Example 4.1.2.

**EXAMPLE 4.2.6.** Suppose we have a collection of $n$ men and $n$ women, and we want to choose $r$ of them for a focus group, but we must include at least one woman. In how many ways can this be done? Use two different methods to count the solutions, and deduce a combinatorial identity.

**SOLUTION.** *The problem.* On this occasion the problem was given to us explicitly, and we will need to come up with the two sides of the combinatorial identity. We will find two methods for counting the number of ways of including at least one woman in an $r$-member focus group, from an initial collection of $n$ men and $n$ women.

*Counting method 1:* Using the same reasoning that we applied in Example 4.1.2, we see that the number of ways of choosing a group that includes at least one woman is the total number of ways of choosing a group of $r$ people from these $2n$ people, less the number of ways that include only men; that is: $\binom{2n}{r} - \binom{n}{r}$.

*Counting method 2:* Alternatively, we can divide the problem up into $r$ cases depending on how many women are to be included in the group (there must be $i$ women, for some $1 \leq i \leq r$). There are $\binom{n}{i}$ ways to choose $i$ women for the group, and for each of these, there are $\binom{n}{r-i}$ ways to choose $r - i$ men to complete the group. Thus, the total number of ways of choosing a group that includes at least one woman, is

$$\sum_{i=1}^{r} \binom{n}{i}\binom{n}{r-i}.$$

*Conclusion.* Since both of these methods counted solutions to the same problem, this argument yields the combinatorial identity

$$\sum_{i=1}^{r} \binom{n}{i}\binom{n}{r-i} = \binom{2n}{r} - \binom{n}{r},$$

which we have thereby proven. □

Our next combinatorial identity was used historically to work out unknown binomial coefficients from known values.

**EXAMPLE 4.2.7.** Although his ultimate goal was to compute how many "words" (letter combinations of various lengths) could be composed from the Arabic alphabet of 28 letters, Ahmad ibn Mun'im al-'Abdarī (11??—1228) began with a different problem.

In Morocco around 1200CE, where Ahmad lived, silk tassels made with a variety of colours of silk threads were a common decoration. He asked: of the 10 available colours, how many ways are there to make tassels that contain exactly three colours? (Since the threads are wound together and get disarranged easily, the order of the colours is not relevant.)

**SOLUTION.** It was known to Ahmad how to calculate $\binom{n}{2}$ for any $n$, but explicit formulas for $\binom{n}{3}$ were not known. He began by ordering the colours:

| | | | |
|---|---|---|---|
| 1) indigo; | 4) silver; | 7) azure; | 10) gold. |
| 2) black; | 5) orange; | 8) red; | |
| 3) emerald; | 6) white; | 9) purple; | |

Now he argued: the problem of choosing three colours can be broken down into 8 cases, depending on the highest numbered colour that is used. (This number must be at least 3 in order to allow two colours with smaller numbers to exist, and cannot be more than 10.) For each choice of highest-numbered colour, he could work out the number of ways to choose the other two colours from colours with lower numbers as follows:

1) you can choose colour 3 (emerald) as the highest colour; in this case you must choose two (both) of the colours with lower numbers (indigo and black). There are $\binom{2}{2} = 1$ ways to do this;
2) you can choose colour 4 (silver) as the highest colour; in this case you must choose two of the three colours with lower numbers. There are $\binom{3}{2} = 3$ ways to do this;
3) you can choose colour 5 (orange) as the highest colour; in this case you must choose two of the four colours with lower numbers. There are $\binom{4}{2} = 6$ ways to do this;
4) you can choose colour 6 (white) as the highest colour; in this case you must choose two of the five colours with lower numbers. There are $\binom{5}{2} = 10$ ways to do this;
5) you can choose colour 7 (azure) as the highest colour; in this case you must choose two of the six colours with lower numbers. There are $\binom{6}{2} = 15$ ways to do this;
6) you can choose colour 8 (red) as the highest colour; in this case you must choose two of the seven colours with lower numbers. There are $\binom{7}{2} = 21$ ways to do this;
7) you can choose colour 9 (purple) as the highest colour; in this case you must choose two of the eight colours with lower numbers. There are $\binom{8}{2} = 28$ ways to do this;
8) you can choose colour 10 (gold) as the highest colour; in this case you must choose two of the nine colours with lower numbers. There are $\binom{9}{2} = 36$ ways to do this.

So in total, there are

$$1 + 3 + 6 + 10 + 15 + 21 + 28 + 36 = 120$$

ways to create tassels with three colours. Notice that using our formula,

$$\binom{10}{3} = \frac{10 \cdot 9 \cdot 8}{6} = 120$$

also.

The argument used by Ahmad was also used by Abraham ibn Ezra (c. 1093—1167) in his study of planetary conjunctions (see Example 3.2.8). However, Ahmad observed this as a general pattern. We leave up to you to generalise this argument into a formal proof of the combinatorial identity that underlies it: that

□

$$\binom{n}{k} = \sum_{r=k-1}^{n-1} \binom{r}{k-1}.$$

One context in which combinatorial proofs arise very naturally, is when we are counting ordered pairs that have some property. That is, for some subset of $X \times Y$, we may wish to count all of the ordered pairs $(x, y)$, where $x \in X$ and $y \in Y$, such that $(x, y)$ has some property. We can do this by first considering every possible value of $x \in X$, and for each such value, counting the number of $y \in Y$ such that $(x, y)$ satisfies the desired property, or by first considering every possible value of $y \in Y$, and for each such value, counting the number of $x \in X$ such that $(x, y)$ satisfies the desired property.

Although this idea may not seem very practical, it is actually the context in which many of the combinatorial proofs in later chapters will arise. We will be looking at a set $X$ of elements, and a set $Y$ that is actually a collection of subsets of elements of $X$, and counting pairs $(x, y)$ for which the element $x$ appears in the subset $y$. By counting these pairs in two ways, we will find a combinatorial identity.

**EXAMPLE 4.2.8.** Let $B$ be the set of city blocks in a small city, and let $S$ be the set of street segments in the city (where a street segment means a section of street that lies between two intersections). Assume that each block has at least three sides. Count the number of pairs $(s, b)$ with $s \in S$ and $b \in B$ such that the street segment $s$ is adjacent to the block $b$ in two ways. Use this to deduce a combinatorial inequality.

**SOLUTION.** Note that since the question asks us to deduce an inequality rather than an equality, one of our counts will not be precise. In other words, for one of our two counting methods, we will be trying to say that the result is *at most* or *at least* some formula, rather than that it is equal to that formula.

*The problem.* We will count the number of pairs $(s, b)$ with $s \in S$ and $b \in B$ such that the street segment $s$ is adjacent to the block $b$ in two ways.

*Counting method 1:* Let $|S| = t$. Each street segment is adjacent to two blocks: the blocks that lie on either side of the street. Therefore, for any given street segment $s$, there are two pairs $(s, b)$ such that $s$ is adjacent to the block $b$. Multiplying this count by $t$ (the number of street segments) tells us that the total number of pairs $(s, b) \in S \times B$ with $s$ adjacent to $b$ is $2t$.

*Counting method 2:* Let $|B| = c$. Each block is adjacent to *at least* 3 street segments (this is where the inequality arises): the street segments that surround the block. Therefore, for any given block $b$ in the city, there are at least 3 pairs $(s, b)$ such that $b$ is adjacent to the street segment $s$. Multiplying this count by $c$ (the number of blocks) tells us that the total number of pairs $(s, b) \in S \times B$ with $s$ adjacent to $b$ is at least $3c$.

*Conclusion.* We deduce that $2t \geq 3c$. □

**EXERCISE 4.2.9.** Let $P$ be the set of people in a group, with $|P| = p$. Let $C$ be a set of clubs formed by the people in this group, with $|C| = c$. Suppose that each club contains exactly $g$ people, and each person is in exactly $j$ clubs. Use two different ways to count the number of pairs $(b, h) \in P \times C$ such that person $b$ is in club $h$, and deduce a combinatorial identity.

**EXERCISES 4.2.10.** Prove the following combinatorial identities, using combinatorial proofs:

1) For any natural numbers $r, n$, with $1 \leq r \leq n$, $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$.
[*Hint:* Consider the number of ways to form a team of $r$ people from a group of $n$ people. Then break the problem into two cases depending on whether or not one specific person is chosen for the team.]
2) For any natural numbers $k, r, n$, with $0 \leq k \leq r \leq n$, $\binom{n}{r}\binom{r}{k} = \binom{n}{k}\binom{n-k}{r-k}$.
[*Hint:* Consider the number of ways to choose $r$ dogs who will enter a competition, from a set of $n$ dogs, and to choose $k$ of those $r$ dogs to become the finalists. Then choose the finalists first, followed by the other dogs who entered the competition.]
3) For any natural number $n$, $\sum_{r=0}^{n} \binom{n}{r}^2 = \binom{2n}{n}$.
4) For $n \geq 1$ and $k \geq 1$, $\frac{n!}{(n-k)!} = n\frac{(n-1)!}{(n-1-(k-1))!}$.
5) For $n \geq 1$, $3^n = \sum_{k=0}^{n} \binom{n}{k} 2^{n-k}$.

**EXERCISES 4.2.11.** Sometimes the hardest part of a combinatorial proof can be the first step: figuring out what problem the given formula provides a solution to. For each of the following formulas, state a counting problem that can be solved by the formula.

1) $n2^{n-1}$.
2) $\sum_{r=0}^{n} r\binom{n}{r}$.
3) $\sum_{k=r}^{n} \binom{n}{k}\binom{k}{r}$.
4) $2^{n-r}\binom{n}{r}$.

## 4.3. The Arithmetic Triangle (Pascal's Triangle)

You may well have been introduced to the following triangle (of which we show only the first six rows here):

$$\begin{array}{ccccccccccc}
 & & & & & \binom{0}{0} & & & & & \\
 & & & & \binom{1}{0} & & \binom{1}{1} & & & & \\
 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & \\
 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & \\
 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & \\
\binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5}
\end{array}$$

In the $n$th row of the triangle (starting with row 0), the binomial coefficients $\binom{n}{k}$ appear, with $k$ going from 0 to $n$ as we proceed across the row.

The combinatorial identities shown in Example 4.2.7 and Exercise 4.2.10.1 provide techniques for calculating the actual values of the entries in a particular row, from the entries in the rows above it (we also note that the first and last entries in any row are 1s). The second

of these identities provides the simplest calculation: each entry is the sum of the entry above and to the left, with the entry above and to the right.

Below we show the calculated values for the part of the triangle listed above.

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

This triangle has been used in various forms for many centuries. The Moroccan mathematician Ahmad ibn Mun'im al-'Abdarī (11??—1228) drew the first 10 rows in his calculations of the number of tassels that could be produced using combinations of 10 colours of silk thread (see Example 4.2.7), in the 1200s. The Indian poet Halayudha (10th century CE) in the 900s drew it as a tool in working out the numbers of various kinds of poetic metres that he was interested in (with a specified number of syllables per line, a particular number of which had to be long or short). In the book *Siyuan Yuchian* ("Jade Mirror of the Four Unknowns"), produced in 1303, the Chinese mathematician Zhu Shijie (1249—1314) drew this triangle (and attributed it to much older sources). Although the first documented use by Blaise Pascal (1623—1662) of the triangle was in 1654, it is still usually (Eurocentrically) generally referred to as "Pascal's triangle". The "Arithmetic triangle" is a more neutral term that is still used by some sources, and was in fact the term used by Pascal in his 1654 work *Treatise on the Arithmetic Triangle*, but this term is less likely to be recognised.

---

**SUMMARY:**

- Counting via bijections
- Combinatorial identities
- Combinatorial proofs
- the Arithmetic Triangle

---

# Chapter 5

# Counting with Repetitions

In counting combinations and permutations, we assumed that we were drawing from a set in which all of the elements are distinct. Of course, it is easy to come up with a scenario in which some of the elements are indistinguishable. We need to know how to count the solutions to problems like this, also.

## 5.1. Unlimited repetition

For many practical purposes, even if the number of indistinguishable elements in each class is not actually infinite, we will be drawing a small enough number that we will not run out. The bagel shop we visited in Example 2.2.1 is not likely to run out of one variety of bagel before filling a particular order. In standard card games, we never deal enough cards to a single player that they might have all of the cards of one suit and still be getting more cards.

This is the sort of scenario we'll be studying in this section. The set-up we'll use is to assume that there are $n$ different "types" of item, and there are enough items of each type that we won't run out. Then we'll choose items, allowing ourselves to repeatedly choose items of the same type as many times as we wish, until the total number of items we've chosen is $r$. Notice that (unlike in Chapter 3), in this scenario $r$ may exceed $n$.

We'll consider two scenarios: the order in which we make the choice matters, or the order in which we make the choice doesn't matter.

**EXAMPLE 5.1.1.** Chris has promised to bring back bagels for three friends they are studying with (as well as one for themself). The bagel shop sells eight varieties of bagel. In how many ways can they choose the bagels to give to Jan, Tom, Olive, and themself?

**SOLUTION.** Here, it matters who gets which bagel. We can model this by assuming that the first bagel Chris orders will be for themself, the second for Jan, the third for Tom, and the last for Olive. Thus, the order in which they ask for the bagels matters.

We actually saw back in Chapter 2 how to solve this problem. It's just an application of the product rule! Chris has eight choices for the first bagel; for each of these, they have eight choices for the second bagel; for each of these, they have eight choices for the third bagel; and for each of these, they have eight choices for the fourth bagel. So in total, they have $8^4$ ways to choose the bagels. □

OK, so if the order in which we make the choice matters, we just use the multiplication rule. What about if order doesn't matter?

**EXAMPLE 5.1.2.** When Chris brought back the bagels, it turned out that they'd done a poor job of figuring out what their friends wanted. They all traded around. Later that night, they sent Chris back to the doughnut store, but this time they told them to just bring back eight doughnuts and they'd figure out who should get which. If the doughnut store has five varieties, how many ways are there for Chris to fill this order?

**SOLUTION.** Let's call the five varieties chocolate, maple, boston cream, powdered, and jam-filled. One way to describe Chris' order would be to make a list in which we first write one **c** for each chocolate doughnut, then one **m** for each maple doughnut, then one **b** for each boston cream doughnut, then one **p** for each powdered doughnut, and finally one **j** for each jam-filled doughnut. Since Chris is ordering eight doughnuts, there will be eight letters in this list. Notice that there's more information provided by this list than we actually need. We know that all of the first group of letters are **c**s, so instead of writing them all out, we could simply put a dividing marker after all of the **c**s and before the first **m**. Similarly, we can put three more dividing markers in to separate the **m**s from the **b**s, the **b**s from the **p**s, and the **p**s from the **j**s. Now we have a list that might look something like this:

**cc||bbb|ppp|**

(Notice in this possible list, Chris chose no maple or jam-filled doughnuts.)

Now, we don't actually need to write down the letters **c**, **m**, **b** and so on, as long as we know how many spaces they take up; we know that any letters that appear before the first dividing marker are **c**s, for example. Thus, the following list gives us the same information as the list above:

__||___|___|

Similarly, if we see the list

|__|__|___|_

we understand that Chris ordered no chocolate doughnuts; two maple doughnuts; two boston cream doughnuts; three powdered doughnuts; and one jam-filled doughnut.

So an equivalent problem is to count the number of ways of arranging eight underlines and four dividing markers in a line. This is something we already understand! We have twelve positions that we need to fill, and the problem is: in how many ways can we fill eight of the twelve positions with underlines (placing dividing markers in the other four positions). We know that this can be done in $\binom{12}{8}$ ways. □

This technique can be used to give us a general formula for counting the number of ways of choosing $r$ objects from $n$ types of objects, where we are allowed to repeatedly choose objects of the same type.

**THEOREM 5.1.3.** *The number of ways of choosing $r$ objects from $n$ types of objects (with replacement or repetition allowed) is*

$$\binom{n+r-1}{r}.$$

**PROOF.** We use the same idea as in the solution to Example 5.1.2, above. Since there are $n$ different types of objects, we will need $n-1$ dividing markers to keep them apart. Since we are choosing $r$ objects, we will need $r$ underlines, for a total of $n+r-1$ positions to be filled. We can choose the $r$ positions in which the objects will go in $\binom{n+r-1}{r}$ ways, and then (in each

case) put dividing markers into the remaining $n-1$ positions. Thus, there are $\binom{n+r-1}{r}$ ways to choose $r$ objects from $n$ types of objects, if repetition or replacement of choices is allowed. □

Again, we will want to have a short form for this value.

**NOTATION 5.1.4.** We use $\left(\!\binom{n}{r}\!\right)$ to denote the number of ways of choosing $r$ objects from $n$ types of objects (with replacement or repetition allowed), so

$$\left(\!\binom{n}{r}\!\right) = \binom{n+r-1}{r}.$$

The reason we say "replacement or repetition" is because there is another natural model for this type of problem. Suppose that instead of choosing eight bagels from five varieties, Chris is asked to put their hand into a bag that contains five different-coloured pebbles, and draw one out; then replace it, repeatedly (with eight draws in total). If they keep count of how many times they draw each of the rocks, the number of possible tallies they'll end up with is exactly the same as the number of doughnut orders in Example 5.1.2.

The following table summarises some of the key things we've learned about counting so far:

**Table 5.1. The number of ways to choose $r$ objects from $n$ objects (or types of objects)**

| | **repetition allowed** | **repetition not allowed** |
|---|---|---|
| **order matters** | $n^r$ | $\dfrac{n!}{(n-r)!}$ |
| **order doesn't matter** | $\left(\!\binom{n}{r}\!\right)$ | $\binom{n}{r}$ |

**EXERCISES 5.1.5.** Solve the following problems.

1) Each of the ten sections in your community band (trombones, flutes, and so on) includes at least four people. The conductor needs a quartet to play at a school event. How many different sets of instruments might end up playing at the event?
2) The prize bucket at a local fair contains six types of prizes. Kim wins 4 prizes; Jordan wins three prizes, and Finn wins six. Each of the kids plans to give one of the prizes they have won to their teacher, and keep the rest. In how many ways can their prizes (including the gifts to the teacher) be chosen? (It is important which gift comes from which child.)
3) There are three age categories in the local science fair: junior, intermediate, and senior. The judges can choose nine projects in total to advance to the next level of competition, and they must choose at least one project from each age group. In how many ways can the projects that advance be distributed across the age groups?

**EXERCISES 5.1.6.** Prove the following combinatorial identities:

1) For $k, n \geq 1$, $\left(\!\binom{n}{k}\!\right) = \left(\!\binom{n-1}{k}\!\right) + \left(\!\binom{n}{k-1}\!\right)$.
2) For $k, n \geq 1$, $\left(\!\binom{n}{k}\!\right) = \binom{n+k-1}{k}$.

3) For $k, n \geq 1$, $\left(\binom{k+1}{n-1}\right) = \binom{n+k-1}{k}$.

4) For $1 \leq n \leq k$, $\left(\binom{n}{k-n}\right) = \binom{k-1}{k-n}$.

**EXERCISES 5.1.7.** Solve the following problems.

1) Find the number of 5-lists of the form $(x_1, x_2, x_3, x_4, x_5)$, where each $x_i$ is a nonnegative integer and $x_1 + x_2 + x_3 + x_4 + 3x_5 = 12$.

2) We will buy 3 pies (not necessarily all different) from a store that sells 4 kinds of pie. How many different orders are possible? List all of the possibilities (using A for apple, B for blueberry, C for cherry, and D for the other one).

3) Suppose Lacrosse balls come in 3 colours: red, yellow, and blue. How many different combinations of colours are possible in a 6-pack of Lacrosse balls?

4) After expanding $(a+b+c+d)^7$ and combining like terms, how many terms are there? [Justify your answer without performing the expansion.]

## 5.2. Sorting a set that contains repetition

In the previous section, the new work came from looking at combinations where repetition or replacement is allowed. For our purposes, we assumed that the repetition or replacement was effectively unlimited; that is, the store might only have 30 cinnamon raisin bagels, but since Chris was only ordering four bagels, that limit didn't matter.

In this section, we're going to consider the situation where there are a fixed number of objects in total; some of them are "repeated" (that is, indistinguishable from one another), and we want to determine how many ways they can be arranged (permuted). This can arise in a variety of situations.

**EXAMPLE 5.2.1.** When Chris gets back from the doughnut store run, they discover that Mohammed, Jing, Karl, and Sara have joined the study session. Chris has bought two chocolate doughnuts, three maple doughnuts, and three boston cream doughnuts. In how many ways can the doughnuts be distributed so that everyone gets one doughnut?

**SOLUTION.** Initially, this looks a lot like a permutation question: we need to figure out the number of ways to arrange the doughnuts in some order, and give the first doughnut to the first student, the second doughnut to the second student, and so on.

The key new piece in this problem is that, unlike the permutations we've studied thus far, the two chocolate doughnuts are indistinguishable (as are the three maple doughnuts and the three boston cream doughnuts). This means that there is no difference between giving the first chocolate doughnut to Tom and the second to Mohammed, and giving the first chocolate doughnut to Mohammed and the second to Tom.

One way to solve this problem is to look at it as a series of combinations of the people, rather than as a permutation question about the doughnuts. Instead of arranging the doughnuts, we can first choose which two of the eight people will receive the two chocolate doughnuts. Once that is done, from the remaining six people, we choose which three will receive maple doughnuts. Finally, the remaining three people receive boston cream doughnuts. Thus, the solution is $\binom{8}{2}\binom{6}{3}$.

Another approach is more like the approach we used to figure out how many $r$-combinations there are of $n$ objects. In this approach, we begin by noting that we would be able to arrange the eight doughnuts in 8! orders if all of them were distinct. For any fixed choice of two people who receive the chocolate doughnuts, there are 2! ways in which those two chocolate doughnuts could have been distributed to them, so in the 8! orderings of the doughnuts, each of these choices for who gets the chocolate doughnuts has been counted 2! times rather than once.

Similarly, for any fixed choice of three people who receive the maple doughnuts, there are 3! ways in which these three maple doughnuts could have been distributed to them, and each of these choices has been counted 3! times rather than once. The same holds true for the three boston cream doughnuts. Thus, the solution is 8!/(2!3!3!).

Since

$$\binom{8}{2}\binom{6}{3} = \frac{8!}{2!6!} \cdot \frac{6!}{3!3!} = \frac{8!}{2!3!3!},$$

we see that these solutions are in fact identical although they look different. □

This technique can be used to give us a general formula for counting the number of ways of arranging $n$ objects some of which are indistinguishable from each other.

**THEOREM 5.2.2.** *Suppose that:*

- *there are $n$ objects;*
- *for each $i$ with $1 \leq i \leq m$, $r_i$ of them are of type $i$ (indistinguishable from each other); and*
- $r_1 + \ldots + r_m = n$.

*Then the number of arrangements (permutations) of these $n$ objects is*

$$\frac{n!}{r_1!r_2!\ldots r_m!}.$$

**PROOF.** We use the same idea as in the solution to Example 5.2.1, above. Either approach will work, but we'll use the first. There will be $n$ positions in the final ordering of the objects. We begin by choosing $r_1$ of these to hold the objects of type 1. Then we choose $r_2$ of them to hold the objects of type 2, and so on. Ultimately, we choose the final $r_m$ locations (in $\binom{r_m}{r_m} = 1$ possible way) to hold the objects of type $m$.

Using the product rule, the total number of arrangements is

$$\begin{aligned}
&\binom{n}{r_1}\binom{n-r_1}{r_2}\cdots\binom{n-r_1-\ldots-r_{m-1}}{r_m} \\
&= \frac{n!}{r_1!(n-r_1)!} \cdot \frac{(n-r_1)!}{r_2!(n-r_1-r_2)!} \cdot \ldots \cdot \frac{(n-r_1-\ldots-r_{m-1})!}{r_m!0!} \\
&= \frac{n!}{r_1!r_2!\ldots r_m!},
\end{aligned}$$

since all of the other terms cancel. □

We have notation for this value also.

**NOTATION 5.2.3.** We use $\binom{n}{r_1,\ldots,r_m}$ to denote the number of arrangements of $n = r_1 + \ldots + r_m$ objects where for each $i$ with $1 \leq i \leq m$ we have $r_i$ indistinguishable objects of type $i$. Thus,

$$\binom{n}{r_1,\ldots,r_m} = \frac{n!}{r_1!\ldots r_m!}.$$

This can sometimes apply in unexpected ways.

**EXAMPLE 5.2.4.** Cathy, Akos, and Dagmar will be going into a classroom of 30 students. They will each be pulling out four students to work with in a small group setting. In how many ways can the groups be chosen?

**SOLUTION.** Even though all of the students in the class are distinct, the order in which they get chosen for the group they end up in doesn't matter. One way of making the selection would be to put the names Cathy, Akos, and Dagmar into a hat (four times each) along with 18 blank slips of paper. Each student could choose a slip of paper and would be assigned to the group corresponding to the name they chose. The four slips with Cathy's name on them are identical, as are the four with Akos' name, the four with Dagmar's name, and the 18 blank slips.

Thus, the solution to this problem is

$$\binom{30}{4,4,4,18} = \frac{30!}{4!4!4!18!}.$$

We could also work this out more directly, by allowing each of Cathy, Akos, and Dagmar to choose four students; Cathy's choice can be made in $\binom{30}{4}$ ways; then Akos' in $\binom{26}{4}$ ways; then Dagmar's in $\binom{22}{4}$ ways, and the product of these is $30!/(4!4!4!18!)$. □

**EXERCISES 5.2.5.** Evaluate the following problems.

1) Charlie's teacher gives him a set of magnetic words. He has to make a "poem" using all of them. The words are: on, the, one, up, a, tree, the, child, on, jumps, feels, the, child, with. How many different "poems" can Charlie create, if any ordering of the words is considered to be a poem?
2) When filling the soccer team's fundraising order, the chocolate company sent six extra boxes of chocolate-covered almonds, three extra boxes of mints, and two extra boxes of plain chocolate. In how many ways can the extras be fairly distributed to the eleven families who ordered chocolates?

**EXERCISE 5.2.6.** Prove the *Multinomial Theorem*: that

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{k_1+k_2+\cdots+k_m=n} \binom{n}{k_1, k_2, \ldots, k_m} \prod_{1\le r\le m} x_r^{k_r}.$$

[*Hint:* Choose arbitrary values for $k_1, k_2, \ldots, k_m$ such that

$$k_1 + k_2 + \cdots + k_m = n,$$

and evaluate the coefficient of $x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m}$ that comes from the product on the left-hand side of the equation.]

**SUMMARY:**

- The number of ways of choosing $r$ objects from $n$ types of objects (with replacement or repetition allowed) is $\left(\binom{n}{r}\right) = \binom{n+r-1}{r}$.
- The number of ways of arranging $n$ objects where $r_i$ of them are of type $i$ (indistinguishable), is $\binom{n}{r_1, r_2, \ldots, r_m}$.
- Notation:
  - $\left(\binom{n}{r}\right)$
  - $\binom{n}{r_1, r_2, \ldots, r_m}$

# Chapter 6

# Induction and Recursion

Some problems can most easily be solved (or counted) with the help of a recursively-defined sequence. We'll begin this chapter by introducing these sequences.

Proofs by induction are an important mathematical technique, and are often used in published papers. This material is being presented under the assumption that you have seen elementary proofs by induction in at least one previous course. The basic explanations and examples are written with the intention of reviewing this background. If you have not encountered induction before you may wish to find supplementary introductory material and exercises. We'll do a quick review of basic proofs by induction, applying them to recursively-defined sequences. Then we'll touch on some slightly more sophisticated uses of induction. Proofs by induction will be a technique we'll use throughout the remainder of the course, in a variety of contexts.

## 6.1. Recursively-defined sequences

You may be familiar with the term "recursion" as a programming technique. It comes from the same root as the word "recur," and is a technique that involves repeatedly applying a self-referencing definition until we reach some initial terms that are explicitly defined, and then going back through the applications to work out the result we want. If you didn't follow that, it's okay; we'll go through the definition and some specific examples that should give you the idea.

**DEFINITION 6.1.1.** A sequence $r_1, r_2, \ldots, r_n, \ldots$ is **recursively defined** if for every $n$ greater than or equal to some bound $b \geq 2$, the value for $r_n$ depends on at least some of the values of $r_1, \ldots, r_{n-1}$. The values for $r_1, \ldots, r_{b-1}$ are given explicitly; these are referred to as the **initial conditions** for the recursively-defined sequence. The equation that defines $r_n$ from $r_1, \ldots, r_{n-1}$ is called the **recursive relation**.

Probably the best-known example of a recursively-defined sequence, is the Fibonacci sequence. In one of the many examples we'll see in this course of inappropriate attribution, this name for the sequence comes from the Italian mathematician Leonardo Pisano ("Fibonacci") (c.1170—c.1250), who introduced the sequence to western culture as an example in a book he wrote in 1202 to advocate for the use of arabic numerals and the decimal system.

Many of the other examples of misattribution that we'll see involve two people of similar identities. This one seems to have a more ethnocentric basis, since the person honoured is the one who introduced the ideas to Europe. European society promptly lost sight of the fact that these ideas had existed prior to that introduction (in much the same way that the "discovery" of

many North American locations has historically been attributed to European explorers). The sequence had been known to Indian mathematicians as early as the 6th century. Nevertheless, to avoid confusion we will use this name.

**DEFINITION 6.1.2.** The **Fibonacci sequence** is the sequence $f_0, f_1, f_2, \ldots$, defined by $f_0 = 1$, $f_1 = 1$, and $f_n = f_{n-1} + f_{n-2}$ for all $n \geq 2$.

So in the Fibonacci sequence, $f_0 = f_1 = 1$ are the initial conditions, and $f_n = f_{n-1} + f_{n-2}$ for all $n \geq 2$ is the recursive relation.

The usual problem associated with recursively-defined sequences, is to find an explicit formula for the $n$th term that does not require calculating all of the previous terms. Clearly, if we want to be able to determine terms that arise later in the sequence, this is critical. If we try to find the millionth term of a recursively-defined sequence directly, it will require a great deal of computing time and might also require a lot of memory.

Every time you were asked in school to look at a sequence of numbers, find a pattern, and give the next number in the sequence, you were probably working out a recurrence relation and applying it.

**EXAMPLE 6.1.3.** Consider the sequence $5, 8, 11, 14$. What number should come next?

**SOLUTION.** We consider the differences between successive pairs: $8 - 5 = 3$; $11 - 8 = 3$; $14 - 11 = 3$. This appears to be an arithmetic sequence, with the constant difference of 3 between successive terms. So the sequence can be defined by $a_1 = 5$ and $a_n = a_{n-1} + 3$, for every $n \geq 2$. We were asked for $a_5$, and we know that $a_4 = 14$, so $a_5 = a_4 + 3 = 14 + 3 = 17$.□

Here's a slightly more complicated example:

**EXAMPLE 6.1.4.** Consider the sequence $3, 6, 11, 18, 27$. What number should come next?

**SOLUTION.** Again, consider the differences between successive terms: $6 - 3 = 3$; $11 - 6 = 5$; $18 - 11 = 7$; $27 - 18 = 9$. These differences aren't constant, but do follow a predictable pattern: they are the odd numbers (starting at 3 and increasing). So the sequence can be defined by $a_1 = 3$ and $a_n = a_{n-1} + (2n - 1)$, for every $n \geq 2$. We were asked for $a_6$, and we know that $a_5 = 27$, so $a_6 = a_5 + 2(6) - 1 = 27 + 11 = 38$. □

This example shows that the recurrence relation can depend on $n$, as well as on the values of the preceding terms. (Although we didn't state this explicitly in our definition, it is implicit because $n - 1$ is the number of previous terms on which $r_n$ depends; we could calculate $n$ as $a_1^0 + a_2^0 + \ldots + a_{n-1}^0 + 1$.)

Let's look at one more example.

**EXAMPLE 6.1.5.** Stavroula's bank pays 1% interest (compounded annually), and charges her a service fee of \$10 per year to maintain the account. The fee is charged at the start of the year, and the interest is calculated on the balance at the end of the year. If she starts with a balance of \$2000, is she making money or losing money? If this account is set up for her by her parents and she's not allowed to touch it, how much money will be in the account after seven years?

**SOLUTION.** We see that the initial term is $r_0 = 2000$. We're going to use $r_0$ as the first term, because then the value of her account after 1 year will be $r_1$; after two years will be $r_2$; and after seven years will be $r_7$. This just makes it a little easier to keep track of what we're aiming to figure out.

If we unpack the financial language, it is telling us that every year, the bank takes \$10 from Stavroula's account at the start of the year. Then at the end of the year, the bank adds 1% of whatever is in Stavroula's account, to her account. This can be represented by the following

recurrence relation: $r_n = r_{n-1} - 10 + .01(r_{n-1} - 10)$ for every $n \geq 1$, which simplifies to $r_n = 1.01(r_{n-1} - 10)$. Logically, she will be making money if the 1% that she earns in interest, exceeds the service fee of \$10, so if she makes money in the first year, she will continue to make money; while if she loses money in the first year, she will continue to lose money after that. So to answer the first question, we'll work out $r_1 = 1.01(r_0 - 10) = 1.01(1990) = 2009.9$. Stavroula is making money.

To answer the second question, unless we've managed to figure out an explicit formula for $r_n$ (which we don't yet know how to do), we need to calculate $r_2$, $r_3$, $r_4$, $r_5$, $r_6$, and $r_7$. It would be reasonable to assume that the bank rounds its calculations to the nearest penny every year, and carries forward with the rounded value, but because this will create an error that will be compounded in comparison with solving our recurrence relation explicitly (which we'll learn later how to do), we'll keep track of the exact values instead. We have

$$\begin{aligned}
r_2 &= 1.01(2009.9 - 10) = 2019.899;\\
r_3 &= 1.01(2009.899) = 2029.99799;\\
r_4 &= 1.01(2019.99799) = 2040.1979699;\\
r_5 &= 1.01(2030.1979699) = 2050.499949599;\\
r_6 &= 1.01(2040.499949599) = 2060.90494909499; \text{ and}\\
r_7 &= 1.01(2050.90494909499) = 2071.4139985859399.
\end{aligned}$$

So at the end of seven years, Stavroula has \$2071.41. □

**EXERCISES 6.1.6.** Solve the following problems about recurrence relations.

1) Consider the sequence $4, 9, 19, 39$. Give a recurrence relation that describes this sequence, and find the next term in the sequence.
2) Use the recurrence relation for the Fibonacci sequence to find $f_6$.
3) If the annual fee on Stavroula's bank account from Example 6.1.5 is \$20 instead of \$10, is she making money or losing money?

## 6.2. Basic induction

Suppose we want to show that $n!$ is at least $2^n - 2$, for every $n \geq 1$ (where $n$ must be an integer). We could start verifying this fact for each of the possible values for $n$:

$$\begin{aligned}
1! &= 1 \geq 2^1 - 2 = 0;\\
2! &= 2 \geq 2^2 - 2 = 2;\\
3! &= 6 \geq 2^3 - 2 = 6;\\
4! &= 24 \geq 2^4 - 2 = 14.
\end{aligned}$$

We could continue verifying the values one at a time, but the process would go on forever, so we'd never be able to complete the proof.

Instead, think about the following method. We know that the inequality holds for $n = 1$. Let's suppose that the inequality holds for some value $n = k$, i.e. that

$$k! \geq 2^k - 2.$$

Now let's use the fact that we can easily calculate $(k+1)!$ from $k!$ together with our supposition, to deduce that the inequality holds when $n = k + 1$, i.e. that

$$(k + 1)! \geq 2^{k+1} - 2.$$

This is enough to prove the inequality for every integer $n \geq 1$, because applying our supposition and deduction enough times will prove the inequality for any value at all that interests us! For example, if we wanted to be sure that the inequality holds for $n = 100$, we could take the fact that we know it holds for 1, to deduce that it holds for 2, then the fact that it holds for 2 allows us to deduce that it holds for 3. By repeating this 97 more times, eventually we see that since it holds for 99, we can deduce that it holds for 100.

**THEOREM 6.2.1 (Principle of Mathematical Induction).** *Let $P(n)$ be an assertion about the integer $n$. If we know that*

*1) the assertion $P(n_0)$ is true for some particular integer $n_0$; and*

*2) for any integer $k \geq n_0$, if $P(k)$ is true then $P(k + 1)$ must also be true,*

*then $P(n)$ is true for every integer $n \geq n_0$.*

**DEFINITION 6.2.2.** In a proof by induction, determining that $P(n_0)$ is true for some particular integer $n_0$ is called the **base case**. Proving the conditional statement that $P(k) \Rightarrow P(k + 1)$ for every $k \geq n_0$ is called the **inductive step**. The assumption we make in the inductive step, that $P(k)$ is true for some arbitrary $k \geq n_0$, is called the **inductive hypothesis**, and can be referred to by (IH) when it is being used in the proof.

Now that we've gone through the formalities, let's write a proper proof by induction for the inequality we used to introduce this idea.

**EXAMPLE 6.2.3.** Prove by induction that $n! \geq 2^n - 2$, for every integer $n \geq 2$. (This inequality is actually true for every $n \geq 0$, but the proof is considerably simpler if we restrict our attention to $n \geq 2$.)

**PROOF..** *Base case:* $n = 2$. We have $n! = 2! = 2$, and

$$2^n - 2 = 2^2 - 2 = 4 - 2 = 2.$$

Certainly $2 \geq 2$, so the inequality holds for $n = 2$. This completes the proof of the base case.

*Inductive step:* We begin with the inductive hypothesis. Let $k \geq 2$ be arbitrary, and suppose that the inequality holds for $n = k$; that is, assume that $k! \geq 2^k - 2$.

Now we want to deduce that

$$(k + 1)! \geq 2^{k+1} - 2.$$

Let's start from the left-hand side of this inequality. By the definition of factorial, we know that

$$(k + 1)! = (k + 1)k!.$$

Now that we have $k!$ in the expression, we're in a position to apply the inductive hypothesis; that is,

$$(k + 1)! = (k + 1)k! \geq (k + 1)(2^k - 2).$$

Since $k \geq 2$, we have $k + 1 \geq 3$, so

$$(k + 1)(2^k - 2) \geq 3(2^k - 2) = 2(2^k) + 2^k - 6 = 2^{k+1} + 2^k - 6.$$

Again, since $k \geq 2$, we have $2^k \geq 4$, so $2^k - 6 \geq -2$. Hence

$$(k+1)! \geq 2^{k+1} + 2^k - 6 \geq 2^{k+1} - 2,$$

which is what we wanted to deduce. This completes the proof of the inductive step.

By the Principle of Mathematical Induction, $n! \geq 2^n - 2$ for every integer $n \geq 2$. □

Proofs by induction work very naturally with recursively-defined sequences, since the recurrence relation gives us information about the $(k+1)$st term of the sequence, based on previous terms.

**EXAMPLE 6.2.4.** Consider the sum of the first $n$ integers. We can think about this as a recursively-defined sequence, by defining $s_1 = 1$, and $s_n = s_{n-1} + n$, for every $n \geq 2$. Thus, $s_2 = 1 + 2$;

$$s_3 = s_2 + 3 = 1 + 2 + 3,$$

and so on. Prove by induction that $s_n = n(n+1)/2$, for every $n \geq 1$.

**PROOF..** *Base case:* $n = 1$. We have $s_n = s_1 = 1$, and

$$n(n+1)/2 = 1(2)/2 = 1,$$

so the equality holds for $n = 1$. This completes the proof of the base case.

*Inductive step:* We begin with the inductive hypothesis. Let $k \geq 1$ be arbitrary, and suppose that the equality holds for $n = k$; that is, assume that $s_k = k(k+1)/2$.

Now we want to deduce that

$$s_{k+1} = (k+1)(k+2)/2.$$

Using the recursive relation, we have $s_{k+1} = s_k + (k+1)$ since $k+1 \geq 2$, and using the inductive hypothesis, we have $s_k = k(k+1)/2$, so putting these together, we see that

$$s_{k+1} = k(k+1)/2 + (k+1).$$

Taking out a common factor of $k+1$ gives

$$s_{k+1} = (k+1)(k/2+1) = (k+1)(k+2)/2,$$

which is what we wanted to deduce. This completes the proof of the inductive step.

By the Principle of Mathematical Induction, $s_n = n(n+1)/2$ for every $n \geq 1$. □

*Caution:* the steps of a proof by induction are precisely defined, and if you leave any of them out, or forget the conditions required, things can go badly wrong. The base case may seem obvious, but can't be left out; also, the hypothesis that $k \geq n_0$ may be critical to the proof, as we saw in Example 6.2.3.

Let's look at an example where, by forgetting to include the base case, we can give a "proof by induction" of something that is clearly false.

**EXAMPLE 6.2.5.** Here is a "proof by induction" (without a base case) that every integer $n$ is at least 1000.

**PROOF..** *Inductive step:* We begin with the inductive hypothesis. Let $k$ be arbitrary, and suppose that $k \geq 1000$.

Now we want to deduce that $k + 1 \geq 1000$. But clearly,

$$k + 1 \geq k \geq 1000$$

(by our inductive hypothesis), which is what we wanted to deduce. This completes the proof of the inductive step.

By the Principle of Mathematical Induction, $n \geq 1000$ for every integer $n$. □

Now it's your turn to try a few. Don't leave out any of the steps!

**EXERCISES 6.2.6.** Use the Principle of Mathematical Induction to prove the following:

1) For the recursively-defined sequence given by $b_1 = 5$ and $b_n = b_{n-1} + 4$ for all $n \geq 2$, prove that for every integer $n \geq 1$, $b_n = 5 + 4(n - 1)$.
2) For the recursively-defined sequence given by $c_1 = 3$ and $c_n = c_{n-1} + 3 \cdot 2^{n-1}$ for all $n \geq 2$, prove that for every integer $n \geq 1$, $c_n = 3(2^n - 1)$.
3) Prove that for every integer $n \geq 0$, $n! \geq n$.
4) Prove that for every integer $n \geq 0$, $4^n - 1$ is divisible by 3.
5) Starting with $n = 2$ and increasing $n$ from there, calculate the first few values for the product

$$t_n = \prod_{j=2}^{n} (1 - \frac{1}{j}).$$

Conjecture a closed formula for $t_n$ based on the values you have calculated, and use induction to prove that your formula is correct.

6) Prove that for every integer $n \geq 1$,

$$\sum_{j=1}^{n} j! \leq \frac{1}{2}(n + 1)!$$

7) Define $c_0 = 1$ and for $n \geq 1$, define $c_n = nc_{n-1} + 1$. Prove by induction: for $n \geq 0$,

$$c_n = \sum_{j=0}^{n} \frac{n!}{(n - j)!}.$$

## 6.3. More advanced induction

Now that we've reviewed the basic form of induction, it's important to consider some more advanced forms that are often used.

The first form we'll look at is strong induction. When we have a recursively-defined sequence that depends on the previous terms, sometimes we need to know not just about the single term that comes immediately before the $n$th term, but about other previous terms. Only by putting all of this information together will we be able to deduce the result we need about the $n$th term.

**EXAMPLE 6.3.1.** Let's define a recursively-defined sequence by $a_1 = 2$ and for every integer $n \geq 2$, we have

$$a_n = \sum_{i=1}^{n-1} a_i.$$

Thus, $a_2 = a_1 = 2$;

$$a_3 = a_1 + a_2 = 2 + 2 = 4;$$
$$a_4 = a_1 + a_2 + a_3 = 2 + 2 + 4 = 8,$$

and so on. Prove by induction that for every $n \geq 2$, we have $a_n = 2^{n-1}$.

**SOLUTION ATTEMPT..** We begin with the base case: when $n = 2$, we have $a_2 = 2 = 2^{2-1}$, so the equality is true for the base case. Now for the inductive hypothesis, we let $k \geq 2$ be arbitrary, and suppose that the equality is true for $n = k$, so $a_k = 2^{k-1}$. Now when $n = k + 1$, we have

$$a_n = a_{k+1} = \sum_{i=1}^{k} a_i,$$

by the recursive relation for this sequence. We know what $a_1$ is, by our initial condition; we know that $a_k = 2^{k-1}$, but what about the values in between? The Principle of Mathematical Induction as we've learned it so far, doesn't allow us to assume anything about (for example) $a_{k-1}$. □

Actually, though, the way the concept of induction works, by the time we're trying to prove something about $a_n$, we've actually already deduced it for *every* value between $n_0$ and $n - 1$ (inclusive). So there is nothing wrong with assuming that $P(i)$ is true for every value between $n_0$ and $k$, rather than just for $k$, in order to deduce that $P(k + 1)$ is true. More concretely, this is saying the following. Suppose that by knowing $P(0)$ we can deduce $P(1)$, and then by knowing $P(0)$ *and* $P(1)$ we can deduce $P(2)$, and so on, so that eventually by knowing that everything from $P(0)$ through $P(k)$ is true, we can deduce that $P(k+1)$ is true. Then $P(n)$ is true for every integer $n \geq 0$. Of course, we don't have to start with 0; we can start with any integer $n_0$. This is the strong form of mathematical induction:

**THEOREM 6.3.2 (Strong Induction).** *Suppose we have a statement $P(n)$ about the integer $n$. If we know that*

*1) the statement $P(n_0)$ is true for some particular integer $n_0$; and*

*2) for any integer $k \geq n_0$, if every $P(i)$ is true for $n_0 \leq i \leq k$, then $P(k + 1)$ must also be true,*

*then $P(n)$ is true for every integer $n \geq n_0$.*

Using this, we can complete the example we started above. We have $a_1 = 2$ (by the initial condition), and the strong induction hypothesis allows us to assume that $a_i = 2^{i-1}$ for every integer $i$ with $2 \leq i \leq k$. So using the recursive relation

$$a_{k+1} = \sum_{i=1}^{k} a_i,$$

we see that

$$a_{k+1} = 2 + \sum_{i=2}^{k} 2^{i-1}.$$

You probably learned in high school how to add up geometric sequences like this; in particular, that

$$\sum_{j=0}^{k} 2^j = 2^{k+1} - 1,$$

and we can re-write what we have as

$$a_{k+1} = 1 + \left(1 + \sum_{j=1}^{k-1} 2^j\right) = 1 + \sum_{j=0}^{k-1} 2^j = 1 + \left(2^{(k-1)+1} - 1\right) = 2^k.$$

This is precisely what we needed to deduce, so this completes the proof.

Let's go over one more example that involves strong induction. In this example, we'll need strong induction for a slightly different reason: we'll need the statement to be true for some values between $n_0$ and $k$, but we're not necessarily sure which ones.

**EXAMPLE 6.3.3.** Shawna is building a tower with lego. Prove that if she has $n$ pieces of lego (where $n \geq 1$), and a "move" consists of sticking two smaller towers together into one (where a tower may consist of one or more pieces of lego), then it will take her $n-1$ moves to complete the tower.

**PROOF..** *Base case:* $n = 1$. Shawna's "tower" is already complete after $n - 1 = 0$ moves. This completes the proof of the base case.

*Induction step:* we begin with the induction hypothesis, that when $1 \leq i \leq k$, it takes Shawna $i - 1$ moves to build a tower that contains $i$ pieces of lego.

Now we want to deduce that when Shawna has $k + 1$ pieces of lego, it takes her $k$ moves to stick them together into a single tower. Notice that when she makes her final move, it must consist of sticking together two smaller towers, one of which contains $j$ pieces of lego, and the other of which contains the remaining $k+1-j$ pieces. Both $j$ and $k+1-j$ must lie between 1 and $k$ (if either of the smaller towers had $k+1$ pieces then the tower would already be complete), so the induction hypothesis applies to each of them. Thus, it has taken Shawna $j - 1$ moves to build the tower that contains $j$ pieces, and $k - j$ moves to build the tower that contains $k - j + 1$ pieces. Together with her final move, then, it must take Shawna

$$(j - 1) + (k - j) + 1$$

moves in total to complete her tower of $k + 1$ pieces. Now,

$$(j - 1) + (k - j) + 1 = k,$$

so it takes Shawna $k$ moves to complete the tower, which is what we wanted to deduce. This completes the proof of the inductive step.

By Strong Induction, it will take Shawna $n - 1$ moves to complete a tower that contains $n$ blocks of lego, for every $n \geq 1$. □

This is pretty amazing. If we tried to go through the full argument for how many moves it takes her to build a tower with four blocks, it would go something like this. First, to build a tower with one block clearly takes 0 moves; to build a tower with two blocks clearly takes 1

move (stick the two blocks together). To build a tower with three blocks, we must use 1 move to stick together a tower of two blocks (which took 1 move to create) with a tower of one block (which took 0 moves to create), meaning that we use 2 moves altogether. Now, a tower of four blocks can be built in two ways: by using 1 move to stick together two towers of two blocks, each of which took 1 move to make, for a total of 3 moves; or by using 1 move to stick together a tower of one block (which took 0 moves to make) with a tower of three blocks (which took 2 moves to make), for a total of 3 moves. So under either method, building a tower of four blocks takes 3 moves. You can see that the argument will get more and more complicated as $n$ increases, but it will always continue to work.

We won't need strong induction as such very much until later in the course, but the idea is useful background for the next kind of induction we'll look at, which is very important when dealing with recurrence relations: induction with multiple base cases.

Induction with multiple base cases is very important for dealing with recursively-defined sequences such as the Fibonacci sequence, where each term depends on more than one of the preceding terms.

Suppose you were asked to prove that the $n$th term of the Fibonacci sequence, $f_n$, is at least $2^{n-2}$. If we try to follow our basic inductive strategy, we'd begin by observing that this is true for $f_0$:

$$f_0 = 1 \geq 2^{-2} = 1/4.$$

Then we'd make the inductive hypothesis that our inequality is true for some arbitrary $k \geq 0$, so $f_k \geq 2^{k-2}$. Now to deduce the inequality for $n = k + 1$, the natural approach is to use the recursive relation, which tells us that

$$f_{k+1} = f_k + f_{k-1}.$$

We can use our inductive hypothesis to make a substitution for $f_k$, but what about $f_{k-1}$? You might (reasonably) argue at this point that we should use strong induction, which will allow us to assume that the result is true for both $f_k$ and $f_{k-1}$, but actually, this doesn't work! Why not? Well, the trouble is that everything we know about the Fibonacci sequence starts with $f_0$, but if $k = 1$ (which is the first time we try to use induction) then $f_{k-1} = f_{-1}$, which we haven't even defined! It is very important to ensure that in the inductive step, we never make our assumption go back *too far*, i.e. to a value below $n_0$.

So, how can we deal with this problem? The solution is to add another base case, for $n = 1$. When $n = 1$, we have

$$f_1 = 1 \geq 2^{1-2} = 1/2.$$

Now if we try induction, at the first step we will be using the fact that the statement is true for $f_0$ and $f_1$ to prove it for $f_2$; then the fact that it's true for $f_1$ and $f_2$ will allow us to deduce it for $f_3$, and so on. The final argument will look like the following.

**EXAMPLE 6.3.4.** Prove by induction that the $n$th term of the Fibonacci sequence, $f_n$, is at least $(3/2)^{n-1}$, for every $n \geq 0$.

**SOLUTION.** Since the recursive relation for the Fibonacci sequence requires the two immediately preceding terms, we will require two base cases.

**Proof.** *Base cases:* When $n = 0$, we have

$$f_0 = 1 \geq (3/2)^{-1} = 2/3,$$

so the inequality holds for $n = 0$. When $n = 1$, we have

$$f_1 = 1 \geq (3/2)^{1-1} = 1,$$

so the inequality holds for $n = 1$. This completes the proof of the base cases.

*Inductive step:* We begin with the (strong) inductive hypothesis. Let $k$ be an arbitrary integer at least as big as our biggest base case, so $k \geq 1$. Assume that for every integer $i$ with $0 \leq i \leq k$, we have $f_i \geq (3/2)^{i-1}$.

Now we want to deduce that

$$f_{k+1} \geq (3/2)^{(k+1)-1} = (3/2)^k.$$

Using the recursive relation, we know that $f_{k+1} = f_k + f_{k-1}$. Since $k \geq 1$, we have $k - 1 \geq 0$, so both $k$ and $k - 1$ satisfy the bounds on $i$ (that $0 \leq i \leq k$), so that we can apply our inductive hypothesis to both $f_k$ and $f_{k-1}$. We therefore have

$$f_{k+1} \geq \left(\frac{3}{2}\right)^{k-1} + \left(\frac{3}{2}\right)^{k-2} = \left(\frac{3}{2} + 1\right)\left(\frac{3}{2}\right)^{k-2} = \frac{5}{2}\left(\frac{3}{2}\right)^{k-2} = \frac{5}{3} \cdot \frac{3}{2}\left(\frac{3}{2}\right)^{k-2} > \left(\frac{3}{2}\right)^{k},$$

since $5/3 > 3/2$. This is what we wanted to deduce. This completes the proof of the inductive step.

By the Principle of Mathematical Induction, $f_n \geq (3/2)^{n-1}$ for every $n \geq 0$. □

**EXERCISES 6.3.5.**

1) Prove by induction that for every $n \geq 0$, the $n$th term of the Fibonacci sequence is no greater than $2^n$.
2) The machine at the coffee shop isn't working properly, and can only put increments of \$4 or \$5 on your gift card. Prove by induction that you can get any amount of dollars that is at least \$12. [*Hint:* You should have four base cases.]
3) Define a recurrence relation by $a_0 = a_1 = a_2 = 1$, and $a_n = a_{n-1} + a_{n-2} + a_{n-3}$ for $n \geq 3$. Prove by induction that $a_n \leq 2^n$ for all $n \geq 0$.

**SUMMARY:**

- Important definitions:
  - recursively-defined sequence
  - initial conditions
  - recursive relation
  - Fibonacci sequence
  - proof by induction
  - base case
  - inductive step
  - inductive hypothesis
  - strong induction
  - induction with multiple base cases
- Notation:
  - (IH)

# Chapter 7

# Generating Functions

Recall that the basic goal with a recursively-defined sequence, is to find an explicit formula for the $n$th term of the sequence. Generating functions will allow us to do this.

## 7.1. What is a generating function?

A generating function is a formal structure that is closely related to a numerical sequence, but allows us to manipulate the sequence as a single entity, with the goal of understanding it better. Here's the formal definition.

**DEFINITION 7.1.1.** For a sequence $a_0, a_1, \ldots, a_n, \ldots$ the corresponding **generating function** $f(x)$ is the series

$$f(x) = a_0 + a_1 x + \ldots + a_n x^n + \ldots = \sum_{i=0}^{\infty} a_i x^i.$$

So $a_n$, the $n$th term of the sequence, is the coefficient of $x^n$ in $f(x)$.

**EXAMPLE 7.1.2.** Here are a number of basic examples.

1) $1, 1, 1, 1, 1, 1, 0, 0, 0, \ldots$ has generating function

$$1 + x + x^2 + x^3 + x^4 + x^5.$$

2) $1, 4, 6, 4, 1, 0, 0, 0, \ldots$ has generating function

$$1 + 4x + 6x^2 + 4x^3 + x^4 = (1 + x)^4.$$

3) $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}, 0, 0, 0, \ldots$ has generating function

$$\binom{n}{0} + \binom{n}{1} x + \ldots + \binom{n}{n} x^n = (1 + x)^n.$$

4) $1, 1, 1, 1, \ldots$ has generating function

$$f(x) = 1 + x + x^2 + x^3 + \ldots = \sum_{i=0}^{\infty} x^i.$$

These generating functions can be manipulated. For example, if $f(x)$ is as in Example 7.1.2(4), suppose we take the product $(1-x)f(x)$. We have

$$\begin{aligned}(1-x)f(x) &= (1-x)(1+x+x^2+x^3+x^4+\ldots)\\ &= (1+x+x^2+x^3+x^4+\ldots)-(x+x^2+x^3+x^4+x^5+\ldots)\\ &= 1\end{aligned}$$

Dividing through by $1-x$, we see that $f(x) = 1/(1-x)$.

This may seem artificial and rather nonsensical since the generating function was defined as a formal object whose coefficients are a sequence that interests us. In fact, although we won't delve into the formalities in this course, algebraic manipulation of generating functions can be formally defined, and gives us exactly these results.

A reasonable question at this point might be, what use is this? Even if we agree that $f(x) = 1/(1-x)$, what we really want is the coefficient of $x^n$ (in order to retrieve $a_n$, the $n$th term of our sequence). If we have an expression like $1/(1-x)$, how can we work out the coefficient of $x^n$? We'll explore answers to this through the next few sections.

**EXERCISES 7.1.3.** For each of the following sequences, give the corresponding generating function.

1) $1, 3, 5, 0, 0, 0, \ldots$.

2) $1, 2, 2^2, 2^3, 2^4, \ldots$.

3) $1, 5, 10, 15, 10, 5, 1, 0, 0, 0, \ldots$.

4) $1, 5, 10, 10, 5, 1, 0, 0, 0, \ldots$.

## 7.2. The Generalised Binomial Theorem

We are going to present a generalised version of the special case of Theorem 3.3.2, the Binomial Theorem. In this generalisation, we will allow the exponent to be negative. Recall that the Binomial Theorem states that

$$(1+x)^n = \sum_{r=0}^{n} \binom{n}{r} x^r.$$

If we have $f(x)$ as in Example 7.1.2.4, we've seen that

$$f(x) = 1/(1-x) = (1-x)^{-1}.$$

So if we were allowed negative exponents in the Binomial Theorem, then a change of variable $y = -x$ would allow us to calculate the coefficient of $x^n$ in $f(x)$.

Of course, if $n$ is negative in the Binomial Theorem, we can't figure out anything unless we have a definition for what $\binom{n}{r}$ means under these circumstances.

**DEFINITION 7.2.1.** The **generalised binomial coefficient**,

$$\binom{n}{r} = \frac{n(n-1)\ldots(n-r+1)}{r!}$$

where $r \geq 0$ but $n$ can be any real number.

Notice that this coincides with the usual definition for the binomial coefficient when $n$ is a positive integer, since

$$n!/(n-r)! = n(n-1)\ldots(n-r+1)$$

in this case.

**EXAMPLE 7.2.2.**

$$\binom{-2}{5} = \frac{(-2)(-3)(-4)(-5)(-6)}{5!} = -6.$$

If $n$ is a positive integer, then we can come up with a nice formula for $\binom{-n}{r}$.

**PROPOSITION 7.2.3.** *If $n$ is a positive integer, then*

$$\binom{-n}{r} = (-1)^r \binom{n+r-1}{r}.$$

**PROOF.** We have

$$\binom{-n}{r} = \frac{-n(-n-1)\dots(-n-r+1)}{r!}.$$

Taking a factor of $(-1)$ out of each term on the right-hand side gives

$$(-1)^r n(n+1)\dots(n+r-1)/(r!).$$

Now,

$$(n+r-1)(n+r-2)\dots n = (n+r-1)!/(n-1)!,$$

so

$$(-1)^r \frac{n(n+1)\dots(n+r-1)}{r!} = (-1)^r \frac{(n+r-1)!}{r!(n-1)!} = (-1)^r \binom{n+r-1}{r},$$

as claimed. □

With this definition, the binomial theorem generalises just as we would wish. We won't prove this.

**THEOREM 7.2.4 (Generalised Binomial Theorem).** *For any $n \in \mathbb{R}$,*

$$(1+x)^n = \sum_{r=0}^{\infty} \binom{n}{r} x^r.$$

**EXAMPLE 7.2.5.** Let's check that this gives us the correct values for the coefficients of $f(x)$ in Example 7.1.2.4, which we already know.

**SOLUTION.** We have

$$f(x) = (1-x)^{-1} = (1+y)^{-1},$$

where $y = -x$. The Generalised Binomial Theorem tells us that the coefficient of $y^r$ will be

$$\binom{-1}{r} = (-1)^r \binom{1+r-1}{r} = (-1)^r,$$

since $\binom{r}{r} = 1$. But we want the coefficient of $x^r$, not of $y^r$, and

$$y^r = (-x)^r = (-1)^r x^r,$$

so we have

$$(-1)^r y^r = (-1)^{2r} x^r = 1^r x^r = x^r.$$

Thus, the coefficient of $x^r$ in $f(x)$ is 1. This is, indeed, precisely the sequence we started with in Example 7.1.2.4. □

**EXAMPLE 7.2.6.** Let's work out $(1+x)^{-3}$.

**SOLUTION.** We need to know what $\binom{-3}{r}$ gives, for various values of $r$. By Proposition 7.2.3, we have

$$\binom{-3}{r} = (-1)^r \binom{3+r-1}{r} = (-1)^r \binom{r+2}{r} = (-1)^r \frac{(r+2)(r+1)}{2}.$$

When $r=0$, this is $(-1)^0 2 \cdot 1/2 = 1$. When $r=1$, this is $(-1)^1 3 \cdot 2/2 = -3$. When $r=2$, this is $(-1)^2 4 \cdot 3/2 = 6$. In general, we see that

$$(1+x)^{-3} = 1 - 3x + 6x^2 - + \ldots + (-1)^n \frac{(n+2)(n+1)}{2} x^n + \ldots$$ □

**EXERCISES 7.2.7.** Calculate the following.

1) $\binom{-5}{7}$

2) The coefficient of $x^4$ in $(1-x)^{-2}$.

3) The coefficient of $x^n$ in $(1+x)^{-4}$.

4) The coefficient of $x^{k-1}$ in

$$\frac{1+x}{(1-2x)^5}.$$

[*Hint:* Notice that $\frac{1+x}{(1-2x)^5} = (1-2x)^{-5} + x(1-2x)^{-5}$. Work out the coefficient of $x^n$ in $(1-2x)^{-5}$ and in $x(1-2x)^{-5}$, substitute $n = k-1$, and add the two coefficients.]

5) The coefficient of $x^k$ in $1/(1-x^j)^n$, where $j$ and $n$ are fixed positive integers.
[*Hint:* Think about what conditions will make this coefficient zero.]

## 7.3. Using generating functions to count things

As you might expect of something that has come up in our study of enumeration, generating functions can be useful in solving problems about counting. We've already seen from the Binomial Theorem, that the coefficient of $x^r$ in $(1+x)^n$ is $\binom{n}{r}$, so the generating function for the binomial coefficients is $(1+x)^n$. In fact, the argument we used to prove the Binomial Theorem explained why this works: if we want the coefficient of $x^r$ in $(1+x)^n$, it must be the number of ways of choosing the $x$ from $r$ of the $n$ factors, while choosing the 1 from the other factors. We can use similar reasoning to solve other counting questions.

**EXAMPLE 7.3.1.** The grocery store sells paper plates in packages of 1, 5, 20, or 75. In how many different ways can Jiping buy a total of 95 paper plates?

**SOLUTION.** We model this with generating functions. The exponent of $x$ will represent the number of paper plates, and the coefficient of $x^n$ will represent the number of ways in which he can buy $n$ paper plates.

We begin by considering the single paper plates that Jiping buys. He could buy 0, or 1, or any other number of these, so we represent this by the generating function

$$1 + x + x^2 + x^3 + x^4 + \ldots = \sum_{i=0}^{\infty} x^i = 1/(1-x).$$

There is exactly one way of choosing any particular number of single paper plates (we are assuming the plates are indistinguishable).

Now, Jiping could also buy any number of packages of 5 paper plates, but the difference is that each package he buys contributes 5 to the exponent, since it represents 5 plates. We represent this by the generating function

$$1 + x^5 + x^{10} + x^{15} + \ldots = \sum_{i=0}^{\infty} x^{5i} = 1/(1-x^5),$$

where $i$ represents the number of packages he buys.

Similarly, Jiping could buy any number of packages of 20 paper plates, and each package he buys contributes 20 to the exponent, since it represents 20 plates. We represent this by the generating function

$$1 + x^{20} + x^{40} + x^{60} + \ldots = \sum_{i=0}^{\infty} x^{20i} = 1/(1-x^{20}),$$

where $i$ represents the number of packages he buys.

Finally, Jiping could buy any number of packages of 75 paper plates, and each package he buys contributes 75 to the exponent, since it represents 75 plates. We represent this by the generating function

$$1 + x^{75} + x^{150} + x^{225} + \ldots = \sum_{i=0}^{\infty} x^{75i} = 1/(1-x^{75}),$$

where $i$ represents the number of packages he buys.

Obviously, for this particular question, Jiping can't actually buy 2 or more of the packages of 75 paper plates, since that would be too many. There are also limits on the number of packages of other sizes that he should buy, since he doesn't want to end up with more than 95 plates. So for this problem, we can assume that the generating function for the full problem actually looks like this:

$$(1 + x + x^2 + \ldots + x^{95})(1 + x^5 + x^{10} + \ldots + x^{95})(1 + x^{20} + x^{40} + x^{60} + x^{80})(1 + x^{75})$$

and we are looking for the coefficient of $x^{95}$.

We could multiply this all out to get our answer. We could be a bit more clever, recognising that we only really care about the coefficient of $x^{95}$, and break the problem down into cases depending on how many of the bigger packages Jiping buys. It should be noted that the generating function hasn't really saved us any work. This approach involves saying, "Well, if Jiping takes the $x^{75}$ from the final factor, then there are only six ways to contribute to the coefficient of $x^{95}$: he could choose an $x^{20}$ from the previous factor and 1s from both of the other factors; or he could choose 1 from the third factor and any of 1, $x^5$, $x^{10}$, $x^{15}$, or $x^{20}$ from the second factor, in each case choosing whichever term from the first factor is needed to bring the

exponent up to 95." This is exactly equivalent to saying, "Well, if Jiping buys a package of 75 plates, then there are only six ways to buy 95 plates in total: he could buy a package of 20 plates and be done; or he could buy 0, 1, 2, 3, or 4 packages of 5 plates, in each case buying as many single plates as are needed to bring the total up to 95."

So what's the advantage of the generating function approach? It comes in a couple of ways. First, it solves multiple problems at once: if we actually multiply out the generating function above, we will be able to read off not only how many ways there are of buying 95 plates, but also how many ways there are of buying every number of plates up to 95. (If we hadn't cut the factors off as we did, we could also work out the answers for any number of plates higher than 95.) So by doing a bunch of multiplication once (and it's easy to feed into a computer algebra system if you don't want to do it by hand), we can simultaneously find out the answer to a lot of closely-related questions.

The other advantage is that the generating function approach can help us solve problems that we don't see how to solve without it, such as finding an explicit formula for the $n$th term of a recursively-defined sequence. □

Here's an example that involves working out the coefficient of a term in a generating function in two different ways.

**EXAMPLE 7.3.2.** Consider the generating function $(1/(1-x))^4 = (1+x+x^2+x^3+\ldots)^4$. As usual, we want to determine the coefficient of $x^r$ in this product.

**SOLUTION.** We must choose a power of $x$ from each of the four factors, in such a way that the sum of the powers we choose must be $n$. This is the same as choosing a total of $r$ items, when the items come in four distinct types (recall for example, Example 5.1.2). The types are represented by the factor the term is chosen from, and the exponent chosen from that factor is the number of items ($x$es) chosen of that type. So we know that the number of ways of doing this is $\left(\binom{4}{r}\right)$.

We have another way of working this out. Our generating function is $(1-x)^{-4}$, and the Generalised Binomial Theorem tells us that the coefficient of $(-x)^r$ in this is $\binom{-4}{r}$, so the coefficient of $x^r$ is

$$(-1)^r(-1)^r\binom{r+3}{r} = \left(\binom{4}{r}\right).$$

□

We'll use the above example to work out a counting question, but first we need an observation.

**PROPOSITION 7.3.3.** *For any positive integer $k$,*

$$1+x+x^2+\cdots+x^k = \frac{1-x^{k+1}}{1-x}.$$

You can prove this by induction on $k$ (this is one of the exercises below), or by multiplying through by $1-x$.

**EXAMPLE 7.3.4.** Trent is playing a dice game, using 12-sided dice. How many ways are there for him to roll a total of 24 on his four dice?

**SOLUTION.** Each die can roll any number between 1 and 12, and there are four dice, so the appropriate generating function is

$$(x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7 + x^8 + x^9 + x^{10} + x^{11} + x^{12})^4.$$

Rolling an $i$ on one of the dice corresponds to choosing $x^i$ from the corresponding factor of this generating function. We are looking for the coefficient of $x^{24}$, this will tell us the number of ways of rolling a total of 24.

It turns out that by manipulating the generating function, we can work this out a bit more easily than by multiplying this out. By taking a common factor of $x$ out of each of the four factors, our generating function can be re-written as

$$x^4(1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7 + x^8 + x^9 + x^{10} + x^{11})^4,$$

and the coefficient of $x^{24}$ in this, will be the same as the coefficient of $x^{20}$ in

$$(1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7 + x^8 + x^9 + x^{10} + x^{11})^4.$$

Using Proposition 7.3.3, we see that this expression can be rewritten as

$$\left(\frac{1 - x^{12}}{1 - x}\right)^4.$$

Using the Binomial Theorem and substituting $y = -x^{12}$, we see that

$$\begin{aligned}(1 - x^{12})^4 &= \binom{4}{0}(-x^{12})^0 + \binom{4}{1}(-x^{12})^1 + \binom{4}{2}(-x^{12})^2 + \binom{4}{3}(-x^{12})^3 + \binom{4}{4}(-x^{12})^4 \\ &= 1 - 4x^{12} + 6x^{24} - 4x^{36} + x^{48}.\end{aligned}$$

Most of these terms can be ignored, as they will not contribute to the coefficient of $x^{20}$. Recall that the function we're interested in is the product of this, with $(1 - x)^{-4}$, and there are only two ways of getting an $x^{20}$ term from this product: by taking the constant term that we've just worked out, and multiplying it by the $x^{20}$ term from $(1 - x)^{-4}$; or by taking the $x^{12}$ term that we've just worked out, and multiplying it by the $x^8$ term from $(1 - x)^{-4}$. In the previous example, we worked out that in $(1 - x)^{-4}$, the coefficient of $x^{20}$ is $\left(\binom{4}{20}\right)$, and the coefficient of $x^8$ is $\left(\binom{4}{8}\right)$.

Thus, the number of ways in which Trent can roll a total of 24 on his four dice is the coefficient of $x^{24}$ in our generating function, which is

$$\left(\binom{4}{20}\right) - 4\left(\binom{4}{8}\right) = 1771 - 660 = 1111. \qquad \square$$

**EXERCISES 7.3.5.**

1) Prove Proposition 7.3.3 by induction on $k$.
2) Find the number of ways Trent can roll a total of 16 on his four dice.
3) If Trent's four dice are 10-sided dice instead of 12-sided, how many ways can he roll a total of 24?
4) If Trent rolls five regular (6-sided) dice, how many ways can he roll a total of 11? What is the probability that he will roll a total of 11?

**SUMMARY:**

- If $n > 0$ is an integer, then
$$\binom{-n}{r} = (-1)^r \binom{n+r-1}{r}.$$
- The Generalised Binomial Theorem
- $1 + x + \ldots + x^k = (1 - x^{k+1})/(1 - x)$
- using generating functions for counting things
- Important definitions:
  - generating function for a sequence
  - generalised binomial coefficients

Chapter 8

# Generating Functions and Recursion

We've seen how the Generalised Binomial Theorem can be used to extract coefficients from a certain sort of generating function. Before we proceed with learning how to use generating functions to find explicit formulas for the $n$th term of a recursively-defined sequence, we need to know how to extract coefficients from some more complicated expressions.

## 8.1. Partial fractions

If a generating function looks like $1/(1+ax^i)^j$, we can use the Generalised Binomial Theorem to find the coefficient of $x^r$. But what can we do if the generating function looks like $1/(a+bx+cx^2)$, for example, or some even more complicated expression?

One tool that can help us extract coefficients from some expressions like this, is the method of partial fractions.

**EXAMPLE 8.1.1.** Suppose we have a generating function

$$f(x) = \frac{1+x}{(1-2x)(2+x)}.$$

How can we work out the coefficient of $x^r$?

**SOLUTION.** Well, if we could separate the factors of the denominator, we would know how to deal with each separately. In fact, this is exactly what we do. We set

$$f(x) = \frac{1+x}{(1-2x)(2+x)} = \frac{A}{1-2x} + \frac{B}{2+x}.$$

As we work this through, you'll see that in working this out, we end up with two equations in the two unknowns $A$ and $B$, which we can therefore solve! So it is possible to "split up" the original generating function, into two separate fractions, each of which has as its denominator one of the factors of the original denominator. This is the method of **partial fractions**.

To solve for $A$ and $B$, we add the fractions $A/(1-2x)$ and $B/(2+x)$ over a common denominator. This gives

$$\frac{A(2+x)+B(1-2x)}{(1-2x)(2+x)} = f(x) = \frac{1+x}{(1-2x)(2+x)}.$$

Clearly, this forces the numerators to be equal, so

$$A(2+x)+B(1-2x) = 1+x.$$

Since these are equal as polynomials in $x$, the constant terms must be equal, and the coefficients of $x$ must be equal, giving us two equations: $2A+B=1$ and $(A-2B)x = x$, so $A-2B=1$. Now there are many ways to algebraically solve for $A$ and $B$; for example, the first equation gives $B = 1-2A$; plugging this into the last equation gives $A-2(1-2A)=1$, so $5A=3$, so $A=3/5$. Now

$$B = 1-2(3/5) = -1/5.$$

Thus, we have

$$f(x) = \frac{3/5}{1-2x} - \frac{1/5}{2+x}.$$

Notice that the $2+x$ is still a bit problematic. We can use the Generalised Binomial Theorem to work out coefficients for something that looks like $(1+ax^i)^j$, but we need that 1, and here instead we have a 2. To deal with this, we observe that

$$2+x = 2(1+(1/2)x).$$

Thus,

$$f(x) = \frac{3/5}{1-2x} - \frac{1/10}{1+(1/2)x}.$$

Now let's expand each of the two summands separately. We have

$$\frac{3}{5}(1-2x)^{-1} = \frac{3}{5}(1+2x+(2x)^2+(2x)^3+\ldots),$$

so the coefficient of $x^r$ in this part is $(3/5)2^r$. Also,

$$\frac{-1}{10}(1+(1/2)x)^{-1} = \frac{-1}{10}(1-\frac{1}{2}x+(\frac{1}{2}x)^2-(\frac{1}{2}x)^3+-\ldots),$$

so the coefficient of $x^r$ in this part is $(-1/10)(-1)^r(1/2)^r$.

Thus, the coefficient of $x^r$ in $f(x)$ is $(3/5)2^r - 1/10(-1/2)^r$. □

The method of partial fractions can be applied to any generating function that has a denominator that can be factored into simpler terms. However, polynomials of degree 3 or higher can become hard to factor, so we'll mostly restrict our attention to applying this either with denominators that are already factored, or with denominators that have degree at most two.

There is an extra trick that you should be aware of. This arises if the denominator is divisible by a square. For example, if we are looking for the coefficient of $x^r$ in

$$g(x) = \frac{1+x}{(1-2x)^2(2+x)}$$

then it doesn't make sense to separate all of the factors out as before, because

$$g(x) = \frac{A}{1-2x} + \frac{B}{1-2x} + \frac{C}{2+x} = \frac{A+B}{1-2x} + \frac{C}{2+x}$$

and when we add this up, the denominator will be $(1-2x)(2+x)$ rather than $(1-2x)^2(2+x)$. This can be dealt with in either of two ways. First, you can include both $1-2x$ and $(1-2x)^2$ as denominators:

$$g(x) = \frac{A}{1-2x} + \frac{B}{(1-2x)^2} + \frac{C}{2+x}.$$

The second option is to include only $(1-2x)^2$ as one of the denominators, but to include an $x$ in the corresponding numerator, in addition to the constant term:

$$g(x) = \frac{Ax+B}{(1-2x)^2} + \frac{C}{2+x}.$$

Either of these methods can be generalised in natural ways to cases where the denominator is divisible by some higher power.

We'll see more examples of partial fractions applied to specific situations, so we'll leave the explanation there for now.

**EXERCISES 8.1.2.** Find the coefficient of $x^r$ in each of the following generating functions, using the method of partial fractions and the Generalised Binomial Theorem.

1) $\dfrac{1}{(1+2x)(2-x)}$

2) $\dfrac{x}{(1+x)^2(1-x)}$

3) $\dfrac{1+2x}{(1-2x)(2+x)(1+x)}$

## 8.2. Factoring polynomials

You should be familiar with the quadratic formula, which allows us to factor any polynomial of degree two, into linear factors. Specifically, it tells us that the roots of $ax^2+bx+c$ are

$$\frac{-b \pm \sqrt{b^2-4ac}}{2a}.$$

Notice that this does *not* tell us immediately how to factor $ax^2+bx+c$, because it's missing a constant factor of $a$. So if we want to factor $ax^2+bx+c$, we actually get

$$ax^2+bx+c = a\left(x - \left(\frac{-b+\sqrt{b^2-4ac}}{2a}\right)\right)\left(x - \left(\frac{-b-\sqrt{b^2-4ac}}{2a}\right)\right).$$

Recall that in order to use the Generalised Binomial Theorem, we need the constant term to be 1. If you are very comfortable with algebraic manipulations, you can use the quadratic formula to factor as above, and then divide each factor by the appropriate value so as to make the constant term 1. This may create a messy constant outside the whole thing, and a messy coefficient of $x$ in each term, but if you are careful, you can get the correct answer this way.

If you are more confident in memorising another formula (closely related to the quadratic formula) for factoring $ax^2+bx+c$, you can also factor a quadratic polynomial directly into the

form we want, using the following formula:

$$ax^2 + bx + c = c\left(1 - \frac{-b + \sqrt{b^2 - 4ac}}{2c}x\right)\left(1 - \frac{-b - \sqrt{b^2 - 4ac}}{2c}x\right).$$

Sometimes a denominator will already be factored in the formula for a generating function, but when it isn't, either of the above methods can be used to factor it.

It may sometimes be the case that $b^2 - 4ac < 0$. In this case $\sqrt{b^2 - 4ac}$ will be a complex number. We will be assuming in some of our examples and exercises that you have some familiarity with complex numbers (enough to be able to use the basic arithmetic operations of addition, subtraction, multiplication, and division on them). If you are not comfortable with this, you may want to review Appendix A or some other background material on complex numbers.

**EXAMPLE 8.2.1.** Factor $3x^2 - 2x + 1$ into linear factors.

**SOLUTION.** We will use the formula given above. We have $a = 3$, $b = -2$, and $c = 1$. Then

$$\begin{aligned} 3x^2 - 2x + 1 &= \left(1 - \frac{2 + \sqrt{4 - 12}}{2}x\right)\left(1 - \frac{2 - \sqrt{4 - 12}}{2}x\right) \\ &= (1 - (1 + i\sqrt{2})x)(1 - (1 - i\sqrt{2})x). \end{aligned}$$

□

It is always a good idea to check your result, by multiplying the factors back out.

When coefficients in the factorisation aren't integers (especially when they are irrational or even complex, as in the example above), you might find the algebra involved in working out the coefficients hard to deal with. Let's work through an example of this, using the factorisation we've just completed.

**EXAMPLE 8.2.2.** Find the coefficient of $x^r$ in $f(x)$, where

$$f(x) = \frac{1}{3x^2 - 2x + 1}.$$

**SOLUTION.** We have determined in the previous example, that

$$3x^2 - 2x + 1 = (1 - (1 + i\sqrt{2})x)(1 - (1 - i\sqrt{2})x),$$

so we need to solve for $A$ and $B$, where

$$\begin{aligned} f(x) &= \frac{1}{3x^2 - 2x + 1} \\ &= \frac{A}{1 - (1 + i\sqrt{2})x} + \frac{B}{1 - (1 - i\sqrt{2})x} \\ &= \frac{A(1 - (1 - i\sqrt{2})x) + B(1 - (1 + i\sqrt{2})x)}{3x^2 - 2x + 1}. \end{aligned}$$

Thus,

$$A(1 - (1 - i\sqrt{2})x) + B(1 - (1 + i\sqrt{2})x) = 1 + 0x,$$

so the constant term gives $A+B = 1$, while the coefficient of $x$ gives $A(1-i\sqrt{2})+B(1+i\sqrt{2}) = 0$. Substituting $B = 1 - A$ into the latter equation, gives

$$A - i\sqrt{2}A + 1 + i\sqrt{2} - A - i\sqrt{2}A = 0,$$

so $1 + i\sqrt{2} = i2\sqrt{2}A$. Hence

$$A = \frac{1 + i\sqrt{2}}{i2\sqrt{2}} = \frac{1}{2\sqrt{2}i} + \frac{1}{2}.$$

We make the denominator of the first fraction rational, by multiplying numerator and denominator by $\sqrt{2}i$, giving

$$A = -\frac{\sqrt{2}i}{4} + \frac{1}{2}.$$

Now since $B = 1 - A$, we have

$$B = \frac{1}{2} + \frac{\sqrt{2}i}{4}.$$

To make things a bit simpler, we'll rewrite $A$ as $(2 - \sqrt{2}i)/4$, and $B = (2 + \sqrt{2}i)/4$.

Thus we have

$$f(x) = \frac{(2 - \sqrt{2}i)/4}{1 - (1 + i\sqrt{2})x} + \frac{(2 + \sqrt{2}i)/4}{1 - (1 - i\sqrt{2})x}.$$

Using the Generalised Binomial Theorem and $y = (1 + i\sqrt{2})x$, we see that the first fraction expands as

$$[(2 - \sqrt{2}i)/4](1 + y + y^2 + y^3 + \ldots),$$

and the coefficient of $x^r$ in this, will be $[(2 - \sqrt{2}i)/4](1 + i\sqrt{2})^r$. Similarly, with $y = (1 - i\sqrt{2})x$, the second fraction expands as

$$[(2 + \sqrt{2}i)/4](1 + y + y^2 + y^3 + \ldots),$$

and the coefficient of $x^r$ in this, will be $[(2 + \sqrt{2}i)/4](1 - i\sqrt{2})^r$.

So the coefficient of $x^r$ in $f(x)$ is

$$[(2 - \sqrt{2}i)/4](1 + i\sqrt{2})^r + [(2 + \sqrt{2}i)/4](1 - i\sqrt{2})^r. \qquad \square$$

You can see from this example that the algebra can get ugly, but the process of finding the coefficient of $x^r$ is nonetheless straightforward.

**EXERCISES 8.2.3.** For each of the generating functions given, factor the denominator and use the method of partial fractions to determine the coefficient of $x^r$.

1) $\dfrac{x}{x^2 + 5x - 1}$

2) $\dfrac{2 + x}{2x^2 + x - 1}$

3) $\dfrac{x}{x^2 - 3x + 1}$

### 8.3. Using generating functions to solve recursively-defined sequences

At last we are ready to apply the mechanics we've introduced in this chapter, to find an explicit formula for the $n$th term of a recursively-defined sequence.

This method is probably most easily understood using examples.

**EXAMPLE 8.3.1.** Consider the recursively-defined sequence: $a_0 = 2$, and for every $n \geq 1$, $a_n = 3a_{n-1} - 1$. Find an explicit formula for $a_n$ in terms of $n$.

**SOLUTION.** The generating function for this sequence is $a(x) = \sum_{i=0}^{\infty} a_i x^i$.

Now, we are going to use the recursive relation. We know that $a_n = 3a_{n-1} - 1$, or, by rearranging this, $a_n - 3a_{n-1} = -1$. Thus, if we could get the coefficient of $x^n$ to look like $a_n - 3a_{n-1}$, we could use the recursive relation to replace this by $-1$. Right now, the coefficient of $x^n$ is $a_n$, and the only place $a_{n-1}$ shows up is as the coefficient of $x^{n-1}$. But if we multiply $a(x)$ by $x$, then $a_{n-1}x^{n-1}$ becomes $a_{n-1}x^n$, so if we also multiply by $-3$, we get a coefficient of $-3a_{n-1}$ for $x^n$ in $-3xa(x)$. Now add this to $a(x)$. This may be easier to see as written below:

$$\begin{array}{rcccccccc} a(x) & = & a_0 & +a_1x & +a_2x^2 & +\ldots & +a_mx^m & +\ldots \\ -3xa(x) & = & & -3a_0x & -3a_1x^2 & -\ldots & -3a_{m-1}x^m & -\ldots \\ \hline (1-3x)a(x) & = & a_0 & -x & -x^2 & -\ldots & -x^m & -\ldots \end{array}$$

We see that this gives

$$(1-3x)a(x) = 3 - (1 + x + x^2 + x^3 + \ldots),$$

and we know that

$$1 + x + x^2 + x^3 + \ldots = 1/(1-x),$$

so $(1-3x)a(x) = 3 - 1/(1-x)$. Dividing through by $1-3x$ gives

$$a(x) = \frac{3}{1-3x} - \frac{1}{(1-x)(1-3x)}.$$

Now it's time to apply what we learned in the preceding sections of this chapter. The denominator is already factored, so we can immediately apply the method of partial fractions to the second fraction. If

$$\frac{-1}{(1-x)(1-3x)} = \frac{A}{1-x} + \frac{B}{1-3x} = \frac{A(1-3x) + B(1-x)}{(1-3x)(1-x)},$$

then $A + B = -1$ and $-3A - B = 0$, so $B = -3A$, which gives $-2A = -1$, so $A = 1/2$ and $B = -3/2$. Thus,

$$a(x) = \frac{3}{1-3x} + \frac{1/2}{1-x} - \frac{3/2}{1-3x} = \frac{1/2}{1-x} + \frac{3/2}{1-3x}.$$

The coefficient of $x^n$ in the first of these terms is $1/2$, while in the second term, the coefficient of $x^n$ is $(3/2)3^n$. Thus, $a_n = 1/2 + (3/2)3^n$. Since our generating function began with $a_0x^0$, this formula applies for every $n \geq 0$.

When going through so much algebra, it's easy to make a mistake somewhere along the way, so it's wise to do some double-checking. For a recursively-defined sequence, if the formula you work out gives the correct answer for the first three or four terms of the sequence, then it's very likely that you've done the calculations correctly. Let's check the first three terms of this one. We know from our initial condition that $a_0 = 2$, and our new formula gives

$$a_0 = 1/2 + (3/2)3^0 = 1/2 + 3/2 = 2.$$

Using the recursive relation, we should have $a_1 = 3(2) - 1 = 5$, and our formula gives

$$a_1 = 1/2 + (3/2)3^1 = 1/2 = 9/2 = 5.$$

Finally, the recursive relation gives $a_2 = 3(5) - 1 = 14$, while our formula gives

$$a_2 = 1/2 + (3/2)3^2 = 1/2 + 27/2 = 14.$$

You can see the benefit to having an explicit formula if you were asked to work out $a_{100}$. Clearly, it's much easier to determine $1/2 + (3/2)3^{100}$ than to apply the recursive relation one hundred times. □

Let's look at one more example, where the recursive relation involves more than one previous term.

**EXAMPLE 8.3.2.** Consider the recursively-defined sequence: $b_0 = 1$, $b_1 = 0$, $b_2 = 1$, and for every $n \geq 3$, $b_n = b_{n-1} - 2b_{n-3}$. Find an explicit formula for $b_n$ in terms of $n$.

**SOLUTION.** The generating function for this sequence is

$$b(x) = \sum_{i=0}^{\infty} b_i x^i.$$

Again, we'll use the recursive relation, which we rearrange as

$$b_n - b_{n-1} + 2b_{n-3} = 0$$

for every $n \geq 3$. We want to end up with a polynomial in which the coefficient of $x^m$ looks like $b_m - b_{m-1} + 2b_{m-3}$, so that we'll be able to use the recursive relation to replace this by 0. In order to do this, we'll take $b(x)$ (to get the $b_m x^m$ piece), minus $xb(x)$ (to get the $-b_{m-1}x^m$ piece), plus $2x^3 b(x)$ (to get the $+2b_{m-3}x^m$ piece).

The result looks like:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $b(x)$ | $=$ | $b_0$ | $+b_1x$ | $+b_2x^2$ | $+b_3x^3$ | $+\ldots$ | $+b_mx^m$ | $+\ldots$ |
| $-xb(x)$ | $=$ | | $-b_0x$ | $-b_1x^2$ | $-b_2x^3$ | $-\ldots$ | $-b_{m-1}x^m$ | $-\ldots$ |
| $+2x^3b(x)$ | $=$ | | | | $+2b_0x^3$ | $+\ldots$ | $+2b_{m-3}x^m$ | $+\ldots$ |
| $(1-x+2x^3)b(x)$ | $=$ | $b_0$ | $+(b_1-b_0)x$ | $+(b_2-b_1)x^2$ | $+0x^3$ | $+\ldots$ | $+0x^m$ | $+\ldots$ |

We see that this gives

$$(1 - x + 2x^3)b(x) = 1 + (-1)x + 1x^2.$$

Dividing through by $1 - x + 2x^3$ gives

$$b(x) = \frac{1 - x + x^2}{1 - x + 2x^3}.$$

If we want to be able to do anything with this, we need to factor the denominator. Although we don't have a general method for factoring cubic polynomials, in this case it's not hard to see that $-1$ is a zero of the polynomial (because $1 - (-1) + 2(-1)^3 = 0$), and hence $x + 1$ is a factor of the polynomial. You will not be expected to factor cubic polynomials yourself in this course, so we won't review polynomial long division, but if you recall polynomial long division (or look it up in some other source), you can use it to determine that

$$1 - x + 2x^3 = (1 + x)(2x^2 - 2x + 1).$$

In any case, you can multiply the right-hand side out to verify that it is true.

Now it's time to use the factoring formula, with $a = 2$, $b = -2$, and $c = 1$, to factor $2x^2 - 2x + 1$. This gives

$$2x^2 - 2x + 1 = 1\left(1 - \frac{2+\sqrt{4-8}}{2}x\right)\left(1 - \frac{2-\sqrt{4-8}}{2}x\right) = (1-(1+i)x)\,(1-(1-i)x)\,.$$

Having factored

$$1 - x + 2x^3 = (1+x)(1-(1+i)x)(1-(1-i)x),$$

we now apply the method of partial fractions to split this up into three separate pieces.

If

$$\begin{aligned}\frac{1-x+x^2}{1-x+2x^3} &= \frac{A}{1+x} + \frac{B}{1-(1+i)x} + \frac{C}{1-(1-i)x} \\ &= \frac{A(2x^2-2x+1) + B(1+x)(1-(1-i)x) + C(1+x)(1-(1+i)x)}{1-x+2x^3},\end{aligned}$$

then this gives us three equations:

$$2A - B(1-i) - C(1+i) = 1$$

(from the coefficient of $x^2$);

$$-2A + B(1-(1-i)) + C(1-(1+i)) = -1$$

(from the coefficient of $x$); and $A + B + C = 1$ (from the constant term). The second of these simplifies to $-2A + iB - iC = -1$.

The algebra can be done in different ways and gets a bit ugly, but these three equations can be solved, resulting in $A = 3/5$, $B = (2-i)/10$, $C = (2+i)/10$.

Thus,

$$\frac{1-x+x^2}{1-x+2x^3} = \frac{3/5}{1+x} + \frac{(2-i)/10}{1-(1+i)x} + \frac{(2+i)/10}{1-(1-i)x}.$$

The coefficient of $x^n$ in the first of these terms is $(3/5)(-1)^n$; in the second, it is $((2-i)/10))(1+i)^n$, and in the third, it is $((2+i)/10)(1-i)^n$.

We conclude that for every $n \geq 0$, we have

$$b_n = \frac{3}{5}(-1)^n + \frac{2-i}{10}\,(1+i)^n + \frac{2+i}{10}\,(1-i)^n\,.$$

It is somewhat surprising that these formulas involving complex numbers will always work out (when $n$ is an integer) to be not only real numbers, but integers! Once again, we should check this formula for several values of $n$ to ensure we haven't made errors in our calculations along the way.

From the initial conditions, $b_0 = 1$. Our formula gives

$$b_0 = 3/5 + (2-i)/10 + (2+i)/10 = 1.$$

From the initial conditions, $b_1 = 0$. Our formula gives

$$b_1 = -3/5 + \frac{2-i}{10}(1+i) + \frac{2+i}{10}(1-i) = -3/5 + (3+i)/10 + (3-i)/10 = 0.$$

Finally, the initial conditions gave $b_2 = 1$, and our formula gives

$$b_2 = \frac{3}{5} + \frac{2-i}{10}(2i) + \frac{2+i}{10}(-2i) = \frac{3}{5} + \frac{2i+1}{5} - \frac{2i-1}{5} = 1.$$

We could continue, but this is sufficient verification to inspire reasonable confidence. □

We now have a general method that we can apply to solve normal linear recursive relations. Before describing this method, we remark that it is important to our method that our recursively-defined sequence begin with the 0th term, as in Examples 8.3.1 and 8.3.2 above where we started out knowing $a_0$ and $b_0$ respectively. If instead we had started with $a_1 = 5$ in Example 8.3.1 and had not been given $a_0$ then in order to use this method we would need to use the recursive relation to work out what $a_0$ should be. We would do this by taking what we do know: $a_1 = 5$ and the recursive relation $a_n = 3a_{n-1} - 1$ to solve for $a_0$. In this case, we see that $a_1 = 5 = 3a_0 - 1$ gives $3a_0 = 6$ so $a_0 = 2$. We would do this even if the recursive relation had only been defined for $n \geq 2$, since we would be using our calculations to *force* the recursive relation to hold for $n = 1$.

**Method**

1) Start with a recurrence relation of the form

$$h_n = a_1 h_{n-1} + a_2 h_{n-2} + \ldots + a_k h_{n-k} + f(n) \text{ for every } n \geq i,$$

for some $i \geq k$, where $f(n)$ is a function in $n$, and with initial terms $h_{i-k}, \ldots, h_{i-1}$.

If $i > k$ then use this information (the recurrence relation and the given initial terms) to find values for $h_0, \ldots, h_{i-k-1}$ that make our recurrence relation true for every $n \geq k$.

2) Rearrange the recurrence relation into the form

$$h_n - a_1 h_{n-1} - a_2 h_{n-2} - \ldots - a_k h_{n-k} = f(n),$$

which by our work in the previous step is now true for every $n \geq k$. Let

$$a(x) = 1 - a_1 x - a_2 x^2 - \ldots - a_k x^k.$$

3) Define the generating function

$$h(x) = h_0 + h_1 x + h_2 x^2 + \ldots.$$

4) Find a linear combination of the generating function so that the coefficient of $x^m$ is $f(m)$ for every $m$ greater than or equal to $k$ as follows:

$$\begin{array}{rcccccccc}
h(x) & = & h_0 & +h_1 x & +h_2 x^2 & +h_3 x^3 & +\ldots & +h_k x^k & +\ldots \\
-a_1 x h(x) & = & & -a_1 h_0 x & -a_1 h_1 x^2 & -a_1 h_2 x^3 & -\ldots & -a_1 h_{k-1} x^k & -\ldots \\
-a_2 x^2 h(x) & = & & & -a_2 h_0 x^2 & -a_2 h_1 x^3 & -\ldots & -a_2 h_{k-2} x^k & +\ldots \\
\vdots & & & & & & & \vdots & \\
-a_k x^k h(x) & = & & & & & & -a_k h_0 x^k & +\ldots \\
\hline
a(x)h(x) & = & h_0 & +(h_1 - a_1 h_0)x & +(h_2 - a_1 h_1 - a_2 h_0)x^2 & +\ldots & \ldots & +f(k)x^k & +\ldots
\end{array}$$

So

$$h(x) = \frac{h_0 + (h_1 - a_1 h_0)x + \ldots + (h_{k-1} - a_1 h_{k-2} + \ldots + a_{k-1} h_0)x^{k-1} + \sum_{n=k}^{\infty} f(n) x^n}{a(x)}.$$

5) Substitute in the values of $h_0, \ldots, h_{k-1}$ and factor $a(x)$ (remember that you can use complex roots). Find a closed form for

$$\sum_{n=k}^{\infty} f(n)x^n.$$

6) Use partial fractions to turn our expression for $h(x)$ into a sum of expressions that we can expand using the generalised binomial theorem.

7) Make variable substitutions if necessary to get summands that look like

$$\frac{A}{(1+y)^n}.$$

8) Use the generalised binomial theorem to find $h_n$, the coefficient of $x^n$ in $h(x)$.

**EXERCISES 8.3.3.** For each of the following recursively-defined sequences, use the method of generating functions to find an explicit formula for the $n$th term of the sequence.

1) $c_0 = 2$, $c_1 = 0$, $c_n = c_{n-1} + 2c_{n-2}$ for every $n \geq 2$.

2) $d_0 = 0$, $d_1 = 1$, $d_n = 2d_{n-2} + 1$ for every $n \geq 2$.

3) $e_0 = 2$, $e_n = 3e_{n-1} - 2$ for every $n \geq 1$.

4) $f_0 = 1$, $f_1 = 3$, and $f_n = 4(f_{n-1} - f_{n-2})$ for every $n \geq 2$.

5) $g_0 = 2$, $g_1 = 0$, and $g_n = 2g_{n-1} - 2g_{n-2}$ for every $n \geq 2$.

6) $h_0 = 1/2$ and $h_n = 3h_{n-1} - 1/2$ for every $n \geq 1$.

7) $i_0 = i_1 = 2$, $i_2 = 0$, and $i_n = 3i_{n-1} - 3i_{n-2} + i_{n-3}$ for every $n \geq 3$.

8) $j_0 = -1$, $j_1 = 0$, and $j_n = 2j_{n-1} + 3j_{n-2}$ for every $n \geq 2$.

9) $k_0 = 10$ and $k_n = 11k_{n-1} - 10$ for every $n \geq 1$.

**EXERCISES 8.3.4.** Solve the following problems.

1) Let $p_n$ denote the number of ways to build a pipe $n$ units long, using segments that are either plastic or metal, and (for each material) come in lengths of 1 unit or 2 units. For example, $p_1 = 2$ since we can use a 1-unit segment that is either plastic or metal, and $p_2 = 6$ since we can use either type of 2-unit segment, or any of the $2^2$ possible ordered choices of 2 segments each having a length of 1 unit. Define $p_0 = 1$. Determine a recurrence relation for $p_n$. Give a combinatorial proof that your recurrence relation does solve this counting problem. Use your recurrence relation and the method of generating functions to find a formula for $p_n$.
[*Hint:* Your final answer should be

$$p_n = \frac{1}{2\sqrt{3}}(1+\sqrt{3})^{n+1} - \frac{1}{2\sqrt{3}}(1-\sqrt{3})^{n+1}$$

for every $n \geq 0$.]

2) Let $s_n$ denote the number of lists of any length that have the fixed sum of $n$, and whose entries come from $\{1, 2, 3\}$. For example, $s_2 = 2$ because $(1, 1)$ and $(2)$ are the only such lists; and $s_4 = 7$ because the lists are $(3, 1)$, $(1, 3)$, $(2, 2)$, $(2, 1, 1)$, $(1, 2, 1)$, $(1, 1, 2)$, and $(1, 1, 1, 1)$. Define $s_0 = 1$. Determine $s_1$, $s_3$, and $s_5$ by finding all possible lists. Give a combinatorial proof that $s_n = s_{n-1} + s_{n-2} + s_{n-3}$ for every $n \geq 3$. Use this recurrence relation to show that the generating function $S(x)$ for $\{s_n\}$ is $\frac{1}{1-x-x^2-x^3}$.

**SUMMARY:**

- Method of partial fractions
- Formula for factoring quadratic polynomials into the required form
- Applying generating functions to recursively-defined sequences

Chapter 9

# Some Important Recursively-Defined Sequences

## 9.1. Derangements

**DEFINITION 9.1.1.** A **derangement** of a list of objects, is a permutation of the objects, in which no object is left in its original position.

A classic example of this is a situation in which you write letters to ten people, address envelopes to each of them, and then put them in the envelopes, but accidentally end up with none of the letters in the correct envelope.

Another example might be a dance class in which five sibling pairs are enrolled. The instructor mixes them up so that no one is dancing with a sibling.

Since we're considering enumeration, it shouldn't surprise you that the question we want answered is: in how many ways can this happen? That is, given $n$ objects, how many derangements of the $n$ objects are there? Let's use $D_n$ to denote the number of derangements of $n$ objects.

We can label the objects with the numbers $\{1, \ldots, n\}$, and think of a derangement as a bijection

$$f\colon \{1, \ldots, n\} \to \{1, \ldots, n\},$$

such that $f$ does not fix any value. There are $n-1$ choices for $f(n)$, since the only restriction is $f(n) \neq n$. Say $f(n) = i$. We consider two possible cases.

**Case 1: $f(i) = n$.** Now, on the other $n-2$ values between 1 and $n$ that are neither $i$ nor $n$, $f$ must map $\{1, \ldots, n-1\} \setminus \{i\}$ to $\{1, \ldots, n-1\} \setminus \{i\}$, and must be a derangement. So there are $D_{n-2}$ derangements that have $f(n) = i$ and $f(i) = n$.

**Case 2: $f(j) = n$ for some $j \neq i$.** In this case, we define another function

$$g\colon \{1, \ldots, n-1\} \to \{1, \ldots, n-1\}$$

as follows. We set $g(j) = i$, and for every other value, $g(a) = f(a)$ (that is, for every $a \in \{1, \ldots, n-1\} \setminus \{j\}$). We had $f(j) = n$ and $f(n) = i$, and we are eliminating $n$ from the derangement while maintaining a bijection, by creating the shortcut $g$ with $g(j) = i$ but $g(a) =$

$f(a)$ for every other $a \in \{1, \ldots, n-1\}$. (So whereas $f$ maps $j$ to $n$ and $n$ to $i$, the map $g$ takes $j$ to $i$ directly and is not defined on $n$.) Since $f$ is a derangement and $j \neq i$, we see that $g$ is also a derangement (this time of $n-1$ objects). So there are $D_{n-1}$ possible derangements $g$, and for a fixed choice of $i$, these are in one-to-one correspondence with derangements $f$ that have $f(j) = n$ and $f(n) = i$, so there are also $D_{n-1}$ of these.

We conclude that $D_n = (n-1)(D_{n-1} + D_{n-2})$.

We also need some initial conditions. We have $D_1 = 0$; there is no way of arranging a single object so that it doesn't end up in the correct place. Also, $D_2 = 1$, since there is exactly one way of deranging two objects (by interchanging them).

If we wanted to solve this recursively-defined sequence, we would need to use *exponential generating functions*, which we'll introduce in this chapter but won't really study in this course. Instead, we'll give the explicit formula for $D_n$ without proof.

**PROPOSITION 9.1.2.** *For any $n \geq 1$, the number of derangements of $n$ objects is*

$$D_n = n! \left( \sum_{i=0}^{n} \frac{(-1)^i}{i!} \right).$$

**EXERCISES 9.1.3.**

1) Use induction to prove Proposition 9.1.2.

2) Which kind of induction did you have to use to prove Proposition 9.1.2?

3) Calculate $D_5$ using the explicit formula given in Proposition 9.1.2.

4) Calculate $D_5$ using the recursive relation.

## 9.2. Catalan numbers

This is an example that shows even more clearly the power of the generating function method.

The Catalan numbers, named for Eugène Charles Catalan (1814—1894) are a sequence that can be defined in a variety of ways, because they arise in a number of different circumstances. We'll use the following definition.

**DEFINITION 9.2.1.** The $n$th **Catalan number**, $C_n$, is the number of different ways in which brackets can be put around $n$ terms, to indicate different orders of combining the terms.

Thus, for example, $C_3 = 2$, since three terms can be combined as either

$$[(\_ \cdot \_) \cdot \_], \text{ or } [\_ \cdot (\_ \cdot \_)].$$

These numbers have something in common with Example 6.3.3, in which Shawna was building towers from lego. If we'd asked in how many different orders she could combine the blocks to build her tower, assuming that the final order for the blocks was decided in advance, we would have been asking for the Catalan number. So we can use logic similar to the logic we used in that example: in order to create an expression with $n$ terms, our final step must involve combining a set of $k$ terms (for which the order of combining them has already been determined) and a set of $n-k$ terms (for which the order of combining them has already been determined). Here, $k$ may take on any value from 1 to $n-1$. This results in the recursive relation:

$$C_n = \sum_{k=1}^{n-1} C_k C_{n-k}.$$

This may be easier to see with an example, so we will use the recursive relation to work out $C_4$.

**EXAMPLE 9.2.2.** We've worked out $C_3$ above; in order to work out $C_4$ using this recursive relation, we also need to know $C_1$ and $C_2$. There is only one way to combine a single term (we don't need brackets at all), so $C_1 = 1$. We also have $C_2 = 1$, since there is only one way to put brackets around a pair of terms: $(\_ \cdot \_)$.

Now, to use brackets to order the operations in a four-term expression, our final operation must either combine a group of three terms with a single term; a group of two terms with another group of two terms; or a single term with a group of three terms (this time, the single term is at the left). The first two expressions below come from combining a group of three terms with a single term; the third comes from combining a group of two terms with another group of two terms; and the last two come from combining a single term with a group of three terms.

$$([(\_ \cdot \_) \cdot \_] \cdot \_), \qquad ([\_ \cdot (\_ \cdot \_)] \cdot \_), \qquad [(\_ \cdot \_) \cdot (\_ \cdot \_)], \qquad (\_ \cdot [(\_ \cdot \_) \cdot \_]), \qquad (\_ \cdot [\_ \cdot (\_ \cdot \_)]).$$

Thus,

$$C_4 = C_3C_1 + C_2C_2 + C_1C_3 = 2 + 1 + 2 = 5.$$

We would like to use generating functions to figure out what we can about the Catalan numbers. Unfortunately, there is a difficulty. Any time we want to use generating functions to solve a recursively-defined sequence, the sequence must start with a 0th term, to be the coefficient of $x^0$. With some recursively-defined sequences (including any linear recurrence of the sort we looked at in Chapter 7), we can simply use the recursive relation "backwards" to solve previous terms, going down to $n = 0$, even if our initial conditions began with much higher terms. For example, if a recursively-defined sequence is given by $h_2 = 1, h_3 = 5$ and $h_n = 8h_{n-2} - h_{n-1}$ for every $n \geq 4$, we can use $n = 3$ in this to get

$$h_3 = 5 = 8h_1 - h_2 = 8h_1 - 1.$$

Solving for $h_1$ gives $h_1 = 3/4$. Then using the recursive relation with $n = 2$ gives

$$h_2 = 1 = 8h_0 - h_1 = 8h_0 - 3/4.$$

Solving for $h_0$ gives $h_0 = 7/32$. This allows us to use generating functions on the sequence.

The recursive relation for the Catalan numbers doesn't have a form that allows us to solve for $C_0$ by knowing other terms of the sequence, so we do what we have to, in order to make things work. Instead of working with the generating function for the Catalan numbers themselves (since we can't), we work with the generating function for the sequence $c_0, c_1, c_2, \ldots$, where $c_i = C_{i+1}$ for every $i \geq 0$. In other words, the $n$th term of our new sequence will be the $(n+1)$st Catalan number.

Adjusting the recursive relation we've determined for the Catalan numbers to this new sequence, gives

$$c_0 = 1 \text{ , and } c_n = \sum_{k=0}^{n-1} c_k c_{n-k-1} \text{ for every } n \geq 1.$$

Notice that

$$\begin{aligned} c(x)c(x) &= (c_0 + c_1x + c_2x^2 + c_3x^3 + \ldots)(c_0 + c_1x + c_2x^2 + c_3x^3 + \ldots) \\ &= c_0c_0 + (c_1c_0 + c_0c_1)x + (c_2c_0 + c_1c_1 + c_0c_2)x^2 + (c_0c_3 + c_1c_2 + c_2c_1 + c_3c_0)x^3 + \ldots \end{aligned}$$

and in general, the coefficient of $x^m$ in $(c(x))^2$, is

$$\sum_{k=0}^{m} c_k c_{m-k}.$$

This should look familiar! In fact, you can see that the coefficient of $x^m$ in $(c(x))^2$, is the same as the coefficient of $x^{m+1}$ in $c(x)$, since the latter is

$$\sum_{k=0}^{m} c_k c_{m-k}$$

also.

Thus, we have an expression for $c(x)$ in terms of $(c(x))^2$, since multiplying $(c(x))^2$ by $x$ gives all of the terms of $c(x)$ except $c_0$: $c(x) = x(c(x))^2 + c_0$. We can rearrange this equation, to see that

$$x[c(x)]^2 - c(x) + 1 = 0.$$

We are about to do something to this generating function that may seem a bit like black magic: we will use the quadratic formula to factor this quadratic equation in $c(x)$, treating $x$ as the coefficient of $(c(x))^2$. Thus, in the quadratic formula, we take $a = x$, $b = -1$, and $c = 1$, and obtain

$$c(x) = \frac{1 \pm \sqrt{1 - 4x}}{2x}.$$

Of course, there are two roots to this, and only one of them will give the correct generating function; we need to work out which one (whether to take the plus or the minus).

Using the Generalised Binomial Theorem, we see that

$$(1-4x)^{1/2} = \binom{1/2}{0} + \binom{1/2}{1}(-4x) + \binom{1/2}{2}(-4x)^2 + \ldots + \binom{1/2}{k}(-4x)^k + \ldots,$$

and

$$\binom{1/2}{k} = \frac{(1/2)(-1/2)(-3/2)\ldots(1/2-k+1)}{k!} = \frac{(-1)^{k-1}1\cdot 3\cdot 5\cdot (2k-3)}{2^k k!}$$

so the coefficient of $x^k$ in $(1-4x)^{1/2}$ is

$$\frac{(-1)^{k-1}1\cdot 3\cdot 5\cdot (2k-3)}{2^k k!}(-4)^k = \frac{(-1)1\cdot 3\cdot 5\cdot (2k-3)2^k}{k!}.$$

Whichever root we use will require this expression, so let's work with it a bit more to get it into a nicer form.

$$2^k k! = 2^k(1\cdot 2\cdot 3\cdot \ldots \cdot k) = 2\cdot 4\cdot 6\cdot \ldots \cdot 2k,$$

so if we multiply the numerator and denominator of the fraction by $k!$ (which does not change the result), we see that we have

$$\frac{(-1)1\cdot 3\cdot 5\cdot (2k-3)2\cdot 4\cdot 6\ldots\cdot 2k}{k!k!} = \frac{(-1)(2k-2)!2k}{k!k!} = \frac{(-1)(2k)!}{(2k-1)k!k!} = \frac{-1}{2k-1}\binom{2k}{k},$$

so

$$(1-4x)^{1/2} = -\sum_{k=0}^{\infty} \frac{1}{2k-1}\binom{2k}{k}x^k.$$

The coefficients shown on the right-hand side of this equation quickly get big and negative. If

$$c(x) = \frac{1+\sqrt{1-4x}}{2x},$$

then for $n > 0$ the coefficient of $x^n$ in $c(x)$ will be half of the coefficient of $x^{n+1}$ in $(1-4x)^{1/2}$, which (when $n$ is large) will be big and negative. But it is easy to see from the recurrence relation that all of the Catalan numbers are positive. To get positive coefficients, we must use

$$c(x) = \frac{1-\sqrt{1-4x}}{2x}.$$

Since in this expression we take the negative of the large negative coefficients, the result will be large positive coefficients (even when we divide by 2, and look for the coefficient of $x^{n+1}$).

Thus,

$$c(x) = \frac{1-(1-4x)^{1/2}}{2x} = \frac{1+\sum_{k=0}^{\infty}\frac{1}{2k-1}\binom{2k}{k}x^k}{2x}.$$

From this, we see that for $n > 0$, the coefficient of $x^n$ in $c(x)$ is half of the coefficient of $x^{n+1}$ in $(1-4x)^{1/2}$, which is

$$\begin{aligned}\frac{1}{2}\binom{2(n+1)}{n+1}\frac{1}{2n+1} &= \frac{(2n+2)!}{2(n+1)!(n+1)!(2n+1)} \\ &= \frac{1}{2}\cdot\frac{2n+2}{n+1}\cdot\frac{(2n)!}{(n+1)!n!}\cdot\frac{2n+1}{2n+1} \\ &= \frac{1}{n+1}\binom{2n}{n}.\end{aligned}$$

So

$$c_n = \frac{1}{n+1}\binom{2n}{n}.$$

Although we derived this expression for $n > 0$ only, we can verify that $c_0 = 1 = \frac{1}{0+1}\binom{0}{0}$ since $0! = 1$, so this expression is true for every $n \geq 0$.

**EXERCISES 9.2.3.**

1) Use induction and the recursive relation for Catalan numbers (as adjusted for the values of $\{c_i\}$, where $c_i = C_{i+1}$) to prove that $c_n > 0$ for every $n \geq 0$.
2) Calculate $c_4$ using the explicit formula that we calculated in this section.
3) Calculate $c_4$ using the recursive relation.

## 9.3. Bell numbers and exponential generating functions

Sometimes a recurrence relation involves factorials, or binomial coefficients. When this happens, it becomes difficult if not impossible to use ordinary generating functions to find an explicit

formula for the $n$th term of the sequence. In some cases, a different kind of generating function, the exponential generating function, may succeed where an ordinary generating function fails.

In this section we will be using techniques from calculus. This section is not required for understanding other parts of this book, and the material can be omitted or skimmed over.

**DEFINITION 9.3.1.** The **exponential generating function** for the sequence $a_0, a_1, \ldots$, is

$$\sum_{i=0}^{\infty} \frac{a_i x^i}{i!}.$$

Obviously, the difference between this and an ordinary generating function comes from the factorial expression in the denominator. Cancellation between this and expressions in the numerator can lead to nicer compact expressions.

**EXAMPLE 9.3.2.** The exponential generating function for the sequence $1, 1, 1, \ldots$ is

$$\frac{1}{0!} + \frac{x}{1!} + \frac{x^2}{2!} + \cdots.$$

This is the Taylor series expansion for $e^x$. Thus, $e^x$ is the exponential generating function for $1, 1, 1, \ldots$.

We will not be using exponential generating functions in this course; we are just introducing the topic. We will go through one example of a sequence for which exponential generating functions are useful: the Bell numbers. As with many other topics in this course, although the Bell numbers are named for Eric Temple Bell (1883—1960), who wrote about them in the 1930s, he does not deserve credit for their discovery. They go back to medieval Japan, and had been much studied by mathematicians prior to Bell's writing.

**DEFINITION 9.3.3.** The **Bell number** $B_n$ is the number of partitions of $\{1, \ldots, n\}$ into subsets.

In case you are not familiar with the term, a **partition** of a set $X$ into subsets is a collection of subsets $X_1, \ldots, X_k$ with the properties that:

- $\bigcup_{i=1}^{k} X_i = X$; and
- for any $1 \leq i, j \leq k$ with $i \neq j$, we have $X_i \cap X_j = \emptyset$.

In other words, every element of $X$ must appear in exactly one of the subsets.

Let's look at the first few Bell numbers.

**EXAMPLE 9.3.4.** There is only one way to partition $\{1\}$ into subsets: $\{1\}$, so $B_1 = 1$.

There are two ways to partition $\{1, 2\}$ into subsets: take the two subsets $\{1\}, \{2\}$, or put everything into the single subset $\{1, 2\}$, so $B_2 = 2$.

There are five ways to partition $\{1, 2, 3\}$ into subsets: $\{1\}, \{2\}, \{3\}$, or $\{1, 2\}, \{3\}$, or $\{1, 3\}, \{2\}$, or $\{2, 3\}, \{1\}$, or $\{1, 2, 3\}$, so $B_3 = 5$.

Probably after seeing the above examples, you don't want to calculate larger Bell numbers directly. However, we can derive a recursive relation for these numbers. For this relation to work properly, we will define $B_0 = 1$.

**PROPOSITION 9.3.5.** *For $n \geq 1$, the nth Bell number*

$$B_n = \sum_{k=1}^{n} \binom{n-1}{k-1} B_{n-k}.$$

**PROOF.** We'll use a combinatorial proof of this statement. We know that $B_n$ is the number of partitions of $\{1, \ldots, n\}$ into subsets.

For the other side of the equation, let's consider the subset that contains the element $n$, and call the cardinality of this subset $k$. Since $n$ is in this subset, $k \geq 1$, and since this is a subset of $\{1, \ldots, n\}$, we have $k \leq n$, so $1 \leq k \leq n$. There are $\binom{n-1}{k-1}$ ways to choose the remaining $k-1$ elements of this subset; that is, for any $1 \leq k \leq n$, there are $\binom{n-1}{k-1}$ ways to choose the subset of $\{1, \ldots n\}$ that contains the element $n$. For each of these ways, there are $n-k$ other elements that must be partitioned, and by the definition of the Bell numbers, there are $B_{n-k}$ ways to partition them into subsets. (Our definition of $B_0 = 1$ deals with the case $k = n$, ensuring that the $\binom{n-1}{n-1} = 1$ way of choosing $n$ to be in a single set of all $n$ elements is counted once and only once.)

Thus, using the product and sum rules, we see that

$$B_n = \sum_{k=1}^{n} \binom{n-1}{k-1} B_{n-k}. \qquad \square$$

Let us try to find the exponential generating function for the Bell numbers. To understand the calculations that follow, you will need to know some calculus (more specifically, you will need to have some background in derivatives). If you don't have that background, you won't understand the details we'll be going over, although you may still be able to get a general sense of what is going on.

When dealing with exponential generating functions, notice that the derivative of $x^n/n!$ is

$$\frac{nx^{n-1}}{n!} = \frac{x^{n-1}}{(n-1)!},$$

so taking derivatives often helps us find a nice expression for the coefficients. You already know a particularly simple example of this: the derivative of $e^x$ is $e^x$, which tells us that all of the coefficients in that exponential generating function are equal.

Define

$$B(x) = \sum_{i=0}^{\infty} B_i \frac{x^i}{i!} = B_0 + B_1 \frac{x}{1!} + B_2 \frac{x^2}{2!} + \ldots + B_n \frac{x^n}{n!} + \ldots.$$

Notice that the derivative of this is

$$\begin{aligned} \frac{d}{dx} B(x) &= B_1 + B_2 \frac{x}{1!} + B_3 \frac{x^2}{2!} + \ldots + B_n \frac{x^{n-1}}{(n-1)!} + \ldots \\ &= \sum_{n=1}^{\infty} B_n \frac{x^{n-1}}{(n-1)!}. \end{aligned}$$

Using our recursive relation from Proposition 9.3.5, we see that this is

$$\begin{aligned}\frac{d}{dx}B(x) &= \sum_{n=1}^{\infty}\left[\sum_{k=1}^{n}\binom{n-1}{k-1}B_{n-k}\right]\frac{x^{n-1}}{(n-1)!}\\ &= \sum_{n=1}^{\infty}\left[\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}B_{n-k}\frac{x^{n-1}}{(n-1)!}\right]\\ &= \sum_{n=1}^{\infty}\left[\sum_{k=1}^{n}\frac{1}{(k-1)!(n-k)!}B_{n-k}x^{n-1}\right]\\ &= \sum_{n=1}^{\infty}\left[\sum_{k=1}^{n}\frac{x^{k-1}}{(k-1)!}B_{n-k}\frac{x^{n-k}}{(n-k)!}\right].\end{aligned}$$

Notice that for each value of $n$, as $k$ goes from 1 to $n$ the values $k-1$ and $n-k$ take on every pair of non-negative integral values that add up to $n$. Thus, as $n$ goes from 1 to infinity, the values $k-1$ and $n-k$ take on every possible pair of non-negative integral values. Therefore, we can rewrite this expression as

$$\begin{aligned}\frac{d}{dx}B(x) &= \sum_{j=0}^{\infty}\left[\sum_{i=0}^{\infty}\frac{x^j}{j!}B_i\frac{x^i}{i!}\right]\\ &= \sum_{j=0}^{\infty}\frac{x^j}{j!}\left[\sum_{i=0}^{\infty}B_i\frac{x^i}{i!}\right]\\ &= \left[\sum_{j=0}^{\infty}\frac{x^j}{j!}\right]\left[\sum_{i=0}^{\infty}B_i\frac{x^i}{i!}\right]\\ &= e^xB(x).\end{aligned}$$

Now, consider the derivative of $e^{-(e^x)}B(x)$. By the product and chain rules, this is

$$e^{-(e^x)}e^xB(x) - B(x)e^{-(e^x)}e^x = 0,$$

so it must be the case that $e^{-(e^x)}B(x)$ is constant, say $e^{-(e^x)}B(x) = c$. Then $B(x) = ce^{(e^x)}$.

Since

$$B(0) = \sum_{n=0}^{\infty}B_n\frac{0^n}{n!} = 1 + \sum_{n=1}^{\infty}0 = 1,$$

(recall that $0^0 = 0$, or if you don't like that, simply use the expansion of $B(0)$), we see that

$$ce^{e^0} = ce^1 = ce = 1,$$

so $c = e^{-1}$. Hence

$$B(x) = e^{-1}e^{(e^x)} = e^{(e^x-1)}.$$

There are techniques to extract coefficients from expressions like this, also, but we will not cover these techniques in this class.

**EXERCISES 9.3.6.**

1) Find $B_4$.

2) What is the exponential generating function for the sequence $a_i = i!$ for every $i \geq 0$? Give the sequence in both an expanded and a closed form.

3) What is the exponential generating function for the sequence $b_i = (i+1)!/2$ for every $i \geq 0$? Give the sequence in both an expanded and a closed form.

---

**SUMMARY:**

- generating functions must start with a 0th term
- Important definitions:
  - derangements
  - Catalan numbers
  - exponential generating function
  - Bell numbers

---

# Chapter 10

# Other Basic Counting Techniques

There are two other elementary techniques that are surprisingly useful even in quite difficult counting problems. We will wrap up our exploration of enumeration by discussing these techniques.

## 10.1. The Pigeonhole Principle

The Pigeonhole Principle is a technique that you can apply when you are faced with items chosen from a number of different categories of items, and you want to know whether or not some of them must come from the same category, without looking at all of the items.

**EXAMPLE 10.1.1.** Suppose I will be teaching an independent study course in graph theory to two students next semester, and I want to use Bondy & Murty's "Graph Theory" text book. It has been issued in two editions, and I don't care which edition we use, but I want both students to have the same edition.

I find a web site on which someone has posted that they have three copies of the text for sale, but they don't say which editions they are. Without any more information, I know that if I buy these texts, I will have suitable texts for my students.

The reasoning is straightforward. The first book could be edition 1 or edition 2. If the second text is the same as the first, then I have what I need, so the only possible problem is if the first two books consist of one copy of edition 1, and one copy of edition 2. But then the third book must match one or the other of the first two, since there are only two editions, so I will have two copies of one or the other of the editions.

This idea can be generalised in several ways. We'll look at the most straightforward generalisation first.

**PROPOSITION 10.1.2 (Pigeonhole Principle).** *If there are $n$ items that fall into $m$ different categories and $n > m$, then at least two of the items must fall into the same category.*

**PROOF.** Amongst the first $m$ items, either two of the items are from the same category (so we are done), or there is exactly one item from each of the $m$ categories. Since $n > m$, there is at least one more item. This item must fall into the same category as one of the previous items, since every category already has an item. □

The name of this principle comes from the idea that it can be stated with the categories being a row of holes, and the items being pigeons who are assigned to these holes.

In Example 10.1.1, the categories were the editions, and the items were the text books.

Example 10.1.1 was a very direct and straightforward application of the Pigeonhole Principle. The Principle can also apply in much more subtle and surprising ways.

**EXAMPLE 10.1.3.** Maria makes a bet with Juan. He must buy her at least one chocolate bar every day for the next 60 days. If, at the end of that time, she cannot point out a span of consecutive days in which the number of chocolate bars he bought for her was precisely 19, then she will pay for all of the chocolate bars and give them back to him. If she can find such a span, then she gets to keep the chocolate bars. To limit the size of the bet, they agree in advance that Juan will not buy more than 100 chocolate bars in total.

Is there a way for Juan to win this bet?

**SOLUTION.** The answer is no. For $1 \leq i \leq 60$, let $a_i$ represent the number of chocolate bars that Juan has bought for Maria by the end of day $i$. Then $1 \leq a_1 < a_2 < \ldots < a_{60} \leq 100$. Maria is hoping that for some $i < j$, she will be able to find that $a_i + 19 = a_j$. We therefore also need to consider the values $a_1 + 19 < a_2 + 19 < \ldots < a_{60} + 19$. By the bounds on $a_1$ and $a_{60}$, we have $a_1 + 19 \geq 20$, and $a_{60} + 19 \leq 119$. Thus, the values $a_1, \ldots, a_{60}, a_1 + 19, \ldots, a_{60} + 19$ are 120 numbers all of which lie between 1 and 119.

By the Pigeonhole Principle, at least two of these numbers must be equal, but we know that the $a_i$s are strictly increasing (as are the $a_i + 19$s), so there must exist some $i < j$ such that $a_i + 19 = a_j$. Maria must point to the span of days from the start of day $i + 1$ to the end of day $j$, since in this span Juan bought her 19 chocolate bars.

In fact, Juan could not win a bet of this nature that lasted more than 56 days, but proving this requires more detailed analysis specific to the numbers involved, and is not really relevant to this course. □

Here is another example that would be hard to prove if you didn't know the Pigeonhole Principle.

**EXAMPLE 10.1.4.** Fix $n$, and colour each point of the plane with one of $n$ colours. Prove that there exists a rectangle whose four corners are the same colour.

**PROOF.** Take a grid of points with $n + 1$ rows and $n\binom{n+1}{2} + 1$ columns. We claim that this grid will contain such a rectangle.

Since $n$ colours have been used, and there are $n+1$ points in each column, by the Pigeonhole Principle each column must contain at least two grid points that are the same colour.

In any column, there are $\binom{n+1}{2}$ possible locations in which a pair of points of the same colour could appear. Thus there are at most $\binom{n+1}{2}$ ways to position two points of colour 1 in a column so that the points do not occupy the same two locations in more than one of these columns. The same is true for each of the $n$ colours. Therefore, we can create a maximum of $n\binom{n+1}{2}$ columns, each having two points of some colour, in such a way as to avoid having the same colour occupy the same two locations in more than one of the columns. Since we have $n\binom{n+1}{2} + 1$ columns, there must exist some pair of columns such that the same colour does occupy the same two locations in both of the columns. These four points form a rectangle whose corners all have the same colour. □

So far we have only thought about guaranteeing that there are at least two items in some category. Sometimes we might want to know that there are at least $k$ items in some category, where $k > 2$. There is a generalisation of the Pigeonhole Principle that applies to such situations.

**PROPOSITION 10.1.5 (Generalised Pigeonhole Principle).** *Given $n$ items that fall into $m$ different categories, if $n > km$ for some positive integer $k$, then at least $k+1$ of the items must fall into the same category.*

**PROOF.** Amongst the first $km$ items, either $k+1$ of the items are from the same category (so we are done), or there are exactly $k$ items from each of the $m$ categories. Since $n > km$, there is at least one more item. This item must fall into the same category as one of the previous items. Since every category already has $k$ items, this means that there will be $k+1$ items in this category. □

Notice that the Pigeonhole Principle is a special case of the Generalised Pigeonhole Principle, obtained by taking $k = 1$.

**EXAMPLE 10.1.6.** The population of West Lethbridge in the 2014 census was 35,377.

Show that at least 97 residents of West Lethbridge share a birthday. If you live in West Lethbridge, how many people can you be sure have the same birthday as you?

**SOLUTION.** For the first part of this question, we apply the generalised pigeonhole principle, with $m = 366$ (for the 366 days of the year, counting February 29 since it is just as legitimate a birthday as any other despite being more uncommon), $k = 96$, and $n = 35,377$. We have

$$n = 35,377 > km = 96 \cdot 366 = 35,136,$$

so the Generalised Pigeonhole Principle tells us that at least $k + 1 = 97$ people must share a birthday.

For the second part of the question, the answer is 0. There is no reason why every single other person in West Lethbridge might not have their birthday on the day after yours (although that particular possibility is quite unlikely). There is certainly no guarantee that any of them has the same birthday as yours. □

Notice that although we have found in the above example that some group of at least 97 people in West Lethbridge must have the same birthday, we have no idea of which 97 people are involved, or of what the joint birthday is. This is rather remarkable, but is an example of a type of proof that is quite common in advanced mathematics. Such proofs are referred to as "non-constructive," since they prove that something exists, without giving you any idea of how to find (or construct) such a thing.

The theorem we are about to present is not in itself central to the rest of the material in this book, but we include it here because of its proof. This proof demonstrates that a relatively straightforward idea such as the Generalised Pigeonhole Principle can be used in subtle and surprising ways to prove results that at first glance do not seem to relate to spreading things out across a variety of categories. If you are interested in combinatorics (or other higher mathematics), reading and understanding the approaches and ideas that are used in this proof may be of considerable value to you. The theorem and its proof are due to Pál Erdős (1913—1996) and György Szekeres (1911—2005).

**THEOREM 10.1.7 (Erdős-Szekeres Theorem).** *For every pair of integers $a, b \geq 1$, if $S$ is a sequence of $ab+1$ distinct real numbers, then there is either an increasing subsequence of length $a+1$ or a decreasing subsequence of length $b+1$ in $S$.*

**PROOF.** Define a function $f$ that maps each element of $S$ to the length of the longest increasing subsequence that begins with that element.

If there exists some $s \in S$ such that $f(s) \geq a + 1$, then we are done. So we may assume that $f(s) \leq a$ for every $s \in S$. Since there is always an increasing sequence of length at least

1 starting at any element of $S$, we in fact have $1 \leq f(s) \leq a$ for every $s \in S$, so there are $a$ possible values for the outputs of $f$. Since $|S| = ab + 1$, and $ab + 1 > ab$, the Generalised Pigeonhole Principle tells us that at least $b+1$ elements of $S$ must have the same output under the function $f$.

We claim that if $x$ is before $y$ in $S$ and $f(x) = f(y)$, then $x > y$. By assumption, $x \neq y$ (all values of $S$ are distinct), so the only other possibility is $x < y$. If $x < y$, then taking $x$ followed by an increasing subsequence of length $f(y)$ that starts at $y$, would give an increasing subsequence of length $f(y) + 1$ that starts at $x$, contradicting $f(x) = f(y)$. This contradiction shows that $x < y$ is not possible, so $x > y$, as claimed.

Let $s_1, s_2, \ldots, s_{b+1}$ be elements of $S$ that have the same output under $f$, and appear in this order. Then by the claim we proved in the previous paragraph, $s_1 > s_2 > \ldots > s_{n+1}$, which is a decreasing subsequence of length $b + 1$. □

For the sake of completeness in our presentation of the Erdős-Szekeres Theorem, we note that their result is best possible. That is, $ab + 1$ is the smallest possible length for $S$ that guarantees the property. that $S$ contains either an increasing subsequence of length $a + 1$ or a decreasing subsequence of length $b + 1$.

To demonstrate this fact, for any $a, b$, we present a sequence of length $ab$ in which the longest increasing sequence has length $a$ and the longest decreasing subsequence has length $b$. One such sequence is

$$b, b-1, \ldots, 1; 2b, 2b-1, \ldots, b+1; \ldots; ab, ab-1, \ldots, (a-1)b+1.$$

Any increasing subsequence can only have one entry from each of the $a$ subsequences of length $b$ that are separated by semicolons, so can only have length $a$. Any decreasing subsequence must be entirely contained within one of the subsequences of length $b$ that are separated by semicolons, so can only have length $b$.

In the Pigeonhole Principle we considered how many items we must have in order to ensure that there are at least two items in some category. In the Generalised Pigeonhold Principle, we expanded on this by considering how many items we must have in order to ensure that there are at least $k + 1$ items in some category, where $k$ can be any positive integer (taking $k = 1$ in this gives the Pigeonhole Principle). Sometimes, applications arise in which not all categories are equally valuable. It may be the case that we're happy if we have at least two items in one specific category, but it will require at least five items in any other category in order to satisfy us. Or each category may have its own specific number of items that we want, and satisfying this requirement for any one category is what we are looking for. This situation is covered by the most general form of the pigeonhole principle.

**PROPOSITION 10.1.8 (Even more generalised pigeonhole principle).** *Let $n_1, n_2, \ldots, n_m$ be positive integers. Given at least*

$$n_1 + n_2 + \ldots + n_m - m + 1$$

*items that fall into $m$ categories, there must be some $1 \leq i \leq m$ such that at least $n_i$ items fall into the $i$th category.*

**PROOF.** Amongst the first

$$n_1 + n_2 + \ldots + n_m - m$$

items, either there is some $1 \leq i \leq m$ such that at least $n_i$ of the items fall into the $i$th category, or there are precisely $n_i - 1$ objects in the $i$th category, for every $1 \leq i \leq m$. Since there is at

least one more item, this item must fall into the $i$th category for some $1 \leq i \leq m$, which means that there will be $n_i$ items in this category. □

Notice that the Generalised Pigeonhole Principle is a special case of the "Even more generalised pigeonhole principle," obtained by taking

$$n_1 = n_2 = \ldots = n_m = k.$$

**EXAMPLE 10.1.9.** Terry wants to bring some of their special stuffed peppers dish to a family potluck, wants to ensure there is enough for everyone who will want it, and wants to use only one colour of peppers to avoid any arguments or disappointment. They also plan to bring a pasta salad dish that everyone likes. They know that of the 12 people who will be at the potluck, everyone will eat the dish if it is made with red peppers; there are 4 people who will not eat it if it is made with yellow peppers (but everyone else will); there are 5 people who will not eat it if it is made with orange peppers (but everyone else will); and there are 6 people who will eat it if it is made with green peppers. The local greenhouse has a great deal on peppers at the farmer's market this week, but you can't request specific colours; you get whatever they give you. How many peppers must Terry buy in order to ensure that they have enough to bring their pepper dish to the potluck?

**SOLUTION.** Terry needs to end up with either 12 red peppers, 8 yellow peppers, 7 orange peppers, or 6 green peppers. So we take $n_1 = 12$, $n_2 = 8$, $n_3 = 7$, and $n_4 = 6$ in the statement of the Even more generalised pigeonhole principle. Notice that we have $m = 4$ categories. Now the statement tells us that given at least

$$n_1 + n_2 + n_3 + n_4 - 4 + 1 = 12 + 8 + 7 + 6 - 4 + 1 = 30$$

peppers, Terry is guaranteed to have either 12 red peppers, 8 yellow peppers, 7 orange peppers, or 6 green peppers.

If you are struggling with this, think again about the worst thing that could happen: Terry could end up with one pepper too few of each colour. This would happen if they had 11 red peppers, 7 yellow peppers, 6 orange peppers, and 5 green peppers, for a total of 29 peppers. As soon as they have one more, they must have enough of some colour. □

**EXAMPLE 10.1.10.** Piet is stocking up on toothbrushes for his autistic daughter for the year. The type of toothbrush his daughter is willing to use is only available online and comes in red, purple, or blue, but you can't choose the colour you will receive. She finds change difficult, so he wants to be able to get her through the year with just one colour of toothbrush. For some reason, the colour affects how long the toothbrush lasts. Red toothbrushes last 2 weeks. Purple toothbrushes last 1 month. Blue toothbrushes last 2 months. Using 52 for the number of weeks and 12 for the number of months in a year, how many toothbrushes must Piet buy to ensure they have enough of one colour to last the whole year?

**SOLUTION.** Piet needs to end up with either 26 red toothbrushes, 12 purple toothbrushes, or 6 blue toothbrushes. So we take $n_1 = 26$, $n_2 = 12$, and $n_3 = 6$ in the statement of the Even more generalised pigeonhole principle. Notice that we have $m = 3$ categories. Now the statement tells us that given at least

$$n_1 + n_2 + n_3 - 3 + 1 = 26 + 12 + 6 - 3 + 1 = 42$$

toothbrushes, Piet is guaranteed to have either 26 red toothbrushes, 12 purple toothbrushes, or 6 blue toothbrushes.

Again you can alternatively think about the worst thing that could happen: Piet could end up with one toothbrush too few of each colour. This would happen if he received 25 red toothbrushes, 11 purple toothbrushes, and 5 blue toothbrushes, for a total of 41 toothbrushes. As long as he has one more than this, he must have enough of some colour. □

**EXAMPLE 10.1.11.** Suppose Ali owes Tomas \$10, and wants to give him a number of identical pieces of currency to pay her debt. Her bank only gives out currency in loonies (which have a value of \$1), twonies (which have a value of \$2), five-dollar bills, or ten-dollar bills, and does not take requests for specific kinds of currency. How much money must Ali request from the teller, if she wants to be sure to have \$10 in identical pieces of currency with which to pay Tomas?

**SOLUTION.** If Ali gets any \$10 bills she can give one of those to Tomas and is done. If she gets at least two \$5 bills, she is done. If she gets at least five twonies, she is done, and if she gets at least 10 loonies she is done. So the most money she can get without being able to give Tomas their \$10 in a single type of currency, is 9 loonies, 4 twonies, and a \$5 bill, for a total of \$22. Therefore, if Ali asks for \$23, she is guaranteed to be able to pay Tomas in a single type of currency. □

Although the above example does not directly use the "Even more generalised pigeonhole principle" because it asks for the value of the currency Ali needs to request rather than the number of items she must request, it uses the same ideas and should be helpful in understanding the concept.

**EXERCISES 10.1.12.**

1) Show that in any positioning of 17 rooks on an 8-by-8 chessboard, there must be at least three rooks none of which threaten each other (i.e. no two of which lie in the same row or column).
2) Sixteen people must sit in a row of eighteen chairs. Prove that somewhere in the row there must be six adjacent chairs all occupied.
3) An artist has produced a large work of art to be carried in a parade. Part of the concept is that it must be carried by people of roughly the same size (i.e., either all adults, or all children). The artist has left it to the last minute to find people to carry this, and is in a bit of a panic. They don't know if they will be able to assemble enough of either adults or children to carry the piece, so they decide to ask everyone they see, until they have enough volunteers. It takes 15 adults to carry the piece, or 23 children. If everyone approached agrees to help, how many people does the artist need to approach before they are sure to have enough people to carry their art in the parade?
4) Let $n$ be odd, let $a$ be even, and let $\pi\colon \{1,\dots,n\} \to \{1,\dots,n\}$ be a permutation. Prove that the product

$$(a+1-\pi(1))(a+2-\pi(2))\cdots(a+n-\pi(n))$$

is even. Is the same conclusion necessarily true if $n$ is even or if $a$ is odd? Give a proof or a counterexamples in each case.

5) Let $n \geq 1$, let $x$ be a positive integer, and let $S$ be a subset of cardinality $n+1$ from $\{x, x^2, \dots, x^{2n}\}$. Prove that there exist two numbers in $S$ whose product is $x^{2n+1}$.
6) Show that in every set of $n+1$ distinct integers, there exist two elements $a$ and $b$ such that $a-b$ is divisible by $n$.

7) A drawer contains socks of 8 different colours. How many socks must you pull out of the drawer to be certain that you have two of the same colour?

8) There are 15 students in a Combinatorics class. Explain how you know that two of them have their birthday in the same month.

9) A pizza restaurant has 8 different toppings. Every day in October, they will put a 2-topping pizza on sale (a pizza that has two distinct toppings on it; the order of the toppings does not matter). Prove that the same pizza will be on sale on two different days.

10) Suppose $A$ is a set of 10 natural numbers between 1 and 100 (inclusive). Show that two different subsets of $A$ have the same sum. For example, if

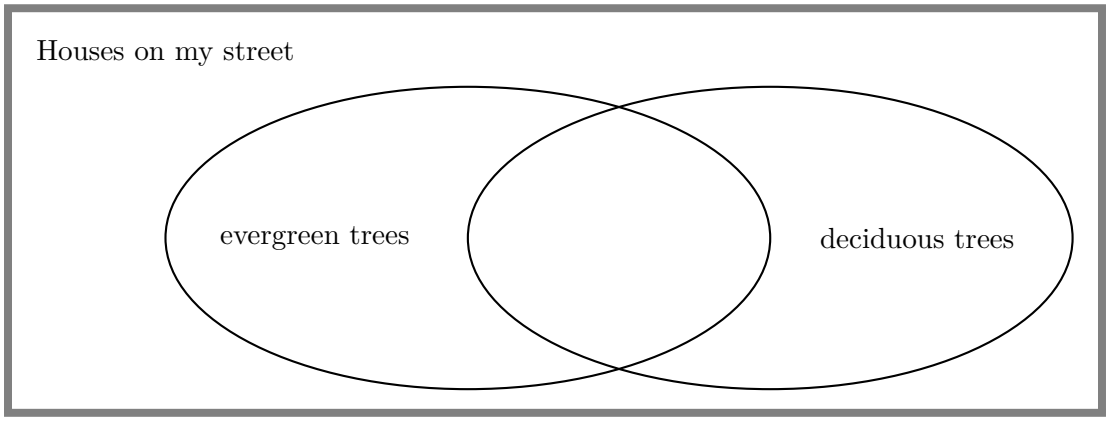
$$A = \{2, 6, 13, 30, 45, 59, 65, 82, 88, 97\},$$

then the subsets $\{6, 13, 45, 65\}$ and $\{2, 30, 97\}$ both add up to 129.
[*Hint:* Compare the answers to two questions: How many subsets of $A$ are there? Since there are only 10 elements of $A$, and each of them is at most 100, how many different possible sums are there?]

11) Consider any set of 5 points in the plane that have integer coordinates. Prove that there is some pair from these 5 points such that the midpoint of the line segment joining this pair of points also has integer coordinates. Give an example of 4 points in the plane that do not have this property, and list all of the midpoints as evidence.

## 10.2. Inclusion-Exclusion

In school, you probably saw Venn diagrams sometimes, showing groups that overlapped with one another. We could draw a very basic Venn diagram showing the kinds of trees that are growing at the various houses on my street:




Looking at the Venn diagram can help us figure out the values of some of the pieces from knowing the values of others. Suppose we know how many houses have deciduous trees, and how many houses have evergreen trees. Naïvely, you might think that adding these together would give us the total number of houses with trees. However, by looking at the Venn diagram, we see that if we simply add the values together, then any houses that have both kinds of trees have been counted twice (once as a house with a deciduous tree, and again as a house with an evergreen tree). So in order to work out the number of houses that have trees, we can add the number that have deciduous trees to the number that have evergreen trees *and then subtract the number that have both kinds of trees.* This is the idea of "inclusion-exclusion."

Specifically, if two sets $A$ and $B$ are disjoint, then $|A \cup B| = |A| + |B|$. However, if $A$ and $B$ are not disjoint, then $|A| + |B|$ counts the elements of $A \cap B$ twice (both as elements of $A$ and as elements of $B$). Subtracting this overcount yields the correct answer:

**PROPOSITION 10.2.1 (Inclusion-Exclusion for 2 sets).** *For any finite sets $A$ and $B$, we have*

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

**PROOF.** Let $A_0 = A \smallsetminus B$ and $B_0 = B \smallsetminus A$, so

- $A$ is the disjoint union of $A_0$ and $A \cap B$,
- $B$ is the disjoint union of $B_0$ and $A \cap B$, and
- $A \cup B = \big(A_0 \cup (A \cap B)\big) \cup \big(B_0 \cup (A \cap B)\big)$ is the disjoint union of $A_0$, $B_0$, and $A \cap B$.

Then

$$\begin{aligned} |A| + |B| &= \big(|A_0| + |A \cap B|\big) + \big(|B_0| + |A \cap B|\big) \\ &= \big(|A_0| + |B_0| + |A \cap B|\big) + |A \cap B| \\ &= |A \cup B| + |A \cap B|. \end{aligned}$$

□

**EXAMPLE 10.2.2.** Let $A = \{\mathsf{p}, \mathsf{r}, \mathsf{o}, \mathsf{n}, \mathsf{g}\}$ and $B = \{\mathsf{h}, \mathsf{o}, \mathsf{r}, \mathsf{n}, \mathsf{s}\}$. Then

$$|A| = 5,\ |B| = 5, \text{ and } |A \cap B| = |\{\mathsf{r}, \mathsf{o}, \mathsf{n}\}| = 3,$$

so Inclusion-Exclusion tells us that

$$|A \cup B| = |A| + |B| - |A \cap B| = 5 + 5 - 3 = 7.$$

This is correct, since

$$|A \cup B| = |\{\mathsf{p}, \mathsf{r}, \mathsf{o}, \mathsf{n}, \mathsf{g}, \mathsf{h}, \mathsf{s}\}| = 7.$$

**EXAMPLE 10.2.3.** Every one of the 4000 students at Modern U owns either a tablet or a smart watch (or both). Surveys show that:

- 3500 students own a tablet, and
- 1000 students own a smart watch.

How many students own *both* a tablet and a smart watch?

**SOLUTION.** Let

- $S$ be the set of all students at Modern $U$,
- $T$ be the set of students who own a tablet, and
- $W$ be the set of students who own a smart watch.

Then, by assumption,

$$|S| = 4000, \qquad |T| = 3500, \qquad |W| = 1000.$$

Since every student owns either a tablet or a smart watch, we have $S = T \cup W$. Therefore, Inclusion-Exclusion tells us that

$$|S| = |T \cup W| = |T| + |W| - |T \cap W|,$$

so

$$|T \cap W| = |T| + |W| - |S| = 3500 + 1000 - 4000 = 500.$$

Hence, there are exactly 500 students who own both a tablet and a smart watch. □

The following exercise provides a formula for the union of three sets $A$, $B$, and $C$. The idea is that $A \cap B$, $A \cap C$ and $B \cap C$ have all been overcounted. However, subtracting all of these will overcompensate, because the elements of $A \cap B \cap C$ have been subtracted too many times, so they need to be added back in.

**EXERCISE 10.2.4.** Suppose $A$, $B$, and $C$ are finite sets. Show

$$\begin{aligned} |A \cup B \cup C| = |A| + |B| + |C| \\ - |A \cap B| - |A \cap C| - |B \cap C| \\ + |A \cap B \cap C|. \end{aligned}$$

[*Hint:* We have formulas for $|(A \cup B) \cup C|$ and $|A \cup B|$. The equality $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ provides another useful formula.]

The following general formula calculates the cardinality of the union of any number of sets, by adding or subtracting the cardinality of every possible intersection of the sets. It is called the *Inclusion-Exclusion* formula, because it works by adding (or "including" ) the cardinalities of certain sets, and subtracting (or "excluding" ) the cardinalities of certain other sets.

**THEOREM 10.2.5 (Inclusion-Exclusion).** *Let $A_1, \ldots, A_n$ be finite sets. Then*

$$|A_1 \cup \ldots \cup A_n| = \left(\sum_{i=1}^{n} |A_i|\right) - \left(\sum_{1 \le i < j \le n} |A_i \cap A_j|\right) + \ldots + \left((-1)^{n+1} |A_1 \cap \ldots \cap A_n|\right).$$

Of course, we can figure out the value of any one of the terms in the inclusion-exclusion formula, if we know the values of all of the other terms.

**EXAMPLE 10.2.6.** Sandy's class is at Calaway Park (an amusement park). There are 21 students in the class. At the end of the day, the teacher asks some questions and determines the following:

- every student rode at least one of the roller coaster, the train, the log ride, or the bumper cars;
- 13 students rode the roller coaster;
- 6 students rode the train;
- 12 students rode the log ride;
- 15 students rode the bumper cars;
- 2 students rode all four of the rides; and
- 10 students rode at least 3 of these 4 rides.

How many students rode exactly two of the four rides?

**SOLUTION.** We begin by establishing some notation. Let $A_1$ be the set of students who rode the roller coaster; $A_2$ will be the set of students who rode the train; $A_3$ will be the set of students who rode the log ride; and $A_4$ will be the set of students who rode the bumper cars.

Ignoring the last piece of information for a moment, the rest of we have been given tells us that:

- $|A_1 \cup A_2 \cup A_3 \cup A_4| = 21$;
- $|A_1| = 13$;
- $|A_2| = 6$;
- $|A_3| = 12$;
- $|A_4| = 15$; and
- $|A_1 \cap A_2 \cap A_3 \cap A_4| = 2$.

The last piece of information is a bit more tricky to encode. Observe that if we take $|A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + |A_1 \cap A_3 \cap A_4| + |A_2 \cap A_3 \cap A_4|$, using ideas similar to inclusion-exclusion (or drawing the Venn diagram) we see that we have found the number of students who rode at least 3 of the 4 rides, except that we have counted the number of students who rode all 4 rides in each of the four summands, instead of counting it only once. So we need to subtract $|A_1 \cap A_2 \cap A_3 \cap A_4|$ off three times to get the number of students who rode at least 3 of the 4 rides. Since we know that $|A_1 \cap A_2 \cap A_3 \cap A_4| = 2$, this tells us that

$$|A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + |A_1 \cap A_3 \cap A_4| + |A_2 \cap A_3 \cap A_4| - 6 = 10,$$

so

$$|A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + |A_1 \cap A_3 \cap A_4| + |A_2 \cap A_3 \cap A_4| = 16.$$

Thus, the inclusion-exclusion formula tells us that

$$21 = (13 + 6 + 12 + 15) - \sum_{1 \le i < j \le 4} |A_i \cap A_j| + 16 - 2,$$

so $\sum_{1 \le i < j \le 4} |A_i \cap A_j| = 39$. Unfortunately, this still isn't quite what we're looking for. The value we want is the number of students who rode exactly two of the four rides. Again, similar reasoning shows that the number of students who rode the roller coaster and the train but neither of the other two rides, will be given by:

$$|A_1 \cap A_2| - |A_1 \cap A_2 \cap A_3| - |A_1 \cap A_2 \cap A_4| + |A_1 \cap A_2 \cap A_3 \cap A_4|.$$

Similar formulas can be worked out for each of the other five pairs that can be formed from the four rides. What we have been asked for, is the sum of these six formulas. This works out to

$$\sum_{1 \le i < j \le 4} |A_i \cap A_j| - 3 \left( \sum_{1 \le i < j < k \le 4} |A_i \cap A_j \cap A_k| \right) + 6|A_1 \cap A_2 \cap A_3 \cap A_4| = 39 - 3(16) + 6(2) = 3.$$

Only three of the students rode exactly two of the four rides. □

This was a very complicated example. You should not expect to have to work out examples that are quite so tricky, but this gives you an idea of the power of inclusion-exclusion. Here is a more straightforward application.

**EXAMPLE 10.2.7.** In the Faculty of Arts and Science, the voting method used is "approval;" that is, regardless of the number of positions available, each voter can mark as many boxes as they wish on their ballot.

Imagine that Prof. Li, Prof. Cheng, and Prof. Osborn were all nominated for two computer science positions on the department's search committee. Barb Hodgson notes the following facts when counting the ballots:

- Prof. Cheng received 18 votes; Prof. Osborn received 15 votes, and Prof. Li received 10 votes.
- Only one ballot had all three boxes marked.
- Five of the ballots were marked for both Prof. Osborn and Prof. Li.
- Ten of the ballots were marked for Prof. Cheng and Prof. Osborn.
- Six of the ballots were marked for Prof. Cheng and Prof. Li.

How many members of the department voted in the election?

**SOLUTION.** Again, we begin by establishing some notation. Let $C$ be the set of ballots that were marked for Prof. Cheng; let $O$ be the set of ballots that were marked for Prof. Osborn; and let $L$ be the set of ballots that were marked for Prof. Li. Then what we want is $|C \cup O \cup L|$: the number of ballots that were marked for at least one of the three candidates; this is the same as the number of people who voted.

Inclusion-Exclusion tells us that

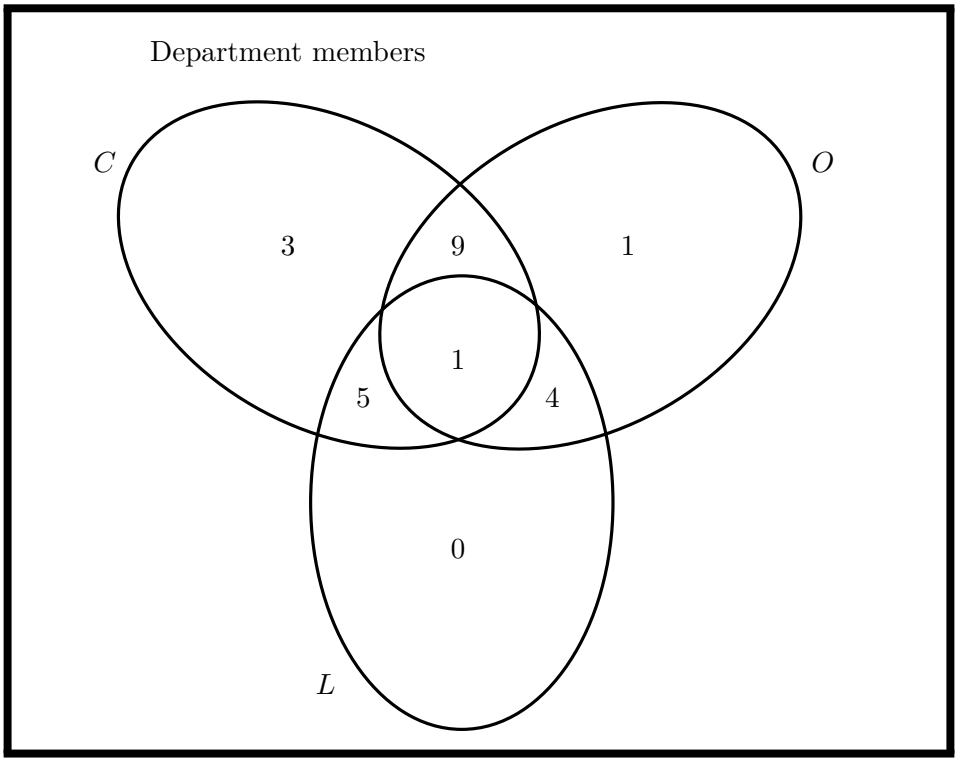
$$|C \cup O \cup L| = |C| + |O| + |L| - |C \cap O| - |C \cap L| - |O \cap L| + |C \cap O \cap L|.$$

We have been given all of the values on the right-hand side of this equation, so we see that

$$|C \cup O \cup L| = 18 + 15 + 10 - 10 - 6 - 5 + 1 = 23.$$

There were 23 department members who voted in the election.

In fact, the information we have been given is enough for us to fill in the values in every piece of the Venn diagram.




The 9 people who voted for Prof. Cheng and Prof. Osborn but not Prof. Li is determined from the fact that 10 people voted for Professors Cheng and Osborn, and only one of those voted for all three professors. Similarly, the 4 people who voted for Prof. Li and Prof. Osborn

but not Prof. Cheng is determined from the fact that 5 people voted for Professors Li and Osborn, and only one of those voted for all three professors; also, the 5 people who voted for Prof. Cheng and Prof. Li but not Prof. Osborn is determined from the fact that 6 people voted for Professors Cheng and Li, and only one of those voted for all three professors.

From the above deductions, we see that of the 18 votes Prof. Cheng received, one ballot was marked for all 3 candidates; 9 were marked for Professors Cheng and Osborn (but not Li); and 5 were marked for Professors Cheng and Li (but not Osborn). The remaining 3 votes must have been for Prof. Cheng alone, allowing us to fill in that spot. Similarly, of the 15 votes Prof. Osborn received, one ballot was marked for all 3 candidates; 9 were marked for Professors Cheng and Osborn (but not Li); and 4 were marked for Professors Osborn and Li (but not Cheng). The remaining vote must have been for Prof. Osborn alone, allowing us to fill in that spot. Finally, all of Prof. Li's 10 votes are accounted for between the 5 who voted for Professors Cheng and Li (but not Osborn), the 4 who voted for Professors Li and Osborn (but not Cheng) and the one who voted for all three, so we put a 0 into the final spot. □

**EXERCISES 10.2.8.**

1) Of 15 students in a stats class, 8 are math majors, 6 are CS majors, and 7 are in education. None are in all three, and none have any other majors. There are two math/CS joint majors, and 3 CS majors who are in education. How many math majors are in education? How many of the math majors are not in either CS or education?
2) Kevin has 165 apps on his phone. Every one of these that is not a game and was not free, requires internet access. Of these, 78 were free. Internet access is necessary for 124 of the apps to function fully. Of the apps on his phone, 101 are games. Kevin has 62 games on his phone that require internet access; 48 of these were free. Out of all of the games on his phone, 58 were free. How many of the free apps on Kevin's phone that aren't games, require internet access?
3) How many integers between 1 and 60 are divisible by at least one of 2, 3, and 5?
4) In the 403 area code, how many of the 10-digit possible phone numbers (where any combination of digits is allowed) contain at least one of each odd digit?
5) Assume $|U| = 15$, $|V| = 12$, and $|U \cap V| = 4$. Find $|U \cup V|$.
6) Assume $|R| = 13$, $|S| = 17$, and $|R \cup S| = 25$. Find $|R \cap S|$.
7) Assume $|J| = 300$, $|J \cup L| = 500$, and $|J \cap L| = 150$. Find $|L|$.
8) At a small university, there are 90 students that are taking either Calculus *or* Linear Algebra (or both). If the Calculus class has 70 students, and the Linear Algebra class has 35 students, then how many students are taking both Calculus *and* Linear Algebra?
9) How many numbers from 1 to 5000 are divisible by either 3 or 17?
10) How many 12-digit numbers (in which the first digit is *not* 0) have either no 0 or no 5?

**SUMMARY:**

- Pigeonhole Principle
- Generalised Pigeonhole Principle
- Even more generalised pigeonhole principle
- Inclusion-Exclusion

# Part II

# Graph Theory

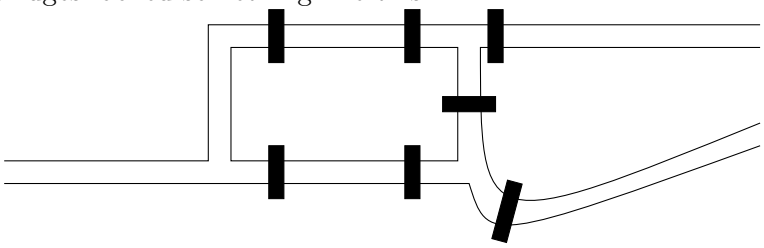
Chapter 11

# Basics of Graph Theory

### 11.1. Background

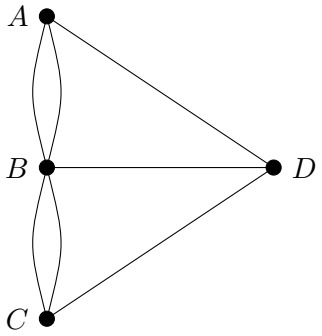
In combinatorics, what we call a *graph* has nothing to do with the $x$ and $y$ axes, and plotting. Here, a graph is the most straightforward way you could think of to model a network. A network could be a computer network, a road network, a telephone network; it doesn't matter what kind of network it is. Conceptually, any network consists of a bunch of things (let's call them nodes) that are being connected in some fashion. To model this, we draw some points for the nodes, and we draw edges between nodes that have a direct connection.

Leonhard Euler (1707—1783) laid the foundations of graph theory in 1735, with his solution to the Königsberg bridge problem. Königsberg, Prussia (now Kaliningrad, Russia) was a city on the river Pregel. The city included two islands in the river, as well as land on both sides of the river, and there were seven bridges connecting the various parts of the city. The lay-out of the city and its bridges looked something like this:



The question had been posed: is it possible for residents of Königsberg, out for Sunday strolls, to cross each of the seven bridges exactly once? Better yet, can they do this *and* end up in the same part of the city where they started?

Euler modeled the problem using a graph, with a vertex for each part of the city (one for each bank, and one for each island), and edges representing the bridges. His model looked like this:




The nodes $A$ and $C$ represent the two banks of the river, while $B$ and $D$ represent the islands. In the model, the question becomes can we trace all of the edges of this graph, without lifting our pen from the paper or going over an edge more than once?

Euler was actually able to find an easy method you can use on any graph, to quickly work out whether or not this can be done for that graph. Also, unlike what we saw in the Pigeonhole Principle (or some of the probabilistic methods we'll mention in Section 11.5), his method is *constructive*: if it can be done, his method shows you how to do it. We'll go over this method later, in Chapter 13.

## 11.2. Basic definitions, terminology, and notation

Now that we have an intuitive understanding of what a graph is, it is time to make a formal definition.

**DEFINITION 11.2.1.** A **graph** $G$ consists of two sets:

- $V$, whose elements are referred to as the **vertices** of $G$ (the singular of vertices is **vertex**); and
- $E$, whose elements are unordered pairs from $V$ (i.e., $E \subseteq \{\{v_1, v_2\} \mid v_1, v_2 \in V\}$). The elements of $E$ are referred to as the **edges** of $G$.

For clarity in situations where more than one graph is being studied, we may use $E(G)$ for $E$ and $V(G)$ for $V$.

According to this definition, Euler's model of the bridges of Königsberg is not actually a graph, because some of the vertices have more than one edge between them (for example, there are two edges between $A$ and $B$), which makes $E$ a multiset rather than a set. This leads naturally to another definition.

**DEFINITION 11.2.2.** For some purposes, we may allow $E$ to be a multiset rather than a set. When we do this, an element that appears more than once in $E$ is called a **multiple edge** or **multiedge**. A graph that includes at least one multiple edge is called a **multigraph**.

Another situation that we might like to allow for some purposes but not allow for others, is the possibility of a connection that goes from a node back to itself.

**DEFINITION 11.2.3.** An edge of the form $\{v, v\}$ for some $v \in V$, is called a **loop**.

A **simple graph** is a graph that has no loops or multiple edges.

For most of the graph theory we cover in this course, we will only consider simple graphs. However, there are some results for which the proof is identical whether or not the graph is simple, and other results that actually become easier to prove if we allow multigraphs and/or loops, than if we only allow simple graphs. It is worthwhile and sometimes important to think about which of our results apply to multigraphs (and/or graphs with loops), and which do not. From this point on, unless otherwise specified, you should assume that *any time the word "graph" is used, it means a simple graph.* However, be aware that many of our definitions and

**Figure 11.2.1.** A loop.

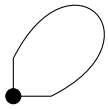
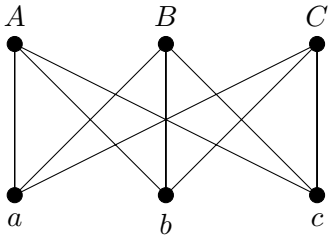


**Figure 11.2.2.** A simple graph.




results generalise to multigraphs and to graphs or multigraphs with loops, even where we don't specify this.

There is still more basic terminology that we need to establish before we can say much about a graph.

**DEFINITION 11.2.4.** If $e = \{u, v\}$ is an edge of a graph (or a multigraph, with or without loops), then we say that $u$ and $v$ are **endvertices** (singular: endvertex) of $e$. We say that $e$ is **incident with** $u$ and $v$ (or vice versa, the vertices are also incident with the edge), and that $u$ and $v$ are **adjacent** since there is an edge joining them, or that $u$ is a **neighbour** of $v$.

**NOTATION 11.2.5.** We use the notation $u \sim v$ to denote that $u$ is adjacent to $v$. We may also denote the edge $e = \{u, v\}$ by $uv$ or by $vu$.

After one more definition, we will go through some examples using the terminology we have established.

**DEFINITION 11.2.6.** If $v \in V$ is a vertex of a graph (simple or multi, with or without loops), then the number of times $v$ appears as the endvertex of some edge is called the **valency** of $v$ in $G$. (Many sources use **degree** rather than valency, but the word degree has many meanings in mathematics, making valency a preferable term for this.) A vertex of valency 0 is called a **isolated vertex**.

*Remark 11.2.7.* In a graph without loops, we can define the valency of any vertex $v$ as the number of edges incident with $v$. For most purposes, this is a good way to think of the valency. However, when a graph has loops, many formulas work out more nicely if we consider each loop to contribute 2 to the valency of its endvertex. This fits the definition we have given, since a vertex $v$ appears twice as the endvertex of any loop incident with $v$.

**NOTATION 11.2.8.** The valency of $v$ is denoted by $\mathrm{val}(v)$ or $\deg(v)$ or $d(v)$ or $d_G(v)$.

**EXAMPLE 11.2.9.** In Figure 11.2.2, the vertices are $a$, $b$, $c$, $A$, $B$, and $C$, and

$$E = \{\{a, A\}, \{a, B\}, \{a, C\}, \{b, A\}, \{b, B\}, \{b, C\}, \{c, A\}, \{c, B\}, \{c, C\}\}\,.$$

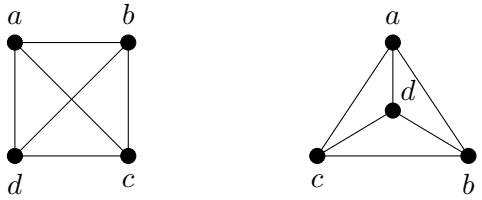
Perhaps you can see already why most people prefer to use a diagram rather than a list of vertices and edges to describe a graph!

The vertex $a$ is adjacent to $A$, $B$ and $C$. The vertex $C$ is not adjacent to $B$. The edge $\{a, C\}$ is incident with the vertex $a$; the vertex $C$ is also an endvertex of this edge. Every vertex in this graph has valency 3, so none of the vertices is isolated. This is a simple graph, as it has no loops or multiple edges.

Although a diagram is a convenient and often helpful way to visualise a graph, it is important to note that because a graph is defined by the sets $V$ and $E$, it is often possible to draw a particular graph in ways that look quite different. Despite the different-looking drawings, as long as $V$ and $E$ are the same, the graph is also the same. In Figure 11.2.3, we see two different drawings of the graph given by $V = \{a, b, c, d\}$ and

$$E = \{\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}\}.$$

**Figure 11.2.3.** Two different drawings of the same graph.




**EXAMPLE 11.2.10.** Let the graph $G$ be defined by $V = \{w, x, y, z\}$ and $E = \{e_1, e_2\}$, where $e_1 = \{w, x\}$ and $e_2 = \{w, y\}$. There are no loops or multiple edges, so $G$ is a simple graph. The edge $e_2$ has endvertices $w$ and $y$. The vertex $w$ is incident with both $e_1$ and $e_2$. The vertices $x$ and $y$ are not adjacent. The vertex $z$ is an isolated vertex, as it has no neighbours. The vertex $y$ has only one neighbour, $w$. The valency of $w$ is 2. The valency of $x$ and the valency of $y$ are both 1. In verifying all of these statements, drawing a diagram of the graph might help you.

**EXERCISES 11.2.11.** For each of the following graphs (which may or may not be simple, and may or may not have loops), find the valency of each vertex. Determine whether or not the graph is simple, and if there is any isolated vertex. List the neighbours of $a$, and all edges with which $a$ is incident.

1) Let $G$ be defined by $V = \{a, b, c, d, e\}$ and $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$ with $e_1 = \{a, c\}$, $e_2 = \{b, d\}$, $e_3 = \{c, d\}$, $e_4 = \{c, e\}$, $e_5 = \{d, e\}$, and $e_6 = \{e, e\}$.
2) Let $G$ be defined by $V = \{a, b, c\}$ and $E = \{e_1, e_2, e_3\}$ with $e_1 = \{a, b\}$, $e_2 = \{a, c\}$, and $e_3 = \{a, c\}$.
3) Let $G$ be defined by $V = \{a, b, c, d\}$ and $E = \{e_1, e_2, e_3\}$ with $e_1 = \{a, b\}$, $e_2 = \{a, c\}$, and $e_3 = \{b, c\}$.

**EXERCISES 11.2.12.**

1) Let $G$ be the graph whose vertices are the 2-element subsets of $\{1, 2, 3, 4, 5\}$, with vertices $\{a, b\}$ and $\{c, d\}$ adjacent if and only if $\{a, b\} \cap \{c, d\} = \emptyset$. Draw $G$.
2) The number of edges in the $k$-dimensional cube $Q_k$ (which is an important structure in network design, but you do not need to know the structure to solve this. We'll discuss $Q_k$ further in Section 12.5) can be found by the recurrence relation:

$$e(Q_0) = 0;\ e(Q_n) = 2e(Q_{n-1}) + 2^{n-1} \text{ for } n \geq 1.$$

Use generating functions to solve this recurrence relation and therefore determine the number of edges in the $k$-dimensional cube.

### 11.3. Subgraphs, complete graphs, and the Handshaking Lemma

We'll begin this section by introducing a basic operation that can change a graph (or a multigraph, with or without loops) into a smaller graph: deletion.

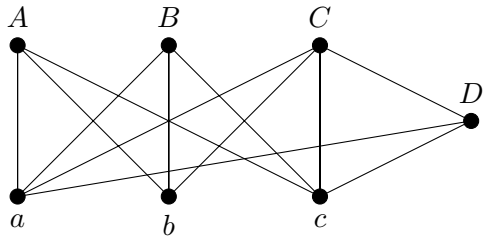
**DEFINITION 11.3.1.** Start with a graph (or multigraph, with or without loops) $G$ with vertex set $V$ and edge set $E$, and some vertex $v \in V$. If we **delete the vertex $v$** from the graph $G$, the resulting graph has vertex set $V \setminus \{v\}$ and edge set

$$E \setminus \{e \mid e \text{ is incident with } v\}.$$

**NOTATION 11.3.2.** The graph obtained by deleting the vertex $v$ from $G$ is denoted by $G \setminus \{v\}$. We can delete more than one vertex; for any set $S \subseteq V$ of vertices of $G$, we use $G \setminus S$ to denote the graph obtained by deleting all of the vertices of $S$ from $G$.

The graph $G \setminus \{v\}$ might be a multigraph, but only if $G$ is. It could have loops, but only if $G$ has loops.

If we begin with the graph




and delete the vertex $D$, then we obtain the graph shown in Figure 11.2.2.

We can also delete edges, rather than vertices.

**DEFINITION 11.3.3.** Start with a graph (or multigraph, with or without loops) $G$ with vertex set $V$ and edge set $E$, and some edge $e \in E$. If we **delete the edge $e$** from the graph $G$, the resulting graph has vertex set $V$ and edge set $E \setminus \{e\}$.

**NOTATION 11.3.4.** The graph obtained by deleting the edge $e$ from $G$ is denoted by $G \setminus \{e\}$. We can delete more than one edge; for any set $T \subseteq E$ of edges of $G$, we use $G \setminus T$ to denote the graph obtained by deleting all of the edges of $T$ from $G$.

The graph $G \setminus \{e\}$ might be a multigraph, but only if $G$ is. It could have loops, but only if $G$ has loops.

Notice that deleting the edges $\{C, D\}$, $\{a, D\}$ and $\{c, D\}$ from the graph drawn above, does *not* result in the graph shown in Figure 11.2.2, because the graph we obtain by deleting these edges still has the vertex $D$ (as an isolated vertex), whereas the graph shown in Figure 11.2.2 has only the six vertices $\{a, b, c, A, B, C\}$.

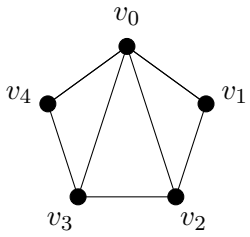
Vertex and edge deletion will be very useful for using proofs by induction on graphs (and multigraphs, with or without loops). It is handy to have terminology for a graph that can be obtained from another graph by deleting vertices and/or edges.

**DEFINITION 11.3.5.** Let $G$ be a graph. If $H$ can be obtained from $G$ by deleting vertices and/or edges, then $H$ is a **subgraph** of $G$. A subgraph $H$ of $G$ is **proper** if $H \neq G$.

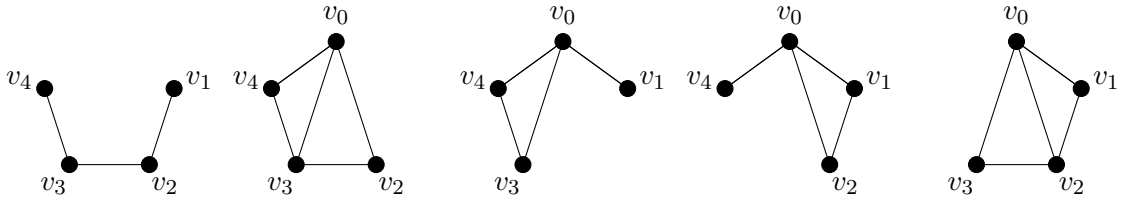
For some later discussions, a particular kind of subgraph is often important.

**DEFINITION 11.3.6.** An **induced subgraph** is a subgraph that can be obtained by deleting only vertices, with no additional edge deletions.

**EXAMPLE 11.3.7.** Consider the following graph. Draw all of its induced subgraphs on 4 vertices.




**SOLUTION.** For an induced subgraph, we can only delete vertices. To end with four vertices, we can only delete one of the vertices. There are five ways to do this (deleting each of $v_0$ through $v_4$ in turn.) The graphs we obtain are:




We now define a very important family of graphs, called *complete graphs*.

**DEFINITION 11.3.8.** A (simple) graph in which every vertex is adjacent to every other vertex, is called a **complete graph**. If this graph has $n$ vertices, then it is denoted by $K_n$.

The notation $K_n$ for a complete graph on $n$ vertices comes from the name of Kazimierz Kuratowski (1896—1980). Although his main area of research was logic, Kuratowski proved an important theorem that involves a complete graph. We'll study his theorem later in the course.

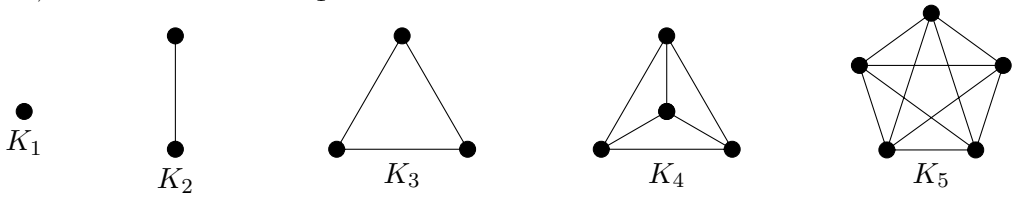
With this set-up, we are ready to prove our first result about graphs.

**PROPOSITION 11.3.9.** *The number of edges of $K_n$ is $\frac{n(n-1)}{2} = \binom{n}{2}$.*

We present two proofs of this proposition: first, a combinatorial proof; then, a proof by induction.

**COMBINATORIAL PROOF.** A complete graph has an edge between any pair of vertices. From $n$ vertices, there are $\binom{n}{2}$ pairs of vertices that must be joined by an edge for the graph to be complete. Thus, there are $\binom{n}{2}$ edges in $K_n$. □

Before giving the proof by induction, let's show a few of the small complete graphs. In particular, we'll need to have $K_1$ in mind as it will be the base case for our induction.




**PROOF BY INDUCTION.** *Base case:* $n = 1$. As we can see above, the graph $K_1$ has 0 edges. Also,

$$n(n-1)/2 = 1(0)/2 = 0.$$

So the equality holds for $n = 1$. This completes the proof of the base case.

*Inductive step:* We begin with the inductive hypothesis. Let $k \geq 1$ be arbitrary, and suppose that $K_k$ has $\binom{k}{2}$ edges.

We want to deduce that $K_{k+1}$ has $\binom{k+1}{2}$ edges. Start with $K_{k+1}$, and let the number of edges of this graph be $t$. Now we delete a vertex $v$ from $K_{k+1}$. By the definition of vertex deletion, we must delete every edge incident with $v$. Since $K_{k+1}$ is complete, $v$ is adjacent to every other vertex, so there are $k$ edges incident with $v$, and it is precisely these edges that we have deleted. There must be $t - k$ edges remaining.

Notice that deleting $v$ does not affect edges that are not incident with $v$. Therefore, if we consider any two vertices in the remaining graph, they will still be adjacent (since they were adjacent in $K_{k+1}$ and the edge between them was not deleted). Thus, the remaining graph is $K_k$.

Using our inductive hypothesis, we know that $K_k$ has $k(k-1)/2$ edges. We have shown that $t - k = k(k-1)/2$, so

$$t = \frac{k(k-1)}{2} + k = k\left(\frac{k-1)}{2} + 1\right) = \frac{k(k+1)}{2} = \binom{k+1}{2},$$

which is what we wanted to deduce. This completes the proof of the inductive step.

By the Principle of Mathematical Induction, $K_n$ has $\binom{n}{2}$ vertices for every $n \geq 1$. □

Although this proof by induction may seem ridiculously long and complicated in comparison with the combinatorial proof, it serves as a relatively simple illustration of how proofs by induction can work on graphs. This can be a very powerful technique for proving results about graphs.

One more definition relates to complete graphs, and will prove useful in a variety of contexts.

**DEFINITION 11.3.10.** The **complement** of $G$ (denoted $G^c$) is the graph with the same vertices as $G$, but whose edges are precisely the *non*-edges of $G$. (That is, $u$ is adjacent to $v$ in the complement of $G$ if and only $u$ is *not* adjacent to $v$ in $G$.) Therefore, if $G^c$ is the complement of $G$, then $E(K_{|V(G)|})$ is the disjoint union of $E(G)$ and $E(G^c)$.

It is worth noting that $\bar{G}$ is also a standard notation for the complement of the graph $G$.

Here is another result, due to Leonhard Euler (1707—1783) that can be proven using either a combinatorial proof, or a proof by induction.

**LEMMA 11.3.11 (Euler's handshaking lemma).** *For any graph (or multigraph, with or without loops)*

$$\sum_{v \in V} d(v) = 2|E|.$$

This is called the handshaking lemma because it is often explained using vertices to represent people, and edges as handshakes between people. In this explanation, the lemma says that if you add up all of the hands shaken by all of the people, you will get twice the number of handshakes that took place. This is an example of using two ways to count pairs $(v, e) \in V \times E$ such that $v$ is incident with $e$, a notion that we discussed briefly when we introduced combinatorial proofs.

**COMBINATORIAL PROOF.** For the left-hand side of the equation, at every vertex we count the number of edges incident with that vertex. To get the right-hand side from this, observe that this process results in every edge having been counted exactly twice (once at each of its two endvertices; or, in the case of a loop, twice at its single endvertex since both ends are there).□

Although from the right perspective the handshaking lemma might seem obvious, it has a very important and useful corollary.

**COROLLARY 11.3.12.** *Every graph has an even number of vertices of odd valency.*

**PROOF.** Since the sum of all of the valencies in the graph is even (by Euler's handshaking lemma), the number of odd summands in this sum must be even. That is, the number of vertices that have odd valency must be even. □

**EXERCISES 11.3.13.**

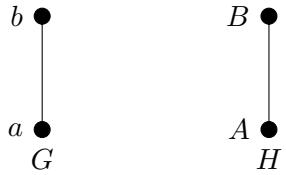
1) Give a proof by induction of Euler's handshaking lemma for simple graphs.
2) Draw $K_7$.
3) Show that there is a way of deleting an edge and a vertex from $K_7$ (in that order) so that the resulting graph is complete. Show that there is a way of deleting an edge and a vertex from $K_7$ (in that order) so that the resulting graph is not complete.
4) Prove Corollary 11.3.12 by induction on the number of edges. (Use edge deletion, and remember that the base case needs to be when there are no edges.)
5) Use graphs to give a combinatorial proof that

$$\sum_{i=1}^{k} \binom{n_i}{2} \leq \binom{n}{2},$$

where $n_1, n_2, \ldots, n_k$ are positive integers with $\sum_{i=1}^{k} n_i = n$. Under what circumstances does equality hold?

## 11.4. Isomorphism of graphs

There is a problem with the way we have defined $K_n$. A graph is supposed to consist of two sets, $V$ and $E$. Unless the elements of the sets are labeled, we cannot distinguish amongst them. Here are two graphs, $G$ and $H$:




Which of these graphs is $K_2$? They can't both be $K_2$ since they aren't the same graph — can they?

The answer lies in the concept of isomorphisms. Intuitively, graphs are isomorphic if they are identical except for the labels (on the vertices). Recall that as shown in Figure 11.2.3, since graphs are defined by the sets of vertices and edges rather than by the diagrams, two isomorphic graphs might be drawn so as to look quite different.

**DEFINITION 11.4.1.** Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are **isomorphic** if there is a bijection (a one-to-one, onto map) $\varphi$ from $V_1$ to $V_2$ such that for any $v, w \in V_1$,

$$\{v, w\} \in E_1 \Leftrightarrow \{\varphi(v), \varphi(w)\} \in E_2.$$

In this case, we call $\varphi$ an **isomorphism** from $G_1$ to $G_2$.

**NOTATION 11.4.2.** When $\varphi$ is an isomorphism from $G_1$ to $G_2$, we abuse notation by writing $\varphi : G_1 \to G_2$ even though $\varphi$ is actually a map on the vertex sets.

We also write $G_1 \cong G_2$ for "$G_1$ is isomorphic to $G_2$."

So a graph isomorphism is a bijection between the vertex sets that preserves edges and non-edges. If you have seen isomorphisms of other mathematical structures in other courses, they would have been bijections that preserved the key defining relation or relations of the structures they were mapping. For graphs, the key relation is which vertices are adjacent to each other. If that is preserved, then the networks being represented are for all intents and purposes, the same.

You may have seen previously that a relation is called an **equivalence relation** if it is a relation that satisfies three properties. It must be:

- **reflexive** (every object must be related to itself);
- **symmetric** (if object $A$ is related to object $B$, then object $B$ must also be related to object $A$); and
- **transitive** (if object $A$ is related to object $B$ and object $B$ is related to object $C$, then object $A$ must be related to object $C$).

The relation "is isomorphic to" is an equivalence relation on graphs. To see this, observe that:

- for any graph $G$, we have $G \cong G$ by the identity map on the vertices;
- for any graphs $G_1$ and $G_2$, we have

$$G_1 \cong G_2 \Leftrightarrow G_2 \cong G_1,$$

since any bijection has an inverse function that is also a bijection, and since

$$\{v, w\} \in E_1 \Leftrightarrow \{\varphi(v), \varphi(w)\} \in E_2$$

is equivalent to

$$\{\varphi^{-1}(v), \varphi^{-1}(w)\} \in E_1 \Leftrightarrow \{v, w\} \in E_2;$$

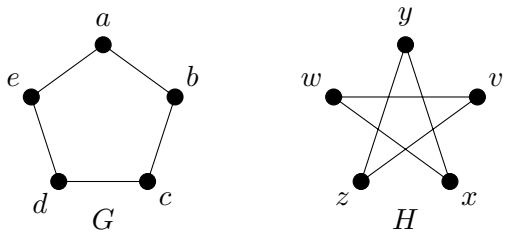
- for any graphs $G_1$, $G_2$, and $G_3$ with $\varphi_1 \colon G_1 \to G_2$ and $\varphi_2 \colon G_2 \to G_3$ being isomorphisms, the composition $\varphi_2 \circ \varphi_1 \colon G_1 \to G_3$ is a bijection, and

$$\{v, w\} \in E_1 \Leftrightarrow \{\varphi_1(v), \varphi_1(w)\} \in E_2 \Leftrightarrow \{\varphi_2(\varphi_1(v)), \varphi_2(\varphi_1(w))\} \in E_3,$$

so $G_1 \cong G_3$.

The answer to our question about complete graphs is that any two complete graphs on $n$ vertices are isomorphic, so even though technically the set of all complete graphs on 2 vertices is an equivalence class of the set of all graphs, we can ignore the labels and give the name $K_2$ to all of the graphs in this class.

**EXAMPLE 11.4.3.** The graphs $G$ and $H$:




are isomorphic. The map $\varphi$ defined by

- $\varphi(a) = v$;

- $\varphi(b) = z$;
- $\varphi(c) = y$;
- $\varphi(d) = x$;
- $\varphi(e) = w$

is an isomorphism. It is straightforward (though perhaps tedious) to check that each of the 5 edges of $G$ maps to one of the five edges of $H$ under $\varphi$.

To prove that two graphs are isomorphic, we must find a bijection that acts as an isomorphism between them. If we want to prove that two graphs are *not* isomorphic, we must show that *no* bijection can act as an isomorphism between them.

Sometimes it can be very difficult to determine whether or not two graphs are isomorphic. It is possible to create very large graphs that are very similar in many respects, yet are not isomorphic. A common approach to this problem has been attempting to find an "invariant" that will distinguish between non-isomorphic graphs. An "invariant" is any function that can be defined on graphs, that must produce the same output for all graphs in any isomorphism class. Thus, if you can find an invariant that is different for two graphs, you know that these graphs must not be isomorphic. We say in this case that this invariant *distinguishes between* these two graphs.

Mathematicians have come up with many, many graph invariants. Unfortunately, so far, for every efficiently-computable invariant it is possible to find two graphs that are not isomorphic, but for which the invariant is the same. In other words, no known efficiently-computable invariant distinguishes between every pair of non-isomorphic graphs. (We are glossing over some details here. What we have said is true if we mean "computable in polynomial time" when we say "efficiently-computable". In 2019, László Babai published an upgrade to the isomorphism test we discuss in the next paragraph, enabling his algorithm to produce an invariant that distinguishes between any pair of nonisomorphic graphs and is computable in quasipolynomial time.)

In case you know what this means (perhaps if you are studying computer science), the graph isomorphism problem is particularly interesting because it is one of a very few natural problems that are known to be in NP but that are not known to be either in P, or to be NP-complete. (Other classical examples are the discrete logarithm and integer factorisation problems.) Although no polynomial time algorithm for solving this problem is known, in 2015 (with a minor patch added in 2017) László Babai (1950—) at the University of Chicago found a quasipolynomial time algorithm for solving the graph isomorphism problem. This was a major development in our understanding of algorithmic complexity.

To give a bit more context for the importance of Babai's result without going into detail, we provide short definitions of various possible algorithmic complexities (running times). An algorithm runs in *polynomial time* if the number of operations it takes is at most $n^c$ for some constant $c$, where $n$ is a parameter describing the size of the input. An algorithm requires *exponential time* if the number of operations it takes is at least $2^{cn}$ for some positive constant $c$, for all sufficiently large $n$. The best known graph isomorphism test prior to Babai's algorithm required time exponential in $\sqrt{n}$ (this is better than being exponential in $n$). An algorithm is *quasipolynomial* if the number of operations it takes is at most $2^{c_1(\log(n))^{c_2}}$ for some positive constants $c_1$ and $c_2$. Note that if we can find a quasipolynomial algorithm with $c_2 = 1$, then this would actually run in polynomial time.

We give a few graph invariants in the following proposition.

**PROPOSITION 11.4.4.** *If $G_1 \cong G_2$ are graphs, then*

*1) $G_1$ and $G_2$ have the same number of vertices;*

*2) $G_1$ and $G_2$ have the same number of edges;*

*3) if we list the valency of every vertex of $G_1$ and do the same for $G_2$, the lists will be the same (though possibly in a different order). (Such a list is called the* **degree sequence** *of the graph.)*

**PROOF.**

1) Since $G_1 \cong G_2$, there is an isomorphism $\varphi\colon V_1 \to V_2$ (where $V_1$ is the vertex set of $G_1$ and $V_2$ is the vertex set of $G_2$). Since $\varphi$ is a bijection, we must have $|V_1| = |V_2|$.

2) Since

$$\{v, w\} \in E_1 \Rightarrow \{\varphi(v), \varphi(w)\} \in E_2,$$

we see that for every edge of $E_1$, there is an edge of $E_2$. Therefore, $|E_2| \geq |E_1|$. Similarly, since

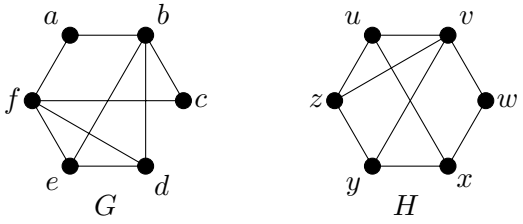
$$\{\varphi(v), \varphi(w)\} \in E_2 \Rightarrow \{v, w\} \in E_1,$$

we see that $|E_1| \geq |E_2|$. So $|E_1| = |E_2|$.

3) If $\varphi(v_1) = v_2$ then $d_{G_1}(v_1) = d_{G_2}(v_2)$, because $u \sim v_1$ if and only if $\varphi(u) \sim v_2$. □

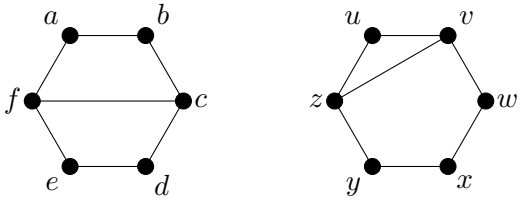
**EXAMPLE 11.4.5.** The graph $G$ of Example 11.4.3 is not isomorphic to $K_5$, because $K_5$ has $\binom{5}{2} = 10$ edges by Proposition 11.3.9, but $G$ has only 5 edges. Notice that the number of vertices, despite being a graph invariant, does not distinguish these two graphs.

The graphs $G$ and $H$:




are not isomorphic. Each of them has 6 vertices and 9 edges. However, the graph $G$ has two vertices of valency 2 ($a$ and $c$), two vertices of valency 3 ($d$ and $e$), and two vertices of valency 4 ($b$ and $f$). Meanwhile, the graph $H$ has one vertex of valency 2 ($w$), four vertices of valency 3 ($u$, $x$, $y$, and $z$), and one vertex of valency 4 ($v$). Although each of these lists has the same values (2s, 3s, and 4s), the lists are not the same since the number of entries that contain each of the values is different. In particular, the two vertices $a$ and $c$ both have valency 2, but there is only one vertex of $H$ (vertex $w$) of valency two. Either $a$ or $c$ could be sent to $w$ by an isomorphism, but either choice leaves no possible image for the other vertex of valency 2. Therefore, an isomorphism between these graphs is not possible.

Observe that the two graphs




both have 6 vertices and 7 edges, and each has four vertices of valency 2 and two vertices of valency 3. Nonetheless, these graphs are not isomorphic. Perhaps you can think of another graph invariant that is not the same for these two graphs.

To prove that these graphs are not isomorphic, since each has two vertices of valency 3, any isomorphism would have to map $\{c, f\}$ to $\{v, z\}$. Now, whichever vertex gets mapped to $u$ must be a mutual neighbour of $c$ and $f$ since $u$ is a mutual neighbour of $v$ and $z$. But $c$ and $f$ have no mutual neighbours, so this is not possible. Therefore there is no isomorphism between these graphs.

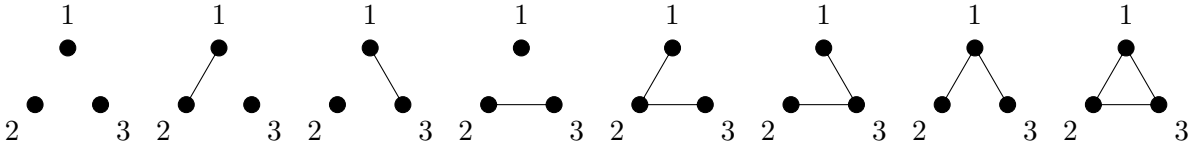
A natural problem to consider is: how many different graphs are there on $n$ vertices? If we are not worrying about whether or not the graphs are isomorphic, we could have infinitely many graphs just by changing the labels on the vertices, and that's not very interesting. To avoid this problem, we fix the set of labels that we use. Label the vertices with the elements of $\{1, \ldots, n\}$. We'll call the number of graphs we find, the number of *labeled* graphs on $n$ vertices.

Any edge is a 2-subset of $\{1, \ldots, n\}$. There are $\binom{n}{2}$ possible edges in total. Any graph is formed by taking a subset of the $n(n-1)/2$ possible edges. In Example 4.1.1, we learned how to count these: there are $2^{n(n-1)/2}$ subsets.

**EXAMPLE 11.4.6.** When $n = 1$, we have $\binom{1}{2} = 0$, and $2^0 = 1$, so there is exactly one labeled graph on 1 vertex. It looks like this:


When $n = 2$, we have $\binom{2}{2} = 1$, and $2^1 = 2$. so there are exactly two labeled graphs on 2 vertices. They look like this:

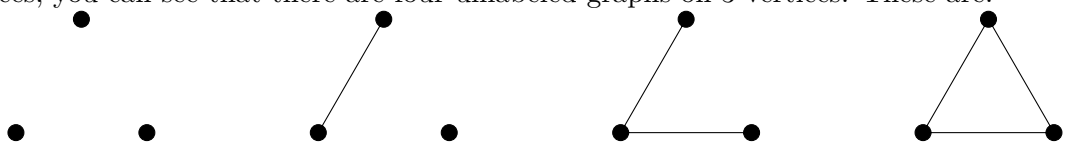

When $n = 3$, we have $\binom{3}{2} = 3$, and $2^3 = 8$, so there are exactly eight labeled graphs on 3 vertices. They look like this:




When $n = 4$, we have $\binom{4}{2} = 6$, and $2^6 = 64$, so there are exactly sixty-four labeled graphs on 4 vertices. We won't attempt to draw them all here.

Although that answer is true as far as it goes, you will no doubt observe that even though we are using a fixed set of labels, some of the graphs we've counted are isomorphic to others. A more interesting question would be, how many isomorphism classes of graphs are there on $n$ vertices? Since we are considering isomorphism classes, the labels we choose for the vertices are largely irrelevant except to tell us which vertices are adjacent to which other vertices, if we don't have a diagram. Thus, if we are drawing the graphs, we usually omit vertex labels and refer to the resulting graphs (each of which represents an isomorphism class) as *unlabeled*. So the question is, how many *unlabeled* graphs are there on $n$ vertices?

We can work out the answer to this for small values of $n$. From the labeled graphs on 3 vertices, you can see that there are four unlabeled graphs on 3 vertices. These are:

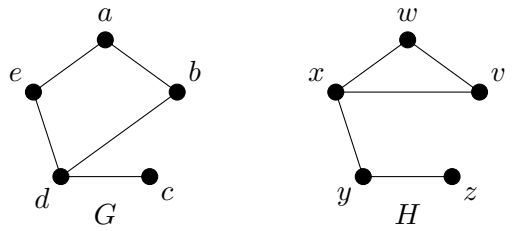


There are 11 unlabeled graphs on four vertices. Unfortunately, since there is no known polynomial-time algorithm for solving the graph isomorphism problem, determining the number of unlabeled graphs on $n$ vertices gets very hard as $n$ gets large, and no general formula is known.

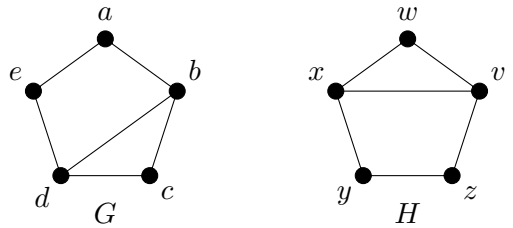
(All of these things could still be hard even if a polynomial time algorithm were found, but without such an algorithm these other questions seem completely out of reach.)

**EXERCISES 11.4.7.** For each of the following pairs of graphs, find an isomorphism or prove that the graphs are not isomorphic.

1)




2)




3) $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ with $V_1 = \{a, b, c, d\}$, $V_2 = \{A, B, C, D\}$, $E_1 = \{ab, ac, ad\}$, $E_2 = \{BC, CD, BD\}$.

**EXERCISES 11.4.8.**

1) Draw five unlabeled graphs on 5 vertices that are not isomorphic to each other.

2) How many labeled graphs on 5 vertices have 1 edge?

3) How many labeled graphs on 5 vertices have 3 or 4 edges?

## 11.5. Random graphs

This section is intended as optional enrichment material. In the current edition, no exercises are included. The main goal of this section is to make you aware of the existence of random graphs, without actually studying their properties for the most part.

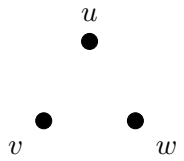
There are two commonly-used models of random graphs on $n$ vertices: the model due to Edgar Nelson Gilbert (1923—2013) (the "Gilbert model") and the model due to Pál Erdős (1913—1996) and Alfréd Rényi (1921—1970) (the "Erdős-Rényi model"). In each case, we consider $n$ (the number of vertices) to be fixed, and use probability (randomness) to determine the edges of our graph.

**DEFINITION 11.5.1.** In the **Gilbert random graph model**, a probability $p$ is chosen with $0 < p < 1$. Let the vertex set be

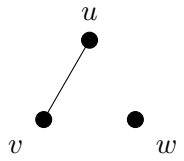
$$V = \{v_1, \ldots, v_n\}.$$

For every pair of vertices $v_i$ and $v_j$ with $1 \leq i < j \leq n$, the edge $v_i v_j$ occurs randomly and independently with probability $p$.

**EXAMPLE 11.5.2.** Suppose we have three vertices $u$, $v$, and $w$, and we fix $p = 1/4$. For each pair of vertices, we flip a coin twice, and draw an edge between the vertices only if we obtain heads both times. We have three pairs of vertices, so flip the coin 6 times, with the first two flips are for $uv$, the next two for $uw$, and the final two are for $vw$. The outcomes on our first attempt are T, T, T, H, T, H. Our result is the empty graph, shown below.




On our second attempt, the outcomes are H, H, T, H, H, T. The result is again shown below.




Under the Gilbert model, the probability of obtaining a given graph that has $m$ edges (where $0 \leq m \leq \binom{n}{2}$) is $p^m(1-p)^{\binom{n}{2}-m}$. If $p$ is large, obtaining a graph that has many edges is much more likely than a graph that has few edges; if $p$ is small, a graph with few edges is more likely. If $p = 1/2$, then any given (labelled) graph will be equally likely to arise.

**DEFINITION 11.5.3.** In the **Erdős-Rényi random graph model**, the number of edges $m$ is also fixed in advance. From all possible (labelled) graphs on $n$ vertices and with $m$ edges, one is chosen at random, with equal probability of choosing any one.

Under the Erdős-Rényi model, since there are $\binom{\binom{n}{2}}{m}$ (labelled) graphs with $m$ edges on $n$ vertices, the probability of choosing a specific graph is $1/\binom{\binom{n}{2}}{m}$.

Probability can be applied to graph theory in some very powerful ways, and Paul Erdős was a pioneer in much of this work. In addition to studying random graphs, one of the techniques he applied with great success was non-constructive probabilistic proofs. That sounds big and complicated, but it's actually a bit like something we saw with the Pigeonhole Principle: it's possible to prove that something exists, without being able to find it.

In the case of probabilistic proofs, the basic idea is this: if you can prove that the probability that a graph has a particular property is not 0, then a graph with that property must exist! Such a proof is "non-constructive" as it does not construct (or tell us how to construct) a graph that has the property.

Although our attention in this course is restricted to finite graphs, random graphs are particularly interesting in the infinite case. If our vertex set is countably infinite and we apply the Gilbert random graph model with any allowed choice of $p$ (so $p$ is not 0 or 1, but can be anything in between), then almost certainly the result we obtain will be isomorphic to a specific graph. ("Almost certainly" has a very precise meaning in this context, that boils down to the probability of its occurrence being 1. In practice this may mean something like the probability of its happening being the limit as $k$ tends to infinity of $1 - p^k$.)

The infinite random graph that almost certainly arises from this process is known as the Rado graph, named for Richard Rado (1906—1989), but it also often just called the random graph, due to its omnipresence. It has many very interesting properties. Perhaps some of the most interesting are that:

- every finite graph appears as an *induced* subgraph of the random graph.
- given any sets $v_1, \ldots, v_n$ and $u_1, \ldots, u_m$ of vertices in the random graph, it is possible to find a vertex $x$ that is adjacent to all of the vertices $v_1, \ldots, v_n$ and not adjacent to any of the vertices $u_1, \ldots, u_m$.

If we allow an uncountably infinite number of vertices, then in contrast to the countable situation, there are many nonisomorphic random graphs.

Despite the apparent universality of the Rado graph, other random graphs on countably infinite sets of vertices can almost certainly arise if other random models are used. In particular, in a 2011 paper, Anthony Bonato (1971—) and Jeannette Janssen (1963—) studied random graphs whose vertices correspond to a dense but countably infinite subset of the real numbers, with one additional condition on the vertex set that is not particularly restrictive, but is too technical to include in this overview. In their model, they fix a distance $D$ and a probability $p$ (with $0 < p < 1$). For any two vertices $x$ and $y$ with $|x - y| < D$, the edge $xy$ occurs with probability $p$ (so there are no edges between vertices that are too far apart). Under this model, they were able to show that a specific graph (up to isomorphism) is almost certainly the outcome. However, this graph is easily seen to be distinct from the random graph, since in the random graph any two nonadjacent vertices have a mutual neighbour, whereas in this graph there are vertices that are not joined by a path of length $n$ for any finite $n$.

In fact, Bonato and Janssen showed that a similar result holds if the vertices are chosen to be dense countably infinite subsets of $\mathbb{R}^n$, as long as the probability of two vertices being adjacent is determined by their maximum difference in any coordinate (that is, for vertices

$$v = (v_1, \ldots, v_n) \text{ and } u = (u_1, \ldots, u_n),$$

we only use our probability to assign an edge if $|u_i - v_i| < D$ for every $1 \leq i \leq n$). Under this model, for any fixed dimension we again obtain a unique graph (up to isomorphism) that depends only on our choice of the dimension $n$. In contrast, they show that if we decide whether or not to use probability to determine the existence of an edge on the basis of the usual Euclidean distance between the corresponding vertices (instead of by the maximum difference in any coordinate) then even in $\mathbb{R}^2$ more than one graph may be obtained.

**SUMMARY:**

- graphs are defined by sets, not by diagrams
- deleting a vertex or edge
- how to use proofs by induction on graphs
- Euler's handshaking lemma
- graph invariants, distinguishing between graphs
- labeled and unlabeled graphs
- random graphs
- Important definitions:
  - graph, vertex, edge
  - loop, multiple edge, multigraph, simple graph
  - endvertex, incident, adjacent, neighbour
  - degree, valency, isolated vertex
  - subgraph, induced subgraph
  - complete graph, complement of a graph
  - isomorphic graphs, isomorphism between graphs
- Notation:
  - $u \sim v$
  - $uv$
  - $\mathrm{val}(v)$, $\deg(v)$, $d(v)$, $d_G(v)$
  - $G \setminus \{v\}$, $G \setminus \{e\}$
  - $K_n, G^c$
  - $G_1 \cong G_2$

Chapter 12

# Moving through graphs

We have some basic concepts and terminology now, but it is important to remember that graphs are models of networks. Networks are all about moving things around, whether those things are cars, data, or whatever. So if graphs are going to be useful models, routes in the network need to correspond to routes in the graph, and we need to be able to describe these routes, and to learn about them.

## 12.1. Directed graphs

Some networks include connections that only allow travel in one direction (one-way roads; transmitters that are not receivers, etc.). These can be modeled using *directed graphs*.

**DEFINITION 12.1.1.** A **directed graph**, or **digraph** for short, consists of two sets:

- $V$, whose elements are the vertices of the digraph; and
- $A$, whose elements are ordered pairs from $V$, so

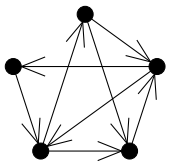
$$A \subseteq \{(v_1, v_2) \mid v_1, v_2 \in V\}.$$

The elements of $A$ are referred to as the **arcs** of the digraph.

When drawing a digraph, we draw an arrow on each arc so that it points from the first vertex of the ordered pair to the second vertex.

Like multigraphs, we will not study digraphs in this course, but you should be aware of the basic definition. Many of the results we will cover in this course, generalise to the context of digraphs.

**EXAMPLE 12.1.2.** A digraph.



We will give one example of generalising a result on graphs, to the context of digraphs. In order to do so, we need a definition.

**DEFINITION 12.1.3.** The **outvalency** or **outdegree** of a vertex $v$ in a digraph is the number of arcs whose first entry is $v$, i.e.,

$$|\{w \in V \mid (v, w) \in A\}|.$$

The **invalency** or **indegree** of a vertex $v$ in a digraph is the number of arcs whose second entry is $v$.

**NOTATION 12.1.4.** The outvalency of vertex $v$ is denoted by $d^+(v)$. The invalency of vertex $v$ is denoted by $d^-(v)$.

**LEMMA 12.1.5 (Euler's handshaking lemma for digraphs).** *For any digraph,*

$$\sum_{v \in V} d^+(v) = \sum_{v \in V} d^-(v) = |A|.$$

**COMBINATORIAL PROOF.** For the left-hand side of the equation, at every vertex we count the number of arcs that begin at that vertex. Since each of these arcs ends at some vertex, we get the same result in the middle part of the equation, where at every vertex we count the number of arcs that end at that vertex. In each case, we have counted every arc precisely once, so both of these values are equal to the right-hand side of the equation, the number of arcs in the digraph. □

**EXERCISES 12.1.6.**

1) Use induction to prove Euler's handshaking lemma for digraphs that have no loops (arcs of the form $(v, v)$ or multiarcs (more than one arc from some vertex $u$ to some other vertex $v$).
2) A digraph isomorphism is a bijection on the vertices that preserves the arcs. Come up with a digraph invariant, and prove that it is an invariant.
3) List the invalency and outvalency of each vertex of the digraph from Example 12.1.2.

## 12.2. Walks and connectedness

Graphs can be connected or disconnected. Intuitively, this corresponds to the network being connected or disconnected — is it possible to travel from any node to any other node? When a graph (or network) is disconnected, it has broken down into some number of separate *connected components* - the pieces that still are connected.

Since this is mathematics, we require more formal definitions, to ensure that the meanings are not open to misunderstanding. Before we can define connectedness, we need the concept of a *walk* in a graph.

**DEFINITION 12.2.1.** A **walk** in a graph $G$ is a sequence of vertices $(u_1, u_2, \ldots, u_n)$ such that for every $1 \leq i \leq n-1$, we have $u_i \sim u_{i+1}$. (That is, consecutive vertices in the walk must be adjacent.)

A $\boldsymbol{u - v}$ **walk** in $G$ is a walk with $u_1 = u$ and $u_n = v$. (That is, a walk that begins at $u$ and ends at $v$.)

Now we can define what it means for a graph to be connected.

**DEFINITION 12.2.2.** The graph $G$ with vertex set $V$ is **connected** if for every $u, v \in V$, there is a $u - v$ walk.

The **connected component** of $G$ that contains the vertex $u$, is

$$\{v \in V \mid \text{ there is a } u - v \text{ walk. }\}.$$

This definition of connected component seems to depend significantly on the choice of the vertex $u$. In fact, though, *being in the same connected component of* $G$ is an equivalence relation on the vertices of $G$, so the connected components of $G$ are a property of $G$ itself, rather than depending on particular choices of vertices. We won't go through a formal proof that being in the same connected component is an equivalence relation (we leave this as an exercise below), but we will go through the proof of a proposition that is closely related.

**PROPOSITION 12.2.3.** *Let $G$ be a graph, and let $u, v, w \in V(G)$. Suppose that $v$ and $w$ are in the connected component of $G$ that contains the vertex $u$. Then $w$ is in the connected component of $G$ that contains the vertex $v$.*

**PROOF.** Since $v$ and $w$ are in the connected component of $G$ that contains the vertex $u$, by definition there is a $u - v$ walk, and a $u - w$ walk. Let $(u = u_1, u_2, \ldots, u_k = w)$ be a $u - w$ walk, and let $(u = v_1, v_2, \ldots, v_m = v)$ be a $u - v$ walk.

We need to show that $w$ is in the connected component of $G$ that contains the vertex $v$; by definition, this is equivalent to showing that there is a $v - w$ walk. Consider the sequence of vertices:

$$(v = v_m, v_{m-1}, \ldots, v_1 = u = u_1, u_2, \ldots, u_k = w).$$

Since $(u = v_1, v_2, \ldots, v_m = v)$ is a $u - v$ walk, consecutive vertices are adjacent, so consecutive vertices in the first part of the given sequence (from $v_m$ through $v_1 = u$) are adjacent. Similarly, since $(u = u_1, u_2, \ldots, u_k = w)$ is a $u - w$ walk, consecutive vertices are adjacent, so consecutive vertices in the last part of the given sequence (from $u = u_1$ through $u_k$) are adjacent. Therefore, the given sequence is in fact a $v - w$ walk, as desired. □

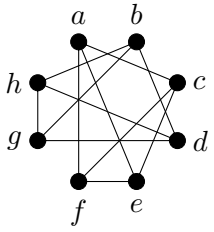
When discussing walks, it is convenient to have standard terminology for describing the length of the walk.

**DEFINITION 12.2.4.** The **length of a walk** is one less than the number of vertices in the walk. Thus, if we think of a walk as a sequence of edges (formed by consecutive pairs of vertices from the walk), the length of the walk is the number of edges in the walk.

Unfortunately, there is some disagreement amongst mathematicians as to whether the *length* of a walk should be used to mean the number of vertices in the walk, or the number of edges in the walk. We will use the latter convention throughout this course because it is consistent with the definition of the length of a cycle (which will be introduced in the next section). You should be aware, though, that you might find the other convention used in other sources.

Sometimes it is obvious that a graph is disconnected from the way it has been drawn, but sometimes it is less obvious. In the following example, you might not immediately notice whether or not the graph is connected.

**EXAMPLE 12.2.5.** Consider the following graph.




Find a walk of length 4 from $a$ to $f$. Find the connected component that contains $a$. Is the graph connected?

**SOLUTION.** A walk from of length 4 from $a$ to $f$ is $(a, c, a, c, f)$. (Notice that the vertices and edges used in a walk need not be distinct.) Remember that the length of this walk is the number of edges used, which is one less than the number of vertices in the sequence!

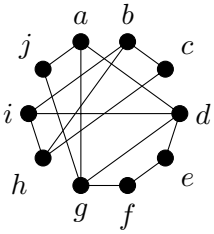
The connected component that contains $a$ is $\{a, c, e, f\}$. There are walks from $a$ to each of these vertices, but there are no edges between any of these vertices and any of the vertices $\{b, d, g, h\}$.

Since there is no walk from $a$ to $b$ (for example), the graph is not connected. □

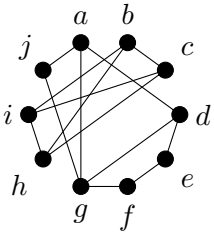
## EXERCISES 12.2.6.

1) Prove that *being in the same connected component of* $G$ is an equivalence relation on the vertices of any graph $G$.

2) Is the following graph connected? Find the connected component that contains $a$. Find a walk of length 5 from $a$ to $f$.




3) Is the following graph connected? Find the connected component that contains $a$. Find a walk of length 3 from $a$ to $d$.




4) Use Euler's handshaking lemma to prove (by contradiction) that if $G$ is a connected graph with $n$ vertices and $n - 1$ edges, and $n \geq 2$, then $G$ has at least 2 vertices of valency 1. [*Hint:* What does $G$ being connected imply about $d(v)$ for any vertex $v$ of $G$?]

5) Fix $n \geq 1$. Prove by induction on $m$ that for any $m \geq 0$, a graph with $n$ vertices and $m$ edges has at least $n - m$ connected components.

## 12.3. Paths and cycles

Recall the definition of a walk, Definition 12.2.1. As we saw in Example 12.2.5, the vertices and edges in a walk do not need to be distinct.

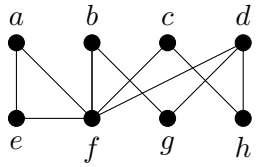
There are many circumstances under which we might not want to allow edges or vertices to be re-visited. Efficiency is one possible reason for this. We have a special name for a walk that does not allow vertices to be re-visited.

**DEFINITION 12.3.1.** A walk in which no vertex appears more than once is called a **path**.

**NOTATION 12.3.2.** For $n \geq 0$, a graph on $n+1$ vertices whose only edges are those used in a path of length $n$ (which is a walk of length $n$ that is also a path) is denoted by $P_n$. (Notice that $P_0 \cong K_1$ and $P_1 \cong K_2$.)

Notice that if an edge were to appear more than once in a walk, then at least one of its endvertices would also have to appear more than once, so a path does not allow vertices or edges to be re-visited.

**EXAMPLE 12.3.3.** In the graph




$(a, f, c, h)$ is a path of length 3. However, $(a, f, c, h, d, f)$ is not a path, even though no edges are repeated, since the vertex $f$ appears twice. Both *are* walks.

**PROPOSITION 12.3.4.** *Suppose that $u$ and $v$ are in the same connected component of a graph. Then any $u - v$ walk of minimum length is a path. In particular, if there is a $u - v$ walk, then there is a $u - v$ path.*

**PROOF.** Since $u$ and $v$ are in the same connected component of a graph, there is a $u - v$ walk.

Towards a contradiction, suppose that we have a $u - v$ walk of minimum length that is not a path. By the definition of a path, this means that some vertex $x$ appears more than once in the walk, so the walk looks like:

$$(u = u_1, \ldots, u_i = x, \ldots, u_j = x, \ldots, u_k = v),$$

and $j > i$. Observe that the following is also a $u - v$ walk:

$$(u = u_1, \ldots, u_i = x, u_{j+1}, u_{j+2}, \ldots, u_k = v).$$

Since consecutive vertices were adjacent in the first sequence, they are also adjacent in the second sequence, so the second sequence is a walk. The length of the first walk is $k - 1$, and the length of the second walk is $k - 1 - (j - i)$. Since $j > i$, the second walk is strictly shorter than the first walk. In particular, the first walk was not a $u - v$ walk of minimum length. This contradiction serves to prove that every $u - v$ walk of minimum length is a path. □

This allows us to prove another interesting fact that will be useful later.

**PROPOSITION 12.3.5.** *Deleting an edge from a connected graph can never result in a graph that has more than two connected components.*

**PROOF.** Let $G$ be a connected graph, and let $uv$ be an arbitrary edge of $G$. If $G \setminus \{uv\}$ is connected, then it has only one connected component, so it satisfies our desired conclusion. Thus, we assume in the remainder of the proof that $G \setminus \{uv\}$ is not connected.

Let $G_u$ denote the connected component of $G \setminus \{uv\}$ that contains the vertex $u$, and let $G_v$ denote the connected component of $G \setminus \{uv\}$ that contains the vertex $v$. We aim to show that $G_u$ and $G_v$ are the only connected components of $G \setminus \{uv\}$.

Let $x$ be an arbitrary vertex of $G$, and suppose that $x$ is a vertex that is not in $G_u$. Since $G$ is connected, there is a $u - x$ walk in $G$, and therefore by Proposition 12.3.4 there is a $u - x$ path in $G$. Since $x$ is not in $G_u$, this $u - x$ path must use the edge $u - v$, so must start with this edge since $u$ only occurs at the start of the path. Therefore, by removing the vertex $u$ from the start of this path, we obtain a $v - x$ path that does not use the vertex $u$. This path cannot use the edge $uv$, so must still be a path in $G \setminus \{uv\}$. Therefore $x$ is a vertex in $G_v$.

Since $x$ was arbitrary, this shows that every vertex of $G$ must be in one or the other of the connected components $G_u$ and $G_v$, so there are at most two connected components of $G \setminus \{uv\}$. Since $uv$ was an arbitrary edge of $G$ and $G$ was an arbitrary connected graph, this shows that deleting any edge of a connected graph can never result in a graph with more than two connected components. □

A cycle is like a path, except that it starts and ends at the same vertex. The structures that we will call cycles in this course, are sometimes referred to as *circuits*.

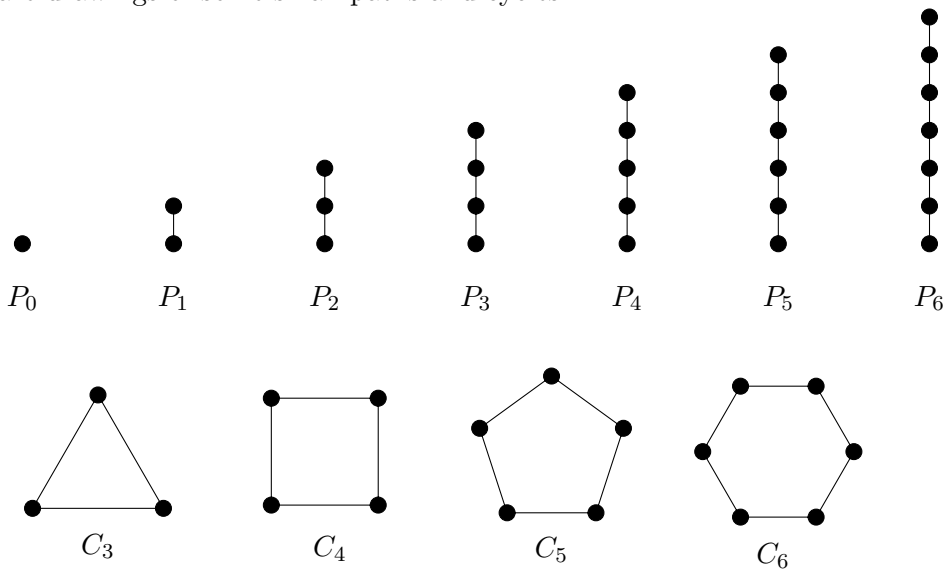
**DEFINITION 12.3.6.** A walk of length at least 1 in which no vertex appears more than once, except that the first vertex is the same as the last, is called a **cycle**.

**NOTATION 12.3.7.** For $n \geq 3$, a graph on $n$ vertices whose only edges are those used in a cycle of length $n$ (which is a walk of length $n$ that is also a cycle) is denoted by $C_n$.

The requirement that the walk have length at least 1 only serves to make it clear that a walk of just one vertex is not considered a cycle. In fact, a cycle in a simple graph must have length at least 3.

**EXAMPLE 12.3.8.** In the graph from Example 12.3.3, $(a, e, f, a)$ is a cycle of length 3, and $(b, g, d, h, c, f, b)$ is a cycle of length 6.

Here are drawings of some small paths and cycles:

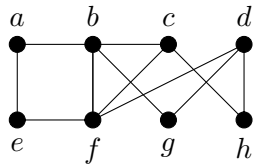



We end this section with a proposition whose proof will be left as an exercise.

**PROPOSITION 12.3.9.** *Suppose that $G$ is a connected graph. If $G$ has a cycle in which $u$ and $v$ appear as consecutive vertices (so $uv$ is an edge of $G$) then $G \setminus \{uv\}$ is connected.*

**EXERCISES 12.3.10.**

1) In the graph




   (a) Find a path of length 3.

   (b) Find a cycle of length 3.

   (c) Find a walk of length 3 that is neither a path nor a cycle. Explain why your answer is correct.

2) Prove that in a graph, any walk that starts and ends with the same vertex and has the smallest possible non-zero length, must be a cycle.

3) Prove Proposition 12.3.9.

4) Prove by induction that if every vertex of a connected graph on $n \geq 2$ vertices has valency 1 or 2, then the graph is isomorphic to $P_n$ or $C_n$.

5) Let $G$ be a (simple) graph on $n$ vertices. Suppose that $G$ has the following property: whenever $u \nsim v$, $d_G(u) + d_G(v) \geq n - 1$. Prove that $G$ is connected.

## 12.4. Trees

A special class of graphs that arise often in graph theory, is the class of trees. If a mathematician suspects that something is true for all graphs, one of the first families of graphs for which they will probably try to prove it, is the family of trees. The strong structure of trees makes them much easier to work with than many other families of graphs.

**DEFINITION 12.4.1.** A **tree** is a connected graph that has no cycles.

A **forest** is a disjoint union of trees. So a forest is a graph that has no cycles (but need not be connected).

A **leaf** is a vertex of valency 1 (in any graph, not just in a tree or forest).

Notice that the graph $P_n$ is a tree, for every $n \geq 1$.

We prove some important results about the structure of trees.

**PROPOSITION 12.4.2.** *Let $T$ be a connected graph with no cycles. Then deleting any edge from $T$ disconnects the graph.*

**PROOF.** If $T$ has no edges, the statement is vacuously true. We may thus assume that $T$ has at least one edge. Let $\{u, v\}$ be an arbitrary edge of $T$. (Since a loop is a cycle, we must have $u \neq v$ even if we were not assuming that our graphs are simple.)

Towards a contradiction, suppose that deleting $\{u, v\}$ from $T$ does not disconnect $T$. Then by the definition of a connected graph, there is a $u - v$ walk in $T \setminus \{uv\}$. By Proposition 12.3.4, the shortest $u - v$ walk in $T \setminus \{uv\}$ must be a $u - v$ path. If we take this same walk in $T$ and add $u$ to the end, this will still be a walk in $T$ since $T$ contains the edge $uv$. Since the walk in $T \setminus \{uv\}$ was a path, no vertices were repeated. Adding $u$ to the end of this walk makes a walk (certainly of length at least 2) in which no vertex is repeated except that the first and last

vertices are the same: by definition, a cycle. Thus, $T$ has a cycle, contradicting our hypothesis. This contradiction serves to prove that deleting any edge from $T$ disconnects the graph. □

Since a tree is a connected graph with no cycles, this shows that deleting any edge from a tree will disconnect the graph.

**PROPOSITION 12.4.3.** *Every tree that has at least one edge, has at least two leaves.*

We will provide two proofs of this result. The first is a direct proof; the second is longer and more complicated, but provides an example of a proof using strong induction on graphs. Since induction on graphs is often a technique that students struggle with, you may find it useful to work at understanding this proof also.

**DIRECT PROOF.** Consider any tree that has at least one edge, and find a longest path in this tree: say this path is $(v_1, v_2, \ldots, v_t)$. Since the tree has at least one edge, $t \geq 2$ (we could use the endvertices of the edge to get a path with two vertices). This means that $v_1$ and $v_t$ are distinct. We claim that $v_1$ and $v_t$ are both leaves.

Towards a contradiction, suppose for a moment that $v_1$ is not a leaf. Then it has a neighbour besides $v_2$. If it has at least one neighbour from $\{v_3, \ldots, v_t\}$, let $i$ be the smallest value greater than 2 such that $v_i$ is a neighbour of $v_1$. Then $(v_1, v_2, \ldots, v_i, v_1)$ is a cycle. This is impossible, since we are in a tree. On the other hand, if $v_1$ has a neighbour $w$ that is not in $\{v_3, \ldots, v_t\}$, then $(w, v_1, \ldots, v_t)$ is a longer path in the tree, contradicting our choice of a longest path. This argument shows that $v_1$ cannot have any neighbours besides $v_2$, so $v_1$ is a leaf.

A very similar argument shows that $v_t$ is a leaf. For the sake of completeness, we will include the full argument here. Towards a contradiction, suppose for a moment that $v_t$ is not a leaf. Then it has a neighbour besides $v_{t-1}$. If it has at least one neighbour from $\{v_1, \ldots, v_{t-2}\}$, let $i$ be the largest value less than $t-2$ such that $v_i$ is a neighbour of $v_t$. Then $(v_t, v_{t-1}, \ldots, v_i, v_t)$ is a cycle. This is impossible, since we are in a tree. On the other hand, if $v_t$ has a neighbour $w$ that is not in $\{v_1, \ldots, v_{t-2}\}$, then $(v_1, \ldots, v_t, w)$ is a longer path in the tree, contradicting our choice of a longest path. This argument shows that $v_t$ cannot have any neighbours besides $v_{t-1}$, so $v_t$ is a leaf.

Since $v_1$ and $v_t$ are distinct vertices and both are leaves, and we started with an arbitrary tree with at least one edge, we conclude that any tree with at least one edge has at least two leaves. □

**PROOF USING STRONG INDUCTION.** We prove this by strong induction on the number of vertices. Notice that a (simple) graph on one vertex must be $K_1$, which has no edges, so the proposition does not apply. Therefore our base case will be when there are 2 vertices.

*Base case:* $n = 2$. Of the two (unlabeled) graphs on 2 vertices, only one is connected: $K_2$ (or $P_1$; these are isomorphic). Both of the vertices have valency 1, so there are two leaves. This completes the proof of the base case.

*Induction step:* We begin with the strong inductive hypothesis. Let $k \geq 2$ be arbitrary. Suppose that for every $2 \leq i \leq k$, every tree with $i$ vertices has at least two leaves. (Since $i \geq 2$ and a tree is a connected graph, every tree on $i$ vertices has at least one edge, so we may omit this part of the hypothesis.)

Let $T$ be a tree with $k+1$ vertices. Since $k+1 > 1$, $T$ has at least one edge. Choose any edge $\{u, v\}$ of $T$, and delete it. By Proposition 12.4.2, the resulting graph is disconnected. By Proposition 12.3.5, it cannot have more than two connected components, so it must have exactly two connected components. Furthermore, by the proof of that proposition, the components are $T_u$ (the connected component that contains the vertex $u$) and $T_v$ (the connected component that contains the vertex $v$).

Since $T$ has no cycles, neither do $T_u$ or $T_v$. Since they are connected components, they are certainly connected. Therefore, both $T_u$ and $T_v$ are trees. Since $u$ is not a vertex of $T_v$ and $v$ is not a vertex of $T_u$, each of these trees has at most $k$ vertices.

If both $T_u$ and $T_v$ have at least two vertices, then we can apply our induction hypothesis to both. This tells us that $T_u$ and $T_v$ each have at least two leaves. In particular, $T_u$ must have some leaf $x$ that is not $u$, and $T_v$ must have some leaf $y$ that is not $v$. Deleting $uv$ from $T$ did not change the valency of either $x$ or $y$, so $x$ and $y$ must also have valency 1 in $T$. Therefore $T$ has at least two leaves. This completes the induction step and therefore the proof, in the case where $T_u$ and $T_v$ each have at least two vertices. We must still consider the possibility that at least one of $T_u$ and $T_v$ has only one vertex.

Since $k+1 \geq 3$, at least one of $T_u$ and $T_v$ must have two vertices, so only one of them can have only one vertex. Without loss of generality (since nothing in our argument so far made any distinction between $u$ and $v$, we can switch $u$ and $v$ if we need to), we may assume that $T_v$ has only one vertex, and $T_u$ has at least two vertices. Applying our induction hypothesis to $T_u$, we conclude that $T_u$ has some leaf $x$ that is not $u$, and that is also a leaf of $T$. Furthermore, since $T_v$ has only one vertex, this means that deleting the edge $uv$ left $v$ as an isolated vertex, so $uv$ was the only edge incident with $v$ in $T$. Therefore, $v$ is a leaf of $T$. Thus, $T$ has at least two leaves: $x$ and $v$. This completes the induction step and therefore the proof, in the case where at least one of $T_u$ and $T_v$ has only one vertex. Since we have dealt with all possibilities, this completes our induction step.

By the Principle of Mathematical Induction, every tree that has at least one edge, has at least two leaves. □

The next result will be left to you to prove.

**PROPOSITION 12.4.4.** *If a leaf is deleted from a tree, the resulting graph is a tree.*

**THEOREM 12.4.5.** *The following are equivalent for a graph $T$ with $n$ vertices:*

*1) $T$ is a tree;*

*2) $T$ is connected and has $n-1$ edges;*

*3) $T$ has no cycles, and has $n-1$ edges;*

*4) $T$ is connected, but deleting any edge leaves a disconnected graph.*

**PROOF.** We will prove that the statements are equivalent by showing that $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 1$. Thus, by using a sequence of implications, we see that any one of the statements implies any other.

**(1 ⇒ 2)** We assume that $T$ is a tree, and we would like to deduce that $T$ is connected and has $n-1$ edges. By the definition of a tree, $T$ is connected. We will use induction on $n$ to show that $T$ has $n-1$ edges.

*Base case:* $n = 1$. There is only one (unlabeled) graph on one vertex, it is $K_1$, so $T \cong K_1$, which has no edges. Since $0 = n-1$, this completes the proof of the base case.

*Inductive step:* We begin with the inductive hypothesis. Let $k \geq 1$ be arbitrary, and suppose that every tree on $k$ vertices has $k-1$ edges.

Let $T$ be an arbitrary tree with $k+1$ vertices. Since $k+1 \geq 2$ and $T$ is connected, $T$ must have at least one edge, so by Proposition 12.4.3, $T$ has at least two leaves. Let $v$ be a leaf of $T$. By Proposition 12.4.4, $T \setminus \{v\}$ is a tree. Also, $T \setminus \{v\}$ has $k$ vertices, so we can apply our induction hypothesis to conclude that $T \setminus \{v\}$ has $k-1$ edges. Since $v$ was a leaf, $T$ has precisely one more edge than $T \setminus \{v\}$, so $T$ must have $k = (k+1) - 1$ edges. This completes our inductive step.

By the Principle of Mathematical Induction, every tree on $n$ vertices has $n-1$ edges.

**(2 ⇒ 3)** We assume that $T$ is connected and has $n-1$ edges. We need to deduce that $T$ has no cycles.

Towards a contradiction, suppose that $T$ has a cycle. Repeatedly delete edges that are in cycles until no cycles remain. By Proposition 12.3.9 (used repeatedly), the resulting graph is connected, so by definition it is a tree. Since we have already proven that $1 \Rightarrow 2$, this tree must have $n-1$ edges. Since we started with $n-1$ edges and deleted at least one (based on our assumption that $T$ has at least one cycle), this is a contradiction. This contradiction serves to prove that $T$ must not have any cycles.

**(3 ⇒ 4)** We assume that $T$ has no cycles, and has $n-1$ edges. We must show that $T$ is connected, and that deleting any edge leaves a disconnected graph. We begin by showing that $T$ is connected; we prove this by induction on $n$.

*Base case:* $n = 1$. Then $T \cong K_1$ is connected.

*Inductive step:* We begin with the inductive hypothesis. Let $k \geq 1$ be arbitrary, and suppose that every graph on $k$ vertices that has $k-1$ edges and no cycles, is connected.

Let $T$ be an arbitrary graph with $k+1$ vertices that has $k$ edges and no cycles. By Euler's handshaking lemma,

$$\sum_{v \in V} d(v) = 2k.$$

If each of the $k+1$ vertices had valency 2 or more, then we would have

$$\sum_{v \in V} d(v) \geq 2(k+1)$$

(this is a lot like the Pigeonhole Principle in concept, but the Pigeonhole Principle itself doesn't apply to this situation). Since $2k < 2(k+1)$, there must be some vertex $v$ that does not have valency 2 or more. Delete $v$. In so doing, we delete at most 1 edge, since $v$ has at most 1 incident edge. Thus, the resulting graph has $k$ vertices and $k$ or $k-1$ edges, and since $T$ has no cycles, neither does $T \setminus \{v\}$.

If $T \setminus \{v\}$ has $k$ edges, then deleting any of the edges results in a graph on $k$ vertices with no cycles and $k-1$ edges, which by our inductive hypothesis must be connected. Therefore $T \setminus \{v\}$ is a connected graph that remains connected after any edge is deleted. By Proposition 12.4.2 (in the contrapositive), this means that $T \setminus \{v\}$ must contain a cycle, but this is a contradiction. This contradiction serves to prove that $T \setminus \{v\}$ cannot have $k$ edges.

Thus, $T \setminus \{v\}$ has $k-1$ edges and $k$ vertices, and no cycles. By our inductive hypothesis, $T \setminus \{v\}$ must be connected. Furthermore, the fact that $T \setminus \{v\}$ has $k-1$ edges means that $v$ is incident to an edge, which must have its other endvertex in $T \setminus \{v\}$. Therefore $T$ is connected. This completes the inductive step.

By the Principle of Mathematical Induction, every graph on $n$ vertices with no cycles and $n-1$ edges is connected.

It remains to be shown that deleting any edge leaves a disconnected graph, but now that we know that $T$ is connected, this follows from Proposition 12.4.2.

**(4 ⇒ 1)** We assume that $T$ is connected, but deleting any edge leaves a disconnected graph. By the definition of a tree, we must show that $T$ has no cycles. This follows immediately from Proposition 12.3.9. □

**EXERCISES 12.4.6.**

1) Prove Proposition 12.4.4.

2) Draw a tree on 6 vertices.

3) There are two non-isomorphic trees on 4 vertices. Find them.

4) There are 11 non-isomorphic graphs on 4 vertices. Draw all 11, and under each one indicate: is it connected? Is it a forest? Is it a tree?
[*Hint:* One has 0 edges, one has 1 edge, two have 2 edges, three have 3 edges, two have 4 edges, one has 5 edges, and one has 6 edges.]

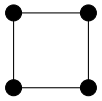
## 12.5. Automorphisms of graphs

This section is intended as optional enrichment material. In the current edition, no exercises are included. The main goal of this section is to make you aware of some interesting results involving automorphisms of graphs, without actually studying them in any detail.

**DEFINITION 12.5.1.** An **automorphism** of a graph $G = (V, E)$ is an isomorphism from $G$ to $G$. More precisely, it is a bijection $\varphi$ from $V$ to $V$ such that for any $v, w \in V$, we have

$$\{v, w\} \in E \Leftrightarrow \{\varphi(v), \varphi(w)\} \in E.$$

**EXAMPLE 12.5.2.** The following graph ($C_4$) has 8 automorphisms. It can be left alone; rotated by $90°, 180°$, or $270°$ or reflected through the vertical, horizontal, or either diagonal axis.

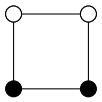


Observe that leaving a graph alone: that is, mapping every vertex to itself, is always an automorphism of any graph. We call this the *trivial automorphism*.

There are many interesting problems and results relating to automorphisms of graphs. In this section we have chosen a few topics to focus on from this broad field. The first topic is the distinguishing number of graphs. Although this idea had appeared previously, broad interest in it (and the origin of this terminology) began with a paper by Michael Owen Albertson (1946—2009) and Karen Linda Collins (1959—) from 1996, in which they provided the basic definitions and some initial results.

**DEFINITION 12.5.3.** Suppose that a colour has been assigned to each vertex of a graph $G$. An automorphism $\varphi$ of $G$ is said to **preserve the colouring** if for every $v \in V(G)$, the colour of $v$ is the same as the colour of $\varphi(v)$.

**EXAMPLE 12.5.4.** Only the trivial automorphism, and reflection through the vertical axis, preserve the colouring of $C_4$ shown below.



**DEFINITION 12.5.5.** The **distinguishing number** of a graph is the smallest number of colours required to colour the vertices of that graph so that the trivial automorphism is the only automorphism that preserves the colouring.

Albertson and Collins posed the problem of finding the distinguishing number of any graph.

One of the very early results in probabilistic graph theory (a technique mentioned in Section 11.5) was a theorem by Pál Erdős (1913—1996) and Alfréd Rényi (1921—1970) that almost every graph has only the trivial automorphism. More precisely, what they showed is that if you choose a finite graph completely at random, then the probability is 1 that the graph you chose has no automorphisms other than the trivial automorphism. Erdős and Rényi's result immediately implies that almost every graph has a distinguishing number of 1: we can colour

every vertex black, and still only the trivial automorphism preserves this colouring, since it is the only automorphism of the graph.

Despite the Erdős-Rényi result, there are many interesting families of graphs (such as the cycles $C_n$) that have plenty of nontrivial automorphisms. Although (by Erdős and Rényi's result) the number of such graphs is vanishingly small as a proportion of all possible graphs, they nonetheless exist, and finding their distinguishing number is an interesting problem.

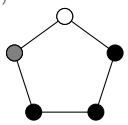
For the cycle $C_n$, the problem of finding the distinguishing number is equivalent to asking if you have a set of $n$ identical keys on a key ring, and you apply coloured covers to some of them, how many colours of covers do you need to ensure that you can tell all of your keys apart from each other? This question has some strong relationships to one of the enumeration questions we considered much earlier in this course: how many beaded necklaces can be made with a collection of coloured beads?

We begin by considering $C_5$.

**EXAMPLE 12.5.6.** The cycle $C_5$ has a distinguishing number of 3.

To see that two colours are not sufficient, note that if two colours are used then one of the colours must be used on 1 or 2 of the vertices, while the other colour is used on the remaining 4 or 3 vertices. In either case, some case-by-case analysis shows that there is a reflection that preserves the colouring.

To see that three colours are sufficient, observe this colouring:



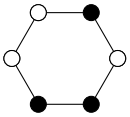
The only nontrivial automorphism that fixes the white vertex is reflection through the vertical axis, and this does not fix the grey vertex.

Now we consider $C_6$.

**EXAMPLE 12.5.7.** The cycle $C_6$ has a distinguishing number of 2.

A single colour is clearly not sufficient, since $C_6$ admits both nontrivial rotations and reflections as automorphisms.

To see that two colours are sufficient, observe this colouring:



Since only one of the white vertices has no white neighbours, any automorphism that preserves the colouring must fix that vertex. The only nontrivial automorphism that fixes this vertex is the reflection through the horizontal axis, and that reflection does not preserve the colouring of one of the other two white vertices. Therefore, only the trivial automorphism preserves this colouring.

Before providing some of the known results on distinguishing numbers, we introduce one interesting family of graphs: the hypercubes.

**DEFINITION 12.5.8.** The **hypercube** of dimension $n$, $Q_n$ has as its vertices all possible binary strings of length $n$. Two vertices are adjacent if the corresponding strings differ in exactly one position.

In Chapter 19 we will look at some of these ideas (binary strings and how many positions they differ in) in more detail. According to the terminology that will be used in that chapter, the vertices of $Q_n$ are the $n$-bit binary strings, and two vertices are adjacent if the corresponding strings have a Hamming distance of 1.

We can now give distinguishing numbers for some families of graphs.

**PROPOSITION 12.5.9.** *The distinguishing number of*

- $C_n$ *is* 2 *when* $n \geq 6$. *It is* 3 *for* $3 \leq n \leq 5$.
- $P_n$ *is* 2 *when* $n \geq 2$. *It is* 1 *for* $n = 1$.
- $K_n$ *is* $n$.
- $K_{n,n}$ *is* $n + 1$.
- $Q_n$ *is* 1 *when* $n = 0$. *It is* 2 *if* $n = 1$ *or if* $n \geq 4$. *It is* 3 *if* $n = 2$ *or* $n = 3$.

We mentioned above that almost all graphs have a distinguishing number of 1. The results we have just listed indicate the pattern that has emerged from research into the distinguishing number of graphs whose distinguishing number is not 1: that is, in almost all such cases it is possible to find a distinguishing colouring that uses just 2 colours.

After doing some work on the distinguishing number herself, Debra Lynn Boutin (1957—) introduced a related concept: the *distinguishing cost*.

**DEFINITION 12.5.10.** For a graph with distinguishing number 2, the **distinguishing cost** is the smallest number of vertices that need to be coloured with the less-commonly-used colour, in order to produce a distinguishing colouring of the graph with 2 colours (that is, a colouring with 2 colours for which only the trivial automorphism preserves the colours).

**EXAMPLE 12.5.11.** The distinguishing cost of $C_6$ is 3. We have already seen in Example 12.5.7 that using three vertices of each colour can provide a distinguishing colouring. To see that one or two vertices of one colour are not sufficient, we observe that any subset of one or two of the vertices of $C_6$ is preserved (as a set) by some reflection automorphism. Thus, no matter which one or two vertices we might choose to colour with the second colour, there would be a nontrivial automorphism preserving the colouring.

In fact, the distinguishing cost of $C_n$ is 3 for every $n \geq 6$, by similar reasoning.

In 2008, Boutin found the following bounds on the distinguishing cost of $Q_n$:

**THEOREM 12.5.12.** *If the distinguishing cost of* $Q_n$ *is denoted by* $dc(Q_n)$, *then*

$$\lceil \log_2(n) \rceil + 1 \leq dc(Q_n) \leq 2\lceil \log_2(n) \rceil - 1.$$

In case you are unfamiliar with this notation, $\lceil r \rceil$ denotes the *ceiling* of the real number $r$: the smallest integer that is greater than or equal to $r$.

**EXAMPLE 12.5.13.** Since $\log_2(16) = 4$, we have

$$\lceil \log_2(16) \rceil + 1 = 4 + 1 = 5$$

and

$$2\lceil \log_2(n) \rceil - 1 = 2(4) - 1 = 7,$$

so $5 \leq dc(Q_{16}) \leq 7$.

While distinguishing is a fascinating topic by itself, we wish to introduce one other family of graphs that relates to automorphisms. We will not provide nearly as much detail on this topic.

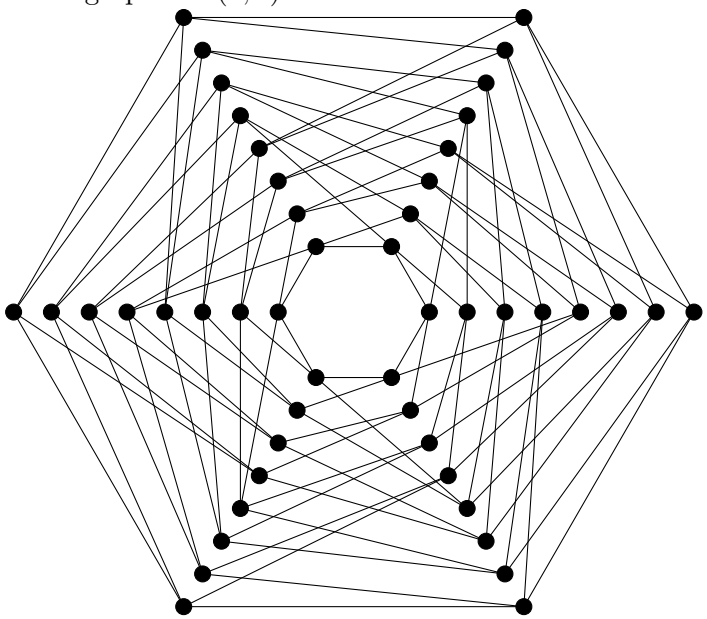
In 1989, Cheryl Elisabeth Praeger (1948—) and Ming-Yao Xu (1941—) introduced a family of graphs that has become known as the *Praeger-Xu* graphs. This is a family of graphs for which every vertex in each of the graphs has valency 4. Without going into detail, each of the graphs in the Praeger-Xu family has far more automorphisms (relative to the number of vertices that it has) than any other graph for which every vertex has valency 4. The Praeger-Xu graphs

have been studied by a number of researchers, and have arisen in several contexts related to automorphisms of graphs. In the same paper, Praeger and Xu actually introduced a collection of related families of graphs where every vertex of every graph in a given family has valency $2p$ for some prime $p$, but the case where $p = 2$ is particularly interesting and this is the family that is known as the Praeger-Xu graphs.

**DEFINITION 12.5.14.** The **Praeger-Xu graph** $PX(n, k)$ has for its vertices the ordered pairs $(i, x)$, where $i \in \{0, \ldots, n-1\}$ and $x$ is a binary string of length $k$. Edges are any pair of vertices of the form $\{(i, ay), (i+1, yb)\}$ where $y$ is a binary string of length $k-1$, $a, b \in \{0, 1\}$, $i \in \{0, \ldots, n-1\}$, and addition in the first entry is performed modulo $n$.

**EXAMPLE 12.5.15.** In $PX(8, 5)$, the vertex $(7, 10111)$ has four neighbours: $(0, 01110)$, $(0, 01111)$, $(6, 01011)$, and $(6, 11011)$.

**EXAMPLE 12.5.16.** The graph $PX(6, 3)$ can be drawn as follows.



The sets of vertices forming a line represent fixed values of the first coordinate, increasing in a clockwise direction. From outside to inside, the second entries of each vertex label are

$$000, 001, 010, 011, 100, 101, 110, 111.$$

**SUMMARY:**

- many definitions and results about graphs can be generalised to the context of digraphs.
- trees can be characterised in a variety of ways.
- Important definitions:
  - digraph
  - arc
  - walk
  - length of a walk
  - connected
  - connected component
  - path, cycle
  - tree, forest, leaf
  - automorphism
- Notation:
  - $P_n$
  - $C_n$

# Chapter 13

# Euler and Hamilton

Some sorts of walks through a graph are particularly important for routing problems. We will be considering some of these in this chapter.

## 13.1. Euler tours and trails

To introduce these concepts, we need to know about some special kinds of walks.

**DEFINITION 13.1.1.** A walk is **closed** if it begins and ends with the same vertex.

A **trail** is a walk in which no two vertices appear consecutively (in either order) more than once. (That is, no edge is used more than once.)

A **tour** is a closed trail.

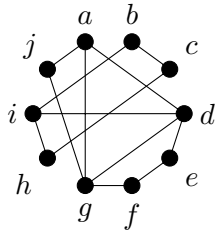
An **Euler trail** is a trail in which every pair of adjacent vertices appear consecutively. (That is, every edge is used exactly once.)

An **Euler tour** is a closed Euler trail.

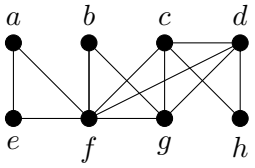
Recall the work of Leonhard Euler (1707—1783) on the bridges of Königsberg. The problem of finding a route that crosses every bridge exactly once, is equivalent to finding an Euler trail in the corresponding graph. If we want the route to begin and end at the same place (for example, someone's home), then the problem is equivalent to finding an Euler tour in the corresponding graph.

Euler tours and trails are important tools for planning routes for tasks like garbage collection, street sweeping, and searches.

**EXAMPLE 13.1.2.** In the graph




$(i, b, c, h, i, d, e, f, g, d, a, g, j, a)$ is an Euler trail. In the graph




$(a, e, f, b, g, f, c, d, h, c, g, d, f, a)$ is an Euler tour.

Here is Euler's method for finding Euler tours. We will state it for multigraphs, as that makes the corresponding result about Euler trails a very easy corollary.

**THEOREM 13.1.3.** *A connected graph (or multigraph, with or without loops) has an Euler tour if and only if every vertex in the graph has even valency.*

**PROOF.** As the statement is if and only if, we must prove both implications.

**(⇒)** Suppose we have a multigraph (possibly with loops) that has an Euler tour,

$$(u_1, u_2, \ldots, u_{e+1} = u_1),$$

where $e = |E|$. Let $u$ be an arbitrary vertex of the multigraph. Every time $u$ appears in the tour, exactly two of the edges incident with $u$ are used: if $u = u_j$, then the edges used are $u_{j-1}u_j$ and $u_ju_{j-1}$ unless $j = 1$ or $j = e + 1$ in which case $u = u_1 = u_{e+1}$ and the edges are $u_eu$ and $uu_2$ (and we consider this as one appearance of $u$ in the tour). Therefore, if $u$ appears $k$ times in the tour, then since by the definition of an Euler tour all edges incident with $u$ are used exactly once, we conclude that $u$ must have valency $2k$. Since $u$ was an arbitrary vertex of the multigraph and $k$ (the number of times $u$ appears in the tour) must be an integer, this shows that the valency of every vertex must be even.

**(⇐)** Suppose we have a connected multigraph in which the valency of every vertex is even. Consider the following algorithm (which will be the first stage of our final algorithm):

Make $u$ (some arbitrary vertex) our active vertex, with a list $L$ of all of the edges of $E$. Make $u$ the first vertex in a new sequence $C$ of vertices. Repeat the following step as many times as possible:

Call the active vertex $v$. Choose any edge $vx$ in $L$ that is incident with $v$. Add $x$ (the other endvertex of this edge) to the end of $C$, and make $x$ the new active vertex. Remove $vx$ from $L$.

We claim that when this algorithm terminates, the sequence $C$ will be a tour (though not necessarily an Euler tour) in the multigraph. By construction, $C$ is a walk, and $C$ cannot use any edge more than once since each edge appears in $L$ only once and is removed from $L$ once it has been used, so $C$ is a trail. We need to show that the walk $C$ is closed.

The only way the algorithm can terminate is if $L$ contains no edge that is incident with the active vertex. Towards a contradiction, suppose that this happens at a time when the active vertex is $y \neq u$. Now, $y$ has valency $2r$ in the multigraph for some integer $r$, so there were $2r$ edges in $L$ that were incident with $y$ when we started the algorithm. Since $y \neq u$, every time $y$ appears in $C$ before this appearance, we removed 2 edges incident with $y$ from $L$ (one in the step when we made $y$ the active vertex, and one in the following step). Furthermore, we removed one additional edge incident with $y$ from $L$ in the final step, when we made $y$ the active vertex again. Thus if there are $t$ previous appearances of $y$ in $C$, we have removed $2t+1$ edges incident with $y$ from $L$. Since $2r$ is even and $2t+1$ is odd, there must still be at least one edge incident with $y$ that is in $L$, contradicting the fact that the algorithm terminated. This contradiction shows that, when the algorithm terminates, the active vertex must be $u$, so the sequence $C$ is a closed walk. Since $C$ is a trail, we see that $C$ must be a tour.

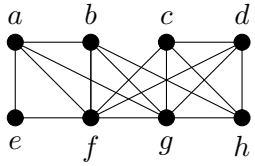
If the tour $C$ is not an Euler tour, let $y$ be the first vertex that appears in $C$ for which there remains an incident edge in $L$. Repeat the previous algorithm starting with $y$ being the active vertex, and with $L$ starting at its current state (not all of $E$). The result will be a tour beginning and ending at $y$ that uses only edges that were not in $C$. Insert this tour into $C$ as follows: if $C = (u = u_1 \ldots, y = u_i, \ldots, u_k = u)$ and the new tour is $(y = v_1, \ldots, v_j = y)$, then the result of inserting the new tour into $C$ will be

$$(u = u_1, \ldots, y = u_i = v_1, v_2, \ldots, v_j = y = u_i, u_{i+1}, \ldots, u_k = u).$$

Replace $C$ by this extended tour.

Repeat the process described in the previous paragraph as many times as possible (this is the second and last stage of our final algorithm). Since $E$ is finite and the multigraph is connected, sooner or later all of the edges of $L$ must be exhausted. At this point, we must have an Euler tour. □

**EXAMPLE 13.1.4.** Use the algorithm described in the proof of the previous result, to find an Euler tour in the following graph.




**SOLUTION.** Let's begin the algorithm at $a$. As $E = L$ is a large set, we won't list the remaining elements every time we choose a new active vertex in the early stages. An easy method for you to keep track of the edges still in $L$ is to colour the edges that are no longer in $L$ (the edges we use) with a different colour as we go.

There are many different possible outcomes for the algorithm since there are often many acceptable choices for the next active vertex. One initial set of choices could be

$$C = (a, b, f, e, a, f, g, a).$$

The first stage of the algorithm terminates at this point since all four edges incident with $a$ have been used. At this point, we have

$$L = \{bg, bh, cd, cf, cg, ch, df, dg, dh, gh\}.$$

The first vertex in $C$ that is incident with an edge in $L$ is $b$. We run the first stage of the algorithm again with $b$ as the initial active vertex and this list for $L$. Again, there are many possible outcomes; one is $(b, g, h, b)$.

We insert $(b, g, h, b)$ into $C$, obtaining a new $C = (a, b, g, h, b, f, e, a, f, g, a)$. At this point, we have

$$L = \{cd, cf, cg, ch, df, dg, dh\}.$$

Now $g$ is the first vertex in $C$ that is incident with an edge in $L$. We run the first stage of the algorithm again with $g$ as the initial active vertex and the current $L$. One possible outcome is $(g, c, f, d, g)$.

Inserting this into $C$ yields a new

$$C = (a, b, g, c, f, d, g, h, b, f, e, a, f, g, a).$$

At this point, we have $L = \{cd, ch, dh\}$. The first vertex in $C$ that is incident with an edge in $L$ is $c$. We run the first stage of the algorithm one final time with $c$ as the initial active vertex and $L = \{cd, ch, dh\}$. This time there are only two possible outcomes: $(c, d, h, c)$ or $(c, h, d, c)$. We choose $(c, d, h, c)$.

Inserting this into $C$ yields our Euler tour:

$$C = (a, b, g, c, d, h, c, f, d, g, h, b, f, e, a, f, g, a).$$ □

**COROLLARY 13.1.5.** *A connected graph (or multigraph, with or without loops) has an Euler trail if and only if at most two vertices have odd valency.*

**PROOF.** Suppose we have a connected graph (or multigraph, with or without loops), $G$. Since the statement is if and only if, there are two implications to prove.

($\Rightarrow$) Suppose that $G$ has an Euler trail. If the trail is closed then it is a tour, and by Theorem 13.1.3, there are no vertices of odd valency. If the trail is not closed, say it is a $u-v$ walk. Add an edge between $u$ and $v$ to $G$, creating a new graph $G^*$ (note that $G^*$ may be a multigraph if $uv$ was already an edge of $G$, even if $G$ wasn't a multigraph), and add $u$ to the end of the Euler trail in $G$, to create an Euler tour in $G^*$. By Theorem 13.1.3, the fact that $G^*$ has an Euler tour means that every vertex of $G^*$ has even valency. Now, the vertices of $G$ all have the same valency in $G^*$ as they have in $G$, with the exception that the valencies of $u$ and $v$ are one higher in $G^*$ than in $G$. Therefore, in this case there are exactly two vertices of odd valency in $G$; namely, $u$ and $v$.

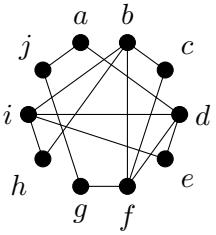
($\Leftarrow$) Now we suppose that $G$ has at most 2 vertices of odd valency. By Corollary 11.3.12 (the corollary to Euler's handshaking lemma), if there are at most two vertices of odd valency, then there are either 0 or 2 vertices of odd valency. We consider these two cases.

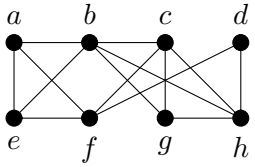
If there are 0 vertices of odd valency, then by Theorem 13.1.3, $G$ has an Euler tour.

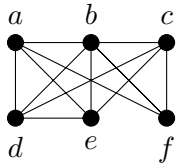
If there are two vertices of odd valency, say $u$ and $v$, add an edge between $u$ and $v$ to $G$, creating a new graph $G^*$ (note that $G^*$ may be a multigraph if $uv$ was already an edge of $G$, even if $G$ wasn't a multigraph). Now in $G^*$ every vertex has even valency, so $G^*$ has an Euler tour. In fact, a careful look at the algorithm given in the proof of Theorem 13.1.3 shows that we may choose $u$ and $v$ (in that order) to be the first two vertices in this Euler tour, so that $uv$ (the edge that is in $G^*$ but not $G$) is the first edge used in the tour. Now if we delete $u$ from the start of this Euler tour, the result is an Euler trail in $G$ that starts at $v$ and ends at $u$. □

## EXERCISES 13.1.6.

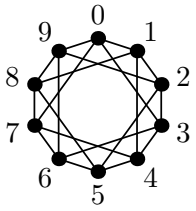
1) For each of the following graphs, is there an Euler tour? Is there an Euler trail? If either exists, find one; if not, explain why not.

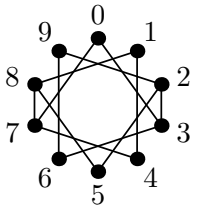
(a) 

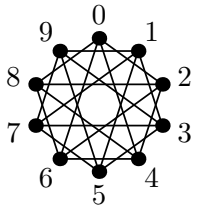

(b) 

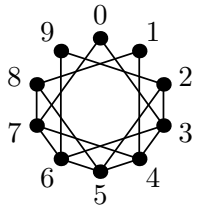

(c) 


2) For each of the following four graphs, determine whether it has an Euler trail, and whether it has an Euler tour. (Explain how you know.)

(a) 


(b) 

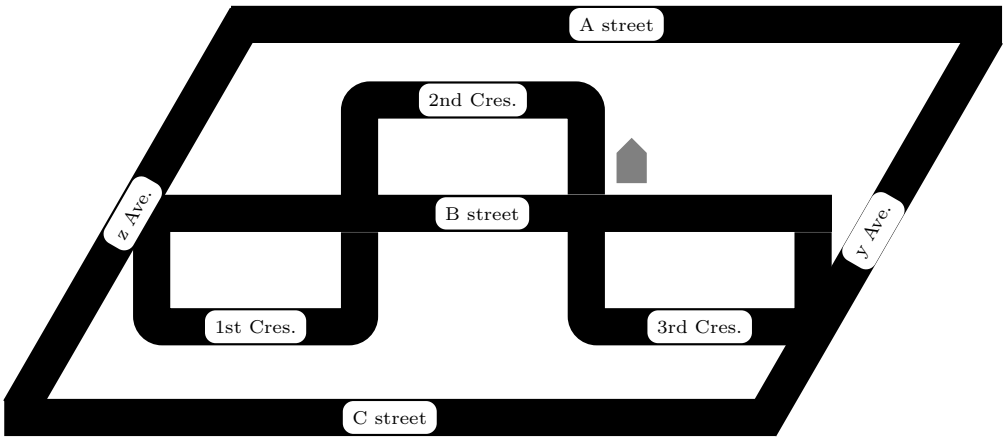

(c) 


(d) 


3) If it is possible, draw a graph that has an even number of vertices and an odd number of edges, that also has an Euler tour. If that isn't possible, explain why there is no such graph.

4) Which complete graphs have an Euler tour? Of the complete graphs that do not have an Euler tour, which of them have an Euler trail?

**EXERCISE 13.1.7.** Sylvia's cat is missing. She wants to look for it in all the nearby streets, but she is tired and doesn't want to walk any farther than she must. Find an efficient route for Sylvia to take through her neighbourhood so that she starts and ends at home and walks through each street exactly once. The location of Sylvia's house is marked with a house-shaped symbol (⌂).




## 13.2. Hamilton paths and cycles

Sometimes, rather than traveling along every connection in a network, our object is simply to visit every node of the network. This relates to a different structure in the corresponding graph.

**DEFINITION 13.2.1.** A **Hamilton cycle** is a cycle that visits **every** vertex of the graph.

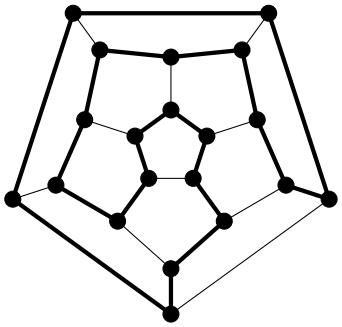
A **Hamilton path** is a path that visits **every** vertex of the graph.

The definitions of path and cycle ensure that vertices are not repeated. Hamilton paths and cycles are important tools for planning routes for tasks like package delivery, where the important point is not the routes taken, but the places that have been visited.

In 1857, Sir William Rowan Hamilton (1805—1865) first presented a game he called the "icosian game." It involved tracing edges of a dodecahedron in such a way as to visit each corner precisely once. In fact, two years earlier Reverend Thomas Penyngton Kirkman (1806—1895) had sent a paper to the Royal Society in London, in which he posed the problem of finding what he called *closed polygons* in polyhedra; a closed polygon he defined as a circuit of edges that passes through each vertex just once. Thus, Kirkman had posed a more general problem prior to Hamilton (and made some progress toward solving it); nonetheless, it is Hamilton for whom these structures are now named. As we'll see later when studying Steiner Triple Systems in design theory, Kirkman was a gifted mathematician who seems to have been singularly unlucky in terms of receiving proper credit for his achievements. As his title indicates, Kirkman was a minister who pursued mathematics on the side, as a personal passion.

Hamilton managed to convince the company of John Jacques and sons, who were manufacturers of toys (including high-quality chess sets) to produce and market the "icosian game." It was produced under the name *Around the World*, and sold in two forms: a flat board, or an actual dodecahedron. In both cases, nails were placed at the corners of the dodecahedron representing cities, and the game was played by wrapping a string around the nails, traveled only along edges, visiting each nail once, and ending at the starting point. Unfortunately, the game was not a financial success. It is not very difficult and becomes uninteresting once solved. John Jacques and sons paid Hamilton 25 pounds for the rights to it.

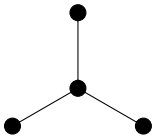
The thick edges form a Hamilton cycle in the graph of the dodecahedron:



Not every connected graph has a Hamilton cycle; in fact, not every connected graph has a Hamilton path.

**Figure 13.2.1.** A graph with a Hamilton path but no Hamilton cycle

**Figure 13.2.2.** A graph with no Hamilton path



Unfortunately, in contrast to Euler's result about Euler tours and trails (given in Theorem 13.1.3 and Corollary 13.1.5), there is no known characterisation that enables us to quickly determine whether or not an arbitrary graph has a Hamilton cycle (or path). This is a hard problem in general. We do know of some necessary conditions (any graph that fails to meet these conditions cannot have a Hamilton cycle) and some sufficient conditions (any graph that meets these must have a Hamilton cycle). However, many graphs meet all of the necessary conditions while failing to meet all of the sufficient conditions. There are also some known conditions that are either necessary or sufficient for the existence of a Hamilton path.

Here is a necessary condition for a graph to have a Hamilton cycle.

**THEOREM 13.2.2.** *If $G$ is a graph with a Hamilton cycle, then for every $S \subset V$ with $S \neq \emptyset, V$, the graph $G \setminus S$ has at most $|S|$ connected components.*

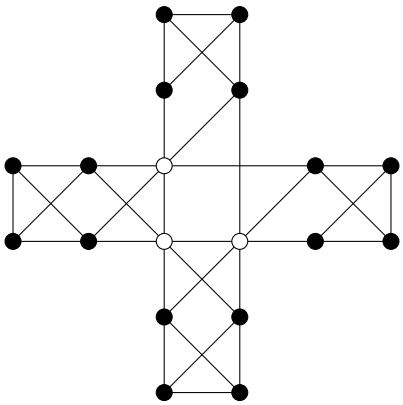
**PROOF.** Let $C$ be a Hamilton cycle in $G$. Fix an arbitrary proper, nonempty subset $S$ of $V$.

One at a time, delete the vertices of $S$ from $C$. After the first vertex is deleted, the result is still connected, but has become a path. When any of the subsequent $|S| - 1$ vertices is deleted, it either breaks some path into two shorter paths (increasing the number of connected components by one) or removes a vertex at an end of some path (leaving the number of connected components unchanged, or reducing it by one if this component was a $P_0$). So $C \setminus S$ has at most $1 + (|S| - 1) = |S|$ connected components.

Notice that if two vertices $u$ and $v$ are in the same connected component of $C \setminus S$, then they will also be in the same connected component of $G \setminus S$. This is because adding edges can only connect things more fully, reducing the number of connected components. More formally, if there is a $u - v$ walk in $C$, then any pair of consecutive vertices in that walk is adjacent in $C$ so is also adjacent in $G$. Therefore the same walk is a $u - v$ walk in $G$. This tells us that the number of connected components of $G \setminus S$ is at most the number of connected components of $C \setminus S$, which we have shown to be at most $|S|$. □

**EXAMPLE 13.2.3.** When a non-leaf is deleted from a path of length at least 2, the deletion of this single vertex leaves two connected components. So no path of length at least 2 contains a Hamilton cycle.

Here's a graph in which the non-existence of a Hamilton cycle might be less obvious without Theorem 13.2.2. Deleting the three white vertices leaves four connected components.



As you might expect, if all of the vertices of a graph have sufficiently high valency, it will always be possible to find a Hamilton cycle in the graph. (In fact, generally the graph will have many different Hamilton cycles.) Before we can formalise this idea, it is helpful to have an additional piece of notation.

**DEFINITION 13.2.4.** The **minimum valency** of a graph $G$ is

$$\min_{v \in V} d(v).$$

The **maximum valency** of a graph $G$ is

$$\max_{v \in V} d(v).$$

**NOTATION 13.2.5.** We use $\delta$ to denote the minimum valency of a graph, and $\Delta$ to denote its maximum valency. If we need to clarify the graph involved, we use $\delta(G)$ or $\Delta(G)$.

The following theorem was proven by Gabriel Andrew Dirac (1925—1984) back in 1952, yet it remains one of the most powerful known methods for determining that a graph has a Hamilton cycle.

**THEOREM 13.2.6 (Dirac, 1952).** *If $G$ is a graph with vertex set $V$ such that $|V| \geq 3$ and $\delta(G) \geq |V|/2$, then $G$ has a Hamilton cycle.*

**PROOF.** Towards a contradiction, suppose that $G$ is a graph with vertex set $V$, that $|V| = n \geq 3$, and that $\delta(G) \geq n/2$, but $G$ has no Hamilton cycle.

Repeat the following as many times as possible: if there is an edge that can be added to $G$ without creating a Hamilton cycle in the resulting graph, add that edge to $G$. When this has been done as many times as possible, call the resulting graph $H$. The graph $H$ has the same vertex set $V$, and since we have added edges we have not decreased the valency of any vertex, so we have $\delta(H) \geq n/2$. Now, $H$ still has no Hamilton cycle, but adding any edge to $H$ gives a graph that does have a Hamilton cycle.

Since complete graphs on at least three vertices always have Hamilton cycles (see Exercise 13.2.12.1), we must have $H \ncong K_n$, so there are at least two vertices of $H$, say $u$ and $v$, that are not adjacent. By our construction of $H$ from $G$, adding the edge $uv$ to $H$ would result in a Hamilton cycle, and this Hamilton cycle must use the edge $uv$ (otherwise it would be a

Hamilton cycle in $H$, but $H$ has no Hamilton cycle). Thus, the portion of the Hamilton cycle that is in $H$ forms a Hamilton path from $u$ to $v$. Write this Hamilton path as

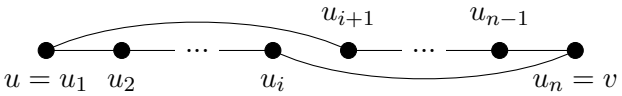
$$(u = u_1, u_2, \ldots, u_n = v).$$

Define the sets

$$S = \{u_i \mid u \sim u_{i+1}\} \text{ and } T = \{u_i \mid v \sim u_i\}.$$

That is, $S$ is the set of vertices that appear immediately before a neighbour of $u$ on the Hamilton path, while $T$ is the set of vertices on the Hamilton path that are neighbours of $v$. Notice that $v = u_n \notin S$ since $u_{n+1}$ isn't defined, and $v = u_n \notin T$ since our graphs are simple (so have no loops). Thus, $u_n \notin S \cup T$, so $|S \cup T| < n$.

Towards a contradiction, suppose that for some $i$, $u_i \in S \cap T$. Then by the definitions of $S$ and $T$, we have $u \sim u_{i+1}$ and $v \sim u_i$, so:




is a Hamilton cycle in $H$, which contradicts our construction of $H$ as a graph that has no Hamilton cycle. This contradiction serves to prove that $|S \cap T| = \emptyset$.

Now we have

$$d_H(u) + d_H(v) = |S| + |T| = |S \cup T| + |S \cap T|$$

(the last equality comes from Inclusion-Exclusion). But we have seen that $|S \cup T| < n$ and $|S \cap T| = 0$, so this gives

$$d_H(u) + d_H(v) < n.$$

This contradicts $\delta(H) \geq n/2$, since

$$d_H(u), d_H(v) \geq \delta(H).$$

This contradiction serves to prove that no graph $G$ with vertex set $V$ such that $|V| \geq 3$ and $\delta(G) \geq |V|/2$ can fail to have a Hamilton cycle. □

In fact, the statement of Dirac's theorem was improved by John Adrian Bondy (1944—) and Václav Chvátal (1946—) in 1976. They began by observing that the proof given above for Dirac's Theorem actually proves the following result.

**LEMMA 13.2.7.** *Suppose that $G$ is a graph on $n$ vertices, $u$ and $v$ are nonadjacent vertices of $G$, and $d(u) + d(v) \geq n$. Then $G$ has a Hamilton cycle if and only if the graph obtained by adding the edge $uv$ to $G$ has a Hamilton cycle.*

With this in mind, they made the following definition.

**DEFINITION 13.2.8.** Let $G$ be a graph on $n$ vertices. The **closure** of $G$ is the graph obtained by repeatedly joining pairs of nonadjacent vertices $u$ and $v$ for which $d(u) + d(v) \geq n$, until no such pair exists.

Before they were able to work with this definition, they had to prove that the closure of a graph is well-defined. In other words, since there will often be choices involved in forming the closure of a graph (if more than one pair of vertices satisfy the condition, which edge do we add first?), is it possible that by making different choices, we might end up with a different graph at the end? The answer, fortunately, is no; any graph has a unique closure, as we will now prove.

**LEMMA 13.2.9.** *Closure is well-defined. That is, any graph has a unique closure.*

**PROOF.** Let $(e_1, \ldots, e_\ell)$ be one sequence of edges we can choose to arrive at the closure of $G$, and let the resulting closure be the graph $G_1$. Let $(f_1, \ldots, f_m)$ be another such sequence, and let the resulting closure be the graph $G_2$. We will prove by induction on $\ell$ that for every $1 \leq i \leq \ell$, $e_i \in \{f_1, \ldots, f_m\}$. We will use $\{u_i, v_i\}$ to denote the endvertices of $e_i$.

*Base case:* $\ell = 1$, so only the edge $\{u_1, v_1\}$ is added to $G$ in order to form $G_1$. Since this was the first edge added, we must have

$$d_G(u_1) + d_G(v_1) \geq n.$$

Since $G_2$ has all of the edges of $G$, we must certainly have

$$d_{G_2}(u_1) + d_{G_2}(v_1) \geq n.$$

Since $G_2$ is a closure of $G$, it has no pair of nonadjacent edges whose valencies sum to $n$ or higher, so $u_1$ must be adjacent to $v_1$ in $G_2$. Since the edge $u_1v_1$ was not in $G$, it must be in $\{f_1, \ldots, f_m\}$. This completes the proof of the base case.

*Inductive step:* We begin with the inductive hypothesis. Let $k \geq 1$ be arbitrary (with $k \leq \ell$), and suppose that

$$e_1, \ldots, e_k \in \{f_1, \ldots, f_m\}.$$

Consider $e_{k+1} = \{u_{k+1}, v_{k+1}\}$. Let $G'$ be the graph obtained by adding the edges $e_1, \ldots, e_k$ to $G$. Since $e_{k+1}$ was chosen to add to $G'$ to form $G_1$, it must be the case that

$$d_{G'}(u_{k+1}) + d_{G'}(v_{k+1}) \geq n.$$

By our induction hypothesis, all of the edges of $G'$ are also in $G_2$, so this means

$$d_{G_2}(u_{k+1}) + d_{G_2}(v_{k+1}) \geq n.$$

Since $G_2$ is a closure of $G$, it has no pair of nonadjacent edges whose valencies sum to $n$ or higher, so $u_{k+1}$ must be adjacent to $v_{k+1}$ in $G_2$. Since the edge $e_{k+1}$ was not in $G$, it must be in $\{f_1, \ldots, f_m\}$.

By the Principle of Mathematical Induction, $G_2$ contains all of the edges of $G_1$. Since there was nothing special about $G_2$ as distinct from $G_1$, we could use the same proof to show that $G_1$ contains all of the edges of $G_2$. Therefore, $G_1$ and $G_2$ have the same edges. Since they also have the same vertices (the vertices of $G$), they are the same graph. Thus, the closure of any graph is unique. □

This allowed Bondy and Chvátal to deduce the following result, which is stronger than Dirac's although as we've seen the proof is not significantly different.

**THEOREM 13.2.10.** *A simple graph has a Hamilton cycle if and only if its closure has a Hamilton cycle.*

**PROOF.** Repeatedly apply Lemma 13.2.7. □

This has a very nice corollary.

**COROLLARY 13.2.11.** *A simple graph on at least* 3 *vertices whose closure is complete, has a Hamilton cycle.*

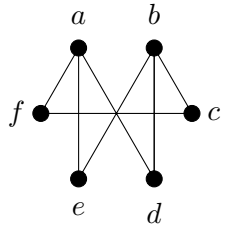
**PROOF.** This is an immediate consequence of Theorem 13.2.10 together with the fact (see Exercise 13.2.12.1) that every complete graph on at least 3 vertices has a Hamilton cycle. □

**EXERCISES 13.2.12.**

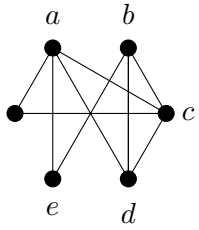
1) Prove by induction that for every $n \geq 3$, $K_n$ has a Hamilton cycle.

2) Find the closure of each of these graphs. Can you easily tell from the closure whether or not the graph has a Hamilton cycle?

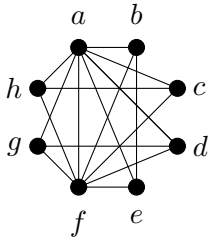
(a)




(b)




3) Use Theorem 13.2.2 to prove that this graph does not have a Hamilton cycle.




4) Prove that if $G$ has a Hamilton path, then for every nonempty proper subset $S$ of $V$, $G - S$ has no more than $|S| + 1$ connected components.

5) For the two graphs in Exercise 2, either find a Hamilton cycle or use Theorem 13.2.2 to show that no Hamilton cycle exists.

---

**SUMMARY:**

- algorithms for finding Euler tours and trails
- Important definitions:
  - closed walk, trail, tour
  - Euler tour, Euler trail
  - Hamilton cycle, Hamilton path
  - minimum valency, maximum valency
  - closure of a graph
- Notation:
  - $\delta, \Delta$

---

# Chapter 14

# Graph Colouring

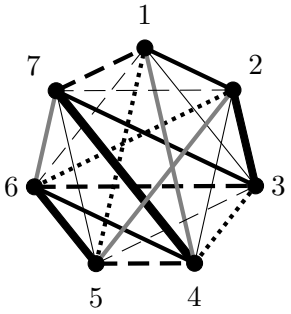
## 14.1. Edge colouring

Suppose you have been given the job of scheduling a round-robin tennis tournament with $n$ players. One way to approach the problem is to model it as a graph: the vertices of the graph will represent the players, and the edges will represent the matches that need to be played. Since it is a round-robin tournament, every player must play every other player, so the graph will be complete. Creating the schedule amounts to assigning a time to each of the edges, representing the time at which that match is to be played.

Notice that there is a constraint. When you have assigned a time to a particular edge $uv$, no other edge incident with either $u$ or $v$ can be assigned the same time, since this would mean that either player $u$ or player $v$ is supposed to play two games at once. Instead of writing times on each edge, we will choose a colour to represent each of the time slots, and colour the edges that are to be played at that time, with that colour.

Here is an example of a possible schedule for the tournament, when $n = 7$.

**EXAMPLE 14.1.1.** The players are numbered from 1 through 7, and we will spread the tournament out over seven days. Games to be played on each day should have a different colour than the games on other days, but, because this text is printed in black-and-white, we will use some line patterns, instead of colours. Games to be played on Monday will be drawn as usual. Games to be played on Tuesday will be thin. Games to be played on Wednesday will be dotted. Games to be played on Thursday will be dashed. Games to be played on Friday will be thick. Games to be played on Saturday will be grey. Games to be played on Sunday will be thin and dashed.




This gives a schedule. In case you have trouble distinguishing the "colours" of the edges, the normal edges are 12, 37, and 46; the thin edges are 13, 24, and 57; the dotted edges are 15, 26, and 34; the dashed edges are 17, 36, and 45; the thick edges are 23, 47, and 56; the grey edges are 14, 25, and 67; and the thin dashed edges are 16, 27, and 35.

**DEFINITION 14.1.2.** A **proper $k$-edge-colouring** of a graph $G$ is a function that assigns to each edge of $G$ one of $k$ colours, such that edges that meet at an endvertex must be assigned different colours.

The constraint that edges of the same colour cannot meet at a vertex turns out to be a useful constraint in a number of contexts.

If the graph is large enough we are liable to run out of colours that can be easily distinguished (and we get tired of writing out the names of colours). The usual convention is to refer to each colour by a number (the first colour is colour 1, etc.) and to label the edges with the numbers rather than using colours.

**DEFINITION 14.1.3.** A graph $G$ is **$k$-edge-colourable** if it admits a proper $k$-edge-colouring. The smallest integer $k$ for which $G$ is $k$-edge-colourable is called the **edge chromatic number**, or **chromatic index** of $G$.

**NOTATION 14.1.4.** The chromatic index of $G$ is denoted by $\chi'(G)$, or simply by $\chi'$ if the context is unambiguous.

Here is an easy observation:

**PROPOSITION 14.1.5.** *For any graph $G$, $\chi'(G) \geq \Delta(G)$.*

**PROOF.** Recall that $\Delta(G)$ denotes the maximum value of $d(v)$ over all vertices $v$ of $G$. So there is some vertex $v$ of $G$ such that $d(v) = \Delta(G)$. In any proper edge-colouring, the $d(v)$ edges that are incident with $v$, must all be assigned different colours. Thus, any proper edge-colouring must have at least $d(v) = \Delta(G)$ distinct colours. This means $\chi'(G) \geq \Delta(G)$. □

**EXAMPLE 14.1.6.** The colouring given in Example 14.1.1 shows that $\chi'(K_7) \leq 7$, since we were able to properly edge-colour $K_7$ using seven colours.

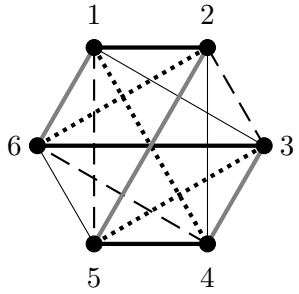
To show that we cannot colour $K_7$ with fewer than 7 colours, notice that because each of the 7 vertices can only be incident with one edge of a given colour, there cannot be more than 3 edges coloured with any given colour (3 edges are already incident with 6 of the 7 vertices, and a fourth edge would have to be incident with two others).

We know that $K_7$ has $\binom{7}{2} = 21$ edges, so if at most 3 edges can be coloured with any given colour, we will require at least 7 colours to properly edge-colour $K_7$. Thus $\chi'(K_7) \geq 7$.

Thus, we have shown that $\chi'(K_7) = 7$.

This shows that $\chi'(K_7) = 7 > \Delta(K_7) = 6$, so it is not always possible to achieve equality in the bound given by Proposition 14.1.5. Our next example shows that it *is* sometimes possible to achieve equality in that bound.

**EXAMPLE 14.1.7.** Here is a proper 5-edge-colouring of $K_6$:




In case the edge colours are difficult to distinguish, the thick edges are 12, 36, and 45; the thin edges are 13, 24, and 56; the dotted edges are 14, 26, and 35; the dashed edges are 15, 23, and 46; and the grey edges are 16, 25, and 34. This shows that $\chi'(K_6) \leq 5$. Since the valency

of every vertex of $K_6$ is 5, Proposition 14.1.5 implies that $\chi'(K_6) \geq 5$. Putting these together, we see that $\chi'(K_6) = \Delta(K_6) = 5$, so equality in the bound of Proposition 14.1.5 is achieved by $K_6$.

The following rather remarkable result was proven by Vadim Georgievich Vizing (1937—2017) in 1964:

**THEOREM 14.1.8 (Vizing's Theorem).** *For any simple graph $G$, $\chi'(G) \in \{\Delta(G), \Delta(G)+1\}$.*

We will not go over the proof of this theorem.

**DEFINITION 14.1.9.** If $\chi'(G) = \Delta(G)$ then $G$ is said to be a **class-one graph**, and if $\chi'(G) = \Delta(G)+1$ then $G$ is said to be a **class-two graph**.

To date, graphs have not been completely classified according to which graphs are class one and which are class two, but it has been proven that "almost every" graph is of class one. Technically, this means that if you choose a random graph out of all of the graphs on at most $n$ vertices, the probability that you will choose a class-two graph approaches 0 as $n$ approaches infinity.

There are, however, infinitely many class-two graphs; the same argument we used to show that $\chi'(K_7) \geq 7$ can also be used to prove that $\chi'(K_{2n+1}) = 2n+1$ for any positive integer $n$, since the number of edges is

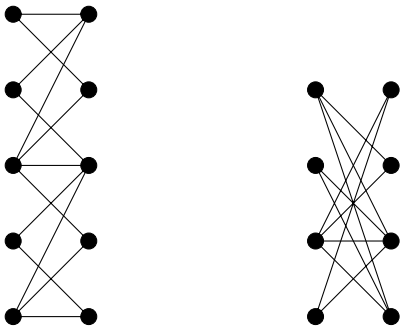
$$(2n+1)(2n)/2 = n(2n+1)$$

and each colour can only be used to colour $n$ of the edges. Since $\Delta(K_{2n+1}) = 2n$, this shows that $K_{2n+1}$ is class two.

Large families of graphs have been shown to be class-one graphs. We will devote most of the rest of this section to proving that all of the graphs in one particular family are class one. First we need to define the family.

**DEFINITION 14.1.10.** A graph is **bipartite** if its vertices can be partitioned into two sets $V_1$ and $V_2$, such that every edge of the graph has one of its endvertices in $V_1$, and the other in $V_2$. The sets $V_1$ and $V_2$ form a **bipartition** of the graph.

**EXAMPLE 14.1.11.** The following graphs are bipartite. Every edge has one endvertex on the left side, and one on the right.



The graph $K_n$ is not bipartite if $n \geq 3$. The first vertex may as well go into $V_1$; the second vertex is adjacent to it, so must go into $V_2$; but the third vertex is adjacent to both, so cannot go into either $V_1$ or $V_2$.

Although the following class of bipartite graphs will not be used in this chapter, they are an important class of bipartite graphs that will come up again later.

**DEFINITION 14.1.12.** The **complete bipartite graph**, $K_{m,n}$, is the bipartite graph on $m+n$ vertices with as many edges as possible subject to the constraint that it has a bipartition into sets of cardinality $m$ and $n$. That is, it has every edge between the two sets of the bipartition.

Before proving that all bipartite graphs are class one, we need to understand the structure of bipartite graphs a bit better. Here is an important theorem.

**THEOREM 14.1.13.** *A graph $G$ is bipartite if and only if $G$ contains no cycle of odd length.*

**PROOF.** This is an if and only if statement, so we have two implications to prove.

($\Rightarrow$) We prove the contrapositive, that if $G$ contains a cycle of odd length, then $G$ cannot be bipartite.

Let

$$(v_1, v_2, \ldots, v_{2k+1}, v_1)$$

be a cycle of odd length in $G$. We try to establish a bipartition $V_1$ and $V_2$ for $G$. Without loss of generality, we may assume that $v_1 \in V_1$. Then we must have $v_2 \in V_2$ since $v_2 \sim v_1$. Continuing in this fashion around the cycle, we see that for every $1 \leq i \leq k$, we have $v_{2i+1} \in V_1$ and $v_{2i} \in V_2$. In particular, $v_{2k+1} \in V_1$, but $v_1 \in V_1$ and $v_1 \sim v_{2k+1}$, contradicting the fact that every edge must have one of its endvertices in $V_2$. Thus, $G$ is not bipartite.

($\Leftarrow$) Let $G$ be a graph that is not bipartite. We must show that there is an odd cycle in $G$.

If every connected component of $G$ is bipartite, then $G$ is bipartite (choose one set of the bipartition from each connected component; let $V_1$ be the union of these, and $V_2$ the set of all other vertices of $G$; this is a bipartition for $G$). Thus there is at least one connected component of $G$ that is not bipartite.

Pick any vertex $u$ from a non-bipartite connected component of $G$, and assign it to $V_1$. Place all of its neighbours in $V_2$. Place all of their neighbours into $V_1$. Repeat this process, at each step putting all of the neighbours of every vertex of $V_1$ into $V_2$, and then all of the neighbours of every vertex of $V_2$ into $V_1$.

Since this component is not bipartite, at some point we must run into the situation that we place a vertex $v$ into $V_j$, but a neighbour $u_1$ of $v$ is also in $V_j$ (for some $j \in \{1, 2\}$). By our construction of $V_1$ and $V_2$, there must be a walk from $u_1$ to $v$ that alternates between vertices in $V_j$ and vertices in $V_{3-j}$. By Proposition 12.3.4, there must in fact be a *path* from $u_1$ to $v$ that alternates between vertices in $V_j$ and vertices in $V_{3-j}$. Since the path alternates between the two sets but begins and ends in $V_j$, it has even length. Therefore, adding $u_1$ to the end of this path yields a cycle of odd length in $G$. □

In order to prove that bipartite graphs are class one, we require a lemma.

**LEMMA 14.1.14.** *Let $G$ be a connected graph that is not a cycle of odd length. Then $G$ the edges of $G$ can be 2-coloured so that edges of both colours are incident with every non-leaf vertex. (Note: this will probably not be a* proper *2-edge-colouring of $G$.)*

**PROOF.** We first consider the case where every vertex of $G$ has even valency.

Choose a vertex $v$ of $G$ subject to the constraint that if any vertex of $G$ has valency greater than 2, then $v$ is such a vertex. Since every vertex of $G$ has even valency, we can find an Euler tour of $G$ that begins and ends at $v$. Alternate edge colours around this tour. Clearly, every vertex that is visited in the middle of the tour (that is, every vertex except possibly $v$) must be incident with edges of both colours, since whichever colour is given to the edge we travel to reach that vertex, the other colour will be given to the edge we travel when leaving that vertex. If any vertex of $G$ has valency greater than 2, then by our choice of $v$, the valency of $v$ must be greater than 2, so $v$ is visited in the middle of the tour, and this colouring has the desired property. If every vertex of $G$ has valency 2, then since $G$ is connected, $G$ must be a cycle (see

Exercise 12.3.10.4. Since $G$ is not a cycle of odd length (by hypothesis), $G$ must be a cycle of even length. Therefore the number of edges of $G$ is even, so the tour will begin and end with edges of opposite colours, both of which are incident with $v$. Again we see that this colouring has the desired property.

Suppose now that $G$ has at least one vertex of odd valency.

Create a new vertex, $u$, and add edges between $u$ and every vertex of $G$ that has odd valency. All of the vertices of $G$ will have even valency in this new graph, and by Corollary 11.3.12 (the corollary to Euler's handshaking lemma), $u$ must also have even valency, so this graph has an Euler tour; in fact, we can find an Euler tour that begins with the vertex $u$. Alternate edge colours around this tour. Delete $u$ (and its incident edges) but retain the colours on the edge of $G$. We claim that this colouring will have the desired property. If a vertex $v$ has even valency in $G$, it must be visited in the middle of the tour and the edges we travel on to reach $v$ and to leave $v$ will have different colours. Neither of these edges is incident with $u$, so both are in $G$. If a vertex $v$ has valency 1 in $G$, then $v$ is a leaf and the colour of its incident edge doesn't matter. If a vertex $v$ has odd valency at least 3 in $G$, then $v$ is visited at least twice in the middle of the tour. Only one of these visits can involve the edge $uv$, so during any other visit, the edges we travel on to reach $v$ and to leave $v$ will have different colours. Neither of these edges is incident with $u$, so both are in $G$. Thus, this colouring has the desired property. □

**NOTATION 14.1.15.** Given a (not necessarily proper) edge-colouring $\mathcal{C}$, we use $c(v)$ to denote the number of distinct colours that have be used on edges that are incident with $v$. Clearly, $c(v) \leq d(v)$.

**DEFINITION 14.1.16.** An edge colouring $\mathcal{C}'$ is an **improvement** on an edge colouring $\mathcal{C}$ if it uses the same colours as $\mathcal{C}$, but $\sum_{v \in V} c'(v) > \sum_{v \in V} c(v)$.

An edge colouring is **optimal** if no improvement is possible.

Notice that since $c(v) \leq d(v)$ for every $v \in V$, if

$$\sum_{v \in V} c(v) = \sum_{v \in V} d(v)$$

then we must have $c(v) = d(v)$ for every $v \in V$. This is precisely equivalent to the definition of a proper colouring.

At last, we are ready to prove that bipartite graphs are class one.

**THEOREM 14.1.17.** *If $G$ is bipartite, then $\chi'(G) = \Delta(G)$.*

**PROOF.** Let $G$ be a bipartite graph. Towards a contradiction, suppose that $\chi'(G) > \Delta(G)$.

Let $\mathcal{C}$ be an optimal $\Delta(G)$-edge-colouring of $G$. By assumption, $\mathcal{C}$ will not be a proper edge colouring, so there must be some vertex $u$ such that $c(u) < d(u)$. By the Pigeonhole Principle, some colour $j$ must be used to colour at least two of the edges incident with $u$, and since there are $\Delta(G) \geq d(u)$ colours in total and only $c(u)$ are used on edges incident with $u$, there must be some colour $i$ that is not used to colour any edge incident with $u$.

Consider only the edges of $G$ that have been coloured with either $i$ or $j$ in the colouring $\mathcal{C}$. Since $G$ is bipartite, these edges cannot include an odd cycle. We apply Lemma 14.1.14 to each connected component formed by these edges to re-colour these edges . Our re-colouring will use only colours $i$ and $j$, and if a vertex $v$ was incident to at least two edges coloured with either $i$ or $j$ in $\mathcal{C}$, then under the re-colouring, $v$ will be incident with at least one edge coloured with $i$ and at least one edge coloured with $j$. Leave all of the other edge colours alone, and call this new colouring $\mathcal{C}'$.

We claim that $\mathcal{C}'$ is an improvement on $\mathcal{C}$. Any vertex $v$ that had at most one incident edge coloured with either $i$ or $j$ under $\mathcal{C}$, will still have exactly the same colours except that the edge coloured $i$ or $j$ might have switched its colour to the other of $i$ and $j$. In any case, we will have $c'(v) = c(v)$. Any vertex $v$ that had at least two incident edges coloured with either $i$ or $j$ under $\mathcal{C}$, will still have all of the same colours except that it will now have incident edges coloured with both $i$ and $j$, so $c'(v) \geq c(v)$. Furthermore, we have $c'(u) > c(u)$ since the edges incident with $u$ now include edges coloured with both $i$ and $j$, where before there were only edges coloured with $j$. Thus,

$$\sum_{v \in V} c'(v) \geq \sum_{v \in V} c(v),$$

so $\mathcal{C}'$ is an improvement on $\mathcal{C}$, as claimed.

We have contradicted our assumption that $\mathcal{C}$ is an optimal $\Delta(G)$-edge-colouring. This contradiction serves to prove that $\chi'(G) = \Delta(G)$. □

**EXERCISES 14.1.18.**

1) Prove that every tree is a class-one graph.

2) Prove that every cycle of odd length is a class-two graph.

3) Find a graph that contains a cycle of odd length, but is a class-one graph.

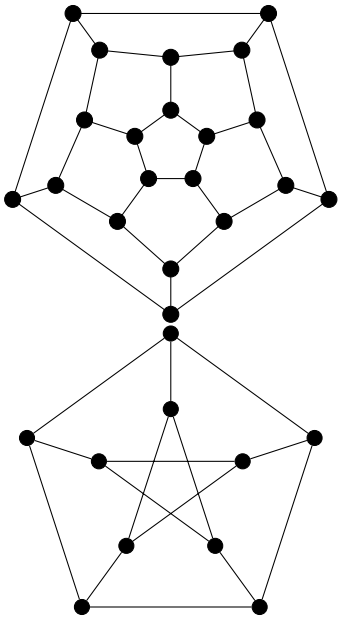
4) For each of the following graphs, find the edge-chromatic number, determine whether the graph is class one or class two, and find a proper edge-colouring that uses the smallest possible number of colours.

   (a) The two graphs in Exercise 13.2.12.2.

   (b) The two graphs in Example 14.1.11.

   (c) The skeleton of a dodecahedron (the upper of the two graphs drawn below).

   (d) The *Petersen graph* (the lower of the two graphs drawn below). You may assume, without proof, that the Petersen graph is class two.



5) Find a systematic approach to colouring the edges of complete graphs that demonstrates that $\chi'(K_{2n-1}) = \chi'(K_{2n}) = 2n - 1$.

6) Find a systematic approach to colouring the edges of complete bipartite graphs that demonstrates that $\chi'(K_{m,n}) = \Delta(K_{m,n}) = \max\{m, n\}$.

**EXERCISES 14.1.19.** The following exercises illustrate some of the connections between Hamilton cycles and edge-colouring.

1) **Definition.** A graph is said to be **Hamilton-connected** if there is a Hamilton path from each vertex in the graph to each of the other vertices in the graph.
Prove that if $G$ is bipartite and has at least 3 vertices, then $G$ is not Hamilton-connected.
[*Hint:* Prove this by contradiction. Consider the length of a Hamilton path and where it can end.]
2) Suppose that $G$ is a bipartite graph with $V_1$ and $V_2$ forming a bipartition. Show that if $|V_1| \neq |V_2|$ then $G$ has no Hamilton cycle.
3) Prove that if every vertex of $G$ has valency 3, and $G$ has a Hamilton cycle, then $G$ is class one.
[*Hint:* Use Corollary 11.3.12 (the corollary to Euler's handshaking lemma), and find a way to assign colours to the edges of the Hamilton cycle.]

## 14.2. Ramsey Theory

Although Ramsey theory is an important part of combinatorics (along with enumeration, graph theory, and design theory), this course will touch on it only very lightly. As noted in Section 1.3, this branch of combinatorics is named for Frank Plumpton Ramsey (1903—1930). The basic idea is that if a very large object is cut into two pieces (or a small number of pieces), then at least one of the pieces must contain a very nice subset. Here is an illustration.

**EXAMPLE 14.2.1.** Suppose each edge of $K_6$ is coloured either red or blue. Show that either there is a triangle whose edges are all red, or there is a triangle whose edges are all blue. That is, $K_6$ contains a copy of $K_3$ that has all of its edges of the same colour. For short, we say that $K_6$ contains a *monochromatic* triangle.

**SOLUTION.** Choose some vertex $v$. Since the 5 edges incident with $v$ are coloured with only two colours, the generalized Pigeonhole Principle implies that three of these edges are the same colour. For definiteness, let us say that three edges $vu_1$, $vu_2$, and $vu_3$ are all red.

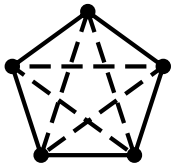
Now, $u_1$, $u_2$, and $u_3$ are the vertices of a copy of $K_3$ that is inside $K_6$. If all three edges of this $K_3$ are blue, then we have our desired monochromatic triangle (specifically, a blue triangle). So we may assume that one of the edges is red; say, $u_1u_2$ is red. Since the edges $vu_1$ and $vu_2$ are also red, we see that $v$, $u_1$, and $u_2$ are the vertices of a monochromatic triangle (specifically, a red triangle). □

**DEFINITION 14.2.2.** Let $k, \ell \in \mathbb{N}^+$.

1) Suppose each edge of $K_n$ is coloured either red or blue. We say **there is a red copy of $K_k$** if there exist $k$ vertices $u_1, \ldots, u_k$, such that the edge $u_iu_j$ is red for all $i$ and $j$ (with $i \neq j$). Similarly, we say **there is a blue copy of $K_\ell$** if there exist $\ell$ vertices $v_1, \ldots, v_\ell$, such that the edge $v_iv_j$ is blue for all $i$ and $j$ (with $i \neq j$).
2) The **Ramsey number** $R(k, \ell)$ is the smallest number $n$, such that whenever each edge of $K_n$ is coloured either red or blue, there is always either a red copy of $K_k$, or a blue copy of $K_\ell$.

**EXAMPLE 14.2.3.**

1) We have $R(k,1) = 1$ for all $k$. This is because $K_1$ has no edges, so, for any colouring of any $K_n$, it is true (vacuously) that all of the edges of $K_1$ are blue.
2) We have $R(k,2) = k$ for all $k$. If some edge is blue, then there is a blue $K_2$, while if there are no blue edges, then the entire graph is a red $K_k$.
3) We have $R(3,3) = 6$. To see this, note that Example 14.2.1 shows $R(3,3) \leq 6$, while there is no monochromatic triangle in the edge-colouring of $K_5$ pictured below (because the only monochromatic cycles are of length 5), so $R(3,3) > 5$.



4) We have $R(k,\ell) = R(\ell,k)$ for all $k$ and $\ell$. If every colouring of $K_n$ has either a red $K_k$ or a blue $K_\ell$, then we see that every colouring of $K_n$ must have either a red $K_\ell$ or a blue $K_k$, just by switching red and blue in the colouring.
5) If $k \leq k'$ and $\ell \leq \ell'$, then $R(k,\ell) \leq R(k',\ell')$. We have a colouring of $K_n$ that contains either a red $K_{k'}$ or a blue $K_{\ell'}$. Since $k \leq k'$ and $\ell \leq \ell'$, we know that any $K_{k'}$ contains a copy of $K_k$, and any $K_{\ell'}$ contains a copy of $K_\ell$.

It is not at all obvious that $R(k,\ell)$ exists: theoretically, $R(4,4)$ might not exist, because it might be possible to colour the edges of a very large $K_n$ in such a way that there is no monochromatic $K_4$. Fortunately, the following extension of the proof of Example 14.2.1 implies that $R(k,\ell)$ does exist for all $k$ and $\ell$. In fact, it provides a bound on how large $R(k,\ell)$ can be (see Exercise 14.2.5.3 below).

**PROPOSITION 14.2.4.** *$R(k,\ell) \leq R(k-1,\ell) + R(k,\ell-1)$ for all $k,\ell \geq 2$.*

**PROOF.** Let $n = R(k-1,\ell) + R(k,\ell-1)$, and suppose each edge of $K_n$ is coloured either red or blue. We wish to show there is either a red $K_k$ or a blue $K_\ell$.

Choose some vertex $v$ of $K_n$. Then the number of edges incident with $v$ is

$$n-1 = R(k-1,\ell) + R(k,\ell-1) - 1 > \big(R(k-1,\ell)-1\big) + \big(R(k,\ell-1)-1\big),$$

so the very generalized Pigeonhole Principle implies that either $R(k-1,\ell)$ of these edges are red, or $R(k,\ell-1)$ of these edges are blue.

For definiteness, let us assume that the edges $vu_1, vu_2, \ldots, vu_r$ are all blue, where $r = R(k,\ell-1)$. Now, $u_1, u_2, \ldots, u_r$ are the vertices of a copy of $K_r$ that is inside $K_n$. Since $r = R(k,\ell-1)$, we know that this $K_r$ contains either a red $K_k$ or a blue $K_{\ell-1}$.

If it contains a red $K_k$, then we have the desired red $K_k$. So we may assume $u_1, \ldots, u_{\ell-1}$ are the vertices of a blue $K_{\ell-1}$. Since the edges $vu_1, vu_2, \ldots, vu_{\ell-1}$ are also blue, we see that $v, u_1, u_2, \ldots, u_{\ell-1}$ are the vertices of the desired blue $K_\ell$. □

**EXERCISES 14.2.5.**

1) Show $R(3,4) > 6$.
2) Using Proposition 14.2.4 and the values of $R(k,\ell)$ given in Example 14.2.3, find the best upper bound you can on $R(k,\ell)$ for $3 \leq k \leq \ell \leq 6$.
3) Show $R(k,\ell) \leq 2^{k+\ell}$ for all $k$ and $\ell$.
   [*Hint:* Use Proposition 14.2.4 and induction on $k+\ell$.]

4) It is known that $40 \leq R(3,10) \leq 42$. Using this information, what can you say about $R(3,11)$?

5) Show there are at least *two* monochromatic triangles in every colouring of the edges of $K_6$ with two colours.
[*Hint:* Show that there must either be two red (say) triangles, or a red triangle and a blue edge whose endvertices are not in the triangle. Then show that any colouring of the edges joining the red triangle with the blue edge creates either a blue triangle or a second red triangle.]

*Remark 14.2.6.* The exact value of $R(k,\ell)$ seems to be extremely difficult to find, except for very small values of $k$ and $\ell$. For example, although it has been proved that $R(4,4) = 18$ and $R(4,5) = 25$, no one has been able to determine the precise value of $R(k,\ell)$ for any situation in which $k$ and $\ell$ are both at least 5. The legendary combinatorist Pál Erdős (1913—1996) said that it would be hopeless to try to calculate the exact value of $R(6,6)$, even with all of the computer resources and brightest minds in the whole world working on the problem for a year. (We do know that $R(6,6)$ is somewhere between 102 and 165.) For more information about the values that have been calculated, see the *Wikipedia* article on *Ramsey's theorem.*

**EXERCISES 14.2.7.** The edges of $K_n$ can also be coloured with more than two colours.

1) Show that every colouring of the edges of $K_{17}$ with 3 colours has a monochromatic triangle.

2) Suppose there is a monochromatic triangle in every colouring of the edges of $K_n$ with $c$ colours. Show that if $N-1 > (c+1)(n-1)$, then every colouring of the edges of $K_N$ with $c+1$ colours has a monochromatic triangle.

Similar arguments (combined with induction on the number of colours) establish the following very general result.

**THEOREM 14.2.8 (Ramsey's Theorem).** *Given $c$ colours and fixed sizes $n_1, \ldots, n_c \geq 1$, there is an integer*

$$r = R(n_1, \ldots, n_c)$$

*such that for any $c$-colouring of the edges of $K_r$, there must be some $i \in \{1, \ldots, c\}$ such that $K_r$ has a subgraph isomorphic to $K_{n_i}$ all of whose edges have been coloured with colour $i$.*

**PROOF.** We will prove this result by induction on $c$, the number of colours.

*Base cases:* When $c = 1$, all edges of $K_r$ are coloured with our single colour, so if we let $r = R(n_1) = n_1$, the whole graph is the $K_{n_1}$ all of whose edges have been coloured with colour 1.

We will also require $c = 2$ to be a base case in our induction. In order to prove this second base case, we perform a second proof by induction, this time on $n_1 + n_2$. To make the proof easier to read, we'll call the two colours in any 2-edge-colouring red and blue, and if all of the edges of a $K_i$ have been coloured with one colour, we'll simply call it a red $K_i$, or a blue $K_i$.

*Base case for the second induction:* We'll actually prove a lot of base cases all at once. Since $n_1$ and $n_2$ are the number of vertices of a complete graph, we must have $n_1, n_2 \geq 1$. A $K_1$ has no edges, so vacuously its edges have whichever colour we desire. Thus if $n_1 = 1$ or $n_2 = 1$, we have $r = R(n_1, n_2) = 1$, since for any 2-edge-colouring of $K_1$, there is a red $K_1$ and a blue $K_1$.

*Inductive step for the second induction:* We begin with the inductive hypothesis. Let $k \geq 2$ be arbitrary. Assume that for every $k_1, k_2 \geq 1$ such that $k_1 + k_2 = k$, there is some integer $r = R(k_1, k_2)$ such that for any 2-edge-colouring of the edges of $K_r$, there is a subgraph that is either a red $K_{k_1}$ or a blue $K_{k_2}$.

Let $n_1, n_2 \geq 1$ such that $n_1 + n_2 = k + 1$. If either $n_1 = 1$ or $n_2 = 1$, then this was one of our base cases and the proof is complete, so we may assume that $n_1, n_2 \geq 2$. Therefore, $n_1 - 1, n_2 - 1 \geq 1$, and $n_1 + n_2 - 1 = k$. Now, by our inductive hypothesis, there is some integer $r_1 = R(n_1, n_2 - 1)$ such that for any 2-edge-colouring of the edges of $K_{r_1}$, there is a subgraph that is either a red $K_{n_1}$ or a blue $K_{n_2-1}$. We can also use our inductive hypothesis to conclude that there is some integer $r_2 = R(n_1 - 1, n_2)$ such that for any 2-edge-colouring of the edges of $K_{r_2}$, there is a subgraph that is either a red $K_{n_1-1}$ or a blue $K_{n_2}$.

We claim that $R(n_1, n_2) \leq r_1 + r_2$. We will show this by proving that any 2-edge-colouring of the edges of $K_{r_1+r_2}$ must have a subgraph that is either a red $K_{n_1}$or a blue $K_{n_2}$.

Consider a complete graph on $r_1 + r_2$ vertices whose edges have been coloured with red and blue. Choose a vertex $v$, and divide the remaining vertices into two sets: $u \in V_1$ if the edge $uv$ has been coloured red, and $u \in V_2$ if the edge $uv$ has been coloured blue. Since this graph has

$$r_1 + r_2 = |V_1| + |V_2| + 1$$

vertices, we must have either $|V_1| \geq r_2$, or $|V_2| \geq r_1$.

Suppose first that $|V_1| \geq r_2$. Since $r_2 = R(n_1 - 1, n_2)$, the subgraph whose vertices are the elements of $V_1$ has a subgraph that is either a red $K_{n_1-1}$ or a blue $K_{n_2}$. In the latter case, this subgraph is also in our original $K_{r_1+r_2}$ and we are done. In the former case, the subgraph whose vertices are the elements of $V_1 \cup \{v\}$ has a red $K_{n_1}$ and we are done.

Suppose now that $|V_2| \geq r_1$ (the proof is similar). Since $r_1 = R(n_1, n_2 - 1)$, the subgraph whose vertices are the elements of $V_2$ has a subgraph that is either a red $K_{n_1}$ or a blue $K_{n_2-1}$. In the former case, this subgraph is also in our original $K_{r_1+r_2}$ and we are done. In the latter case, the subgraph whose vertices are the elements of $V_2 \cup \{v\}$ has a blue $K_{n_2}$ and we are done.

By the Principle of Mathematical Induction, for every $n_1, n_2 \geq 1$, there is some integer $r = R(n_1, n_2)$ such that for any colouring of the edges of $K_r$, there is a subgraph that is either a red $K_{n_1}$ or a blue $K_{n_2}$.

This second proof by induction completes the proof of the second base case for our original induction on $c$, the number of colours. We are now ready for the inductive step for our original proof by induction.

*Inductive step:* We begin with the inductive hypothesis. Let $m \geq 2$ be arbitrary. Assume that for every $k_1, \ldots, k_m \geq 1$, there is an integer $r = R(k_1, \ldots, k_m)$ such that for any $m$-colouring of the edges of $K_r$, there must be some $i \in \{1, \ldots, m\}$ such that $K_r$ has a subgraph isomorphic to $K_{k_i}$, all of whose edges have been coloured with colour $i$.

Let $n_1, \ldots, n_{m+1}$ be arbitrary. Take a complete graph on

$$r = R(n_1, \ldots, n_{m-1}, R(n_m, n_{m+1}))$$

vertices, and colour its edges with $m + 1$ colours. Temporarily consider the colours $m$ and $m + 1$ to be the same, resulting in a colouring of the edges with $m$ colours. By our inductive hypothesis, there must either be some $i \in \{1, \ldots, m - 1\}$ such that our $K_r$ has a subgraph isomorphic to $K_{n_i}$, all of whose edges have been coloured with colour $i$, or $K_r$ has a subgraph isomorphic to $K_{R(n_m,n_{m+1})}$ all of whose edges have been coloured with the $m$th colour (where this $m$th colour is really the combination of the colours $m$ and $m + 1$).

If there is some $i \in \{1, \ldots, m - 1\}$ such that our $K_r$ has a subgraph isomorphic to $K_{n_i}$, all of whose edges have been coloured with colour $i$, then we are done. The possibility remains that our $K_r$ has a subgraph isomorphic to $K_{R(n_m,n_{m+1})}$ all of whose edges have been coloured with either colour $m$ or colour $m + 1$. But by our base case for $c = 2$, this graph must have either a subgraph isomorphic to $K_{n_m}$ all of whose edges have been coloured with colour $m$, or a subgraph isomorphic to $K_{n_{m+1}}$ all of whose edges have been coloured with colour $m + 1$. This completes the inductive step.

By the Principle of Mathematical Induction, for every $c \geq 1$ and fixed sizes $n_1, \ldots, n_c \geq 1$, there is an integer $r = R(n_1, \ldots, n_c)$ such that for any $c$-colouring of the edges of $K_r$, there must be some $i \in \{1, \ldots, c\}$ such that $K_r$ has a subgraph isomorphic to $K_{n_i}$ all of whose edges have been coloured with colour $i$. □

**EXERCISES 14.2.9.**

1) Find $R(2, 2, 3)$.

2) Find $R(2, 4)$.

3) Find a 2-edge-colouring of $K_6$ that does not have a $K_4$ of either colour.

**EXERCISE 14.2.10.** Let $c \in \mathbb{N}^+$, and let $N = R(3, \ldots, 3)$ where there are $c$ entries (all equal to 3). If $\{A_1, A_2, \ldots, A_c\}$ is any partition of $\{1, 2, \ldots, N\}$ into $c$ subsets, show that some $A_i$ contains three integers $x$, $y$, and $z$, such that $x + y = z$. This result is known as Schur's Theorem, named for Issai Schur (1875—1941).
[*Hint:* The vertices of $K_N$ are $1, 2, \ldots, N$. Put colour $i$ on each edge $uv$ with $|u - v| \in A_i$. If $u, v, w$ are the vertices of a monochromatic triangle of colour $i$, with $u > v > w$, then $\{u - v, v - w, u - w\} \subseteq A_i$, and we have $(u - v) + (v - w) = u - w$.]

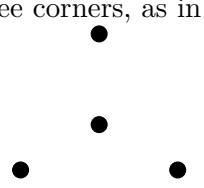
To provide a somewhat broader perspective on Ramsey theory, we conclude this section with a couple of results on other problems that are related to this branch of combinatorics. We begin with a theorem from the 1930s.

**THEOREM 14.2.11 (Happy Ending Theorem).** *Any set of* 5 *points in the plane such that no three are collinear, form the corners of a convex quadrilateral.*

Eszter Klein (1910—2005) asked the more general question: how many points in the plane are required (assuming that no three are collinear) to ensure that some subset of $n$ of them form the corners of a convex $n$-gon? She solved this question for the case $n = 4$ in the above theorem, but in the more general context only upper (and lower) bounds are known; Erdős and György Szekeres (1911—2005) were the first to publish such bounds. Erdős dubbed this the "Happy Ending" problem (and the special case is sometimes known as the Happy Ending Theorem) because this collaboration developed Klein's relationship with Szekeres, who she later married. The Erdős-Szekeres bounds for Klein's general question used Ramsey's Theorem, together with the Happy Ending Theorem.

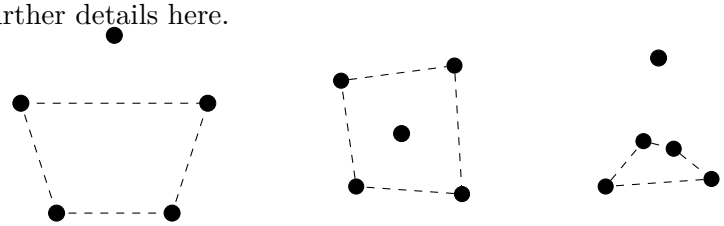
We show in the following example that four points would not be sufficient.

**EXAMPLE 14.2.12.** Recall that in order for a polygon to be convex, the straight line segment connecting any points that lie within or on the boundary of the polygon cannot pass outside the polygon. A convex quadrilateral therefore cannot include a corner that lies in the interior of the triangle formed by the other three corners, as in this configuration:



The remainder of the proof of the Happy Ending Theorem consists of showing that 5 points suffices. There are only three possible configurations for 5 points in the plane: they may form a convex 5-gon; four of them may form a convex quadrilateral with the remaining point in the interior; or three of them may form a triangle with the remaining two points in the interior. These configurations and the convex quadrilaterals that can be found in each case are drawn below. Our final drawing is in fact the only way the last configuration can arise. This means that there is always a convex quadrilateral, but determining exactly how the points can lie in

this final configuration requires some additional consideration of the geometry involved. We will not include further details here.



The other result we present here looks more similar to Ramsey's Theorem. It arises from the general question posed by Kazimierz Zarankiewicz (1902—1959): Given integers $m, n, s$, and $t$, what is the largest number of edges that a bipartite graph whose bipartition sets contain $m$ and $n$ vertices, without containing a complete bipartite $K_{s,t}$ subgraph? The value of the answer to this question is denoted by $z(m, n; s, t)$.

In about 1951 (published in 1954), Tamás Kővári (1930—2010), Vera Sós (1930—), and Pál Turán (1910—1976) proved the following upper bound on this value.

**THEOREM 14.2.13.**

$$z(m, n; s, t) < (s-1)^{1/t}(n-t+1)m^{1-1/t} + (t-1)m.$$

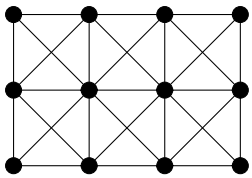
## 14.3. Vertex colouring

Suppose you have been given the task of assigning broadcast frequencies to transmission towers. You have been given a list of frequencies that you are permitted to assign. There is a constraint: towers that are too close together cannot be assigned the same frequency, since they would interfere with each other.

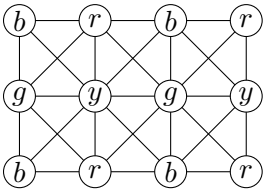
One way to approach this problem is to model it as a graph. The vertices of the graph will represent the towers, and the edges will represent towers that can interfere with each other. Your job is to assign a frequency to each of the vertices. Instead of writing a frequency on each vertex, we will choose a colour to represent that frequency, and use that colour to colour the vertices to which you assign that frequency.

Here is an example of this.

**EXAMPLE 14.3.1.** This graph represents the towers and their interference patterns.



This represents a possible assignment of 4 colours to the vertices. The colour of each vertex (red, green, blue, or yellow) is indicated by writing the first letter of the colour's name on the vertex).




Notice that this colouring obeys the constraint that interfering towers are not assigned the same frequencies.

**DEFINITION 14.3.2.** A **proper $k$-vertex-colouring** (or just $k$-colouring) of a graph $G$ is a function that assigns to each vertex of $G$ one of $k$ colours, such that adjacent vertices must be assigned different colours.

As with edge-colouring, the constraint that adjacent vertices receive different colours turns out to be a useful constraint that arises in many contexts. We often represent the $k$ colours by the numbers $1, \ldots, k$, and label the vertices with the appropriate numbers rather than colouring them.

**DEFINITION 14.3.3.** A graph $G$ is **$k$-colourable** if it admits a proper $k$-(vertex-)colouring. The smallest integer $k$ for which $G$ is $k$-colourable is called the **chromatic number** of $G$.

**NOTATION 14.3.4.** The chromatic number of $G$ is denoted by $\chi(G)$, or simply by $\chi$ if the context is unambiguous.

We leave the proof of the following as an exercise (see Exercise 14.3.14.2).

**PROPOSITION 14.3.5.** *For every $n \geq 1$, $\chi(K_n) = n$.*

**EXAMPLE 14.3.6.** Prove that for a graph $G$, $\chi(G) = 2$ if and only if $G$ is a bipartite graph that has at least one edge.

**PROOF.** ($\Rightarrow$) Suppose that $\chi(G) = 2$. Take a proper 2-colouring of $G$ with colours 1 and 2. Let $V_1$ denote the set of vertices of colour 1, and let $V_2$ denote the set of vertices of colour 2. Since the colouring is proper, there are no edges both of whose endvertices are in $V_1$ (as these would be adjacent vertices both coloured with colour 1). Similarly, there are no edges both of whose endvertices are in $V_2$. Thus, the sets $V_1$ and $V_2$ form a bipartition of $G$, so $G$ is bipartite. Since 2 colours were required to properly colour $G$, $G$ must have at least one edge.

($\Leftarrow$) Suppose that $G$ is bipartite, and that $V_1$ and $V_2$ form a bipartition of $G$. Colour the vertices in $V_1$ with colour 1, and colour the vertices of $V_2$ with colour 2. By the definition of a bipartition, no pair of adjacent vertices can have been assigned the same colour. Thus, this is a proper 2-colouring of $G$, so $\chi(G) \leq 2$. Since $G$ has at least one edge, the endpoints of that edge must be assigned different colours, so $\chi(G) \geq 2$. Thus $\chi(G) = 2$. □

**EXAMPLE 14.3.7.** Show that for any $n \geq 1$, $\chi(C_{2n+1}) = 3$.

**SOLUTION.** Since this graph has an edge whose endvertices must be assigned different colours, we see that $\chi(C_{2n+1}) \geq 2$. Since a cycle of odd length is not bipartite (see Theorem 14.1.13), Example 14.3.6 shows that $\chi(C_{2n+1}) \neq 2$, so $\chi(C_{2n+1}) \geq 3$. Let the cycle be $(u_1, u_2, \ldots, u_{2n+1}, u_1)$. Since the only edges in the graph are between consecutive vertices in this list, if we assign colour 1 to $u_1$, colour 2 to $u_{2i}$ for $1 \leq i \leq n$, and colour 3 to $u_{2i+1}$ for $1 \leq i \leq n$, this will be a proper 3-colouring. Thus, $\chi(C_{2n+1}) = 3$. □

**DEFINITION 14.3.8.** A graph $G$ is **$k$-critical** if $\chi(G) = k$, but for every proper subgraph $H$ of $G$, $\chi(H) < \chi(G)$.

**PROPOSITION 14.3.9.** *Any $k$-critical graph is connected.*

**PROOF.** Towards a contradiction, suppose that $G$ is a disconnected $k$-critical graph, and let $G_1$ and $G_2$ be (nonempty) subgraphs of $G$ such that every vertex of $G$ is in either $G_1$ or $G_2$, and there is no edge from any vertex in $G_1$ to any vertex in $G_2$. By the definition of $k$-critical, $\chi(G_1) < \chi(G)$ and $\chi(G_2) < \chi(G)$. But if we colour $G_1$ with $\chi(G_1)$ colours and $G_2$ with $\chi(G_2)$

colours, since there is no edge from any vertex of $G_1$ to any vertex of $G_2$, this produces a proper colouring of $G$ with

$$\max(\chi(G_1), \chi(G_2)) < \chi(G)$$

colours. This contradiction serves to prove that every $k$-critical graph is connected. □

**THEOREM 14.3.10.** *If $G$ is $k$-critical, then $\delta(G) \geq k - 1$.*

**PROOF.** Towards a contradiction, suppose that $G$ is $k$-critical and has a vertex $v$ of valency at most $k-2$. By the definition of $k$-critical, $G \setminus \{v\}$ must be $(k-1)$-colourable. Now, since $v$ has no more than $k-2$ neighbours, its neighbours can be assigned at most $k-2$ distinct colours in this colouring. Therefore, amongst the colours used in the $(k-1)$-colouring of $G \setminus \{v\}$, there must be a colour that is not assigned to any of the neighbours of $v$. If we assign this colour to $v$, the result is a proper $(k-1)$-colouring of $G$, contradicting $\chi(G) = k$. This contradiction serves to prove that every $k$-critical graph has minimum valency at least $k - 1$. □

**COROLLARY 14.3.11.** *For any graph $G$, $\chi(G) \leq \Delta(G) + 1$.*

**PROOF.** Let $G$ be an arbitrary graph. By deleting as many edges and vertices as it is possible to delete without reducing the chromatic number (we can never increase the chromatic number by deleting vertices or edges, see Exercise 14.3.14.1), we see that $G$ must have a subgraph $H$ that is $\chi(G)$-critical. By Theorem 14.3.10, we see that

$$\delta(H) \geq \chi(G) - 1.$$

Thus, every vertex of $H$ has valency at least $\chi(G) - 1$, so in $G$, these same vertices still have valency at least $\chi(G) - 1$. For any such vertex $v$, we have

$$\Delta(G) \geq d(v) \geq \chi(G) - 1,$$

so $\chi(G) \leq \Delta(G) + 1$. □

We have already seen two families of graphs for which this bound is attained: for complete graphs, we have

$$\Delta(K_n) + 1 = (n-1) + 1 = n = \chi(K_n)$$

(see Proposition 14.3.5); and for cycles of odd length, we have

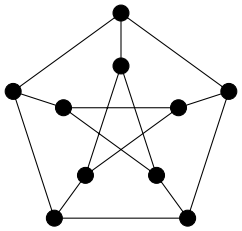
$$\Delta(C_{2n+1}) + 1 = 2 + 1 = 3 = \chi(C_{2n+1})$$

(see Example 14.3.7). In fact, Rowland Leonard Brooks (1916—1993) proved in 1941 that these are the only connected graphs for which this bound is obtained.

**THEOREM 14.3.12 (Brooks' Theorem).** *If $G$ is connected and for every $n \geq 1$, $G \ncong C_{2n+1}$ and $G \ncong K_n$, then $\chi(G) \leq \Delta(G)$.*

We will not include the proof of this result in this course. This theorem does allow us to determine the chromatic number of some graphs with very little work.

**EXAMPLE 14.3.13.** The following very famous graph is called the **Petersen graph**, named for Peter Christian Julius Petersen (1839—1910). It is an exceptional graph in many ways, so when mathematicians are trying to come up with a proof or a counterexample in graph theory, it is often one of the first examples they will check. Find its chromatic number.



**SOLUTION.** We have $\Delta = 3$, and since this graph is neither a complete graph nor a cycle of odd length, by Brooks' Theorem this shows that $\chi \leq 3$. We can find a cycle of length 5 around the outer edge of the graph, so this graph is not bipartite but has an edge. Therefore (by Example 14.3.6), $\chi > 2$. Hence $\chi = 3$. □

**EXERCISES 14.3.14.**

1) Prove that if $H$ is a subgraph of $G$ then $\chi(G) \geq \chi(H)$.

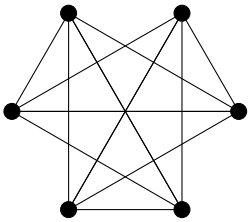
2) Prove Proposition 14.3.5 by induction.

3) Prove Corollary 14.3.11 by induction for every graph on at least one vertex.

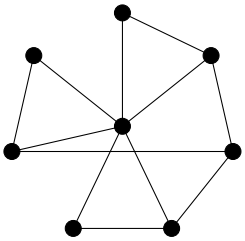
4) For each $i, j \in \{4, 5, 6\}$, suppose you are given a graph $G$ that contains a subgraph isomorphic to $K_i$ and no vertex has more than $j$ neighbours. What (if anything) can you say about $\chi(G)$? Can you say more if you know that $G$ is connected and is neither a complete graph nor a cycle of odd length?

**EXERCISES 14.3.15.** For each of the following graphs, determine its chromatic number by using theoretical arguments to provide a lower bound, and then producing a colouring that meets the bound. Do the same for the edge-chromatic number.

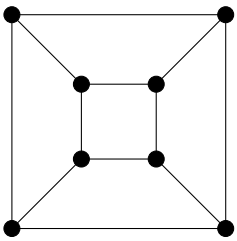
1)



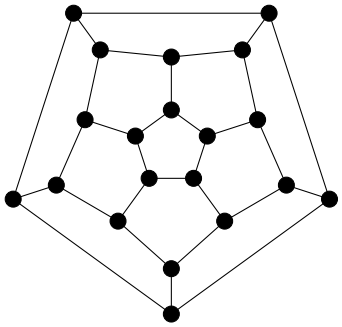
2)



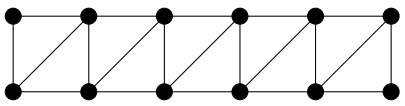
3)



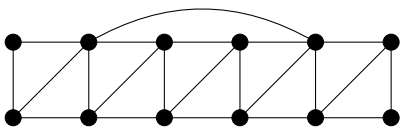
4)



5)



6)



We have seen that the largest complete subgraph of a graph provides a lower bound on the chromatic number of the graph. In general, this isn't a very good lower bound, and we've certainly seen examples where it is not achieved.

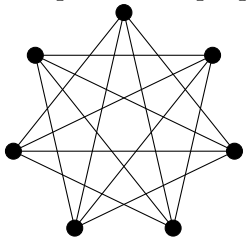
**DEFINITION 14.3.16.** A graph $G$ is a **perfect graph** if for every induced subgraph $H$ of $G$, the chromatic number $\chi(H)$ is equal to the number of vertices in the largest complete subgraph of $H$.

A characterisation of perfect graphs had long been conjectured, but was finally proved in 2006. Although Maria Chudnovsky (1977—), George Neil Robertson (1938—), Paul Seymour (1950—), and Robin Thomas (1962—2020) were involved in this proof and are credited with the theorem, it is the main result in the Ph.D. thesis of Chudnovsky.

**THEOREM 14.3.17 (Strong Perfect Graph Theorem).** *A graph is perfect if and only if no induced subgraph is either a cycle of odd length at least* 5*, or the complement of a cycle that has odd length at least* 5*.*

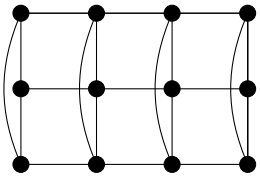
Note that any cycle of odd length has chromatic number 3, but its largest complete subgraph is $K_2$, so such a graph can certainly not be perfect, and therefore by definition any graph having such an induced subgraph cannot be perfect. This is less obvious in the case of the complement of an odd cycle. We will not try to prove in general that the complement of an odd cycle is not perfect, but we provide an illustrative example.

**EXAMPLE 14.3.18.** The following graph is the complement of a cycle of length 7. Its largest complete subgraph is a $K_3$. However, if we colour any of its triangles with three colours and determining all of the colours that would be forced by that beginning in a proper 3-colouring, it is not hard to see that 4 colours are required in a proper colouring.



We also provide an example of a perfect graph.

**EXAMPLE 14.3.19.** This graph is perfect.



Its largest complete subgraph has three vertices (any column of vertices). It can be properly coloured with 3 colours.

**SUMMARY:**

- Vizing's Theorem
- graphs are bipartite if and only if they contain no cycle of odd length
- Ramsey's Theorem
- graphs are bipartite if and only if they are 2-colourable
- Brooks' Theorem
- Petersen graph
- Strong Perfect Graph Theorem
- Important definitions:
  - proper $k$-edge-colouring, $k$-edge-colourable
  - edge chromatic number, chromatic index
  - class one graph, class two graph
  - bipartite, bipartition
  - complete bipartite graph
  - proper $k$-colouring, $k$-colourable
  - chromatic number
  - $k$-critical
  - perfect graph
- Notation:
  - $\chi'(G)$
  - $K_{m,n}$
  - $R(n_1, \ldots, n_c)$
  - $\chi(G)$

Chapter 15

# Planar graphs

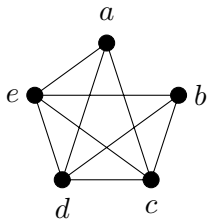
## 15.1. Planar graphs

Visually, there is always a risk of confusion when a graph is drawn in such a way that some of its edges cross over each other. Also, in physical realisations of a network, such a configuration can lead to extra costs (think about building an overpass as compared to building an intersection). It is therefore helpful to be able to work out whether or not a particular graph can be drawn in such a way that no edges cross.

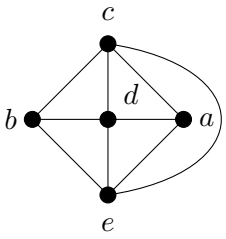
**DEFINITION 15.1.1.** A graph is **planar** if it can be drawn in the plane ($\mathbb{R}^2$) so edges that do not share an endvertex have no points in common, and edges that do share an endvertex have no other points in common.

Such a drawing is called a **planar embedding** of the graph.

**EXAMPLE 15.1.2.** The following graph is planar:




Here is a planar embedding:




**THEOREM 15.1.3.** *The graph $K_5$ is not planar.*

**PROOF.** Label the vertices of $K_5$ as $v_1, \ldots, v_5$. Consider the 3-cycle $(v_1, v_2, v_3, v_1)$. The vertex $v_4$ must lie either inside or outside the boundary determined by this 3-cycle. Furthermore, since there is an edge between $v_4$ and $v_5$, the vertex $v_5$ must lie on the same side (inside or outside) as $v_4$.

Suppose first that $v_4$ and $v_5$ lie inside the boundary. The edges $v_1v_4$, $v_2v_4$, and $v_3v_4$ divide the area inside the boundary into three regions, and $v_5$ must lie inside one of these three regions.

One of $v_1$, $v_2$, and $v_3$ is not a corner of this region, and in fact lies outside of it while $v_5$ lies inside of it, making it impossible to draw the edge from this vertex to $v_5$.

The proof is similar if $v_4$ and $v_5$ lie on the outside of the boundary determined by the 3-cycle $(v_1, v_2, v_3, v_1)$. □

**THEOREM 15.1.4.** *The complete bipartite graph $K_{3,3}$ is not planar.*

**PROOF.** Label the vertices in one of the bipartition sets as $v_1, v_2, v_3$, and the vertices in the other part as $u_1, u_2, u_3$. Consider the 4-cycle

$$(v_1, u_1, v_2, u_2, v_1).$$

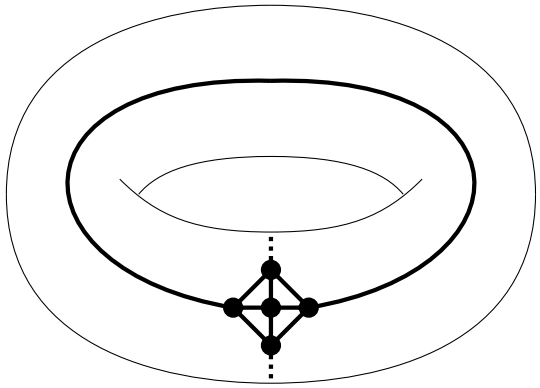
The vertex $v_3$ must lie either inside or outside the boundary determined by this 4-cycle. Furthermore, since there is an edge between $v_3$ and $u_3$, the vertex $u_3$ must lie on the same side (inside or outside) as $v_3$.

Suppose first that $v_3$ and $u_3$ lie inside the boundary. The edges $v_3u_1$ and $v_3u_2$ divide the area inside the boundary into two regions, and $u_3$ must lie inside one of these two regions. One of $v_1$ and $v_2$ does not lie on the boundary of this region, and in fact lies outside of it while $u_3$ lies inside of it, making it impossible to draw the edge from this vertex to $u_3$.

The proof is similar if $v_3$ and $u_3$ lie on the outside of the boundary determined by the 4-cycle $(v_1, u_1, v_2, u_2, v_1)$. □

However, both $K_5$ and $K_{3,3}$ can be embedded onto the surface of what we call a torus (a doughnut shape), with no edges meeting except at mutual endvertices. Embeddings are shown in Figures 15.1.1 and Figure 15.1.2.

**Figure 15.1.1.** $K_5$ embedded on a torus



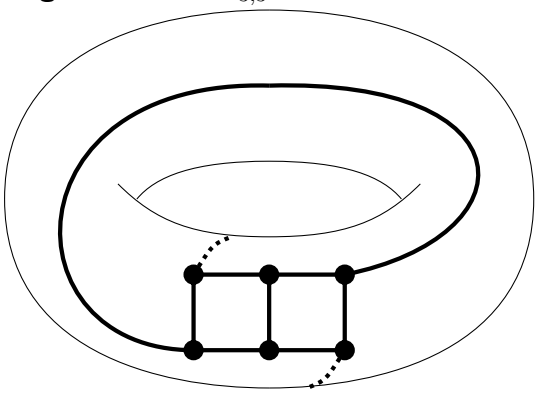
The dotted edge wraps around through the hole in the torus.

You might think at this point that every graph can be embedded on the torus without edges meeting except at mutual endvertices, but this is not the case. In fact, for any surface there are graphs that cannot be embedded in that surface (without any edges meeting except at mutual endvertices).

For any embedding of a planar graph, there is another embedded planar graph that is closely related to it, which we will now describe. Notice that a planar embedding partitions the plane into regions.

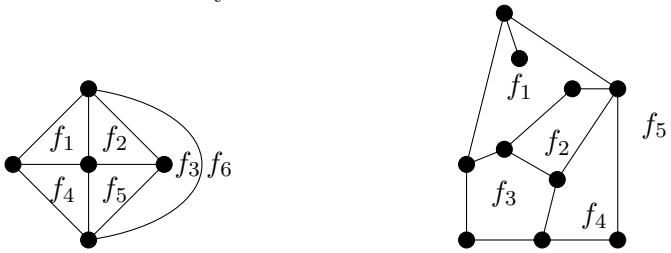
**DEFINITION 15.1.5.** The regions into which a planar embedding partitions the plane, are called the **faces** of the planar embedding.

**Figure 15.1.2.** $K_{3,3}$ embedded on a torus



The dotted edge wraps around through the hole in the torus.

**EXAMPLE 15.1.6.** In these drawings, we have labeled the faces of the two planar embeddings with $f_1$, $f_2$, etc., to show them clearly.



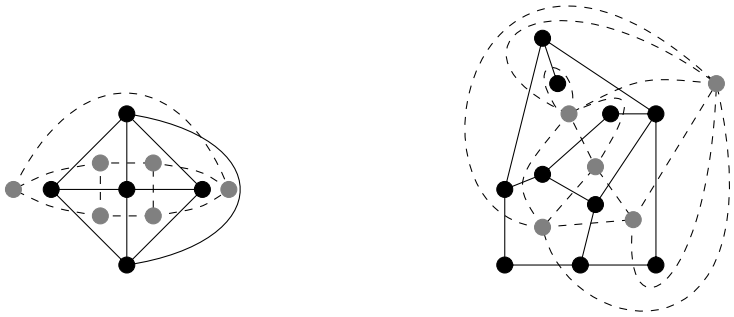

**NOTATION 15.1.7.** We use $F(G)$ (or simply $F$ if the graph is clear from the context) to denote the set of faces of a planar embedding.

**DEFINITION 15.1.8.** We say that an edge is **incident with a face** of a planar embedding, if it lies on the boundary of the face (or inside the face).

For a planar embedding of $G$, the **dual graph** or **planar dual**, $G^*$, is defined by $V(G^*) = F(G)$, and $f_i \sim f_j$ if and only if there is an edge of $G$ that is incident with both $f_i$ and $f_j$.

It is possible that the dual graph of a planar embedding will not be a simple graph, even if the original graph was simple.

**EXAMPLE 15.1.9.** Here we show how to find the planar duals of the embeddings given in Example 15.1.6. We include the original embedding as above; the grey vertices and dashed edges are the vertices and edges of the dual graph.



Note that the second graph has loops and multiedges. Note also that although $f_1$ and $f_5$ meet at a vertex in the embedding of the first graph, they are not adjacent in the dual since they do not share a common edge.

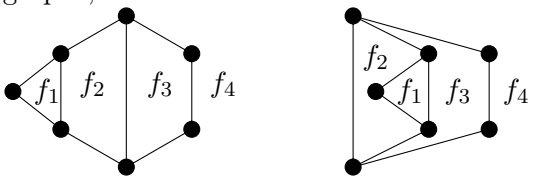
Some other useful observations:

- $|E(G)| = |E(G^*)|$, and every dashed edge crosses exactly one black edge;
- the valency of the vertex $f_i$ in $G^*$ is equal to the number of edges you trace, if you trace around the perimeter of the face $f_i$ in $G$ (so edges that dangle inside the face get counted twice).

**PROPOSITION 15.1.10.** *The dual graph of a planar embedding has a natural planar embedding, so is a planar graph. Furthermore,* $(G^*)^* = G$.

Both of these facts follow fairly directly from the definitions.

**EXAMPLE 15.1.11.** Be careful! — Two different planar embeddings of the same graph may have nonisomorphic dual graphs, as we show here.




In the planar dual of the embedding on the left, $f_1$ will have valency 3; $f_2$ and $f_3$ will have valency 4; and $f_4$ will have valency 7. In the planar dual of the embedding on the right, $f_1$ will have valency 3; $f_2$ will have valency 5; $f_4$ will have valency 4, and $f_3$ will have valency 6. Since the lists $3, 4, 4, 7$ and $3, 4, 5, 6$ are not permutations of each other, the planar duals cannot be isomorphic.

Before moving on to other related topics, we present a classification of planar graphs. This is a theorem by Kazimierz Kuratowski (1896—1980) (from whose name the notation for complete graphs is taken). He proved this result in 1930.

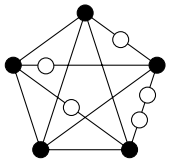
We need one new definition.

**DEFINITION 15.1.12.** An edge $uv$ can be **subdivided** by placing a vertex somewhere along its length. Technically, this means deleting $uv$, adding a new vertex $x$, and adding the edges $ux$ and $vx$.

A **subdivision** of a graph is a graph that is obtained by subdividing some of the edges of the graph.

**EXAMPLE 15.1.13.** An example is shown in Figure 15.1.3. The white vertices are the new ones.

**Figure 15.1.3.** A subdivision of $K_5$



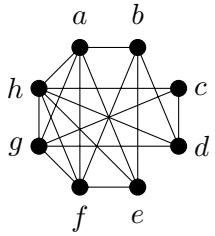
**THEOREM 15.1.14 (Kuratowski's Theorem).** *A graph $G$ is planar if and only if no subgraph of $G$ is a subdivision of $K_5$ or $K_{3,3}$.*

One direction of the proof is fairly straightforward, since we have already proven that $K_5$ and $K_{3,3}$ are not planar. However, we won't try to prove this theorem in this course.

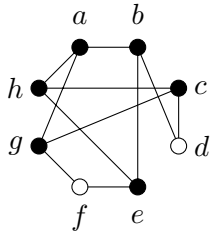
A subdivision of $K_5$ or of $K_{3,3}$ will sometimes be very difficult to find, but efficient algorithms do exist.

Typically, to prove that a graph is planar you would find a planar embedding. To prove that a graph is not planar, you would find a subgraph that is a subdivision of either $K_5$ or $K_{3,3}$.

**EXAMPLE 15.1.15.** Find a subgraph that is a subdivision of $K_5$ or $K_{3,3}$ in this graph, to show that it is not planar.



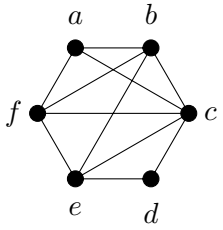

**SOLUTION.** Here is a subdivision of $K_{3,3}$ in the given graph. The white vertices are the vertices that are subdividing edges. Unnecessary edges have been deleted. The bipartition consists of $\{a, c, e\}$ and $\{b, g, h\}$.



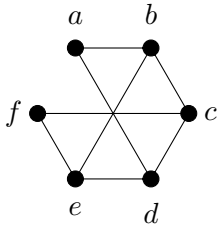

**EXERCISES 15.1.16.**

1) Prove that if a graph $G$ has a subgraph $H$ that is not planar, then $G$ is not planar. Deduce that for every $n \geq 6$, $K_n$ is not planar.

2) Find a planar embedding of the following graph, and find the dual graph of your embedding:



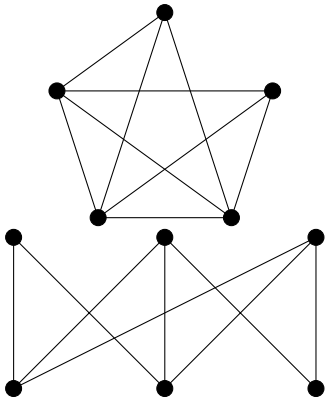

3) Find a planar embedding of the following graph, and find the dual graph of your embedding:




4) The graph in Example 15.1.15 also has a subgraph that is a subdivision of $K_5$. Find such a subgraph.

5) Prove that the Petersen graph is not planar. [*Hint:* Use Kuratowski's Theorem.]

6) Find planar embeddings of the two graphs pictured below. (These graphs are obtained by deleting an edge from $K_5$ and deleting an edge from $K_{3,3}$, respectively.)



Rather than embedding graphs onto a different surface (such as a torus) to avoid edges that cross, it sometimes makes more sense to ask what is the fewest number of possible edge-crossings, for any embedding of the graph into the plane. Certainly for many real-life applications, this may be more useful, as we are looking at the smallest number of overpasses that may be required, or places where we need to add insulation into a chip so that wires that cross do not interfere with each other.

Even this number can get very large, and for some purposes it is more useful to consider how many times any given edge must participate in edge-crossings. Under this model, it is better to spread around the edge crossings so that no edge participates in more than one (for example), than to have one edge appear in all of the crossings, even if this lowers the total number of crossings.

With this in mind, we define a graph to be ***k*-planar** if it can be drawn in the plane so that no edge participates in more than $k$ edge-crossings.

It is possible to test for planarity in linear time, using Kuratowski's Theorem or Wagner's Theorem (which we will discuss later in this chapter), and even to find a planar embedding if the graph is planar, or a minimal subgraph obstructing planarity if the graph is non-planar. One of the best of the known algorithms is the "edge addition method" due to John M. Boyer (1968—) and Wendy Joanne Myrvold (1961—) from 2004. In contrast, testing whether or not a graph is 1-planar is known to be NP-complete. In fact, in a 2020 preprint (accepted but not yet published at the time of this edit), John Cameron Urschel (1991—) and Jake Wellens (1992—) have shown that testing whether or not a graph is $k$-planar for any $k \geq 1$ is NP-complete.

## 15.2. Euler's Formula

Leonhard Euler (1707—1783) came up with a formula that holds true for any planar embedding of a connected graph.

**THEOREM 15.2.1 (Euler's Formula).** *If $G$ is a planar embedding of a connected graph (or multigraph, with or without loops), then*

$$|V| - |E| + |F| = 2.$$

**PROOF 1..** We will prove this formula by induction on the number of faces of the embedding. Let $G$ be a planar embedding of a connected graph (or multigraph, with or without loops).

*Base case:* If $|F| = 1$ then $G$ cannot have any cycles (otherwise the interior and exterior of the cycle would be 2 distinct faces). So $G$ must be a connected graph that has no cycles, i.e., a tree. By Theorem 12.4.5 we know that we must have $|E| = |V| - 1$, so

$$|V| - |E| + |F| = |V| - (|V| - 1) + 1 = 2.$$

A tree cannot have any loops or multiple edges, as these form cycles.

*Inductive step:* We begin by stating our inductive hypothesis. Let $k \geq 1$ be arbitrary, and assume that for any planar embedding of a connected graph (or multigraph, with or without loops) with $k$ faces, $|V| - |E| + |F| = 2$.

Let $G$ be a planar embedding of a connected graph with $k + 1 \geq 2$ faces. Since trees have only one face, $G$ must have a cycle. Choose any edge $e$ that is in a cycle of $G$, and let $H = G \setminus \{e\}$. Clearly, we have

$$|E(H)| = |E(G)| - 1$$

and $|V(H)| = |V(G)|$. Also,

$$|F(H)| = |F(G)| - 1 = k$$

since the edge $e$ being part of a cycle must separate two faces of $G$, which are united into one face of $H$. Furthermore, since $e$ was in a cycle and $G$ is connected, by Proposition 12.3.9 $H$ is connected, and $H$ has a planar embedding induced by the planar embedding of $G$. Therefore our inductive hypothesis applies to $H$, so

$$\begin{aligned} 2 &= |V(H)| - |E(H)| + |F(H)| \\ &= |V(G)| - (|E(G) - 1) + (|F(G)| - 1) \\ &= |V(G)| - |E(G)| + |F(G)| \end{aligned}$$

This completes the inductive step.

By the Principle of Mathematical Induction, $|V| - |E| + |F| = 2$ for any planar embedding of a connected graph (or multigraph, with or without loops). □

The above proof is unusual for a proof by induction on graphs, because the induction is not on the number of vertices. If you try to prove Euler's formula by induction on the number of vertices, deleting a vertex might disconnect the graph, which would mean the induction hypothesis doesn't apply to the resulting graph.

However, there is a different graph operation that reduces the number of vertices by 1, and keeps the graph connected. Unfortunately, it may turn a graph into a multigraph, so it can only be used to prove a result that holds true for multigraphs as well as for graphs. This operation is called *edge contraction*.

**DEFINITION 15.2.2.** Let $G$ be a graph with an edge $uv$. The graph $G'$ obtained by **contracting the edge $uv$** has vertices

$$V(G') = (V(G) \setminus \{u, v\}) \cup \{u'\},$$

where $u'$ is a new vertex. The edges are

$$E(G') = \big([E(G) \setminus \{ux : ux \in E(G)\}] \setminus \{vx : vx \in E(G)\}\big) \cup \{u'y \mid uy \in E(G) \text{ or } vy \in E(G)\}.$$

If you think of adjacent vertices $u$ and $v$ as being joined by a very short elastic that has been stretched out in $G$, then you can think of $G'$ as the graph you get if you allow the elastic to contract, combining the vertices $u$ and $v$ into a "new" vertex $u'$.

Notice that if $G$ is connected, then the graph obtained by contracting any edge of $G$ will also be connected. However, if $uv$ is the edge that we contract, and $u$ and $v$ have a mutual neighbour $x$, then in the graph obtained by contracting $uv$, there will be a multiple edge between $u'$ and $x$. (By our definition, however, even though $u$ and $v$ were adjacent, we do *not* introduce a loop at the new vertex $u'$.) Also, if $G$ has a planar embedding, then after contracting any edge there will still be a planar embedding. If $u \neq v$, then contracting $uv$ reduces the number of vertices by one, reduces the number of edges by one, and does not change the number of faces.

Now we can use this operation to prove Euler's formula by induction on the number of vertices.

**PROOF 2..** Let $G$ be a planar embedding of a connected graph (or multigraph, with or without loops).

*Base case:* If $|V| = 1$ then $G$ has one vertex. Furthermore, every edge is a loop. Every loop involves 1 edge, and encloses 1 face. This graph will therefore have one more face than it has loops (since it has one face even if there are no loops). Thus,

$$|V| - |E| + |F| = 1 - e + (e + 1) = 2.$$

*Inductive step:* We begin by stating our inductive hypothesis. Let $k \geq 1$ be arbitrary, and assume that for any planar embedding of a connected graph (or multigraph, with or without loops) with $k$ vertices, $|V| - |E| + |F| = 2$.

Let $G$ be a planar embedding of a connected graph with $k+1 \geq 2$ vertices. Since the graph is connected and has at least two vertices, it has at least one edge $uv$, with $u \neq v$. Let $G'$ be the graph we obtain by contracting $uv$. Then $G'$ is a planar embedding of a connected graph (or multigraph, with or without loops) on $k$ vertices, so our inductive hypothesis applies to $G'$. Therefore,

$$\begin{aligned} 2 &= |V(G')| - |E(G')| + |F(G')| \\ &= (|V(G)| - 1) - (|E(G) - 1) + |F(G)| \\ &= |V(G)| - |E(G) + |F(G)| \end{aligned}$$

This completes the inductive step.

By the Principle of Mathematical Induction, $|V| - |E| + |F| = 2$ for any planar embedding of a connected graph (or multigraph, with or without loops). □

Contraction of edges has some other very important uses in graph theory. Before looking at some corollaries of Euler's Formula, we'll explain one well-known theorem that involves edge contraction and planar graphs.

**DEFINITION 15.2.3.** Let $G$ be a graph. Then $H$ is a **minor** of $G$ if we can construct $H$ from $G$ by deleting or contracting edges, and deleting vertices.

In 1937, Klaus Wagner (1910—2000) proved a theorem quite similar to Kuratowski's.

**THEOREM 15.2.4 (Wagner's Theorem).** *A graph is planar if and only if it has no minor isomorphic to $K_5$ or $K_{3,3}$.*

It is possible to prove Wagner's Theorem as an easy consequence of Kuratowski's Theorem, since if $G$ has a subgraph that is a subdivision of $K_5$ or $K_{3,3}$ then contracting all but one piece of each subdivided edge gives us a minor that is isomorphic to $K_5$ or $K_{3,3}$. Nonetheless, Wagner's Theorem is important in its own right, as the first example of the much more recent and very powerful work by George Neil Robertson (1938—) and Paul Seymour (1950—) on graph minors.

A family is said to be *minor-closed* if given any graph in the family, any minor of the graph is also in the family. Planar graphs are an example of a minor-closed family, since the operations of deletion (of edges or vertices) and contraction of edges preserve a planar embedding. Robertson and Seymour proved the remarkable result that if a family of graphs is minor-closed, then the family can be characterised by a *finite* set of "forbidden minors." That is, for any such family $\mathcal{F}$, there is a finite set $\mathcal{L}$ of graphs, such that $G \in \mathcal{F}$ if and only if no minor of $G$ appears in $\mathcal{L}$. Wagner's Theorem tells us that when $\mathcal{F}$ is the family of planar graphs, $\mathcal{L} = \{K_5, K_{3,3}\}$.

Historically, mathematicians who study other branches of mathematics have often looked down on graph theory as "easy". The technical and impressive structural proof about forbidden minors by Robertson and Seymour was instrumental in beginning to change that stereotype.

Euler's Formula has some important corollaries.

**COROLLARY 15.2.5.** *Let $G$ be a connected graph. Then every planar embedding of $G$ has the same number of faces.*

**PROOF.** We have $|V| - |E| + |F| = 2$. Since $|V|$ and $|E|$ do not depend on the choice of embedding, we have $|F| = 2 + |E| - |V|$ cannot depend on the choice of embedding. □

**COROLLARY 15.2.6.** *If $G$ is a simple connected planar graph and $|V| \geq 3$, then*

$$|E| \leq 3|V| - 6.$$

*If in addition, $G$ has no cycles of length less than 4, then $|E| \leq 2|V| - 4$.*

**COMBINATORIAL PROOF..** Fix a planar embedding of $G$. We move around each face, counting the number of edges that we encounter, and work out the result in two ways.

First, we look at every face in turn and count how many edges surround that face. Since the graph is simple, every face must be surrounded by at least 3 edges unless there is only one face. If there is only one face and when moving around this face we do not count at least 3 edges, then the graph is a tree that has at most one edge, so $|V| \leq 2$. Therefore, our count will come to at least $3|F|$.

Every edge either separates two faces, or dangles into a face. In the former case, it will be counted once each time we move around one of the two incident faces. In the latter case, it will be counted twice as we move around the face it dangles into: once when we move inwards

along this dangling part, and once when we move back outward. Thus, every edge is counted exactly twice, so our count will come to exactly $2|E|$.

Combining these, we see that $2|E| \geq 3|F|$, so $|F| \leq 2|E|/3$. If $G$ has no cycles of length less than 4, then every face must be surrounded by at least 4 edges, so the same argument gives $2|E| \geq 4|F|$, so $|F| \leq |E|/2$.

By Euler's Formula, $|V| - |E| + |F| = 2$, so

$$|V| - |E| + 2|E|/3 \geq 2.$$

Multiplying through by 3 and moving the $|E|$ terms to the right-hand side, gives

$$3|V| \geq |E| + 6,$$

which can easily be rearranged into the form of our original statement. In the case where $G$ has no cycles of length less than 4, we obtain instead

$$|V| - |E| + |E|/2 \geq 2,$$

so $2|V| \geq |E| + 4$, which again can easily be rearranged into the form given in the statement of this corollary. □

**COROLLARY 15.2.7.** *If $G$ is a simple connected planar graph, then $\delta(G) \leq 5$.*

**PROOF.** Towards a contradiction, suppose that $G$ is a simple connected planar graph, and for every $v \in V$, $d(v) \geq 6$. Then

$$\sum_{v \in V} d(v) \geq 6|V|.$$

By Euler's handshaking lemma, this gives

$$\sum_{v \in V} d(v) = 2|E| \geq 6|V|.$$

Therefore,

$$|E| \geq 3|V| > 3|V| - 6,$$

but this contradicts Corollary 15.2.6. □

Euler's Formula (and its corollaries) give us a much easier way to prove that $K_5$ and $K_{3,3}$ are non-planar.

**COROLLARY 15.2.8.** *The graph $K_5$ is not planar.*

**PROOF.** In $K_5$ we have $|E| = \binom{5}{2} = 10$, and $|V| = 5$. So

$$3|V| - 6 = 15 - 6 = 9 < 10 = |E|.$$

By Corollary 15.2.6, $K_5$ must not be planar. □

**COROLLARY 15.2.9.** *The graph $K_{3,3}$ is not planar.*

**PROOF.** In $K_{3,3}$ we have $|E| = 9$, and $|V| = 6$. So

$$2|V| - 4 = 12 - 4 = 8 < 9 = |E|.$$

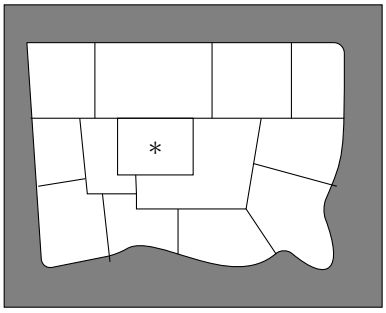
Since $K_{3,3}$ is bipartite, it has no cycles of length less than 4, so by Corollary 15.2.6, $K_{3,3}$ must not be planar. □

**EXERCISES 15.2.10.**

1) Use induction to prove an Euler-like formula for planar graphs that have exactly two connected components.
2) Euler's formula can be generalised to disconnected graphs, but has an extra variable for the number of connected components of the graph. Guess what this formula will be, and use induction to prove your answer.
3) Find and prove a corollary to Euler's formula for disconnected graphs, similar to Corollary 15.2.6. (Use your answer to Exercise 2.)
4) For graphs embedded on a torus, $|V| - |E| + |F|$ has a different (but constant) value, as long as all of the faces "look like" discs. (To be more precise for the sake of anyone familiar with topology, the faces must be embeddable into a plane, rather than looking like a torus. So putting a planar embedding of a graph down on one side of a torus doesn't count.) What is this value?
5) **Definition.** We say that a planar embedding of a graph is **self-dual** if it is isomorphic to its planar dual.
   Prove that if a planar embedding of the connected graph $G$ is self-dual, then $|E| = 2|V| - 2$.
6) Show that if $G$ is a simple planar graph with at least eleven vertices, then the complement of $G$ is not planar.
7) Find a planar graph $G$ with $|V| = 8$ whose complement is also planar.
8) For each of the following sets of conditions, either draw a connected, simple graph $G$ in the plane that satisfies the conditions, or explain how you know that there isn't one.
   (a) The graph has 15 vertices and 12 edges.
   (b) The graph has 10 vertices and 33 edges.
   (c) The graph has 5 vertices and 8 edges.
   (d) The graph has 6 vertices and 9 edges, and the embedding has 6 faces.

## 15.3. Map colouring

Suppose we have a map of an island that has been divided into states.



Traditionally, map-makers colour the different states so as to ensure that if two states share a border, they are not coloured with the same colour. (This makes it easier to distinguish the borders.) If two states simply meet at a corner, then they may be coloured with the same colour.

Using additional colours used to add to the cost of producing the map. Also, if there are too many colours they become harder and harder to distinguish amongst. The question is, without knowing in advance what the map will look like, is there a bound on how many colours will be required to colour it? If so, what is that bound? In other words, what is the largest number of colours that might be required to colour some map?

Well over a century ago, mathematicians observed that it never seemed to require more than 4 colours to colour a map. The map shown above does require 4 colours, since the central rectangular state (marked with an asterisk) and the three states that surround it must all receive different colours (each shares a border with each of the others). Unfortunately, they couldn't prove that no more would ever be required, although a number of purported proofs were published and later found to have errors.

Although the bound of 4 eluded many attempts at proof, in 1890 Percy John Heawood (1861—1955) successfully proved that 5 colours suffice to colour any map. His method was based on an incorrect proof of the Four-Colour Theorem by Sir Alfred Bray Kempe (1849—1922), from 1879. This result is known as the Five-Colour Theorem. Its proof is slightly technical but not difficult, and we will give it in a moment. First we will give a very short proof that 6 colours suffice.

Notice that if we turn the map into a graph by placing a vertex wherever borders meet, and an edge wherever there is a border, this problem is equivalent to finding a proper vertex colouring of the planar dual of this graph. Thus, what we will actually prove is that the vertices of any planar graph can be properly coloured using 6 (or in the subsequent result, 5) colours. There is a detail that we are skimming over here: the planar dual could have loops, which would make it impossible to colour the graph. However, this can only happen if there is a face of the original map that meets itself along a border, which would never happen in a map. The planar dual might also have multiedges, but this does not affect the number of colours required to properly colour the graph, so we can delete any multiedges and assume that we are dealing with a simple planar graph.

**PROPOSITION 15.3.1.** *Every planar graph is properly* 6*-colourable.*

**PROOF.** Towards a contradiction, suppose that there is a planar graph that is not properly 6-colourable. By deleting edges and vertices, we can find a subgraph $G$ that is a 7-critical planar graph.

By Corollary 15.2.7, we must have $\delta(G) \leq 5$ since $G$ is planar. But by Theorem 14.3.10, we must have

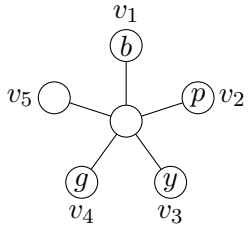
$$\delta(G) \geq 7 - 1 = 6$$

since $G$ is 7-critical. This contradiction serves to prove that every planar graph is properly 6-colourable. □

**THEOREM 15.3.2 (Five-Colour Theorem).** *Every planar graph is properly* 5*-colourable.*

**PROOF.** Towards a contradiction, suppose that there is a planar graph that is not properly 5-colourable. By deleting edges and vertices, we can find a subgraph $G$ that is a 6-critical planar graph. Since $G$ is planar, Corollary 15.2.7 tells us that $\delta(G) \leq 5$. We also know from Theorem 14.3.10 that $\delta(G) \geq 6 - 1 = 5$ since $G$ is 6-critical. Let $v$ be a vertex of valency $\delta(G) = 5$.

By the definition of a $k$-critical graph, $G \setminus \{v\}$ can be properly 5-coloured. Since $G$ itself cannot be properly 5-coloured, the neighbours of $v$ must all have been assigned different colours in the proper 5-colouring of $G \setminus \{v\}$. Let's label the neighbours of $v$ as $v_1$, $v_2$, $v_3$, $v_4$, and $v_5$ as they appear clockwise around $v$. We will call the colour of $v_1$ blue, the colour of $v_2$ purple, the colour of $v_3$ yellow, and the colour of $v_4$ green, as shown in the picture. Here is a picture (where, in order to enable viewing in black-and-white, we have put the first letter of a colour onto a vertex, instead of actually colouring the vertex).




Consider the subgraph consisting of the vertices coloured blue or yellow (and all edges between such vertices). If $v_1$ and $v_3$ are not in the same connected component of this subgraph, then in the connected component that contains $v_1$, we could interchange the colours yellow and blue. Since we are doing this to everything in a connected component of the yellow-blue subgraph, the result will still be a proper colouring, but $v_1$ now has colour yellow, so $v$ can be coloured with blue. This contradicts the fact that $G$ is 6-critical, so it must be the case that $v_1$ and $v_3$ are in the same connected component of the yellow-blue subgraph. In particular, there is a walk from $v_1$ to $v_3$ that uses only yellow and blue vertices. By Proposition 12.3.4, there is in fact a path from $v_1$ to $v_3$ that uses only yellow and blue vertices.

Similarly, if we consider the subgraph consisting of the vertices coloured purple or green (and all edges between such vertices), we see that there must be a path from $v_2$ to $v_4$ that uses only purple or green vertices.

There is no way to draw the yellow-blue path from $v_1$ to $v_3$ and the purple-green path from $v_2$ to $v_4$, without the two paths crossing each other. Since the graph is planar, they must cross each other at a vertex, $u$. Since $u$ is on the yellow-blue path, it must be coloured either yellow or blue. Since $u$ is on the purple-green path, it must be coloured either purple or green. It's not possible to satisfy both of these conditions on the colour of $u$. This contradiction serves to prove that no planar 6-critical graph exists, so every planar graph is properly 5-colourable. □

In fact, Kenneth Ira Appel (1932—2013) and Wolfgang Haken (1928—) proved the Four-Colour Theorem in 1976.

**THEOREM 15.3.3 (Four-Colour Theorem).** *Every planar graph is properly* 4*-colourable.*

Their proof involved considering a very large number of cases — so many that they used a computer to analyse them all. Although computers are often used in mathematical work now, this was the first proof that could not reasonably be verified by hand. It was viewed with suspicion for a long time, but is now generally accepted.

One of the methods by which mathematicians attempted unsuccessfully to prove the Four-Colour Theorem seemed particularly promising, and has led to a lot of interesting work in its own right. We require a couple of definitions to explain the connection.

**DEFINITION 15.3.4.** A **cubic graph** is a graph for which all of the vertices have valency 3.

**DEFINITION 15.3.5.** A **bridge** in a connected graph is an edge whose deletion disconnects the graph.

**THEOREM 15.3.6.** *The problem of 4-colouring a planar graph is equivalent to the problem of 3-edge-colouring a cubic graph that has no bridges.*

We'll prove one direction of the equivalence stated in this theorem; the other direction is a bit more complicated.

**PROOF.** Suppose that every planar graph can be properly 4-coloured, and that $G$ is a (simple) bridgeless cubic graph, embedded in the plane. We'll show that there is a proper 3-edge-colouring of $G$. Since $G$ is bridgeless, we don't run into the problem of a loop in the planar dual, so the Four-Colour Theorem applies to the faces of $G$. Properly colour the faces of $G$ with colours red, green yellow, and black. Every edge of $G$ lies between faces of two distinct colours, by the definition of a proper colouring of a map. Colour the edges of $G$ according to the following table: if the colours of the faces separated by the edge $e$ are the colours listed in the left-hand column, then use the colour listed in the right-hand column to colour $e$.

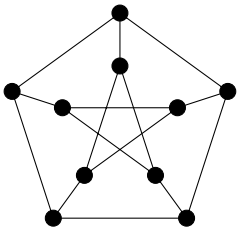
| Face colours | Edge colour |
|---|---|
| green, black | dashed |
| yellow, black | dotted |
| red, black | solid |
| green, yellow | solid |
| green, red | dotted |
| red, yellow | dashed |

Let $v$ be an arbitrary vertex. We will show that the three edges incident with $v$ must all receive different colours. Since 3 edges meet at $v$, three faces also meet at $v$, and every pair of these faces share an edge. Thus the three faces that meet at $v$ must all receive different colours. There are four different cases, depending on which colour is not used for a face at $v$. We show what happens in the following picture, using R, G, Y, and B to indicate the face colours, and colouring the edges according to the table above in each case.


In each case, the three edges incident with $v$ are assigned different colours, so this is a proper 3-edge colouring of $G$. □

This theorem was proven by Peter Guthrie Tait (1831—1901) in 1880; he thought that every cubic graph with no bridges must be 3-edge-colourable, and thus that he had proven the Four-Colour Theorem. In fact, Vizing's Theorem tells us that any cubic graph can be 4-edge-coloured, so we would only need to reduce the number of colours by 1 in order to prove the Four-Colour Theorem. The problem therefore boils down to proving that there are no bridgeless planar cubic graphs that are class two.

In 1881, Peter Christian Julius Petersen (1839—1910) published the Petersen graph that we saw previously in Example 14.3.13.



This graph is cubic and has no bridges, but is not 3-edge-colourable (this can be proved using a case-by-case analysis). Thus, there exist bridgeless cubic graphs that are class two!

Many people have tried to find other examples, as classifying these could provide a proof of the Four-Colour Theorem.

For many years (regularly from 1956 – 1981, with a few additional sporadic articles until 1986), Martin Gardner (1914—2010) wrote a "Mathematical Games" column in the *Scientific American* about interesting math problems and puzzles. As the Four-Colour Theorem is easy to explain without technical language, it was a topic he wrote about in 1976, in an article entitled "Snarks, Boojums and other conjectures related to the four-color-map theorem". When writing about the importance of bridgeless cubic class-two graphs, he decided they needed a more appealing name. Since they seemed rare and elusive, he called them *snarks*, after Lewis Carroll's poem "The Hunting of the Snark." The name has stuck.

**DEFINITION 15.3.7.** A **snark** is a bridgeless cubic class-two graph.

Two infinite families and a number of individual snarks are known. There is no reason to believe that these are all of the snarks that exist. By the Four-Colour Theorem, we know that there are no planar snarks; if we could find a direct proof that there are no planar snarks, this would provide a new proof of the Four-Colour Theorem.

**EXERCISES 15.3.8.**

1) Prove that if a cubic graph $G$ has a Hamilton cycle, then $G$ is a class-one graph.
2) Properly 4-colour the faces of the map given at the start of this section.
3) The map given at the start of this section can be made into a cubic graph, by placing a vertex everywhere two borders meet (including the coast as a border) and edges where there are borders. Use the method from the proof of Theorem 15.3.6 to properly 3-edge-colour this cubic graph, using your 4-colouring of the faces.
4) Prove that a graph $G$ that admits a planar embedding has an Euler tour if and only if every planar dual of $G$ is bipartite.
5) Prove that if a graph $G$ that admits a planar embedding in which every face is surrounded by exactly 3 edges, $G$ is 3-colourable if and only if it has an Euler tour.

**SUMMARY:**

- Kuratowski's Theorem, Wagner's Theorem
- Euler's Formula
- $|E| \leq 3|V| - 6$ for a planar graph
- colouring maps
- the Five Colour Theorem
- the Four Colour Theorem
- Important definitions:
  - planar graph, planar embedding
  - face
  - dual graph, planar dual
  - subdividing an edge, subdivision of a graph
  - edge contraction, contracting an edge
  - minor
  - cubic graph
  - bridge
  - snark

# Part III

# Design Theory

# Chapter 16

# Latin squares

## 16.1. Latin squares and Sudokus

You can think of a Latin square as a Sudoku puzzle that can be of any (square) size, and does not have the requirement that every value appear in each of the outlined smaller subsquares.

**DEFINITION 16.1.1.** A **Latin square** of order $n$ is an $n \times n$ array whose entries are elements of a set $N$ of cardinality $n$, with the property that every element of $N$ appears exactly once in each row and each column.

**EXAMPLE 16.1.2.** Here is a Latin square of order 4:

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \\ 3 & 4 & 1 & 2 \\ 2 & 3 & 4 & 1 \end{array}$$

Notice that in the above example, we placed the numbers $1, 2, 3$, and 4 in the first row, in that order. For each subsequent row, we shifted the numbers one place to the right (wrapping around). This same technique (placing the numbers from 1 through $n$ across the first row) will work to construct a Latin square of order $n$.

So you might think (with reason) that Latin squares aren't very interesting. However, even knowing that there is a Latin square of every possible order and they are easy to construct, there remain some interesting related questions.

Some of these questions are related to Sudokus. If we fill in some entries of a Latin square, are there conditions on these entries that guarantee that this can be completed to a full Latin square? Are there conditions under which we can be sure that a partial Latin square has a unique completion to a full Latin square?

Some of these questions have easy answers that are not what we are really looking for. For example, if we give you all but one entry of a Latin square (or Sudoku), then if it can be completed at all, the completion will be unique. However, some interesting mathematical work has been done on these problems, both for Latin squares and for Sudokus.

It has recently been proved (using computers) that a Sudoku puzzle with 16 or fewer squares pre-filled cannot be completed uniquely. This is a theorem by Gary McGuire (1967—), Bastian Tugemann, and Gilles Civario (1972—), published in 2014 in *Experimental Mathematics*. The published version is available online by subscription to the journal, but a preprint is freely available on the widely-used "ar$\chi$iv" scientific preprint server hosted by Cornell University. However, there are tens of thousands of (non-isomorphic) ways of pre-filling 17 entries of a Sudoku puzzle, that have a unique completion. Gordon F. Royle (1962—) built a web page

with a complete list of those that were known at the time. His work on this was used by McGuire, Tugemann, and Civario in their paper. The original web page was inaccessible at the most recent editing of this book, but there is a copy available through the Wayback Machine's archive of the internet.

Looking only at "non-isomorphic" examples is important, because there are many ways of creating Latin squares (or Sudoku puzzles) that are essentially the same. The following operations take a Latin square to another Latin square that is structurally essentially the same:

- Permuting of the symbols used in the set $N$. For example, changing every 1 to a 2 and every 2 to a 1.
- Interchanging any two rows.
- Interchanging any two columns.
- Making all of the rows into columns, and all of the columns into rows.

**EXERCISES 16.1.3.**

1) Prove that interchanging two rows of a Latin square, yields a Latin square.

2) Complete the following Latin square. Is the completion unique?

$$\begin{array}{cccc} 1 & - & 4 & - \\ - & 1 & - & - \\ - & - & - & 3 \\ - & - & - & - \end{array}$$

3) Use the method described at the start of this chapter to create a Latin square of order 5. What 3 of the operations listed above that change a Latin square to an isomorphic Latin square, are required to arrive at the following result?

$$\begin{array}{ccccc} 5 & 1 & 4 & 2 & 3 \\ 1 & 3 & 5 & 4 & 2 \\ 4 & 5 & 2 & 3 & 1 \\ 2 & 4 & 3 & 1 & 5 \\ 3 & 2 & 1 & 5 & 4 \end{array}$$

4) Show there are exactly two different Latin squares of order 3 whose first row is $1, 2, 3$.

5) Show there are exactly twelve different Latin squares of order 3 whose entries are the numbers $1, 2, 3$. [*Hint:* Use Exercise 16.1.3.4.]

6) There are four different Latin squares of order 4 whose first row is $1, 2, 3, 4$ and whose first column is also $1, 2, 3, 4$. That is, there are only four ways to complete the following Latin square:

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 2 & - & - & - \\ 3 & - & - & - \\ 4 & - & - & - \end{array}$$

Find all four.

[*Hint:* Each of the possibilities for the second entry of the second row can be completed in only one or two ways.]

## 16.2. Mutually orthogonal Latin squares (MOLS)

Most of design theory is concerned with creating nice structures in which different combinations of elements occur equally often. This is the general structure of all of the design theory we will be covering here, and in this context, orthogonal Latin squares are the natural thing to learn about. Before giving a formal definition, we'll explain the problem that introduced this concept.

Once again, Leonhard Euler (1707—1783) was involved in the origins of this problem. In fact, the name Latin square comes from his terminology. In 1782, he posed the problem of arranging 36 officers into a $6 \times 6$ square. The officers come from 6 different regiments (which he denoted with the Latin characters $a$, $b$, $c$, $d$, $e$, and $f$) and each holds one of 6 possible ranks (which he denoted with the Greek characters $\alpha$, $\beta$, $\gamma$, $\delta$, $\varepsilon$, and $\zeta$). No two officers from the same regiment hold the same rank. The question he posed was, is it possible to organise the officers into the square so that in each row and each column, there is precisely one officer from each regiment, and precisely one officer of each rank? Since he was using Greek and Roman letters to denote the classes, he called this a "Graeco-Latin square." He chose the first step to consist of arranging the regiments, i.e. for each regiment to set aside 6 positions in the square to be filled with officers from that regiment. Subsequently, he would try to assign ranks to the officers in these 6 positions. Since the regiments were denoted by Latin characters, he called this first step a "Latin square." The Graeco-Latin square of his question is a pair of orthogonal Latin squares of order 6. We'll explain this in more detail shortly, but the key idea is that after the second step is completed, each entry consists of a Greek letter and a Roman letter, and each Greek letter should appear exactly once with each Roman letter since there is to be one officer from each regiment who holds each of the possible ranks.

Euler could not find a solution to this problem, although he was able to produce pairs of orthogonal Latin squares for any $n \not\equiv 2 \pmod 4$. Since there is also no pair of orthogonal Latin squares of order 2 (and possibly for other reasons), he conjectured that there is no pair of orthogonal Latin squares of order $n$ for any $n \equiv 2 \pmod 4$. Although Euler was correct that there is no pair of orthogonal Latin squares of order 6, his conjecture was not true. In $1959 - 1960$, Raj Chandra Bose (1901—1987), Sharadchandra Shankar Shrikhande (1917—2020), and Ernest Tilden Parker (1926—1991) first found constructions for pairs of orthogonal Latin squares of orders 22 and 10, and then found a general construction that can produce a pair of orthogonal Latin squares of order $n$ for every $n > 6$ with $n \equiv 2 \pmod 4$.

**DEFINITION 16.2.1.** Two Latin squares $S_1$ and $S_2$ of order $n$ are **orthogonal** if when we look at each position in turn and consider the ordered pair formed by the entry of $S_1$ in that position, and the entry of $S_2$ in that position, every possible ordered pair appears.

In the context of Euler's problem, this definition proposes that the problem be worked on slightly differently: first create a Latin square with the Roman letters (as Euler did). Then, rather than trying to assign a Greek letter to each officer (and trying to satisfy the other conditions), we try to form a second Latin square with just the Greek letters. Once we have both squares, we check to see if they are "orthogonal". If they are, it means that if we look at each of the 36 positions in turn and write the Roman letter from that position in the first square followed by the Greek letter from that position in the second square as an ordered pair, each of the 36 possible ordered pairs occurs exactly once in our list. Translating this back into Euler's terms, each regiment will contain exactly one officer of each rank.

In our definition, we are looking at positions in Latin squares of order $n$, and trying to ensure that every ordered pair appears in some position. Notice that since the set $N$ has $n$ elements, the total number of ordered pairs possible is $n^2$ (there are $n$ choices for the first entry and $n$ choices for the second entry). A Latin square of order $n$ has $n^2$ positions since it has $n$

rows and $n$ columns. Thus, if every possible ordered pair appears in some position, then each ordered pair must appear exactly once.

**EXAMPLE 16.2.2.** Here is a pair of orthogonal Latin squares of order 3:

$$\begin{array}{ccc} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{array} \qquad \begin{array}{ccc} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{array}$$

We see that the ordered pairs $(1,1)$, $(2,2)$ and $(3,3)$ appear in the first row; the pairs $(3,2)$, $(1,3)$, and $(2,1)$ appear in the second row; and the pairs $(2,3)$, $(3,1)$, and $(1,2)$ appear in the third row. Every possible ordered pair whose entries lie in $\{1,2,3\}$ has appeared.

There is a nice pattern to the squares given in this example. The first follows the general construction we mentioned at the start of this chapter. For the second, each row has been shifted one place to the left (rather than to the right) from the one above it. This construction does actually work for $n$ odd, but never for $n$ even. For example, when $n = 4$, it would give

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \\ 3 & 4 & 1 & 2 \\ 2 & 3 & 4 & 1 \end{array} \qquad \begin{array}{cccc} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{array}$$

You can see that the ordered pair $(1,1)$ occurs in two positions: row 1, column 1, and row 3, column 3. So this pair of Latin squares is definitely not orthogonal. In fact, the first of these squares has no Latin square that is orthogonal to it. However, there is a pair of orthogonal Latin squares of order 4:

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \\ 2 & 1 & 4 & 3 \end{array} \qquad \begin{array}{cccc} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \end{array}$$

**DEFINITION 16.2.3.** A set of Latin squares is **mutually orthogonal** if **every** distinct pair of Latin squares in the set are orthogonal. We call such a set, a set of MOLS (for Mutually Orthogonal Latin Squares).

There is a very nice method of representing MOLS. The key idea is that it is not necessary to use the *same* set of symbols for each square. Furthermore, we can write the whole structure as one square, with an ordered pair in each position, as Euler did for his Graeco-Roman problem. So Euler might have represented our orthogonal Latin squares of order 3 as the single square

$$\begin{array}{ccc} (a,\alpha) & (b,\beta) & (c,\gamma) \\ (c,\beta) & (a,\gamma) & (b,\alpha). \\ (b,\gamma) & (c,\alpha) & (a,\beta) \end{array}$$

But there's no reason to stick to Greek and Roman letters. Why not have colours instead of the Greek letters? If we change $\alpha$ to red, $\beta$ to blue, and $\gamma$ to green, we have instead:

$$\begin{array}{ccc} (a,\text{red}) & (b,\text{blue}) & (c,\text{green}) \\ (c,\text{blue}) & (a,\text{green}) & (b,\text{red}) \\ (b,\text{green}) & (c,\text{red}) & (a,\text{blue}) \end{array}.$$

In fact, since the second entry of the ordered pair can now be thought of as another characteristic (colour), we can represent this by simply listing the first entry and varying the characteristic.

So instead of (3, blue) we can simply show a blue 3. However, varying the colours is not feasible here (since we are keeping essential elements in black-and-white for the sake of accessibility and ease of printing). Instead, let us use "tilted left" (for blue), "straight up" (for red), and "tilted right" (for green). So (for example) since in the second row, third column our square of ordered pairs had the entry (2, blue), and blue corresponds to a left tilt, we place a 2 that is tilted to the left in that location in our new representation. Here is the complete representation:

$$\begin{array}{ccc} \diagdown & 2 & \diagup \\ 3 & \diagup & \diagdown \\ \diagup & 3 & 1 \end{array}$$

By the property of orthogonality, every combination of tilting and number must appear in exactly one position! Even more amazing, if we have a larger set of MOLS and vary different parameters for each of the squares, the fact that the squares are all mutually orthogonal will mean that every combination of the parameters appears in exactly one position. For example, if we have a set of five MOLS, we could place a coloured shape behind each coloured symbol, and have different numbers of copies of the symbol. For any possible colour of any possible shape appearing behind any possible number of any possible colour of any possible symbol, you would be able to find a position in which that combination appears!

This approach to MOLS is essentially the context in which they first arose, as we can see from Euler's example of the officers.

A natural question that arises in the context of the concepts we have been introducing in this section is, how many Latin squares can there be in a set of MOLS?

Before we attempt to answer this question, notice that if we have a pair of orthogonal Latin squares and we permute the symbols used in the set $N$ independently for each of the squares (resulting in new Latin squares that are nonetheless essentially the same, as discussed in Section 16.1), the resulting pair of Latin squares will still be orthogonal. If in the first square the symbol $x$ maps to the symbol $y$, and in the second square the symbol $u$ maps to the symbol $v$, then in the new pair of Latin squares the ordered pair $(y, v)$ will appear precisely once, since the ordered pair $(x, u)$ appeared precisely once in the original pair of Latin squares. This is true for any pair of entries $(y, v)$, so every pair of entries must appear precisely once. Now we are ready to partially answer the question of how many MOLS of order $n$ there can be:

**THEOREM 16.2.4.** *If $S$ is a set of MOLS of order $n$, then $|S| \leq n - 1$.*

**PROOF.** We may assume that $N = \{1, \ldots, n\}$. In each of the Latin squares in $S$, we can independently permute the symbols of $N$. As was noted above, the result will still be a set of MOLS. We permute the symbols so that the first row of each of the Latin squares has the entries $1, 2, \ldots, n$ in that order.

Now, if we take any $i \in N$ and consider any pair of the Latin squares, the ordered pair $(i, i)$ appears somewhere in the first row. Consider the first entry of the second row in each square of $S$. None of these entries can be 1, since 1 has already appeared in the first column of each of the Latin squares. No two of the Latin squares can have the same entry $j$ in this position, since the ordered pair $(j, j)$ has already appeared in the $j$th position of the first row of this pair of squares, so can't appear again in the first position of the second row. So there cannot be more squares in $S$, than the $n-1$ distinct entries from $N \setminus \{1\}$ that could go into this position. Thus, $|S| \leq n - 1$, as claimed. □

The next natural question is, is it possible to achieve $n - 1$ MOLS of order $n$? We have already seen that the answer is yes in one very small case, since we found 2 MOLS of order 3. In fact, there are infinitely many values of $n$ for which there are $n - 1$ MOLS of order $n$.

The following result can be generalised to prime powers using some basic field theory that you should understand if you have taken a course that includes any field theory. However, for

the purposes of this course, we will avoid the explicit field theory and prove the result only for primes.

We do require a bit of modular arithmetic for this result. As modular arithmetic will also be useful for some of our later results, here is a quick review of some key points.

**DEFINITION 16.2.5.** Performing calculations **modulo $n$** means replacing the result with the remainder you would get upon dividing that result by $n$. In other words, if the result of a computation is $n$ or larger, replace the result by its remainder upon division by $n$.

**NOTATION 16.2.6.** If $a$ and $b$ have the same remainder upon division by $n$, then we write $a \equiv b \pmod{n}$.

There are two key facts from modular arithmetic that we will require. The first is that if $a \equiv b \pmod{n}$ and $0 \leq a, b < n$, then we must have $a = b$.

The other is that if $qa \equiv qb \pmod{n}$ and $n$ and $q$ have a greatest common divisor of 1, then $a \equiv b \pmod{n}$. In the special case where $n$ is prime, as long as $q$ is not a multiple of $n$ then $n$ and $q$ will always have a greatest common divisor of 1.

**THEOREM 16.2.7.** *For any prime $p$, there are $p-1$ MOLS of order $p$.*

**PROOF.** We will use $N = \{0, \ldots, p-1\}$. In order to ensure that the results of our computations will be in $N$, all of the calculations given in this result should be taken modulo $p$.

The squares will be $\{S_1, \ldots, S_{p-1}\}$. For $k \in \{1, \ldots, p\}$,

$$S_k = \begin{bmatrix} 0 & 1 & \ldots & p-1 \\ k & k+1 & \ldots & k+(p-1) \\ 2k & 2k+1 & \ldots & 2k+(p-1) \\ \vdots & \vdots & & \vdots \\ (p-1)k & (p-1)k+1 & \ldots & (p-1)k+(p-1) \end{bmatrix}$$

We first verify that each $S_k$ is a Latin square. The entries in each row are easily seen to be distinct. If the entries in the first column are distinct, then we can see that the entries in every other column will be distinct. Suppose that $0 \leq i, j \leq p-1$ and that $ik \equiv jk \pmod{p}$. Then since every $k \in \{1, \ldots, p-1\}$ has a greatest common divisor of 1 with $p$, we see that $i \equiv j \pmod{p}$. Since $0 \leq i, j \leq p-1$, this forces $i = j$. So the entries in the first column of $S_k$ are all distinct. Thus, every $S_k$ is a Latin square.

Now we verify that any two squares $S_i$ and $S_j$ are orthogonal. We will do this by taking an arbitrary pair of squares $S_i$ and $S_j$ with $i \neq j$, and supposing that the ordered pair that arises from the entries in position $(k_1, m_1)$ of these two squares is the same as the ordered pair that arises from the entries in position $(k_2, m_2)$. If this can happen, then unless the two positions are actually the same, the squares would not be orthogonal. So our goal will be to deduce that in fact the two positions are the same: that is, $k_1 = k_2$ and $m_1 = m_2$.

Suppose that for some $1 \leq i, j \leq p-1$, the squares $S_i$ and $S_j$ have the same ordered pair in two positions: row $k_1$, column $m_1$, and row $k_2$, column $m_2$. Then by the formulas given for the entries of each Latin square, we must have

$$\begin{gathered} (k_1-1)i + m_1 - 1 \equiv (k_2-1)i + m_2 - 1 \pmod{p}, \\ \text{and } (k_1-1)j + m_1 - 1 \equiv (k_2-1)j + m_2 - 1 \pmod{p}. \\ \text{Thus, } k_1 i + m_1 \equiv k_2 i + m_2 \pmod{p}, \\ \text{and } k_1 j + m_1 \equiv k_2 j + m_2 \pmod{p}. \\ \text{Therefore, } m_2 - m_1 \equiv k_2 i - k_1 i = (k_2 - k_1)i \equiv k_2 j - k_1 j = (k_2 - k_1)j \pmod{p}. \end{gathered}$$

Since $(k_2 - k_1)i \equiv (k_2 - k_1)j \pmod{p}$, and $1 \leq i, j \leq p-1$, and we are assuming $i \neq j$, we must have $k_2 - k_1 \equiv 0 \pmod{p}$. Since $k_1$ and $k_2$ are row numbers, they are between 1 and $p$ so this forces $k_1 = k_2$. Furthermore, in this case we must also have $m_2 - m_1 \equiv 0 \pmod{p}$, and we see that this also forces $m_1 = m_2$. This is what we wanted to show.

This shows that $\{S_k \mid 1 \leq k \leq p-1\}$ is indeed a set of $p-1$ MOLS. □

**EXAMPLE 16.2.8.** Here are the first 8 of the 10 MOLS of order 11, found using the formula given in the proof above.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 |
| 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 |
| 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 |
| 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 |
| 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 |
| 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 |
| 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 |
| 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 |
| 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 |
| 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 |
| 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 |
| 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 |
| 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 |
| 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 |
| 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 |
| 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 |
| 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 |
| 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 |
| 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 |
| 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 |
| 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 |
| 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 |
| 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 |
| 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 |
| 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 |
| 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 |
| 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 |
| 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 |
| 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 |
| 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 |
| 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 |
| 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 |
| 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 | 3 | 4 | 5 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 0 | 1 | 2 |

We've now seen that it is possible to find $p-1$ MOLS of order $p$ for any prime $p$, and have noted that the proof can be generalised to prime powers. However, as we've already discussed in relation to Euler's original problem, there are orders for which the bound of $n-1$ MOLS of order $n$ cannot be attained: in fact, for order 6 it is not possible even to find a pair of orthogonal Latin squares.

If you are interested in or familiar with some finite geometry, the existence of $n-1$ MOLS of order $n$ is equivalent to the existence of a projective plane of order $n$. Projective planes, in turn, are a special kind of design. For an interesting article about some of these relationships, see `https://www.maa.org/sites/default/files/pdf/upload_library/22/Ford/Lam305-318.pdf`. This article is by Clement Wing Hong Lam (1949—), who (with coauthors Larry Henry Thiel (1945—) and Stan Swiercz (1955—)) in 1989 used a computer to prove that there is no projective plane of order 10 (and equivalently, that there are not 9 MOLS(10)). There is also some information about affine planes, projective planes, and their existence in Sections 18.3 and 18.4. Whether or not a projective plane of order 12 exists is unknown. The nonexistence of a plane of order 10 was independently verified in 2011 and again in 2020 using a different (but still computer-based) approach.

**EXERCISES 16.2.9.**

1) Find the two MOLS of order 11 that are not included in Example 16.2.8, but are orthogonal to each other and to the squares listed there.

2) Find a third Latin square of order 4 that is orthogonal to both of the orthogonal Latin squares of order 4 that were given earlier in this section.

3) Here is a Latin square of order 8, and some entries for a second Latin square of order 8. Complete the second square so as to obtain a pair of orthogonal Latin squares.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2 | 1 | 4 | 3 | 6 | 5 | 8 | 7 |
| 3 | 4 | 1 | 2 | 7 | 8 | 5 | 6 |
| 4 | 3 | 2 | 1 | 8 | 7 | 6 | 5 |
| 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 |
| 6 | 5 | 8 | 7 | 2 | 1 | 4 | 3 |
| 7 | 8 | 5 | 6 | 3 | 4 | 1 | 2 |
| 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | – | – | – | 8 | 5 | 6 |
| 5 | – | 7 | – | – | 2 | 3 | – |
| 7 | – | – | 6 | – | – | – | – |
| 4 | – | – | 1 | 8 | – | – | – |
| 2 | – | – | – | – | 5 | – | – |
| 8 | – | – | – | – | – | – | 1 |
| 6 | – | – | 7 | – | – | – | 3 |

4) Write down the six mutually orthogonal Latin squares $S_1, \ldots, S_6$ of order 7 that are constructed by letting $p = 7$ in the proof of Theorem 16.2.7.

### 16.3. Systems of distinct representatives

Suppose we start filling in a Latin square, one row at a time, at each step ensuring that no element has yet appeared more than once in a column (or in a row). Under what conditions will it be impossible to complete this to a Latin square? Although it may not be immediately obvious, the answer to this question can be found in a well-known theorem published by Philip Hall (1904—1982) in 1935, about systems of distinct representatives.

**DEFINITION 16.3.1.** Let $T_1 \ldots, T_n$ be sets. If there exist $a_1, \ldots, a_n$ all distinct such that for every $1 \leq i \leq n$, $a_i \in T_i$, then $\{a_1, \ldots a_n\}$ form a **system of distinct representatives (SDR)** for $T_1, \ldots, T_n$.

**EXAMPLE 16.3.2.** The university is striking a student committee on the subject of tutorials. For each of the 5 Faculties, they ask students to elect one representative who is taking classes from that faculty. They do not want one student trying to represent more than one faculty. The candidates are:

- Joseph, who is taking courses in Arts and Science, Fine Arts, Management, and Education;
- René, who is taking courses in Health Sciences;
- Claire, who is taking courses in Education and Health Sciences;
- Sandra, who is taking courses in Management, Fine Arts, and Health Sciences;
- Laci, who is taking courses in Education and Health Sciences; and
- Jing, who is taking courses in Education.

Can the committee be filled?

**SOLUTION.** The answer is no. For the three Faculties of Arts & Science, Fine Arts, and Management, there are only two possible student representatives: Joseph (who could represent any of the three), and Sandra (who could represent either Fine Arts or Management). So it is not possible to elect one student to represent each of the five Faculties, without allowing one of these students to fill two roles. □

In Example 16.3.2, we observed that we could find a collection of the sets to be represented, that collectively had fewer possible representatives than there are sets in the collection. It is easy to see that if this happens, there cannot be a system of distinct representatives for the sets.

What Philip Hall proved is the converse: unless we have an obstruction of this type, it is always possible to find a system of distinct representatives.

**THEOREM 16.3.3 (Hall's Theorem).** *The collection of sets $T_1, \ldots, T_n$ has a system of distinct representatives if and only if for every $1 \leq k \leq n$, the union of any $k$ of the sets has cardinality at least $k$.*

This theorem is often referred to as "Hall's Marriage Theorem," as one of the problems it solves can be stated as follows. Suppose we have a collection of men and a collection of women. Each of the women has a list of men she likes (from the collection). When is it possible to marry each of the women to a man that she likes? (The context is historical, and the assumption of the time was that every woman would want to marry a man.)

We have seen that one of the two implications of Hall's Theorem is easy to prove. We will not try to prove the other implication here, but will focus on using the result. Here are some examples and exercises. For completeness, we provide a proof of Hall's Theorem at the end of this chapter.

**EXAMPLE 16.3.4.** Let $A_1 = \{a, b, c, d\}$, $A_2 = \{b, c\}$, $A_3 = \{b\}$, $A_4 = \{b, c\}$. Does this collection of sets have a system of distinct representatives? If so, find one; if not, explain why.

**SOLUTION.** The answer is that this collection of sets has no system of distinct representatives, because the union of three sets $A_2$, $A_3$, and $A_4$ has only two elements: $b$ and $c$. □

**EXAMPLE 16.3.5.** Let $A_1 = \{a, d\}$, $A_2 = \{a, c\}$, $A_3 = \{b, c\}$, $A_4 = \{c, d\}$. Does this collection of sets have a system of distinct representatives? If so, find one; if not, explain why.

**SOLUTION.** This collection of sets does have a system of distinct representatives: take $a$ for $A_1$, $c$ for $A_2$, $b$ for $A_3$, and $d$ for $A_4$. For clarity, we underline the representatives in the following list of the sets: $A_1 = \{\underline{a}, d\}$, $A_2 = \{a, \underline{c}\}$, $A_3 = \{\underline{b}, c\}$, $A_4 = \{c, \underline{d}\}$.

(In fact, it also has another system of distinct representatives: take $d$ for $A_1$, $a$ for $A_2$, $b$ for $A_3$, and $c$ for $A_4$: $A_1 = \{a, \underline{d}\}$, $A_2 = \{\underline{a}, c\}$, $A_3 = \{\underline{b}, c\}$, $A_4 = \{\underline{c}, d\}$. ) □

**EXERCISES 16.3.6.** For each collection of sets, determine whether or not it has a system of distinct representatives. If so, find one; if not, explain why.

1) $A_1 = \{x\}$, $A_2 = \{y, z\}$, $A_3 = \{x, y\}$.
2) $A_1 = \{u, v, w, x, y, z\}$, $A_2 = \{v, w, y\}$, $A_3 = \{w, x, y\}$, $A_4 = \{v, w, x, y\}$, $A_5 = \{v, x, y\}$, $A_6 = \{v, y\}$.
3) $A_1 = \{x\}$, $A_2 = \{y\}$, $A_3 = \emptyset$.
4) $A_1 = \{x, z\}$, $A_2 = \{y\}$, $A_3 = \{x, y, z\}$.
5) $T_1 = \{a, b, c, d\}$, $T_2 = \{a, b, c\}$, $T_3 = \{a\}$, $T_4 = \{c\}$.
6) $U_1 = \{x, y\}$, $U_2 = \{y, z\}$, $U_3 = \emptyset$.
7) $V_1 = \{e, f\}$, $V_2 = \{e, g\}$, $V_3 = \{e, h\}$, $V_4 = \{f, g\}$, $V_5 = \{h, i\}$.
8) $W_1 = \{+, -, \times, \div, 0\}$, $W_2 = \{+, -, \times\}$, $W_3 = \{+, \times\}$, $W_4 = \{\times, -\}$, $W_5 = \{+, -\}$.

Let's return to our original question about Latin squares. To answer this, we first give an important general consequence of Hall's Theorem.

**PROPOSITION 16.3.7.** *Suppose $T_1, \ldots, T_n$ is a collection of sets each of which contains exactly $r$ elements. Further suppose that no element appears in more than $r$ of the sets. Then this collection has a system of distinct representatives.*

**PROOF.** By Hall's Theorem, we must show that for every $1 \leq k \leq n$, the union of any $k$ of the sets $T_1, \ldots, T_n$ has cardinality at least $k$. Let $k \in \{1, \ldots, n\}$ be arbitrary, and arbitrarily choose $k$ of the sets $T_{j_1}, \ldots T_{j_k}$. If $k \leq r$ then since each $T_{j_i}$ has $r$ elements, their union must have at least $r \geq k$ elements, as desired.

Suppose on the other hand that $k > r$. Amongst the $k$ sets of $r$ elements, a total of $kr$ elements appear (counting each element every time it appears). Since each element appears in at most $r$ of the sets, it must be the case that at least $k$ distinct elements appear. This completes the proof. □

We can now answer the question about Latin squares.

**THEOREM 16.3.8.** *Suppose that $m$ rows of a Latin square of order $n$ have been filled, where $m < n$, and that to this point no entry appears more than once in any row or column. Then another row can be added to the Latin square, maintaining the condition that no entry appears more than once in any row or column.*

**PROOF.** For $1 \leq i \leq n$, let $T_i$ be the set of elements that have not yet appeared in column $i$ (from the entries in the first $m$ rows). So $T_i$ can be thought of as the set of allowable entries for the $i$th column of the new row. Notice that each $T_i$ has cardinality $n - m$ (the number of rows that are still empty). The task of finding a new row all of whose entries are distinct, and whose $i$th entry comes from the set $T_i$ of allowable entries for that column, is equivalent to finding a system of distinct representatives for the sets $T_1, \ldots, T_n$. Thus, we must show that the collection $T_1, \ldots, T_n$ has a system of distinct representatives.

Notice also that every element has appeared once in each of the first $m$ rows, and thus has appeared in precisely $m$ of the columns. Therefore, there are exactly $n - m$ of the columns in which it has not yet appeared. In other words, each element appears in exactly $n - m$ of the sets.

We can now apply Proposition 16.3.7, with $r = n - m$ to see that our sets do have a system of distinct representatives. This can be used to form a new row for the Latin square. □

**COROLLARY 16.3.9.** *Suppose that $m$ rows of a Latin square of order $n$ have been filled, where $m < n$, and that to this point no entry appears more than once in any row or column. This structure can always be completed to a Latin square.*

**PROOF.** As long as $m < n$, we can repeatedly apply Theorem 16.3.8 to deduce that it is possible to add a row. Once you actually find a row that can be added (note that the statement of Hall's Theorem does not explain how to do this), do so. Eventually this process will result in a complete square. □

Hall's Theorem can also be used to prove a special case of a result we proved previously, Theorem 14.1.17. The special case we can prove with Hall's Theorem, is the case where every vertex has the same valency.

**THEOREM 16.3.10.** *If $G$ is a bipartite graph in which every vertex has the same valency, then any bipartition sets $V_1$ and $V_2$ have the same cardinality, and there is a set of $|V_1|$ edges that can be properly coloured with the same colour.*

**PROOF.** Let $k$ be the valency of every vertex, and for some arbitrary bipartition sets $V_1$ and $V_2$, let $n = |V_1|$. By a slight adaptation of Euler's handshaking lemma, taking into account the fact that every edge has exactly one of its endvertices in $V_1$, we see that $nk = |E|$. By the same argument, $k|V_2| = |E|$, which forces $|V_2| = n = |V_1|$. Since the bipartition was arbitrary, the first statement follows.

Let $V_1 = \{v_1, \ldots, v_n\}$. For every $1 \leq i \leq n$, let

$$T_i = \{u \in V \mid u \sim v_i\}.$$

A system of distinct representatives for the sets $T_1, \ldots, T_n$ will produce $n = |V_1|$ edges that can be properly coloured with the same colour.

Observe that every set $T_i$ has cardinality $k$ (since this is the valency of every vertex), and every vertex appears in exactly $k$ of the sets (again because this is the valency of the vertex). Therefore, by Proposition 16.3.7, there is a system of distinct representatives for $T_1, \ldots, T_n$. Hence we can find $n = |V_1|$ edges that can be properly coloured with the same colour. □

**COROLLARY 16.3.11.** *If $G$ is a bipartite graph in which every vertex has the same valency $k$, then its edges can be properly coloured using $k$ colours.*

**PROOF.** Repeat the following step $k$ times: find a set of edges that can be properly coloured. Colour them with a new colour, and delete them from the graph.

At each stage, Theorem 16.3.10 tells us providing that every vertex has the same valency, we can find a set of edges that are properly coloured, one of which is incident with every vertex of the graph. Since the valency of every vertex is reduced by exactly one when we delete such a set of edges, at each stage every vertex will have the same valency. □

To conclude the chapter, we provide a proof of Hall's Theorem. As previously noted, one direction is obvious, so we prove only the other direction.

**PROOF OF HALL'S THEOREM..** We will prove this by strong induction on the number of sets, $n$.

*Base case:* $n = 1$. If a single set has one element, then that set has a representative. This completes the proof of the base case.

*Inductive step:* We begin with the inductive hypothesis. Let $m \geq 1$ be arbitrary. Suppose that whenever $1 \leq i \leq m$, and a collection of $i$ sets $T_1, \ldots, T_i$ has the property that for every $1 \leq k \leq i$, the union of any $k$ of the sets has cardinality at least $k$, then that collection of sets has a system of distinct representatives.

We want to deduce that whenever we have a collection of $m+1$ sets with the property that for every $1 \leq k \leq m+1$, the union of any $k$ of the sets has cardinality at least $k$, then that collection of sets has a system of distinct representatives. Let $T_1, \ldots, T_{m+1}$ be such a collection of sets.

We look at the subcollection of sets $T_1, \ldots, T_m$, and consider two cases. First, suppose that for every $1 \leq k \leq m$, the union of any $k$ of these sets has cardinality at least $k+1$. In this case, take any element $t \in T_{m+1}$ (which by hypothesis is nonempty) to be the representative for $T_{m+1}$, and remove this element from each of the other sets. Due to the case we are in, for every $1 \leq k \leq m$, the union of any $k$ of the sets $T_1 - \{t\}, \ldots, T_m - \{t\}$ still has cardinality at least $k$, (we have removed only the element $t$ from this union, which previously had cardinality at least $k+1$). Thus we can apply our induction hypothesis to find a system of distinct representatives for $T_1, \ldots, T_m$ that does not include the element $t$. This completes a system of distinct representatives for our full collection of sets.

The other case is that there is some $1 \leq k \leq m$ and some collection of $k$ of the sets $T_1, \ldots, T_m$, whose union has cardinality precisely $k$. By our induction hypothesis, there is a system of distinct representatives for these $k$ sets. From the other $m+1-k$ sets (observe that $1 \leq m+1-k \leq m$), remove the $k$ elements that were in the union of the original $k$ sets. Consider any $k'$ of these adjusted sets, with $1 \leq k' \leq m+1-k$. Observe that the union of the $k+k'$ sets consisting of the original $k$ sets together with these $k'$ sets must have contained at least $k+k'$ distinct elements by hypothesis, so after removing the $k$ representatives of the original $k$ sets (which are the only elements in those sets), these $k'$ sets must still have at least $k'$ distinct elements in their union. Therefore, we can apply our induction hypothesis to these other $m+1-k$ adjusted sets, and see that they too have a system of distinct representatives, none of which are amongst the $k$ representatives for the original $k$ sets. Combining these two systems of distinct representatives yields a system of distinct representatives for the full collection of sets. □

**EXERCISES 16.3.12.**

1) Onyx is doing some research on what people learn from visiting different countries. They have a set of questions that are to be answered by someone who has visited the country. Before fully launching the study, Onyx wants to try the questions out on some of their friends. Since the questions are the same for different countries, they don't want one person to answer the questions for more than one country, as that could

bias the results. Of their close friends, the following people have visited the following countries:

- England: Adam, Ella, Justin
- Wales: Adam, Justin, Faith, Cayla
- Scotland: Bryant, Justin, Ella
- Ireland: Adam, Bryant, Justin
- Germany: Cayla, Bryant, Justin, Faith, Denise
- France: Ella, Justin, Bryant
- Italy: Adam, Ella, Bryant

Prove that Onyx cannot find seven different friends, each of whom has visited a different one of these countries.

2) Can the following be completed to a $4 \times 4$ Latin square? Does Hall's Theorem apply to this? If not, why not?

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 4 | 1 | 2 | 3 |
| 2 | 4 | | |

3) Show (by example) that it is possible to have a bipartite graph in which the bipartition sets have the same cardinality $k$ and the valency of every vertex is either 3 or 4, but no set of $k$ edges can be properly coloured with a single colour.

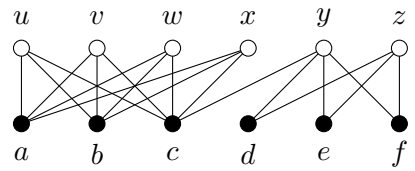
4) How do you know (without actually finding a completion) that the following can be completed to a Latin square of order 7 ?

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 4 | 7 | 6 | 1 | 5 | 3 |
| 3 | 7 | 4 | 2 | 6 | 1 | 5 |
| 4 | 6 | 2 | 5 | 3 | 7 | 1 |

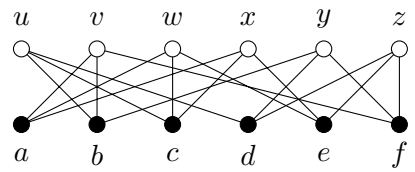
5) Find a completion of the partial Latin square in Exercise 4.

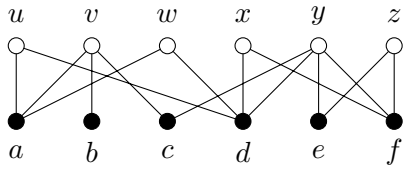
6) For each of these bipartite graphs, let $V_1 = \{a, b, c, d, e, f\}$, and determine whether there is a set of $|V_1|$ disjoint edges. (In other words, determine whether there is a set of $|V_1|$ edges that can be properly coloured with the same colour.)

(a)




(b)



(c) 


**SUMMARY:**

- Hall's (Marriage) Theorem
- a partial Latin square containing $m$ rows can always be completed.
- Important definitions:
  - Latin square
  - orthogonal Latin squares
  - MOLS (mutually orthogonal Latin squares)
  - system of distinct representatives (SDR)

# Chapter 17

# Designs

## 17.1. Balanced Incomplete Block Designs (BIBD)

Suppose you have 7 possible treatments for a disease, that you hope may work well singly or in combination. You would like to try out every possible combination of them, but the number of (non-empty) subsets of the set of 7 treatments is $2^7 - 1 = 127$, and you have only 7 mice in the lab who have this disease. You believe that using a pair of the treatments together will have a more significant impact than adding more of the treatments, but even trying every pair of treatments on a different mouse would require $\binom{7}{2} = 21$ mice.

Here is a strategy you could try: give each of the mice 3 of the treatments, according to the following scheme. For $1 \leq i \leq 7$, the treatments given to mouse $i$ will be the elements of the $i$th set in the following list:

$$\{1,2,3\},\{1,4,5\},\{1,6,7\},\{2,4,6\},\{2,5,7\},\{3,4,7\},\{3,5,6\}.$$

Careful perusal of this scheme will show that every pair of treatments is used together on precisely one of the mice.

**DEFINITION 17.1.1.** A **design** is a pair $(V, \mathcal{B})$, where $V$ is a finite set of points (also called varieties) and $\mathcal{B}$ is a collection of subsets of $V$, called **blocks** of the design.

This definition of a design is too broad to be of much interest without additional constraints, but a number of different constraints have been studied.

**NOTATION 17.1.2.** We use $v$ to denote $|V|$ and $b$ to denote $|\mathcal{B}|$.

**DEFINITION 17.1.3.** A **regular** design is a design in which every point appears in the same number of blocks, $r$.

A **uniform** design is a design in which every block contains the same number of points, $k$.

A **balanced** design is a design in which every pair of points appear together in the same number of blocks, $\lambda$.

**EXAMPLE 17.1.4.** In our disease treatments for the mice, we have $V = \{1,2,3,4,5,6,7\}$,

$$\mathcal{B} = \{\{1,2,3\},\{1,4,5\},\{1,6,7\},\{2,4,6\},\{2,5,7\},\{3,4,7\},\{3,5,6\}\}.$$

This is a regular, uniform, balanced design. In this design, $v = 7$, $b = 7$, $r = 3$, $k = 3$, and $\lambda = 1$.

A regular, uniform, balanced design may be referred to as a $(b, v, r, k, \lambda)$-design. So our disease treatments for the mice formed a $(7, 7, 3, 3, 1)$-design.

A $(b, v, r, k, \lambda)$-design is called *complete* if $v = k$. In this case, every block will contain every point. A $(b, v, r, k, \lambda)$-design is called *trivial* if $k = 1$. In this case, every block consists of a single point, and no points appear together in a block. The case $k = 2$ is also fairly trivial since in this case the blocks of the design must consist of the $\binom{v}{2}$ pairs of elements of $v$, each occurring $\lambda$ times.

**DEFINITION 17.1.5.** A **balanced incomplete block design (BIBD)** is a regular, uniform, balanced design that is not complete. So it is a $(b, v, r, k, \lambda)$-design with $k < v$.

Although this definition includes the possibility $k = 1$ or $k = 2$, these are not interesting cases, and can usually be ignored.

**EXAMPLE 17.1.6.** Here is another BIBD. This one has parameters $(20, 16, 5, 4, 1)$.

$$\begin{array}{llllllll} \{a,b,c,d\}, & \{e,f,g,h\}, & \{i,j,k,l\}, & \{m,n,o,p\}, & \{a,e,i,m\}, & \{a,g,l,n\}, & \{a,h,j,o\}, \\ \{a,f,k,p\}, & \{b,f,j,n\}, & \{b,h,k,m\}, & \{b,g,i,p\}, & \{b,e,l,o\}, & \{c,g,k,o\}, & \{c,e,j,p\}, \\ \{c,f,l,m\}, & \{c,h,i,n\}, & \{d,h,l,p\}, & \{d,f,i,o\}, & \{d,e,k,n\}, & \{d,g,j,m\} & \end{array}$$

**THEOREM 17.1.7.** *If a $(b, v, r, k, \lambda)$-design exists, then $bk = vr$ and*

$$r(k-1) = \lambda(v-1).$$

**COMBINATORIAL PROOF..** For the first equation, we count the total number of appearances of each point in the design (including repetitions) in two ways. This is another example of counting ordered pairs from a cartesian product, as we have discussed previously.

First, there are $b$ blocks, each of which has $k$ points in it. So the answer will be $bk$.

Second, there are $v$ points, each of which appears $r$ times. So the answer will be $vr$.

Thus, $vr = bk$.

For the second equation, we fix a point $p$, and count the number of points with which $p$ appears in a block, in two ways.

First, $p$ appears in $r$ blocks. In each of these, there are $k-1$ points besides $p$. So the answer will be $r(k-1)$.

Second, for every point $p' \in V$ with $p' \neq p$, the point $p'$ appears with $p$ in $\lambda$ different blocks. Since there are $v-1$ choices for $p'$, the answer will be $\lambda(v-1)$.

Thus, $r(k-1) = \lambda(v-1)$. □

With a bit of calculation, the results of Theorem 17.1.7 tell us that:

$$r = \frac{\lambda(v-1)}{k-1} \tag{17.1.1}$$

and

$$b = \frac{vr}{k} = \frac{\lambda v(v-1)}{k(k-1)} \tag{17.1.2}$$

Thus, if we know that a design is regular, uniform, and balanced, then the parameters $r$ and $b$ can be determined from the parameters $v$, $k$, and $\lambda$. We therefore often shorten our notation and refer to a BIBD$(v, k, \lambda)$.

**THEOREM 17.1.8.** *A BIBD$(v, k, \lambda)$ is equivalent to colouring the edges of the multigraph $\lambda K_v$ (the multigraph in which each edge of $K_v$ has been replaced by $\lambda$ copies of that edge) so that the edges of any colour form a $K_k$.*

**PROOF.** Given an edge-colouring of $\lambda K_v$ as described, define the points of the design to be the set of vertices of the multigraph, and for each colour, create a block whose vertices are the vertices of the $K_k$ that has that colour. All of these blocks will have cardinality $k$. Every vertex has valency $\lambda(v-1)$, and every $K_k$ of one colour that contains that vertex will use $k-1$ of the edges incident with that vertex, so every vertex will appear in

$$r = \lambda(v-1)/(k-1)$$

blocks. Now, any edge of the $\lambda K_v$ must appear in some $K_k$ (the one coloured with the colour of that edge). Thus for any pair of points, these vertices are joined by $\lambda$ edges each of which appears in some $K_k$, so these points must appear together in $\lambda$ of the $K_k$ subgraphs, i.e. $\lambda$ of the blocks.

Similarly, given a BIBD$(v,k,\lambda)$, and a multigraph $\lambda K_v$, label the vertices of $K_v$ with the points of the design. For each block of the design, use a new colour to colour the edges of a $K_k$ that connects the points in that block. There will be enough uncoloured edges joining these points, since every pair of points appear together in exactly $\lambda$ blocks, and there are $\lambda$ edges joining the corresponding vertices. In fact, careful counting can show that this will result in colouring every edge of the multigraph. □

This is nicest in the case where $\lambda = 1$, when the BIBD corresponds to an edge-colouring of $K_v$.

A colouring of the edges of a graph (or multigraph) is often referred to as a *decomposition* of the graph (or multigraph), since we can think of the colour classes as sets of edges whose union forms the entire edge set of the graph.

These provide alternate ways of thinking of designs that may be more intuitive, and are certainly more visual.

Equations (17.1.1) and (17.1.2) lead to numerical conditions on $v$, $k$, and $\lambda$ that must be satisfied in order for a BIBD$(v,k,\lambda)$ to exist.

**THEOREM 17.1.9.** *A BIBD$(v,k,\lambda)$ cannot exist unless*

$$\lambda\frac{v-1}{k-1} \quad \textit{and} \quad \lambda\frac{v(v-1)}{k(k-1)}$$

*are integers.*

**PROOF.** By Equation (17.1.1), every point of the design must appear in

$$\lambda(v-1)/(k-1)$$

blocks. Since a point can only appear in an integral number of blocks, the first result follows.

Similarly, By Equation (17.1.2), there must be

$$\frac{\lambda v(v-1)}{k(k-1)}$$

blocks in the design. Since there can't be a fractional number of blocks, the second result follows. □

Although these conditions are necessary to the existence of a BIBD, there is no guarantee that a BIBD with specified parameters will exist, even if those parameters satisfy these conditions.

**EXAMPLE 17.1.10.** The parameters $v = 15$, $k = 5$, $\lambda = 2$ satisfy the conditions of Theorem 17.1.9, but there is no BIBD$(15, 5, 2)$.

We will not prove that such a design does not exist as the proof would be tedious and unenlightening. We will verify that the parameters satisfy the necessary conditions.

We have

$$\lambda(v-1)/(k-1) = 2(14)/4 = 7,$$

and

$$\frac{\lambda v(v-1)}{k(k-1)} = 2 \cdot 15 \cdot 14/20 = 21.$$

Both of these are integers, so if a design were to exist, each point would appear in 7 blocks, and there would be 21 blocks. A computer search can verify that no such design exists.

**EXERCISES 17.1.11.**

1) Show that for any BIBD$(v, k, \lambda)$, the number of edges of $\lambda K_v$ is equal to the number of edges of $K_k$ times the number of blocks of the design.

2) Suppose there is a BIBD$(16, 6, 3)$. How many blocks does it have? In how many of those blocks does each point appear?

3) Find an edge-colouring of $K_5$ so that the edges of any colour form a $K_2$. What are the parameters of the design to which this corresponds?

4) Here are the blocks of a BIBD with $\lambda = 1$:

$$\begin{array}{llll} B_1 = \{1,2,3\} & B_2 = \{1,4,7\} & B_3 = \{1,5,9\} & B_4 = \{1,6,8\} \\ B_5 = \{4,5,6\} & B_6 = \{2,5,8\} & B_7 = \{2,6,7\} & B_8 = \{2,4,9\} \\ B_9 = \{7,8,9\} & B_{10} = \{3,6,9\} & B_{11} = \{3,4,8\} & B_{12} = \{3,5,7\}. \end{array}$$

   (a) What are the values of $v$, $b$, $k$, and $r$ for this BIBD?

   (b) How many of the blocks contain the element 7 ?

   (c) How many of the blocks contain *both* 2 and 7 ?

   (d) Which blocks contain the element 5 ?

   (e) Which blocks contain *both* 5 and 8 ?

5) Assume $\mathcal{B}$ is a BIBD with $v = 16$, $k = 4$, and $r = 3$.

   (a) What are the values of $b$ and $\lambda$?

   (b) Can you tell whether such a BIBD exists or not?

6) A prize draw allows you to enter by picking 3 numbers from $\{1, \ldots, 14\}$. You will win a prize if you choose two of the three numbers that they will draw. Show that it is possible to guarantee a win by having 14 entries in the draw. Explain whether or not a similar strategy would work if the numbers were chosen from $\{1, \ldots, 21\}$.
   [*Hint:* Use the $(7, 7, 3, 3, 1)$ design.]

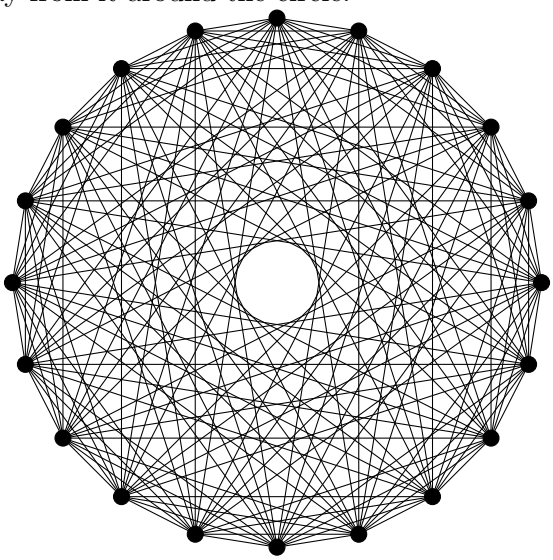
A variety of graphs arise naturally from designs. One example is the **block-intersection graph** for a design. The vertices of this graph correspond to the blocks of the design, and two vertices are adjacent if the corresponding blocks have a non-empty intersection.

**EXAMPLE 17.1.12.** The design of Example 17.1.4 has $K_7$ as its block-intersection graph, because there are 7 blocks (so 7 vertices) and every pair of blocks has nonempty intersection (so every pair of vertices is adjacent).

The design of Example 17.1.6 has the following block intersection graph. Starting from any vertex, label the vertices clockwise in order by

$$\begin{array}{lllll} \{a,b,c,d\}, & \{a,e,i,m\}, & \{a,g,l,n\}, & \{a,h,j,o\}, & \{a,f,k,p\}, \\ \{e,f,g,h\}, & \{b,f,j,n\}, & \{b,h,k,m\}, & \{b,g,i,p\}, & \{b,e,l,o\}, \\ \{i,j,k,l\}, & \{c,g,k,o\}, & \{c,e,j,p\}, & \{c,f,l,m\}, & \{c,h,i,n\}, \\ \{m,n,o,p\}, & \{d,h,l,p\}, & \{d,f,i,o\}, & \{d,e,k,n\}, & \{d,g,j,m\}. \end{array}$$

Each vertex has exactly three other vertices to which it is not adjacent: the vertices that are 5, 10, and 15 vertices away from it around the circle.



Block-intersection graphs have many very nice properties. An example of such a property was given in a 2013 paper by Atif Aliyan Abueida (1966—) and David Angus Pike (1968—). They showed that if you consider any non-Hamilton cycle in the block-intersection graph of any BIBD, the cycle can be "extended": that is, there is a cycle that includes all of the vertices of the original cycle, and one additional vertex. For example, in the block-intersection graph of Example 17.1.12 the cycle

$$(\{a,b,c,d\},\{a,e,i,m\},\{c,f,l,m\})$$

can be extended to the cycle

$$(\{a,b,c,d\},\{c,f,l,m\},\{d,f,i,o\},\{a,e,i,m\}).$$

This cycle in turn can be extended to

$$(\{a,b,c,d\},\{d,f,i,o\},\{a,e,i,m\},\{c,e,j,p\},\{c,f,l,m\}),$$

and so on. The new cycle need not necessarily be achieved by simply inserting a vertex between two consecutive vertices of the previous cycle; it just has to include all of those vertices (and one more) in it. Abueida and Pike also produced a polynomial-time algorithm for finding cycles of every possible length in the block-intersection graph of a BIBD.

## 17.2. Constructing designs, and existence of designs

There are a number of nice methods for constructing designs. We will discuss some of these methods in this section. For some of them, you must start with one design, and use it to create a different design.

**Method 1: Repeating blocks**

This is probably the easiest, and (not surprisingly) the least useful of our construction methods.

Start with a BIBD$(v, k, \lambda)$. For each block of the design, create $t$ copies of that block. The result will be a BIBD$(v, k, t\lambda)$.

**Method 2: Taking the complement**

This method also requires starting with a design. Start with $(V, \mathcal{B})$, a BIBD$(v, k, \lambda)$.

Replace each block $B \in \mathcal{B}$ with its complementary block, $B^c = V \setminus B$. Then

$$(V, \{B^c \mid B \in \mathcal{B}\})$$

is a design.

**DEFINITION 17.2.1.** We call the design constructed by this method, the **complementary design** or **complement** of the design we started with.

**PROPOSITION 17.2.2.** *The complement of a BIBD$(v, k, \lambda)$ is a BIBD$(v, v-k, b-2r+\lambda)$.*

The proof of this proposition is left to the reader, as Exercise 17.2.8.1.

**EXAMPLE 17.2.3.** The complement of the design given in Example 17.1.6, is the following:

$$\begin{array}{ll}
\{e,f,g,h,i,j,k,l,m,n,o,p\}, & \{a,b,c,d,i,j,k,l,m,n,o,p\} \\
\{a,b,c,d,e,f,g,h,m,n,o,p\}, & \{a,b,c,d,e,f,g,h,i,j,k,l\} \\
\{b,c,d,f,g,h,j,k,l,n,o,p\}, & \{b,c,d,e,f,h,i,j,k,m,o,p\} \\
\{b,c,d,e,f,g,i,k,l,m,n,p\}, & \{b,c,d,e,g,h,i,j,l,m,n,o\} \\
\{a,c,d,e,g,h,i,k,l,m,o,p\}, & \{a,c,d,e,f,g,i,j,l,n,o,p\} \\
\{a,c,d,e,f,h,j,k,l,m,n,o\}, & \{a,c,d,f,g,h,i,j,k,m,n,p\} \\
\{a,b,d,e,f,h,i,j,l,m,n,p\}, & \{a,b,d,f,g,h,i,k,l,m,n,o\} \\
\{a,b,d,e,g,h,i,j,k,n,o,p\}, & \{a,b,d,e,f,g,j,k,l,m,o,p\} \\
\{a,b,c,e,f,g,i,j,k,m,n,o\}, & \{a,b,c,e,g,h,j,k,l,m,n,p\} \\
\{a,b,c,f,g,h,i,j,l,m,o,p\}, & \{a,b,c,e,f,h,i,k,l,n,o,p\}
\end{array}$$

Its parameters are $b = 20$, $v = 16$, $r = 15$, $k = 12$, $\lambda = 11$.

**Method 3: Cyclic designs**

This method may be easiest to think of in terms of the associated graph colouring. There are various more complicated versions of this construction that enable us to construct additional designs, but for the purposes of this course, we will focus on almost the most basic version. This most basic version unfortunately gets confusing when $v$ is even, so we will only use it here when $v$ is odd.

**DEFINITION 17.2.4.** Fix an odd integer $n$. A collection of sets $D_1, \ldots, D_m \subseteq \{1, \ldots, n\}$ is a **difference collection** for $n$ (with some fixed multiplicity $\lambda$ — our reasons for using the symbol $\lambda$ for this quantity will become clear shortly), if taking the differences $j - i$ for every pair $i \neq j$ with $i, j \in D_k$ for each set $D_k$, attains each of the values $\pm 1, \ldots, \pm(n-1)/2$, exactly $\lambda$ times, when computations are performed modulo $n$. If $m = 1$ then $D_1$ is called a **difference set**.

**EXAMPLE 17.2.5.** The collection $\{1, 2, 5\}, \{1, 3, 10\}, \{1, 7, 15\}$ is a difference collection for $n = 19$ (with multiplicity 1). The differences we attain appear in the following table.

| Difference set | $i$ | $j$ | $j-i, i-j$ |
|---|---|---|---|
| $D_1$ | 1 | 2 | $\pm 1$ |
| $D_1$ | 1 | 5 | $\pm 4$ |
| $D_1$ | 2 | 5 | $\pm 3$ |
| $D_2$ | 1 | 3 | $\pm 2$ |
| $D_2$ | 1 | 10 | $\pm 9$ |
| $D_2$ | 3 | 10 | $\pm 7$ |
| $D_3$ | 1 | 7 | $\pm 6$ |
| $D_3$ | 1 | 15 | $\pm 14 \equiv \pm 5 \pmod{19}$ |
| $D_3$ | 7 | 15 | $\pm 8$ |

Suppose we have a difference collection for $v$ with multiplicity $\lambda$, in which each set $D_1, \ldots, D_m$ has the same cardinality. Use $D_i + \ell$ to denote the set

$$\{d + \ell \pmod{v} \mid d \in D_i\},$$

performing the modular arithmetic so as to ensure that $D_i + \ell \subseteq \{1, \ldots, v\}$. Then the sets

$$\{D_i + \ell \mid 1 \leq i \leq m, 0 \leq \ell \leq v - 1\}$$

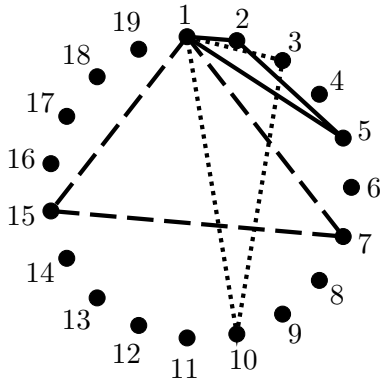
form a BIBD$(v, |D_1|, \lambda)$ (hence our use of the symbol $\lambda$ for the multiplicity).

In the above example, taking the 57 sets $\{1, 2, 5\} + \ell$, $\{1, 3, 10\} + \ell$, and $\{1, 7, 15\} + \ell$, where $0 \leq \ell \leq 18$, gives a BIBD$(19, 3, 1)$.

Let's go over this construction again, thinking about the graph version of the problem. For simplicity, we'll look only at the special case $\lambda = 1$. So our object is to colour the edges of the complete graph $K_v$ so as to ensure that every colour class is a $K_k$. If we draw the vertices of the graph in a circle, and think of the *length* of an edge as being one more than the number of vertices between its endvertices as you travel around the circle in whichever direction is shorter, then for every possible length between 1 and $(v-1)/2$, $K_v$ has $v$ edges of that length. (This is where the trouble arises if $v$ is even: there are only $v/2$ edges of length $v/2$.) Furthermore, if we rotate any edge by one step around the graph (i.e. move both of its endpoints one step in the same direction) repeatedly, after $v$ such rotations we will have moved the edge onto every other edge of that length.

These ideas demonstrate that if we can come up with a set of $K_k$s, such that every edge length appears in exactly one of the $K_k$s, then by taking each one of these as well as every possible rotation of each one of these, as a colour class, we find our desired edge-colouring of $K_v$.

A picture is worth a thousand words. The example above is equivalent to edge-colouring $K_{19}$ so that every colour class forms a $K_3$, since $v = 19$ and $k = 3$. The edge lengths in $K_{19}$ are $\{1, \ldots, 9\}$. We will show three $K_3$s, such that every edge length from $\{1, \ldots, 9\}$ appears in exactly one of them. By rotating each of them, giving each rotation a new colour, we obtain 57 $K_3$s that use every edge of $K_{19}$ exactly once. We've labeled the vertices 1 through 19 to make the edge lengths easier to work out. To enable this to be understood in black-and-white, instead of drawing actual colours on the edges we will draw solid edges for blue, dotted edges for red, and dashed edges for green.




The solid (or blue) triangle has edges of lengths 1, 3, and 4; the dotted (or red) triangle has edges of lengths 2, 7, and 9; and the dashed (or green) triangle has edges of lengths 6, 8, and 5.

A design created using this method is called a *cyclic design*, since a small number of "starter blocks" are being rotated cyclically (in the graph) to find the remaining blocks of the design.

**EXAMPLE 17.2.6.** The collection $\{0,1,3\},\{0,1,3\},\{0,2,5\},\{0,4,5\}$ is a difference collection for $n = 9$ (with multiplicity 3, so the resulting design will have $\lambda = 3$). The differences we attain appear in the following table.

| Difference set | $i$ | $j$ | $j-i, i-j$ |
|---|---|---|---|
| $D_1$ | 0 | 1 | $\pm 1$ |
| $D_1$ | 0 | 3 | $\pm 3$ |
| $D_1$ | 1 | 3 | $\pm 2$ |
| $D_2$ | 0 | 1 | $\pm 1$ |
| $D_2$ | 0 | 3 | $\pm 3$ |
| $D_2$ | 1 | 3 | $\pm 2$ |
| $D_3$ | 0 | 2 | $\pm 2$ |
| $D_3$ | 0 | 5 | $\pm 5 \equiv \pm 4 \pmod 9$ |
| $D_3$ | 2 | 5 | $\pm 3$ |
| $D_4$ | 0 | 4 | $\pm 4$ |
| $D_4$ | 0 | 5 | $\pm 5 \equiv \pm 4 \pmod 9$ |
| $D_4$ | 4 | 5 | $\pm 1$ |

Notice that for a cyclic design to exist, since each set in the difference collection leads to $v$ blocks in the final design, $b$ must be a multiple of $v$.

Although these methods can successfully create designs with many different sets of parameters, they are not nearly enough to allow us to determine the parameters for which BIBDs exist. We noted previously that the necessary conditions given in Theorem 17.1.9 are not sufficient to guarantee the existence of a BIBD with a particular set of parameters. However, there is a very powerful result along these lines, known as Wilson's Theorem after its discoverer, Richard Michael Wilson (1945—). It tells us that if we fix $k$, there are only finitely many values for $v$ that satisfy the necessary conditions but for which no BIBD$(v,k,1)$ exists. Then by Method 1 (repeating blocks), if a BIBD$(v,k,1)$ exists, then so does a BIBD$(v,k,\lambda)$ for any $\lambda$. Here is a formal statement of Wilson's Theorem.

**THEOREM 17.2.7 (Wilson's Theorem).** *Given $k$, there is an integer $v(k)$ such that for every $v > v(k)$ that satisfies the three conditions:*

- $v \in \mathbb{Z}$*;*
- $v(v-1)/[k(k-1)] \in \mathbb{Z}$*; and*
- $(v-1)/(k-1) \in \mathbb{Z}$*,*

*a BIBD$(v, k, 1)$ exists.*

We will not give a proof of this theorem.

**EXERCISES 17.2.8.**

1) Prove that the complement of a BIBD is indeed a design, and that it has the parameters we claimed in Proposition 17.2.2.
[*Hint:* Use inclusion-exclusion to determine how many blocks of the original design contain neither point from an arbitrary pair.]

2) Find the complement of the BIBD$(8, 4, 3)$ given by $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and

$$\mathcal{B} = \begin{Bmatrix} \{1,2,3,4\}, & \{5,6,7,8\}, & \{1,2,5,6\}, & \{1,2,7,8\}, & \{3,4,5,6\}, & \{3,4,7,8\}, & \{2,4,6,8\}, \\ \{1,3,5,7\}, & \{1,3,6,8\}, & \{2,4,5,7\}, & \{1,4,5,8\}, & \{1,4,6,7\}, & \{2,3,5,8\}, & \{2,3,6,7\} \end{Bmatrix}.$$

3) By adding two more sets to the sets $\{1, 3, 7\}$ and $\{1, 6, 13\}$, you can create a difference collection for 25 in which each of the sets has 3 elements, and thus a cyclic BIBD$(25, 3, 1)$. Find two sets to add.

4) Use a difference set to construct a cyclic $(11, 11, 5, 5, 2)$ design.

5) Show that the collection $\mathcal{C} = \big\{\{0, 1, 3\}, \{0, 4, 5\}, \{0, 4, 7\}, \{0, 5, 7\}\big\}$ is a difference collection for 13. Construct the design and give its parameters.

6) Determine whether the given set $D$ is a difference set for the given value of $n$. If it is a difference set, find the parameters of the resulting cyclic BIBD.
   (a) $D = \{1, 2, 4, 10\}$ for $n = 13$.
   (b) $D = \{2, 4, 5, 6, 10\}$ for $n = 21$.

7) Prove that in any cyclic design, there exists an integer $c$ such that $b = cv$, $ck(k-1) = \lambda(v-1)$, and $r = ck$. What is the significance of $c$ in terms of the design?

8) Explain why a BIBD with $v = 6$, $b = 10$, $k = 3$, $r = 5$, and $\lambda = 2$ cannot be cyclic.

9) Does the condition you proved in Exercise 7 show that a BIBD with $v = 61$, $b = 305$, $k = 4$, $r = 20$, and $\lambda = 1$ cannot be cyclic?

## 17.3. Fisher's Inequality

There is one more important inequality that is not at all obvious, but is necessary for the existence of a BIBD$(v, k, \lambda)$. This is known as Fisher's Inequality, since it was proven by Sir Ronald Aylmer Fisher (1890—1962). The proof we will give is somewhat longer than the standard proof. This is because the standard proof uses linear algebra, which we do not expect to be required background for this course.

**THEOREM 17.3.1 (Fisher's Inequality).** *For any BIBD$(v, k, \lambda)$, we must have $b \geq v$.*

Before proving this fact, let's observe the consequences in terms of the usual parameters: $v$, $k$, and $\lambda$. We know from Equation (17.1.2) that

$$b = \frac{\lambda v(v-1)}{k(k-1)},$$

so $b \geq v$ implies

$$\frac{\lambda v(v-1)}{k(k-1)} \geq v.$$

Since $v$ is the number of points of a design, it must be positive, so dividing through by $v$ does not reverse the inequality. Thus,

$$\frac{\lambda(v-1)}{k(k-1)} \geq 1.$$

Since $k$ is the number of points in each block, both $k$ and $k-1$ must be positive (we are ignoring the trivial case $k = 1$), so multiplying through by $k(k-1)$ does not reverse the inequality. Thus,

$$\lambda(v-1) \geq k(k-1).$$

**PROOF OF FISHER'S INEQUALITY.** Suppose we have an arbitrary BIBD$(v, k, \lambda)$. Let $B$ be an arbitrary block of this design. For each value of $i$ between 0 and $k$ (inclusive), let $n_i$ denote the number of blocks $B' \neq B$ such that $|B' \cap B| = i$. (When we say $B' \neq B$ we allow the blocks to be equal as sets if the block $B$ is a repeated block of the design; we are only insisting that $B'$ not be the exact same block of the design as $B$.)

The following equations involving $n_i$ are consequences of easy combinatorial proofs, together with the definition of $n_i$:

$$\sum_{i=0}^{k} n_i = b - 1, \tag{17.3.1}$$

because both sides of this equation count every block except $B$.

$$\sum_{i=0}^{k} i n_i = k(r-1), \tag{17.3.2}$$

because both sides of this equation count the number of times elements of $B$ appear in some other block of the design.

$$\sum_{i=2}^{k} i(i-1)n_i = k(k-1)(\lambda-1),$$

because both sides of this equation count the number of times all of the ordered pairs of elements from $B$ appear together in some other block of the design. Note that when $i = 0$ or $i = 1$, we have $i(i-1)n_i = 0$, so in fact

$$\sum_{i=0}^{k} i(i-1)n_i = \sum_{i=2}^{k} i(i-1)n_i = k(k-1)(\lambda-1). \tag{17.3.3}$$

Adding Equations (17.3.2) and (17.3.3) gives

$$\sum_{i=0}^{k} i^2 n_i = k(k-1)(\lambda-1) + k(r-1). \tag{17.3.4}$$

Now comes the part of the proof where something mysterious happens, and for reasons that are not at all apparent, the result we want will emerge. To fully understand a proof like this one requires deeper mathematics, but even seeing a proof is useful to convince ourselves that the result is true.

Take the polynomial in $x$ given by

$$\sum_{i=0}^{k} (x-i)^2 n_i = \sum_{i=0}^{k} (x^2 - 2xi + i^2)n_i = x^2 \sum_{i=0}^{k} n_i - 2x \sum_{i=0}^{k} i n_i + \sum_{i=0}^{k} i^2 n_i.$$

Using Equations (17.3.1), (17.3.2), and (17.3.4), we see that this is equal to

$$x^2(b-1) - 2xk(r-1) + k(k-1)(\lambda-1) + k(r-1).$$

Notice that the format in which this polynomial started was a sum of squares times non-negative integers, so its value must be non-negative for any $x \in \mathbb{R}$.

Using the quadratic formula, $ax^2 + b'x + c = 0$ has roots at

$$\frac{-b' \pm \sqrt{(b')^2 - 4ac}}{2a}.$$

If a quadratic polynomial has two real roots, then there is a region in which its values are negative. Since this polynomial is non-negative for every $x \in \mathbb{R}$, it can have at most one real root, so $(b')^2 - 4ac \leq 0$. Substituting the actual values from our polynomial, this means that

$$(-2k(r-1))^2 - 4(b-1)(k(k-1)(\lambda-1) + k(r-1)) \leq 0.$$

Hence,

$$k^2(r-1)^2 - k(b-1)((k-1)(\lambda-1) + r - 1) \leq 0.$$

Let's rewrite the $b$ in terms of $v, r$, and $k$. By Theorem 17.1.7, we have $bk = vr$, so

$$k(b-1) = bk - k = vr - k.$$

Hence

$$k^2(r-1)^2 - (vr-k)((k-1)(\lambda-1) + r - 1) \leq 0.$$

Expand the second term slightly, and multiply both sides of the inequality by $v - 1$:

$$k^2(r-1)^2(v-1) - (vr-k)(k-1)(\lambda-1)(v-1) - (vr-k)(r-1)(v-1) \leq 0.$$

In the middle expression, we have $(\lambda-1)(v-1)$. By Theorem 17.1.7, we know that $\lambda = r(k-1)/(v-1)$, so

$$\lambda - 1 = \frac{r(k-1) - (v-1)}{v-1}.$$

Therefore,

$$(\lambda-1)(v-1) = r(k-1) - v + 1.$$

Thus, we have

$$k^2(r-1)^2(v-1) - (vr-k)(k-1)(rk-r-v+1) - (vr-k)(r-1)(v-1) \leq 0.$$

The next step is a lot of work to do by hand. Fortunately there is good math software that can perform routine tasks like this quickly. If we expand this inequality fully, remarkably it has a nice factorisation:

$$r(k-r)(v-k)^2 \leq 0.$$

Now, $r > 0$ for any design, and $(v-k)^2$ is a square, so must be nonnegative. Therefore, this inequality forces $k - r \leq 0$, so $k \leq r$. Hence $r/k \geq 1$. Using Theorem 17.1.7, we have

$$b = vr/k \geq v,$$

as desired. □

Notice that if $k$ is fixed, then only finitely many values of $v$ do not meet Fisher's Inequality, so satisfying this inequality did not need to be added as a condition to Wilson's Theorem.

**EXERCISES 17.3.2.**

1) Find values for $v$, $k$ and $\lambda$ that satisfy Theorem 17.1.9 but do not satisfy Fisher's Inequality. What can you say about the existence of a design with these parameters?
2) Suppose that $\lambda = 1$ and $k = 20$. How big must $v$ be to satisfy Fisher's Inequality? What is the smallest value for $v$ that satisfies all of the necessary conditions?
3) Suppose that $\lambda = 2$ and $k = 20$. How big must $v$ be to satisfy Fisher's Inequality? What is the smallest value for $v$ that satisfies all of the necessary conditions?
4) Explain how you know there does not exist a BIBD with $v = 46$, $b = 23$, and $k = 10$.
5) Explain how you know there does not exist a BIBD with $v = 8$, $b = 10$, $k = 4$, and $r = 5$.
6) If $\mathcal{B}$ is a BIBD with $v = 22$, then what can you say about the value of $b$?

---

**SUMMARY:**

- equivalence between designs and (multi)graph colouring/decomposition problem
- necessary conditions for a BIBD
- construction methods for designs
- Wilson's Theorem
- Fisher's Inequality
- Important definitions:
  - design
  - blocks
  - balanced, regular, uniform
  - BIBD
  - complementary design
  - difference collection
- Notation:
  - $b$, $v$, $r$, $k$, $\lambda$

---

# Chapter 18

# More designs

## 18.1. Steiner and Kirkman triple systems

In 1844, Wesley Stoker Barker Woolhouse (1809—1893), editor of the *Ladies and Gentleman's Diary*, posed that publication's annual prize problem:

> Determine the number of combinations that can be made of $n$ symbols, $p$ symbols in each, with this limitation, that no combination of $q$ symbols which may appear in any one of them shall be repeated in any other.

If we take $q = 2$ then this is essentially a design with $k = p$ and $v = n$, although it need not be balanced; some pairs might appear once while other pairs do not appear. Although some responses were printed in 1845, they were not satisfactory, and in 1846 Woolhouse repeated the question for the special case $q = 2$ and $p = 3$.

In 1847, Reverend Thomas Penyngton Kirkman (1806—1895) (previously mentioned in Chapter 13) found a complete solution to this problem in the case where the design is balanced (with $\lambda = 1$), and made some progress towards solving the complete problem. In our terminology, his solution completely determined the values of $v$ for which a BIBD$(v, 3, 1)$ exists.

Although Jakob Steiner (1796—1863) did not study triple systems until 1853, he came up with Kirkman's result independently, and his work was more broadly disseminated in mathematical circles, so these structures still carry his name. Despite this, as we shall see later in this section, there is a related problem that has been named after Kirkman.

**DEFINITION 18.1.1.** A **triple system** is a (regular) balanced design in which every block has cardinality 3; that is, a BIBD$(v, 3, \lambda)$.

A **Steiner triple system** is a triple system with $\lambda = 1$.

**EXAMPLE 18.1.2.** The cyclic BIBD$(19, 3, 1)$ given in Example 17.2.5 is a Steiner triple system. So is the design on 7 points given by

$$\{1, 2, 3\}, \{1, 4, 5\}, \{1, 6, 7\}, \{2, 4, 6\}, \{2, 5, 7\}, \{3, 4, 7\}, \{3, 5, 6\}.$$

**NOTATION 18.1.3.** Since the only variable in a Steiner triple system is $v$, for such a system on $v$ points we use the notation STS$(v)$.

Triple systems might seem like a very special case of designs, and it would be reasonable to wonder why we have chosen to single these out for special study and attention. The answer is that triple systems can be thought of as the smallest interesting examples of designs, since (as noted previously) if $k = 1$ there are no pairs in any block and the design is trivial; and if $k = 2$ then the blocks are simply copies of every possible pair from the set $V$. So triple systems are a

natural starting point when we are learning about designs: they include examples that are are not too big or complicated to understand, but are non-trivial.

If you are now convinced that triple systems are worth studying, you might still be wondering about Steiner triple systems in particular. We've seen that the method of repeating blocks allows us to construct a triple system for any $\lambda$ if we first have a triple system with $\lambda = 1$, so Steiner triple systems will be the rarest kind of triple system, and are therefore of particular interest.

In the remainder of this section, we will prove Kirkman's result characterising the values of $v$ for which an STS($v$) exists. To do this, we will require two results about the existence of special kinds of Latin squares. There are many connections in combinatorics!

The special kinds of Latin squares we require will be *symmetric*: that is, the entry in row $i$ and column $j$ must equal the entry in row $j$, column $i$. We will also specify the entries on the *main diagonal*: the positions for which the row number and the column number are the same.

**LEMMA 18.1.4.** *For every odd $n$, there is a symmetric Latin square of order $n$ with 1, 2, …, $n$ appearing in that order down the main diagonal.*

**PROOF.** Make the entries of the first row

$$1, (n+3)/2, 2, (n+5)/2, 3, \ldots, (n-1)/2, n, (n+1)/2.$$

For $i \geq 2$, the entries of row $i$ will be the entries of row $i-1$ shifted one position to the left.

Clearly all of the entries in any row are distinct. Also, since it takes $n$ shifts to the left to return to the starting point, all of the entries in any column must be distinct.

Since the entry in row $a$, column $b$ moves to the entry in row $a+1$ column $b-1 \pmod n$, the positions in which this entry appears will be precisely the positions $(x, y)$ for which $x + y \equiv a + b \pmod n$. Since $i + j = j + i$, the entry in column $i$ of row $j$ will be the same as the entry in column $j$ of row $i$. Thus, the Latin square is symmetric.

The same argument shows that the entry in row $i$ and column $i$ will be the entry in position $2i - 1 \pmod n$ of row 1. But this is precisely where $i$ has been placed, so this entry will be $i$, as desired. □

**EXAMPLE 18.1.5.** When $n = 11$, here is the (symmetric) Latin square constructed in the proof of Lemma 18.1.4:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 | 2 | 8 | 3 | 9 | 4 | 10 | 5 | 11 | 6 |
| 7 | 2 | 8 | 3 | 9 | 4 | 10 | 5 | 11 | 6 | 1 |
| 2 | 8 | 3 | 9 | 4 | 10 | 5 | 11 | 6 | 1 | 7 |
| 8 | 3 | 9 | 4 | 10 | 5 | 11 | 6 | 1 | 7 | 2 |
| 3 | 9 | 4 | 10 | 5 | 11 | 6 | 1 | 7 | 2 | 8 |
| 9 | 4 | 10 | 5 | 11 | 6 | 1 | 7 | 2 | 8 | 3 |
| 4 | 10 | 5 | 11 | 6 | 1 | 7 | 2 | 8 | 3 | 9 |
| 10 | 5 | 11 | 6 | 1 | 7 | 2 | 8 | 3 | 9 | 4 |
| 5 | 11 | 6 | 1 | 7 | 2 | 8 | 3 | 9 | 4 | 10 |
| 11 | 6 | 1 | 7 | 2 | 8 | 3 | 9 | 4 | 10 | 5 |
| 6 | 1 | 7 | 2 | 8 | 3 | 9 | 4 | 10 | 5 | 11 |

You can see that this construction will not work when $n$ is even, since the values of some of the entries given in the formulas would not be integers. In fact, it is not possible to construct a symmetric Latin square of order $n$ with the entries $1, \ldots n$ down the main diagonal when $n$ is even. Fortunately, it is possible to construct something similar that will achieve what we will require.

**LEMMA 18.1.6.** *For every even $n$, there is a symmetric Latin square of order $n$ with the values $1, \ldots, n/2, 1, \ldots, n/2$ appearing in that order down the main diagonal.*

**PROOF.** Make the entries of the first row

$$1, (n+2)/2, 2, (n+4)/2, 3, \ldots, (n-2)/2, n.$$

For $i \geq 2$, the entries of row $i$ will be the entries of row $i-1$ shifted one position to the left.

The same arguments as in the proof of Lemma 18.1.4 show that this is a symmetric Latin square. The entry in row $i$ and column $i$ will be the entry in position $2i-1 \pmod n$ of row 1. These are the entries $1, 2, \ldots, n/2$ and since position

$$2(n+2j)/2 - 1 \equiv 2j - 1 \pmod n,$$

the entries in positions $(j, j)$ and $((n+2j)/2, (n+2j)/2)$ will be the same, so each of these entries will be repeated in the same order. □

**EXAMPLE 18.1.7.** When $n = 10$, here is the (symmetric) Latin square constructed in the proof of Lemma 18.1.6:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 6 | 2 | 7 | 3 | 8 | 4 | 9 | 5 | 10 |
| 6 | 2 | 7 | 3 | 8 | 4 | 9 | 5 | 10 | 1 |
| 2 | 7 | 3 | 8 | 4 | 9 | 5 | 10 | 1 | 6 |
| 7 | 3 | 8 | 4 | 9 | 5 | 10 | 1 | 6 | 2 |
| 3 | 8 | 4 | 9 | 5 | 10 | 1 | 6 | 2 | 7 |
| 8 | 4 | 9 | 5 | 10 | 1 | 6 | 2 | 7 | 3 |
| 4 | 9 | 5 | 10 | 1 | 6 | 2 | 7 | 3 | 8 |
| 9 | 5 | 10 | 1 | 6 | 2 | 7 | 3 | 8 | 4 |
| 5 | 10 | 1 | 6 | 2 | 7 | 3 | 8 | 4 | 9 |
| 10 | 1 | 6 | 2 | 7 | 3 | 8 | 4 | 9 | 5 |

We are now ready to characterise the values of $v$ for which there is an STS($v$).

**THEOREM 18.1.8.** *An STS($v$) exists if and only if $v \equiv 1, 3 \pmod 6$.*

**PROOF.** We prove the two implications separately.

**(⇒)** Suppose that an STS($v$) exists. Then by Theorem 17.1.9, the values

$$\lambda\frac{v-1}{k-1} = \frac{v-1}{2} \quad \text{and} \quad \lambda\frac{v(v-1)}{k(k-1)} = \frac{v(v-1)}{6}$$

must be integers. The first of these conditions tells us that $v$ must be odd, so must be 1, 3, or 5 (mod 6). If $v \equiv 5 \pmod 6$, then $v = 6q + 5$ for some $q$, so

$$v(v-1)/6 = (6q+5)(6q+4)/6.$$

Since neither $6q+5$ nor $6q+4$ is a multiple of 3, this will not be an integer. Thus, the second condition eliminates the possibility $v \equiv 5 \pmod 6$. Therefore, $v \equiv 1, 3 \pmod 6$.

**(⇐)** Suppose that $v \equiv 1, 3 \pmod 6$. We give separate constructions of Steiner triple systems on $v$ points, depending on the congruence class of $v$. We will use the graph theoretic approach to the problem, so our goal is to find colour classes for the edges of $K_v$ such that each colour class consists of a $K_3$.

***v* ≡ 3 (mod 6).** Say $v = 6q + 3$. Label the vertices of $K_{6q+3}$ with

$$u_1, \ldots, u_{2q+1}; v_1, \ldots, v_{2q+1}; \text{ and } w_1, \ldots, w_{2q+1}.$$

By Lemma 18.1.4, there is a Latin square of order $2q+1$ in which for every $1 \leq i \leq 2q+1$, the entry $i$ appears in position $(i,i)$.

For $1 \leq i,j \leq 2q+1$ with $i \neq j$, if the entry in position $(i,j)$ of this Latin square is $\ell$, then use new colours to colour the edges that join the vertices in each of the following sets:

$$\{u_i, u_j, v_\ell\}; \{v_i, v_j, w_\ell\}; \text{ and } \{w_i, w_j, u_\ell\}.$$

Since the Latin square was symmetric, both position $(i,j)$ and position $(j,i)$ give rise to the same colour classes. Since we consider every pair $i \neq j$, every edge of the form $u_iu_j$, $v_iv_j$, or $w_iw_j$ has been coloured (we must have $i \neq j$ for such an edge to exist). Since the square is Latin, every possible entry $\ell$ occurs somewhere in row $i$, so every edge of the form $u_iv_\ell$, $v_iw_\ell$, or $w_iu_\ell$ has been coloured, except that we did not look at the entry of the Latin square in the position $(i,j)$, where $i = j$. We know that the entry in position $(i,i)$ is $i$, so the edges of the form $u_iv_i$, $v_iw_i$, and $w_iu_i$ are the only edges that have not yet been coloured.

For each $1 \leq i \leq 2q+1$, make the edges joining $u_i$, $v_i$, and $w_i$ into one colour class.

Now all of the edges of $K_{6q+3}$ have been coloured, and each colour class forms a $K_3$, so we have constructed a Steiner triple system.

***v* ≡ 1 (mod 6).** Say $v = 6q+1$. Label the vertices of $K_{6q+1}$ with

$$u_1, \ldots, u_{2q}; v_1, \ldots, v_{2q}; \text{ and } w_1, \ldots, w_{2q}; x.$$

By Lemma 18.1.6, there is a Latin square of order $2q$ in which for every $1 \leq i \leq q$, the entry $i$ appears in position $(i,i)$, and for every $q+1 \leq i \leq 2q$, $i-q$ appears in position $(i,i)$.

For $1 \leq i,j \leq 2q$ with $i \neq j$, if the entry in position $(i,j)$ of this Latin square is $\ell$, then use new colours to colour the edges that join the vertices in each of the following sets:

$$\{u_i, u_j, v_\ell\}; \{v_i, v_j, w_\ell\}; \text{ and } \{w_i, w_j, u_\ell\}.$$

Since the Latin square was symmetric, both position $(i,j)$ and position $(j,i)$ give rise to the same colour classes. Since we consider every pair $i \neq j$, every edge of the form $u_iu_j$, $v_iv_j$, or $w_iw_j$ has been coloured (we must have $i \neq j$ for such an edge to exist). Since the square is Latin, every possible entry $\ell$ occurs somewhere in row $i$, so every edge of the form $u_iv_\ell$, $v_iw_\ell$, or $w_iu_\ell$ has been coloured, except that we did not look at the entry of the Latin square in the position $(i,j)$, where $i = j$. We know that the entry in position $(i,i)$ is $i$ (if $i \leq q$) or $i-q$ (if $i > q$), so the only edges that have not yet been coloured are the edges of the form $u_iv_i$, $v_iw_i$, and $w_iu_i$ when $i \leq q$ and $u_iv_{i-q}$, $v_iw_{i-q}$, and $w_iu_{i-q}$ when $i > q$, as well as every edge incident with $x$.

For each $1 \leq i \leq q$, make the edges joining $u_i$, $v_i$, and $w_i$ into one colour class. Observe that amongst the remaining edges that are not incident with $x$, every vertex other than $x$ is an endvertex of precisely one of the edges. For example, if $i \geq q$ then $u_iv_{i-q}$ is one of these edges, while if $i < q$, $w_{i+q}u_i$ is one of these edges, so either way, $u_i$ is an endvertex of precisely one of these edges. Therefore, if for every $q+1 \leq i \leq 2q$ we use new colours to colour the edges that join the vertices in each of the following sets:

$$\{u_i, v_{i-q}, x\}; \{v_i, w_{i-q}, x\}; \text{ and } \{w_i, u_{i-q}, x\},$$

every edge incident with $x$ (as well as all of our other remaining edges) will have been coloured.

Now all of the edges of $K_{6q+1}$ have been coloured, and each colour class forms a $K_3$, so we have constructed a Steiner triple system. □

**EXAMPLE 18.1.9.** We will use the method of Theorem 18.1.8 to construct an STS(15).

We have $15 = 6(2) + 3$, so $q = 2$. The points of our design will be $u_i$, $v_i$, and $w_i$ for $1 \leq i \leq 2q+1 = 5$, and we will require a symmetric Latin square of order 5. Here is the square:

$$\begin{array}{ccccc} 1 & 4 & 2 & 5 & 3 \\ 4 & 2 & 5 & 3 & 1 \\ 2 & 5 & 3 & 1 & 4 \\ 5 & 3 & 1 & 4 & 2 \\ 3 & 1 & 4 & 2 & 5 \end{array}$$

Here are the blocks we form from this Latin square:

$$\begin{array}{ccccccc} \{u_1,u_2,v_4\}, & \{v_1,v_2,w_4\}, & \{w_1,w_2,u_4\}, & & \{u_1,u_3,v_2\}, & \{v_1,v_3,w_2\}, & \{w_1,w_3,u_2\}, \\ \{u_1,u_4,v_5\}, & \{v_1,v_4,w_5\}, & \{w_1,w_4,u_5\}, & & \{u_1,u_5,v_3\}, & \{v_1,v_5,w_3\}, & \{w_1,w_5,u_3\}, \\ \{u_2,u_3,v_5\}, & \{v_2,v_3,w_5\}, & \{w_2,w_3,u_5\}, & & \{u_2,u_4,v_3\}, & \{v_2,v_4,w_3\}, & \{w_2,w_4,u_3\}, \\ \{u_2,u_5,v_1\}, & \{v_2,v_5,w_1\}, & \{w_2,w_5,u_1\}, & & \{u_3,u_4,v_1\}, & \{v_3,v_4,w_1\}, & \{w_3,w_4,u_1\}, \\ \{u_3,u_5,v_4\}, & \{v_3,v_5,w_4\}, & \{w_3,w_5,u_4\}, & & \{u_4,u_5,v_2\}, & \{v_4,v_5,w_2\}, & \{w_4,w_5,u_2\}, \end{array}$$

$$\{u_1,v_1,w_1\}, \{u_2,v_2,w_2\}, \{u_3,v_3,w_3\}, \{u_4,v_4,w_4\}, \{u_5,v_5,w_5\}$$

After solving Woolhouse's problem, Kirkman noticed that his construction of an STS(15) had a very nice property. He challenged others to come up with this solution in the following problem that he published in the 1850 edition of the *Ladies and Gentleman's Diary*:

> Fifteen young ladies in a school walk out three abreast for seven days in succession: it is required to arrange them daily so that no two shall walk twice abreast.

This has become known as Kirkman's Schoolgirl Problem.

Although this problem begins by requiring an STS(15), it has the additional requirement that it must be possible to partition the blocks of this design (the rows of young ladies) into seven groups (of five blocks each) so that every point (young lady) appears exactly once in each group. This extra requirement comes from the fact that each of the young ladies must walk out every day, and can only be in one row in any given day.

**DEFINITION 18.1.10.** A BIBD is a **resolvable design** if the blocks of the design can be partitioned into sets, each of which forms a partition of the point set of the BIBD.

A **Kirkman triple system** is a resolvable Steiner triple system.

Notice that since each block has 3 points in it, a Kirkman triple system is only possible if $v/3$ is an integer. Since a Kirkman triple system is also a Steiner triple system, this means that we must have $v \equiv 3 \pmod 6$.

There are seven non-isomorphic Kirkman triple systems of order 15.

Kirkman triple systems are also known to exist whenever $v \equiv 3 \pmod 6$. This result was presented at a conference in 1968 by Dwijendra Kumar Ray-Chaudhuri (1933—) and Richard Michael Wilson (1945—), and published in 1971 in the proceedings of that conference. Lu Jiaxi (1935—1983) had actually written a proof of this result in 1961, but he was unknown at the time and through misunderstandings and mistakes, his work was rejected as ``not really new."

**THEOREM 18.1.11 (Lu, 1961; Ray-Chaudhuri and Wilson, 1971).** *There is a Kirkman triple system whenever* $v \equiv 3 \pmod 6$.

**EXERCISES 18.1.12.**

1) For $v = 37$, give the Latin square you would have to use in order to construct a Steiner triple system using the method described in the proof of Theorem 18.1.8.

2) For $v = 39$, give the Latin square you would have to use in order to construct a Steiner triple system using the method described in the proof of Theorem 18.1.8.

3) For $v = 19$, use the method described in the proof of Theorem 18.1.8 to construct a Steiner triple system.

4) Is the STS(15) constructed in Example 18.1.9 a Kirkman triple system? Explain your answer. [*Hint:* Think of Pigeonhole-type arguments.]

5) Find all values of $\lambda$ for which a triple system on six varieties exists. For each such value of $\lambda$, either give a design or explain how to construct it.
[*Hint:* Begin by showing that if such a design exists, its parameters are $(2r, 6, r, 3, 2r/5)$. Then determine what values $r$ can take on. Finally use some results about how to construct designs.]

6) Construct an STS(13) design. Show your work.

7) Let $v = 21$.
   (a) Write down the Latin square that you would use to construct a Steiner triple system (for this value of $v$) using the method described in the proof of Theorem 18.1.8.
   (b) For the resulting Steiner triple system, which triples contain:
      (i) both $u_1$ and $u_3$?
      (ii) both $v_2$ and $w_7$?
      (iii) both $u_3$ and $w_4$?
      (iv) both $v_5$ and $w_5$?
      (v) $w_3$?

8) Let $v = 27$.
   (a) Write down the Latin square that you would use to construct a Steiner triple system (for this value of $v$) using the method described in the proof of Theorem 18.1.8.
   (b) For the resulting Steiner triple system, which triples contain:
      (i) both $v_3$ and $u_8$?
      (ii) both $u_4$ and $w_4$?
      (iii) both $u_5$ and $w_4$?
      (iv) $v_2$?
      (v) both $v_2$ and $v_5$?

## 18.2. $t$-designs

In a BIBD, every *pair* appears together $\lambda$ times. In the notation of Woolhouse's problem, $q = 2$. What about larger values of $q$? (We'll still only consider the case where every $q$-set appears an equal number of times $\lambda$, so the design must be balanced, but we will include the more general situation that $\lambda \geq 1$.)

**DEFINITION 18.2.1.** A ***t*-$(v, k, \lambda)$ design** is a design on $v$ points with blocks of cardinality $k$, such that every $t$-subset of $V$ appears in exactly $\lambda$ blocks.

So far, all we have looked at have been 2-designs.

**THEOREM 18.2.2.** *In a $t$-$(v, k, \lambda)$ design,*

$$\binom{k-i}{t-i} \text{ is a divisor of } \lambda\binom{v-i}{t-i} \text{ for every } 0 \leq i \leq t-1.$$

**PROOF.** We first consider the special case where $i = 0$. Notice that in each of the $b$ blocks, there are $\binom{k}{t}$ subsets of cardinality $t$ that appear in that block. So in the entire design, $b\binom{k}{t}$ subsets of cardinality $t$ appear.

There exist $\binom{v}{t}$ subsets of cardinality $t$ from the $v$ points of $V$, and each appears in $\lambda$ blocks, so in the entire design, $\lambda\binom{v}{t}$ subsets of cardinality $t$ appear.

Thus, $b\binom{k}{t} = \lambda\binom{v}{t}$.

Similarly, if we fix any set of $i$ varieties, there are $\binom{v-i}{t-i}$ subsets of cardinality $t$ that include these $i$ varieties. Each such subset appears in $\lambda$ blocks. However, for each of the blocks that contains these $i$ elements (the number of these will be our quotient), we can complete our $i$-set to a set of cardinality $t$ that lies within this block, in $\binom{k-i}{t-i}$ ways. Thus, we have counted any such block $\binom{k-i}{t-i}$ times in the preceding count. So $\binom{k-i}{t-i}$ must be a divisor of $\lambda\binom{v-i}{t-i}$, as claimed. □

**EXAMPLE 18.2.3.** Show that there is no $5-(16,7,1)$ design.

**SOLUTION.** We check the necessary conditions given in Theorem 18.2.2. Using the condition when $i = 0$, we see that $b\binom{7}{5} = 1\binom{16}{5}$, so $21b = 4368$. Therefore $b = 208$. This condition is satisfied.

When $i = 1$ we have $\binom{k-i}{t-i} = \binom{6}{4} = 15$, and $\lambda\binom{v-i}{t-i} = \binom{15}{4} = \frac{15\cdot14\cdot13\cdot12}{4\cdot3\cdot2} = 15\cdot7\cdot13$, which is divisible by 15. This condition is satisfied.

When $i = 2$ we have $\binom{k-i}{t-i} = \binom{5}{3} = 10$, and $\lambda\binom{v-i}{t-i} = \binom{14}{3} = \frac{14\cdot13\cdot12}{3\cdot2} = 14\cdot13\cdot2$, which is not divisible by 10 since it is not a multiple of 5. This condition fails. Thus, there is no $5-(16,7,1)$ design. □

Suppose we have a 3-$(10,4,1)$ design. By Theorem 18.2.2, it will have

$$b\binom{4}{3} = 1\binom{10}{3},$$

so $4b = 10\cdot9\cdot8/6 = 120$. Thus, $b = 30$. Also, since $bk = vr$, we have $30\cdot4 = 10r$, so $r = 12$.

**EXAMPLE 18.2.4.** Here is a 3-$(10,4,1)$ design.

| | | | | | |
|---|---|---|---|---|---|
| $\{1,5,6,10\}$, | $\{1,2,8,9\}$, | $\{2,3,6,7\}$, | $\{3,4,9,10\}$, | $\{4,5,7,8\}$, | $\{1,3,4,7\}$, |
| $\{2,4,5,10\}$, | $\{1,3,5,8\}$, | $\{1,2,4,6\}$, | $\{2,3,5,9\}$, | $\{4,6,8,9\}$, | $\{1,7,9,10\}$, |
| $\{3,6,8,9\}$, | $\{5,6,7,9\}$, | $\{2,7,8,10\}$, | $\{1,2,3,10\}$, | $\{1,2,5,7\}$, | $\{1,4,5,9\}$, |
| $\{1,3,6,9\}$, | $\{1,6,7,8\}$, | $\{1,4,8,10\}$, | $\{2,3,4,8\}$, | $\{2,4,7,9\}$, | $\{2,5,6,8\}$, |
| $\{2,6,9,10\}$, | $\{3,4,5,6\}$, | $\{3,5,7,10\}$, | $\{3,7,8,9\}$, | $\{4,6,7,10\}$, | $\{5,8,9,10\}$ |

Notice: For $t \geq 3$, a $t$-design is also a $(t-1)$-design. If every $t$-set appears in exactly $\lambda$ blocks, then any $(t-1)$-set must appear in exactly

$$\lambda(v-t+1)/(k-t+1)$$

blocks. This is because if we fix a $(t-1)$-set, it can be made into a $t$-set by adding any one of the $v-t+1$ other elements of $V$. Each of these $t$-sets appears in $\lambda$ of the blocks. However, some of these blocks are the same; in fact, we have counted each block containing this $(t-1)$-set once for every other element of the block (since every other element of the block forms a $t$-set when put together with the $(t-1)$-set). So every block that contains this $(t-1)$-set has been counted $k-(t-1)$ times. The result follows. (From the above formula we can see that $k-t+1$ is a divisor of $\lambda(v-t+1)$; this is exactly the condition that Theorem 18.2.2 gives when we take $i = t-1$.)

Therefore, since

$$1(10-3+1)/(4-3+1) = 4,$$

the $3-(10,4,1)$ design that we gave above, is also a $2-(10,4,4)$ design. In more generality, a $t-(v,k,\lambda)$ design with $t>2$ is also a $(t-1)-(v,k,\lambda(v-t+1)/(k-t+1))$ design.

The name of Jakob Steiner (1796—1863) is also used in this more general context, and without the constraint on the block sizes.

**DEFINITION 18.2.5.** A **Steiner system** is a $t$-design with $\lambda = 1$.

The 3-$(10,4,1)$ design above is a Steiner system.

Our ability to construct Steiner systems when $k>3$, or when $t>2$, except in trivial cases, is almost nonexistent. In fact, there are no known constructed Steiner systems with $t>5$, with the exception that taking every possible $t$-subset of a $v$-set is always a (trivial) $t-(v,t,1)$ design.

Despite this, in 2014 a remarkable theorem was proved by Peter Keevash (1978—), along similar lines to Wilson's Theorem, but applying to this more general context.

The necessary conditions given in Theorem 18.2.2 are not sufficient to guarantee the existence of a BIBD with a particular set of parameters. However, Keevash's Theorem tells us that if we fix $k$ and $t$, there are only finitely many values for $v$ that satisfy the necessary conditions but for which no $t-(v,k,1)$ design exists. His proof was probabilistic, so does not produce constructions for any designs. Here is a formal statement of Keevash's Theorem.

**THEOREM 18.2.6 (Keevash's Theorem).** *Given $k$ and $t$, there is an integer $v(k,t)$ such that for every $v > v(k,t)$ that satisfies the conditions:*

- *$v \in \mathbb{Z}$; and*
- *for every $0 \le i \le t-1$,*

$$\binom{k-i}{t-i} \text{ is a divisor of } \lambda\binom{v-i}{t-i}.$$

*a $t-(v,k,1)$ design exists.*

We will not give a proof of this theorem.

**EXERCISES 18.2.7.**

1) Substituting $t=2$ into the equations of Theorem 18.2.2 doesn't immediately look like either of the equations in Theorem 17.1.7. Use the equations of Theorem 17.1.7 to deduce that

$$\binom{k}{2} \text{ is a divisor of } \lambda\binom{v}{2}.$$

2) If $v=15$ and $\lambda=1$, what are all possible values of $k$ and $t \ge 2$ for which $t$-designs might exist? Do not include any trivial $t-(v,t,1)$ design, so you may assume $v>k>t$.

3) Is it possible for a 3-$(16,6,1)$ design to exist? If so, how many blocks will it have? What will the value of $r$ be?

4) Let $\mathcal{B}$ be the $(7,3,1)$-design. Define a new design $D$ as follows, on the varieties $\{1,\dots,8\}$. It has 14 blocks, of two types:

   I. the blocks of $\mathcal{B}$ but with variety 8 added to each; and

   II. the blocks of the complementary design to $\mathcal{B}$.

Prove that this is a Steiner system with $t = 3$, $k = 4$ and $v = 8$. Use the structure of $\mathcal{B}$ and its complement to show that $\lambda = 1$; do not check all $\binom{8}{3}$ possible 3-subsets of $\{1, \ldots, 8\}$.

5) Define a design as follows. Label the edges of the complete graph $K_6$; these will be the varieties of the design. The blocks are of two types:

   I. any set of three edges from $K_6$ that can be properly coloured with the same colour; and

   II. any set of three edges that form a triangle in $K_6$.

   Determine the parameters of this $t$-design (including the highest value of $t$ for which this is a $t$-design, and justifying each value you determine), and show that this is a Steiner system.

6) Might a $3 - (20, 5, 8)$ design exist according to the necessary conditions we have determined (Theorem 18.2.2)? State the formulas that must be satisfied and show your work.

## 18.3. Affine planes

You are probably familiar with at least some of the axioms of geometry introduced by Euclid (c.325BCE—c.265BCE). The following are amongst Euclid's axioms (we have not used the same terms Euclid used in his *Elements*, but commonly-used statements that are equivalent to Euclid's):

**Key Euclidean Axioms**

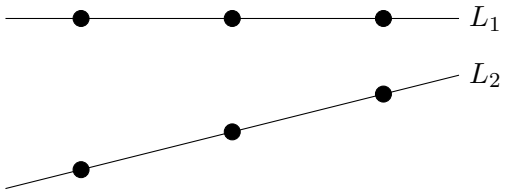
- Any two points determine (and so lie together on) a unique line.
- (Parallel postulate) For any line $L$, and any point $p$ that does not lie on the line $L$, there is a unique line $L'$ through $p$ that is parallel to $L$; that is, $L$ and $L'$ have no points in common.

If you haven't taken geometry classes in university, you may not know that we can apply these axioms to finite sets of points, and discover structures that we call *finite Euclidean geometries*, or more commonly, *affine planes*. To avoid some trivial situations, we also require that the structure has at least three points, and that not all of the points lie on a single line.

The following definition probably seems obvious.

**DEFINITION 18.3.1.** We say that two lines are **parallel** if no point lies on both lines.

We've made special note of this definition because in the finite case, "parallel" lines might be drawn in such a way that they don't look parallel according to our usual understanding of the term. Since each line has only a finite number of points on it, two lines $L_1$ and $L_2$ are parallel as long as none of the points on $L_1$ also lies on $L_2$, even if in a particular drawing in appears that these lines will meet if we extend them. In the figure below, lines $L_1$ and $L_2$ have three points each, and the lines are parallel.




**DEFINITION 18.3.2.** A (finite) **affine plane** consists of a (finite) set of points, a (finite) set of lines, and an incidence relation between the points and the lines. The incidence relation must satisfy these Euclidean axioms:

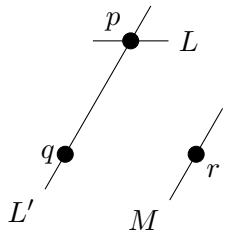
- Any two points lie together on a unique line.
- For any line $L$, and any point $p$ that does not lie on the line $L$, there is a unique line $L'$ that passes through $p$ and is disjoint from $L$ (that is, it is parallel to $L$).
- There are at least three points that are not all on the same line.

For the purposes of this book, we will only consider finite affine planes, so assume from now on that the set of points is finite. It is not very obvious, but the parallel postulate (together with the final axiom) ensure that it is not possible to have a line that doesn't contain any points, so the set of lines will also be finite.

There is a very nice bijective argument that can be used to show that the number of points on any two lines is equal. Before presenting this, we show that there cannot be a line that contains only one point.

**PROPOSITION 18.3.3.** *In a finite affine plane, no line contains only one point.*

**PROOF.** The following diagram may be a helpful visual aid as you read through the proof below.

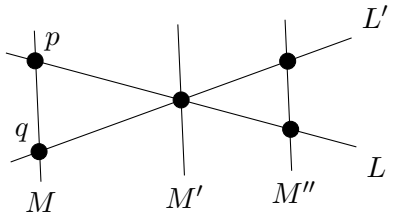



Towards a contradiction, suppose that there were a line $L$ that contained only a single point, $p$. By the final axiom, there are at least two other points in the finite affine plane, $q$ and $r$, that do not both lie on a line with $p$. By the first axiom, there is a line $L'$ that contains both $p$ and $q$ (but by our choice of $q$ and $r$, $L'$ does not contain $r$). Now by the parallel postulate, there is a line $M$ through $r$ that is parallel to $L'$. Furthermore, there is a unique line through $p$ that is parallel to $M$. But we know that $M$ is parallel to $L'$; in addition, $M$ does not contain $p$, so it is also parallel to $L$; this is a contradiction. □

We can now show that every line contains the same number of points.

**PROPOSITION 18.3.4.** *In a finite affine plane, if one line contains exactly $n$ points, then every line contains exactly $n$ points.*

**PROOF.** Again, we begin with a diagram that may be a helpful visual aid to understanding this proof.




Let $L$ be a line that contains exactly $n$ points, and let $L'$ be any other line of the finite affine plane. By Proposition 18.3.3, we know that $n > 1$, and that $L'$ also contains at least two points (we observed above that no line can be empty of points). Since any two points lie together on a unique line, if $L$ and $L'$ meet, they meet in a single point, so there is a point $p$ of $L$ that is not in $L'$, and a point $q$ of $L'$ that is not in $L$. By the first axiom, there is a line $M$ that contains the points $p$ and $q$.

We define a map $\psi$ from points of $L$ to points of $L'$ as follows. Let $\psi(p) = q$. For any other point $p'$ of $L$ (with $p' \neq p$), by the parallel postulate there is a unique line $M'$ through $p'$ that is parallel to $M$. Since $M$ passes through $q$ and is parallel to $M'$, it is the unique line with this property, so in particular, $L'$ cannot be parallel to $M'$. Therefore, $M'$ has a unique point of intersection, say $q'$, with $L'$. Define $\psi(p') = q'$. Since $M'$ and $q'$ were uniquely determined, the map $\psi$ is well-defined (that is, there is no ambiguity about which point of $L'$ is found to be $\psi(p')$).

We claim that $\psi$ is a bijection between the points of $L$ and the points of $L'$; proving this will complete the proof. We first show that $\psi$ is one-to-one. Suppose that $\psi(p_1) = q_1$, $\psi(p_2) = q_2$, and $q_1$ and $q_2$ are actually the same point of $L'$. Then by the definition of $\psi$, $q_1$ is on some line $M_1$ that is parallel to $M$, and contains $p_1$, while $q_2$ is on some line $M_2$ that is parallel to $M$, and contains $p_2$. Since $q_1 = q_2$, the parallel postulate tells us that we must have $M_1 = M_2$. This line can only meet $L$ in a single point, so we must have $p_1 = p_2$. Thus, $\psi$ is one-to-one.

To show that $\psi$ is onto, let $q''$ be any point of $L'$ with $q'' \neq q$ (we already know that $q$ has a pre-image, $p$). By the parallel postulate, there is a unique line $M''$ through $q''$ that is parallel to $M$. Since $M$ is the unique line through $p$ that is parallel to $M''$, we see that $L$ is not parallel to $M''$, so $L$ must meet $M''$ at some point that we will call $p''$. Now by the definition of $\psi$, we have $\psi(p'') = q''$. Thus, $q''$ has a pre-image, so $\psi$ is onto. □

We refer to the number of points on each line of a finite affine plane as the *order* of the plane. We can now figure out how many points are in a finite affine plane of order $n$.

**PROPOSITION 18.3.5.** *A finite affine plane of order $n$ has $n^2$ points.*

**PROOF.** Since the plane has at least three points, not all of which lie on the same line, it has at least two lines $L$ and $L'$ that intersect at a point $q$ but are not equal. We know that each of these lines contains $n$ points. By the parallel postulate, for each of the $n-1$ points on $L'$ that is not on $L$, there is a line through that point that is parallel to $L$. Now, $L$ and these $n-1$ lines that are parallel to $L$ each contain $n$ points, and since they are all parallel (it is an exercise, see below, to prove that if $M$ is parallel to $L$ and $N$ is parallel to $L$, then $N$ is parallel to $M$), these points are all distinct. Therefore the plane has at least $n^2$ points.

Consider any point $p$ that is not on $L$. By the parallel postulate, there must be a line $P$ through $p$ that is parallel to $L$. Now, $L$ is the unique line through $q$ that is parallel to $P$, so in particular, $L'$ is not parallel to $P$. Therefore, $L'$ and $P$ have a point of intersection, which is one of the $n-1$ points of $L'$ that is not on $L$. So $P$ was one of the $n-1$ lines that we found in our first paragraph, meaning that $p$ is one of the $n^2$ points that we found there. Thus, the plane has exactly $n^2$ points. □

You might be wondering by now why we are spending so much time looking at affine planes, when they are a geometric structure. Despite the fact that they come from geometry, finite affine planes can be thought of as a special kind of design.

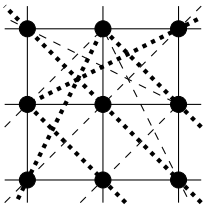
Think of the points of a finite affine plane as points of a design, and the lines as blocks, with a point being in a block if it is incident with (on) the corresponding line. The first axiom for the incidence relation guarantees that every pair of points appear together in exactly one block, so our design has $\lambda = 1$. By Proposition 18.3.4, we see that an affine plane of order $n$ is uniform, with $k = n$. By Proposition 18.3.5, an affine plane of order $n$ has $v = n^2$. Although we have not included a proof of this, it can also be shown that this design is regular, so it is in fact a BIBD$(n^2, n, 1)$.

Using

$$b\binom{k}{t} = \lambda\binom{v}{t}$$

from Theorem 18.2.2, we see that in a finite affine plane of order $n$ looked at as a design, we have $b\binom{n}{2} = \binom{n^2}{2}$, so $bn(n-1) = n^2(n^2-1)$. Hence $b = n(n+1)$. In other words, a finite affine plane of order $n$ has $n(n+1)$ lines.

**EXAMPLE 18.3.6.** A finite affine plane of order 3 has $3^2 = 9$ points. Each line has 3 points, and there are $3(4) = 12$ lines. Since each line has three points, every line lies in a parallel class consisting of three mutually parallel lines, so there are four such parallel classes. We can choose two of the parallel classes of lines to be "horizontal" and "vertical" lines. The other two classes will be the two types of diagonal lines. We can't draw all of these as straight lines, so we have drawn one parallel class of lines as sets of three points joined by dashes, and the other as sets of three points joined by dots.



The dashes and dots that join sets of points that aren't in a straight line may not provide a very clear image of what's going on; it is probably clearer to think of the diagonal lines as "wrapping around" when they go off the bottom, top, or either side of the image, and reappearing on the opposite side.

There is a nice connection between affine planes and mutually orthogonal Latin squares.

**THEOREM 18.3.7.** *There is an affine plane of order $n > 1$ if and only if there are $n-1$ mutually orthogonal Latin squares of order $n$.*

**PROOF. ($\Rightarrow$)** An affine plane of order $n$ has $n+1$ classes of parallel lines (each class containing $n$ lines). Consider two of these sets as the horizontal lines and the vertical lines: $H_1, \ldots, H_n$ and $V_1, \ldots, V_n$. Label a point as $(i, j)$ if it lies at the intersection of $V_i$ and $H_j$. Since each of the $n^2$ points lies on a vertical line and on a horizontal line, and since every pair of points lie together on only one line, this is actually a bijection between $\{1, \ldots, n\} \times \{1, \ldots, n\}$ and the points of the affine plane; that is, this provides $n^2$ coordinates that uniquely determine the $n^2$ points of the plane.

Consider any one of the remaining parallel classes of $n$ lines, $L_1, \ldots, L_n$. Observe that every point of the affine plane lies on precisely one of these lines. Create a Latin square from this parallel class by placing $k$ in position $(i, j)$ of the Latin square if and only if the point $(i, j)$ of the affine plane lies on line $L_k$. Since every line of $L_1, \ldots, L_n$ meets every line of $H_1, \ldots, H_n$ exactly once (by the axioms of an affine plane), each entry will appear exactly once in each row. Similarly, since every line of $L_1, \ldots, L_n$ meets every line of $V_1, \ldots, V_n$ exactly once (by the axioms of an affine plane), each entry will appear exactly once in each column. So we have indeed created a Latin square.

We will now show that the $n-1$ Latin squares created by this method (using the $n-1$ parallel classes of lines that remain after excluding the ones we have designated as horizontal and vertical lines) are mutually orthogonal. We'll do this by considering two arbitrary Latin squares, $L$ (coming from the lines $L_1, \ldots, L_n$) and $M$ (coming from the lines $M_1, \ldots, M_n$). In position $(i, j)$, the entry of $L$ being $i'$ means that line $L_{i'}$ passes through the point $(i, j)$ (which is the intersection of lines $V_i$ and $H_j$). Similarly, this entry of $M$ being $j'$ means that line $M_{j'}$ passes through the point $(i, j)$ (which is the intersection of lines $V_i$ and $H_j$). Since the lines $L_{i'}$ and $M_{j'}$ have a unique point of intersection, there cannot be any other positions in which the entry of $L$ is $i'$ while the entry of $M$ is $j'$. Thus, each ordered pair $(i', j') \in \{1, \ldots, n\} \times \{1, \ldots, n\}$

must appear in exactly one position as the entries of $L$ and $M$ (in that order), and hence $L$ and $M$ are orthogonal. Since they were arbitrary, we have $n-1$ mutually orthogonal Latin squares.

($\Leftarrow$) The converse of this proof uses the same idea, in the opposite direction. Given $n-1$ mutually orthogonal $n$ by $n$ Latin squares, take the $n^2$ coordinate positions to be the points of our affine plane. Define two parallel classes of lines (each containing $n$ lines) to be the points whose first coordinate is equal (so all of the points with first coordinate 1 form one line, and all of the points with first coordinate 2 form a second line, etc.), and the points whose second coordinate is equal. Each of the $n-1$ Latin squares determines an additional parallel class of $n$ lines: namely, each line consists of the points for which the entry of the Latin square has some fixed value. Since $n > 1$, there are clearly at least three points that are not all on the same line. We leave it as an exercise to prove that any two points lie together in a unique line, and that the parallel postulate is satisfied. □

**EXAMPLE 18.3.8.** Use the formula from the proof of Theorem 16.2.7 to construct 6 MOLS of order 7. Use the construction given in the proof of Theorem 18.3.7 to construct an affine plane of order 7 from your squares.

**SOLUTION.** The squares will be:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 2 | 3 | 4 | 5 | 6 | 0 | | 2 | 3 | 4 | 5 | 6 | 0 | 1 |
| 2 | 3 | 4 | 5 | 6 | 0 | 1 | | 4 | 5 | 6 | 0 | 1 | 2 | 3 |
| 3 | 4 | 5 | 6 | 0 | 1 | 2 | | 6 | 0 | 1 | 2 | 3 | 4 | 5 |
| 4 | 5 | 6 | 0 | 1 | 2 | 3 | | 1 | 2 | 3 | 4 | 5 | 6 | 0 |
| 5 | 6 | 0 | 1 | 2 | 3 | 4 | | 3 | 4 | 5 | 6 | 0 | 1 | 2 |
| 6 | 0 | 1 | 2 | 3 | 4 | 5 | | 6 | 0 | 1 | 2 | 3 | 4 | 5 |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 3 | 4 | 5 | 6 | 0 | 1 | 2 | | 4 | 5 | 6 | 0 | 1 | 2 | 3 |
| 6 | 0 | 1 | 2 | 3 | 4 | 5 | | 1 | 2 | 3 | 4 | 5 | 6 | 0 |
| 2 | 3 | 4 | 5 | 6 | 0 | 1 | | 5 | 6 | 0 | 1 | 2 | 3 | 4 |
| 5 | 6 | 0 | 1 | 2 | 3 | 4 | | 2 | 3 | 4 | 5 | 6 | 0 | 1 |
| 1 | 2 | 3 | 4 | 5 | 6 | 0 | | 6 | 0 | 1 | 2 | 3 | 4 | 5 |
| 4 | 5 | 6 | 0 | 1 | 2 | 3 | | 3 | 4 | 5 | 6 | 0 | 1 | 2 |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 5 | 6 | 0 | 1 | 2 | 3 | 4 | | 6 | 0 | 1 | 2 | 3 | 4 | 5 |
| 3 | 4 | 5 | 6 | 0 | 1 | 2 | | 5 | 6 | 0 | 1 | 2 | 3 | 4 |
| 1 | 2 | 3 | 4 | 5 | 6 | 0 | | 4 | 5 | 6 | 0 | 1 | 2 | 3 |
| 6 | 0 | 1 | 2 | 3 | 4 | 5 | | 3 | 4 | 5 | 6 | 0 | 1 | 2 |
| 4 | 5 | 6 | 0 | 1 | 2 | 3 | | 2 | 3 | 4 | 5 | 6 | 0 | 1 |
| 2 | 3 | 4 | 5 | 6 | 0 | 1 | | 1 | 2 | 3 | 4 | 5 | 6 | 0 |

The affine plane will have $7^2 = 49$ points, and we will denote these as ordered pairs $(a, b)$, where $a, b \in \{1, \ldots, 7\}$ and consider them to represent the 49 positions in a 7 by 7 Latin square. There will be $7(8) = 56$ lines, in 8 parallel classes of seven lines each. Although we could draw the affine plane, you've already seen from the affine plane of order 3 that a black-and-white image all of which is pre-drawn can be more confusing than helpful, so instead we will list each of the 56 lines as a set of 7 points.

The first parallel class will represent the horizontal rows:

$\{(1,1),(2,1),(3,1),(4,1),(5,1),(6,1),(7,1)\}$, $\{(1,2),(2,2),(3,2),(4,2),(5,2),(6,2),(7,2)\}$,
$\{(1,3),(2,3),(3,3),(4,3),(5,3),(6,3),(7,3)\}$, $\{(1,4),(2,4),(3,4),(4,4),(5,4),(6,4),(7,4)\}$,
$\{(1,5),(2,5),(3,5),(4,5),(5,5),(6,5),(7,5)\}$, $\{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6),(7,6)\}$,
$\{(1,7),(2,7),(3,7),(4,7),(5,7),(6,7),(7,7)\}$

and similarly the second parallel class will represent the vertical rows:

$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(1,7)\}$, $\{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(2,7)\}$,
$\{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(3,7)\}$, $\{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(4,7)\}$,
$\{(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(5,7)\}$, $\{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6),(6,7)\}$,
$\{(7,1),(7,2),(7,3),(7,4),(7,5),(7,6),(7,7)\}$.

The remaining six parallel classes will each represent one of the Latin squares. In the next parallel class, the first line consists of all of the points where the entry of the first Latin square is 0; the second consists of all the points where the entry is 1, and so on.

$\{(1,1),(7,2),(6,3),(5,4),(4,5),(3,6),(2,7)\}$, $\{(2,1),(1,2),(7,3),(6,4),(5,5),(4,6),(3,7)\}$,
$\{(3,1),(2,2),(1,3),(7,4),(6,5),(5,6),(4,7)\}$, $\{(4,1),(3,2),(2,3),(1,4),(7,5),(6,6),(5,7)\}$,
$\{(5,1),(4,2),(3,3),(2,4),(1,5),(7,6),(6,7)\}$, $\{(6,1),(5,2),(4,3),(3,4),(2,5),(1,6),(7,7)\}$,
$\{(7,1),(6,2),(5,3),(4,4),(3,5),(2,6),(1,7)\}$.

The next parallel class comes from the second Latin square (reading across, so the second square is the second one in the first line):

$\{(1,1),(6,2),(4,3),(2,4),(7,5),(5,6),(3,7)\}$, $\{(2,1),(7,2),(5,3),(3,4),(1,5),(6,6),(4,7)\}$,
$\{(3,1),(1,2),(6,3),(4,4),(2,5),(7,6),(5,7)\}$, $\{(4,1),(2,2),(7,3),(5,4),(3,5),(1,6),(6,7)\}$,
$\{(5,1),(3,2),(1,3),(6,4),(4,5),(2,6),(7,7)\}$, $\{(6,1),(4,2),(2,3),(7,4),(5,5),(3,6),(1,7)\}$,
$\{(7,1),(5,2),(3,3),(1,4),(6,5),(4,6),(2,7)\}$.

From the third Latin square:

$\{(1,1),(5,2),(2,3),(6,4),(3,5),(7,6),(4,7)\}$, $\{(2,1),(6,2),(3,3),(7,4),(4,5),(1,6),(5,7)\}$,
$\{(3,1),(7,2),(4,3),(1,4),(5,5),(2,6),(6,7)\}$, $\{(4,1),(1,2),(5,3),(2,4),(6,5),(3,6),(7,7)\}$,
$\{(5,1),(2,2),(6,3),(3,4),(7,5),(4,6),(1,7)\}$, $\{(6,1),(3,2),(7,3),(4,4),(1,5),(5,6),(2,7)\}$,
$\{(7,1),(4,2),(1,3),(5,4),(2,5),(6,6),(3,7)\}$.

From the fourth Latin square:

$\{(1,1),(4,2),(7,3),(3,4),(6,5),(2,6),(5,7)\}$, $\{(2,1),(5,2),(1,3),(4,4),(7,5),(3,6),(6,7)\}$,
$\{(3,1),(6,2),(2,3),(5,4),(1,5),(4,6),(7,7)\}$, $\{(4,1),(7,2),(3,3),(6,4),(2,5),(5,6),(1,7)\}$,
$\{(5,1),(1,2),(4,3),(7,4),(3,5),(6,6),(2,7)\}$, $\{(6,1),(2,2),(5,3),(1,4),(4,5),(7,6),(3,7)\}$,
$\{(7,1),(3,2),(6,3),(2,4),(5,5),(1,6),(4,7)\}$.

From the fifth Latin square:

$\{(1,1),(3,2),(5,3),(7,4),(2,5),(4,6),(6,7)\}$, $\{(2,1),(4,2),(6,3),(1,4),(3,5),(5,6),(7,7)\}$,
$\{(3,1),(5,2),(7,3),(2,4),(4,5),(6,6),(1,7)\}$, $\{(4,1),(6,2),(1,3),(3,4),(5,5),(7,6),(2,7)\}$,
$\{(5,1),(7,2),(2,3),(4,4),(6,5),(1,6),(3,7)\}$, $\{(6,1),(1,2),(3,3),(5,4),(7,5),(2,6),(4,7)\}$,
$\{(7,1),(2,2),(4,3),(6,4),(1,5),(3,6),(5,7)\}$.

And finally,

$\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6),(7,7)\}$, $\{(2,1),(3,2),(4,3),(5,4),(6,5),(7,6),(1,7)\}$,
$\{(3,1),(4,2),(5,3),(6,4),(7,5),(1,6),(2,7)\}$, $\{(4,1),(5,2),(6,3),(7,4),(1,5),(2,6),(3,7)\}$,
$\{(5,1),(6,2),(7,3),(1,4),(2,5),(3,6),(4,7)\}$, $\{(6,1),(7,2),(1,3),(2,4),(3,5),(4,6),(5,7)\}$,
$\{(7,1),(1,2),(2,3),(3,4),(4,5),(5,6),(6,7)\}$.

comes from the sixth Latin square. □

Every affine plane that we know of, has as its order some prime power. We have previously seen (through the connection to MOLS) that there are affine planes of every prime order (in fact, we mentioned without proof that this is true for every prime power order). Many design theorists have tried to answer the question of whether or not the order of an affine plane must always be a prime power, but the answer is not yet known. It was only in 1989 that the nonexistence of an affine plane of order 10 was proven. In fact, it is not currently known whether or not there is an affine plane of order 12. (See Section 16.2 for more about these facts.)

**EXERCISES 18.3.9.**

1) Prove that if $L$, $M$, and $N$ are lines of an affine plane, and $L$ is parallel to both $M$ and $N$, then $M$ is parallel to $N$.
2) Draw a finite affine plane of order 5. How many lines does it have?
3) How many points, and how many lines are in a finite affine plane of order 19?
4) Prove the omitted details from the proof of Theorem 18.3.7: that is, that the given construction yields a structure that satisfies the axioms of an affine plane.
5) Draw an affine plane of order 5. Use the construction given in the proof of Theorem 18.3.7 to produce 4 mutually orthogonal Latin squares of order 5 from your plane.

## 18.4. Projective planes

A projective plane is another geometric structure (closely related to affine planes). In a finite projective plane, the set of points (and therefore the set of lines) must be finite. Like finite affine planes, finite projective planes can be thought of as a special kind of design.

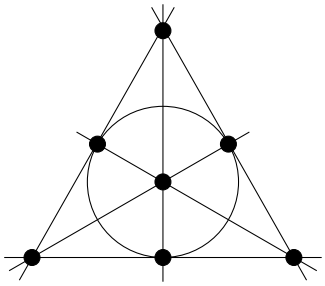
**DEFINITION 18.4.1.** A **projective plane** consists of a set of points, a set of lines, and an incidence relation between the points and the lines. The incidence relation must satisfy the following conditions:

- for any two points, there is a unique line that is incident with both of them;
- for any two lines, there is a unique point that is incident with both of them;
- there exist four points such that no three are incident with a single line.

As in the case of affine planes, the final axiom has been developed to avoid some trivial situations.

Think of the points of a finite projective plane as points of a design, and the lines as blocks, with a point being in a block if it is incident with the corresponding line. Then the first condition on the incidence relation for a projective plane guarantees that every pair of points appear together in exactly one block.

**EXAMPLE 18.4.2.** The Fano plane, named for Gino Fano (1871—1952), is the most well-known finite projective plane (and also the smallest). Here is a drawing of it. It has 7 points and 7 lines, one of which is the circle around the middle.



You have seen this structure already in this course; it is the same as the BIBD$(7, 3, 1)$ that appeared in Example 17.1.4.

The following is a very interesting connection. We will not try to present the proof here, but it is a natural extension of the similar result that we proved for affine planes.

**THEOREM 18.4.3.** *There is a finite projective plane with $n+1$ points on each line, if and only if there is a complete set of $n-1$ MOLS of order $n$.*

**EXERCISES 18.4.4.**

1) Is every design with $\lambda = 1$ a projective plane? If not, what condition could fail?
2) Which (if any) of the designs we have seen in this course, are projective planes?
3) From our results on MOLS, for what values can you be sure that a projective plane exists?
4) From our results on MOLS, for what values can you be sure that a projective plane does not exist?
5) What can you determine about the parameters of a design that corresponds to a projective plane?

---

## SUMMARY:

- construction of Steiner triple systems
- structure of affine planes
- connection between affine planes and MOLS
- Important definitions:
  - triple system, Steiner triple system
  - resolvable design, Kirkman triple system
  - $t$-design
  - affine plane
  - projective plane
- Notation:
  - STS$(v)$

---

# Chapter 19

# Designs and Codes

## 19.1. Introduction

When information is transmitted, it may get garbled along the way. Error-correcting codes can make it possible for the recipient of a garbled message to figure out what the sender intended to say.

**ASSUMPTION 19.1.1.** For definiteness, we assume the message to be sent is a string (or "word") of **bits** (0s and 1s). (Information stored in a computer is always converted to such a string, so this is not a serious limitation.)

**EXAMPLE 19.1.2.** Perhaps the word 0110 tells an automated factory to close the 2nd and 3rd valves. If we send that message over a wireless network, interference (or some other issue) might change one of the bits, so the factory receives the message 0010. As a result, the factory closes only the 3rd valve, and leaves the 2nd valve open. This could have disastrous consequences, so we would like to do something to avoid such problems.

**EXAMPLE 19.1.3.** One simple solution is to append a *check-bit* to the end of the message. To do this, we set three rules:

1) We require all messages to have a certain length. (For example, let's say that all messages must have exactly 5 bits.)
2) We require all messages to have an even number of 1s.
3) We agree that the final bit of the message (called a "parity check-bit" ) will not convey any information, but will be used only to guarantee that Rule 2 is obeyed. (Thus, each message we send will have 4 bits of information, plus the check bit.)

In particular, if we wish to send the message 0110 (which already has an even number of 1s), then we append 0 to the end, and send the message 01100. If, say, the 2nd bit gets changed in transmission, so the factory receives the message 00100, then the factory's computer control can see that this cannot possibly be the intended message, because it has an odd number of 1s. So the factory can return an error message, asking us to send our instructions again.

*Remark 19.1.4.* As a real-life example, bar-code scanners used by cashiers employ the above principle: if the check-bit is not correct, then the scanner does not beep, so the cashier knows that the item needs to be rescanned.

**EXERCISE 19.1.5.** Under the rules of Example 19.1.3, which of the following strings are allowed to be sent as a message?

$$00110,\ 10101,\ 00000,\ 11011.$$

It is sometimes not feasible to have a message re-sent (for example, if it spends a long time in transit), so it would be much better to have a system that enables the recipient to *correct* transmission errors, not just *detect* them.

**EXAMPLE 19.1.6 (Triple-repetition code).** We could agree to send each bit of our message 3 times. For example, if we want to send the message 0110, then we would transform (or "encode" ) it as 000111111000. If, say, the 2nd bit gets garbled, so the factory receives 010111111000, then it knows there was a problem in the transmission of the first 3 bits, because they are not all the same. Furthermore, since most of these bits are 0, the factory can figure out that we probably meant to say 000, and correctly decode the entire message as 0110.

The triple-repetition code works, but it is very inefficient, because only one-third of the bits we send are conveying information — most of the bits are check-bits that were added to correct the possible errors. After developing some theory, we will see some codes that are able to correct every single-bit error, but use far fewer check bits.

**EXERCISE 19.1.7.** Let $\mathbb{T}^n$ be the set of all ternary sequences of length $n$ (so every entry is 0, 1, or 2). Write a recurrence relation for $c_n$, the number of code words from $\mathbb{T}^n$ that have no 2 consecutive zeros. Use generating functions to solve your recurrence relation, deriving an explicit formula for $c_n$.

## 19.2. Error-correcting codes

In order to be able to correct errors in transmission, we agree to send only strings that are in a certain set $\mathcal{C}$ of **codewords**. (So the information we wish to send will need to be "encoded" as one of the codewords.) In our above examples, $\mathcal{C}$ was

- the set of all words of length 5 that have an even number of 1s, or
- the set of words of length 12 that consist of four strings of three equal bits.

The set $\mathcal{C}$ is called a **code**. Choosing the code cleverly will enable us to successfully correct transmission errors.

When a transmission is received, the recipient will assume that the sender transmitted the codeword that is "closest" to the string that was received. For example, if $\mathcal{C}$ were the set of 5-letter words in the English language, and the string "*fruiz*" were received, then the recipient would assume (presumably correctly), that the sender transmitted the word "*fruit*", because that is only off by one letter. (This is how we ordinarily deal with the typographical errors that we encounter.)

By the "closest" codeword, we mean the codeword at the smallest distance, in the following sense:

**DEFINITION 19.2.1.** Suppose $x$ and $y$ are two bit strings of the same length. The **Hamming distance** from $x$ to $y$ (denoted $d(x, y)$) is the number of bits in which they differ.

The Hamming distance takes its name from Richard Wesley Hamming (1915—1998).

**EXAMPLE 19.2.2.** For clarity, we underline the bits in the second string that differ from the corresponding bit in the first string:

- $d(11111, 111\underline{0}\underline{1}) = 1$
- $d(11100, \underline{0}1\underline{0}0\underline{1}) = 3$
- $d(10101, \underline{01010}) = 5$
- $d(10101, 10101) = 0$

*Remark 19.2.3.* When two bits are "transposed" (or "switched"), meaning that a string 01 gets changed to 10 (or vice-versa), this counts as two bits being different, because a 0 is changed to a 1 and a 1 is changed to a 0, even though you might think of the switch as being only a single operation.

**EXERCISE 19.2.4.** Compute the Hamming distance between the following pairs of words:

$$\{110, 011\},\ \{000, 010\},\ \{\text{brats}, \text{grass}\},\ \{11101, 00111\}.$$

**EXERCISE 19.2.5.** Find each of the following:

1) an English word whose Hamming distance from "math" is 0;
2) an English word whose Hamming distance from "math" is 1;
3) an English word whose Hamming distance from "math" is 2;
4) an English word whose Hamming distance from "math" is 3.
5) Can you find more than one of each?

The Hamming distance satisfies the axioms of a "metric" or "distance function":

**EXERCISE 19.2.6.** Prove that the Hamming function satisfies each of the following properties, which define a metric.

Let $x$, $y$, and $z$ be words of the same length. Then:

1) $d(x, y) \geq 0$.
2) $d(x, y) = 0 \iff x = y$.
3) $d(x, y) = d(y, x)$.
4) $d(x, z) \leq d(x, y) + d(y, z)$.

*Remark 19.2.7.* Exercise 19.2.6.4 is called the **triangle inequality**, because it says that the length of one side of a triangle is always less than (or equal to) the sum of the lengths of the other two sides.

**DEFINITION 19.2.8.** The **minimum distance** of a code $\mathcal{C}$ (denoted $d(\mathcal{C})$) is the smallest Hamming distance between two distinct elements of $\mathcal{C}$.

**EXAMPLE 19.2.9.**

1) $d\big(\{100, 010, 011\}\big) = 1$ (because $d(010, 01\underline{1}) = 1$).
2) $d\big(\{000, 011, 101, 110\}\big) = 2$.
3) $d\big(\{10001, 11111\}\big) = 3$.

**EXERCISE 19.2.10.** What is the minimum distance of each of the following codes?

1) { tell , tale , sale , date }
2) { mon , tue , wed , thu , fri , sat , sun }
3) $\{00000, 01011, 10101, 10110, 10011\}$

Making the minimum distance of a code $\mathcal{C}$ large is the key to ensuring that it can detect (or correct) large errors. We will make this idea very explicit in our next two results:

**THEOREM 19.2.11.** *A code $\mathcal{C}$ can* detect *all possible errors affecting at most $k$ bits if and only if $d(\mathcal{C}) > k$.*

**PROOF.** We will prove the contrapositive:

$$d(\mathcal{C}) \leq k \iff \begin{array}{c}\text{there exists an error that cannot be detected,}\\ \text{and affects only } k \text{ (or fewer) bits.}\end{array}$$

(⇐) Suppose there is a situation in which:

- a codeword $x$ is sent,
- a message $y$ is received,
- with only $k$ incorrect bits, and
- the receiver does not realize that there were any errors.

Since the receiver did not realize there were any errors, the message that was received must be a codeword. In other words, $y \in \mathcal{C}$. Since there are $k$ errors in the received message, we also know that $d(x, y) = k$. Since $x, y \in \mathcal{C}$, this implies $d(\mathcal{C}) \leq k$.

(⇒) By assumption, there exist $x, y \in \mathcal{C}$ with $x \neq y$, but $d(x, y) \leq k$. Now, suppose codeword $x$ is sent. Since $d(x, y) \leq k$, changing $k$ (or fewer) bits can change $x$ to $y$, so $y$ can be the message that is received, even if errors in transmission affect only $k$ bits. Since $y \in \mathcal{C}$, the recipient does not realize an error was made, and assumes that $y$ was the intended message. So the $k$ (or fewer) errors were not detected. □

Although a minimum distance of $k$ allows us to detect errors that affect at most $k$ bits, it isn't sufficient to allow us to correct all such errors. For the purposes of correcting errors, we require the minimum distance to be twice this large.

**THEOREM 19.2.12.** *A code $\mathcal{C}$ can* correct *all possible errors affecting at most $k$ bits if and only $d(\mathcal{C}) > 2k$.*

**PROOF.** We will prove the contrapositive:

$$d(\mathcal{C}) \leq 2k \iff \begin{array}{c}\text{there exists an error that is not properly corrected,}\\ \text{and affects only } 2k \text{ (or fewer) bits.}\end{array}$$

(⇐) Suppose there is a situation in which:

- a codeword $x$ is sent,
- a message $y$ is received,
- with only $2k$ incorrect bits, and
- the receiver decodes the message incorrectly, as some codeword $z \neq x$.

It must be the case that $z$ is the closest codeword to $y$ (or, at least, it ties for being the closest), so $d(z, y) \leq d(x, y) = k$. Then, using Exercise 19.2.6, we have

$$d(x, z) \leq d(x, y) + d(y, z) = d(x, y) + d(z, y) \leq k + k = 2k.$$

So $d(\mathcal{C}) \leq 2k$ (because $x, z \in \mathcal{C}$).

(⇒) By assumption, there exist $x, y \in \mathcal{C}$ with $x \neq y$, but $d(x, y) \leq 2k$. Let $r = \lceil d(x, y)/2 \rceil \leq \lceil 2k/2 \rceil = k$. (In other words, $r$ is obtained by rounding $d(x, y)/2$ *up* to the nearest integer.)

Now suppose codeword $x$ is sent. Since $d(x, y) \leq 2k$, the message $y$ could be received, with no more than $2k$ incorrect bits. Construct $z$ from $x$ by changing only $r$ of the $d(x, y)$ bits that are incorrect in $y$, so

$$d(x, z) = r \text{ and } d(z, y) = d(x, y) - r.$$

By the definition of $r$, we have $r \leq d(x, y) \leq 2r$, so

$$d(z, y) = d(x, y) - r \leq 2r - r = r \leq d(x, y).$$

Therefore $z$ is at least as close to $y$ as $x$ is, so the recipient has no way of knowing that $x$ is the message that was sent. So it was not possible to correct the $2k$ (or fewer) errors. □

**EXERCISES 19.2.13.**

1) Suppose that a code $C$ has a minimum distance of 7.
   (a) How many errors can it detect?
   (b) How many errors can it correct?
2) Suppose that a code $C$ has a minimum distance of 6.
   (a) How many errors can it detect?
   (b) How many errors can it correct?

**EXERCISE 19.2.14.** Let $\mathbb{B}^n$ represent the set of binary strings of length $n$. Prove that a code from $\mathbb{B}^{10}$ that has more than 2 words, cannot correct 3 errors. Hypothesize a generalisation of this result to codes on $\mathbb{B}^n$ with more than 2 words.

## 19.3. Using the generator matrix for encoding

This section and Section 19.4 do assume some familiarity with elementary linear algebra: specifically, knowing what vectors and matrices are, and being able to perform matrix multiplication. These two sections can be omitted by anyone who does not have this background, without affecting your understanding of the rest of the book.

**NOTATION 19.3.1.** It is convenient to represent the binary string $x_1x_2 \dots x_n$ as a **column vector**:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

This allows us to use matrix multiplication to append check bits to a string:

**EXAMPLE 19.3.2.** Appending a parity check-bit to the string 010 yields 0101. The same result can be obtained by multiplying the column vector corresponding to 010 by the following **generator matrix**:

$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} I_3 \\ \hline A \end{bmatrix} \text{ where } I_k \text{ is the } k \times k \text{ identity matrix, and } A = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}.$$

Namely (performing all arithmetic modulo 2), we have

$$G \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

In fact, multiplying any 3-bit string by $G$ yields the same string with its parity check-bit appended.

**PROOF.** We see that $G\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \begin{bmatrix}x_1\\x_2\\x_3\\x_1+x_2+x_3\end{bmatrix}$, and $x_1+x_2+x_3 \pmod 2$ is 0 if there are an even number of 1s among $x_1$, $x_2$, and $x_3$, and it is 1 if there are an odd number of 1s among $x_1$, $x_2$, and $x_3$. □

**GENERAL METHOD.** For some $k, r \in \mathbb{N}^+$,

- choose an $r \times k$ matrix $A$ of 0s and 1s, and
- let $G = \left[\frac{I_k}{A}\right]$.

Multiplying a $k$-bit string by $G$ yields the same string, with $r$ check bits appended at the end. We let $\mathcal{C}$ be the set of all possible strings $Gx$, and we call $G$ the **generator matrix** of this code.

*Remark 19.3.3.* In the next section, we will see how to choose $G$ so that the resulting code $\mathcal{C}$ can correct errors.

Although many important error-correcting codes are constructed by other methods, we will only discuss the ones that come from generator matrices (except in Section 19.5).

**DEFINITION 19.3.4.** Any code that comes from a generator matrix $G$ (by the General Method described above) is said to be a **binary linear code**.

**EXAMPLE 19.3.5.** Find all the codewords of the binary linear code $\mathcal{C}$ corresponding to the generator matrix

$$G = \left[\frac{I_3}{A}\right], \text{ with } A = \begin{bmatrix}1&0&1\\0&1&1\end{bmatrix}.$$

**SOLUTION.** We have

$$G = \left[\frac{I_3}{A}\right] = \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\1&0&1\\0&1&1\end{bmatrix}.$$

We use this matrix to encode each of the $2^3 = 8$ binary strings of length 3:

$$G\begin{bmatrix}0\\0\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix}, \quad G\begin{bmatrix}0\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\1\\1\\1\end{bmatrix}, \quad G\begin{bmatrix}0\\1\\0\end{bmatrix} = \begin{bmatrix}0\\1\\0\\0\\1\end{bmatrix}, \quad G\begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}0\\1\\1\\1\\0\end{bmatrix},$$

$$G\begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}1\\0\\0\\1\\0\end{bmatrix}, \quad G\begin{bmatrix}1\\0\\1\end{bmatrix} = \begin{bmatrix}1\\0\\1\\0\\1\end{bmatrix}, \quad G\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}1\\1\\0\\1\\1\end{bmatrix}, \quad G\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}1\\1\\1\\0\\0\end{bmatrix}.$$

So $\mathcal{C} = \{00000, 00111, 01001, 01110, 10010, 10101, 11011, 11100\}$. □

**EXERCISES 19.3.6.** Encode each of the given words by using the generating matrix $G = \begin{bmatrix} I_k \\ A \end{bmatrix}$ associated to the given matrix $A$.

1) $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$. Words to encode: 0101, 0010, 1110.

2) $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. Words to encode: 110, 011, 111, 000.

The generator matrix provides an easy way to *encode* messages for sending, but it is hard to use it to *decode* a message that has been received. For that, the next section will introduce a slightly different matrix. From this new matrix, it will be easy to determine whether the corresponding code can correct every single-bit error.

## 19.4. Using the parity-check matrix for decoding

**NOTATION 19.4.1.** A binary linear code is of **type $(n, k)$** (or we say $\mathcal{C}$ is an **$(n, k)$ code**) if its generator matrix $G = \begin{bmatrix} I_k \\ A \end{bmatrix}$ is an $n \times k$ matrix. In other words, $G$ encodes messages of length $k$ as codewords of length $n$, which means that the number of check bits is $n-k$. We usually use $r$ to denote the number of check bits, so $r = n - k$. Then $A$ is an $r \times k$ matrix.

**EXERCISE 19.4.2.** How many codewords are there in a binary linear code of type $(n, k)$?

**DEFINITION 19.4.3.** If $G = \begin{bmatrix} I_k \\ A \end{bmatrix}$ is the generator matrix of a binary linear code $\mathcal{C}$, and $A$ is an $r \times k$ matrix (so $\mathcal{C}$ is of type $(k + r, k)$), then the **parity-check matrix** of $\mathcal{C}$ is

$$P = [A\ I_r].$$

**EXAMPLE 19.4.4.**

1) For the code $\mathcal{C}$ of Example 19.3.5, the matrix $A$ is $2 \times 3$, so $r = 2$. Therefore, the parity-check matrix of $\mathcal{C}$ is

$$P = [A\ I_r] = [A\ I_2] = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

2) For a single parity check-bit, as in Example 19.3.2, we have $A = [1\ 1\ 1]$. This is a $1 \times 3$ matrix, so $r = 1$. Therefore, the parity-check matrix of the code is

$$P = [A\ I_r] = [A\ I_1] = [1\ 1\ 1\ 1]$$

(since $I_1 = [1]$).

**EXERCISE 19.4.5.** Suppose the generator matrix of the binary linear code $\mathcal{C}$ is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}.$$

What is the parity-check matrix of this code?

The parity-check matrix can be used to check whether a message we received is a valid codeword:

**PROPOSITION 19.4.6.** *A column vector $x$ is a codeword if and only if $Px = 0$.*

**PROOF.** ($\Rightarrow$) Since $x$ is a codeword, we have $x = Gm$ for some ($k$-bit) message $m$. This means that

$$x = Gm = \begin{bmatrix} I_k \\ A \end{bmatrix} m = \begin{bmatrix} m \\ Am \end{bmatrix}.$$

Then

$$Px = [A\ I_r] \begin{bmatrix} m \\ Am \end{bmatrix} = \begin{bmatrix} Am + Am \end{bmatrix} = \begin{bmatrix} 2Am \end{bmatrix} \equiv 0 \pmod 2.$$

($\Leftarrow$) Suppose $Px = 0$. Write $x = \begin{bmatrix} m \\ y \end{bmatrix}$, where

- $m$ is the first $k$ rows of $x$, and
- $y$ is the remaining $r = n - k$ rows of $x$.

Then

$$0 = Px = [A\ I_r] \begin{bmatrix} m \\ y \end{bmatrix} = \begin{bmatrix} Am + y \end{bmatrix}.$$

This means $y = -Am = Am \pmod 2$, so

$$x = \begin{bmatrix} m \\ y \end{bmatrix} = \begin{bmatrix} m \\ Am \end{bmatrix} = \begin{bmatrix} I_k \\ A \end{bmatrix} m = Gm,$$

so $x \in \mathcal{C}$. □

**EXAMPLE 19.4.7.** Here is a simple illustration of Proposition 19.4.6. For the code in which every codeword is required to have an even number of 1s, Example 19.4.4.2 tells us that the parity-check matrix is $P = [1\ 1\ 1\ 1]$. Hence, for any 4-bit string $x_1x_2x_3x_4$, we have

$$Px = [1\ 1\ 1\ 1] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = [x_1 + x_2 + x_3 + x_4].$$

This is 0 (mod 2) if and only if there are an even number of 1s in $x$, which is what it means to say that $x$ is a codeword.

**EXAMPLE 19.4.8.** Use the parity-check matrix to determine whether each of these words is in the code $\mathcal{C}$ of Example 19.3.5:

$$11111,\ 10101,\ 00000,\ 11010.$$

**SOLUTION.** From Example 19.4.4.1, we know that the parity-check matrix of this code is

$$P = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

We have:

- $P \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 0\cdot 1 + 1\cdot 1 + 1\cdot 1 + 0\cdot 1 \\ 0\cdot 1 + 1\cdot 1 + 1\cdot 1 + 0\cdot 1 + 1\cdot 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, so 11111 is not a codeword.
- $P \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 0\cdot 0 + 1\cdot 1 + 1\cdot 0 + 0\cdot 1 \\ 0\cdot 1 + 1\cdot 0 + 1\cdot 1 + 0\cdot 0 + 1\cdot 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, so 10101 is a codeword.
- $P \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1\cdot 0 + 0\cdot 0 + 1\cdot 0 + 1\cdot 0 + 0\cdot 0 \\ 0\cdot 0 + 1\cdot 0 + 1\cdot 0 + 0\cdot 0 + 1\cdot 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, so 00000 is a codeword.
- $P \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 0\cdot 1 + 1\cdot 0 + 1\cdot 1 + 0\cdot 0 \\ 0\cdot 1 + 1\cdot 1 + 1\cdot 0 + 0\cdot 1 + 1\cdot 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, so 11010 is not a codeword.

(These answers can be verified by looking at the list the elements of $\mathcal{C}$ in the solution of Example 19.3.5.) □

It is evident from the parity-check matrix whether a code corrects every single-bit error:

**THEOREM 19.4.9.** *A binary linear code $\mathcal{C}$ can correct every single-bit error if and only if the columns of its parity-check matrix are all distinct and nonzero.*

**PROOF.** Suppose a codeword $x$ is transmitted, but the $i$th bit gets changed, so a different string $y$ is received. Let $e_i$ be the string that is all 0s, except that the $i$th bit is 1, so $y = x + e_i$. Then

$$Py = P(x + e_i) = Px + Pe_i = 0 + Pe_i = Pe_i$$

is the $i$th column of $P$.

Therefore, if all the columns of $P$ are nonzero, then $Py$ is nonzero, so the receiver can detect that there was an error. If, in addition, all of the columns of $P$ are distinct, then $Py$ is equal to the $i$th column of $P$, and not equal to any other column, so the receiver can conclude that the error is in the $i$th bit. Changing this bit corrects the error.

Conversely, if either the $i$th column of $P$ is zero, or the $i$th column is equal to the $j$th column, then either $Pe_i = 0$ or $Pe_i = Pe_j$. Therefore, when the codeword $00\ldots0$ is sent, and an error changes the $i$th bit, resulting in the message $e_i$ being received, either $Pe_i = 0$, so the receiver does not detect the error (and erroneously concludes that the message $e_i$ is what was sent), or cannot tell whether the error is in the $i$th bit (and message 0 was sent) or the error is in the $j$ th bit (and message $e_i + e_j$ was sent). In either case, this is a single-bit error that cannot be corrected. □

**EXERCISE 19.4.10.** The parity-check matrix of the binary linear code $\mathcal{C}$ is

$$P = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Can $\mathcal{C}$ correct all single-bit errors?

The proof of Theorem 19.4.9 shows how to correct any single-bit error (when it is possible):

**GENERAL METHOD.** Assume the word $y$ has been received. Calculate $Py$.

- If $Py = 0$, then $y$ is a codeword. Assume there were no errors, so $y$ is the codeword that was sent.
- Now suppose $Py \neq 0$.
  - If $Py$ is equal to the $i$th column of $P$, then let $x = y + e_i$. (In other words, create $x$ by changing the $i$th bit of $y$ from 0 to 1 or vice-versa.) Then $x$ is a codeword. Assume it is the codeword that was sent.
  - If $Py$ is not equal to any of the columns of $P$, then at least two of the bits of $y$ are wrong. Do not try to correct the error.

**EXAMPLE 19.4.11.** Suppose the parity-check matrix of a binary linear code is

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Decode each of the following received words:

$$111000,\ 101001,\ 001101.$$

**SOLUTION.** Let $P$ be the given parity-check matrix. Then:

- $P \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. This is the 4th column of $P$, so changing the 4th bit corrects the error. This means that the received word 111000 decodes as 111100.

- $P \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. This is 0, so there is no error. This means that the received word 101001 decodes as 101001.

- $P\begin{bmatrix}0\\0\\1\\1\\0\\1\end{bmatrix} = \begin{bmatrix}0\\1\\1\end{bmatrix}$. This is not any of the columns of $P$, so there are at least two errors. Therefore, we cannot decode the received word 001101.

**EXERCISES 19.4.12.**

1) The parity-check matrix of a certain binary linear code is

$$\begin{bmatrix}1&0&1&0&1&0&0&0\\1&0&0&1&0&1&0&0\\0&1&0&1&0&0&1&0\\0&1&1&1&0&0&0&1\end{bmatrix}.$$

(a) Decode each of the following received words: 10001111, 11110000, 01111101.

(b) Find the generator matrix of the code.

2) The parity check matrix of a certain binary linear code is

$$\begin{bmatrix}1&1&0&1&0&0\\0&1&1&0&1&0\\1&0&1&0&0&1\end{bmatrix}.$$

(a) Can the code correct all single-bit errors?

(b) Decode each of the following received words: 001001, 110011, 000110.

**EXAMPLE 19.4.13.** Let

$$A = \begin{bmatrix}1&1&1&1&1&1&1&0&0&0&0\\1&1&1&1&0&0&0&1&1&1&0\\1&1&0&0&1&1&0&1&1&0&1\\1&0&1&0&1&0&1&1&0&1&1\end{bmatrix}.$$

This is a $4 \times 11$ matrix whose columns list all of the binary vectors of length 4 that have at least two 1s. The corresponding $4 \times 15$ parity-check matrix $P = [A\ I_4]$ lists all $2^4 - 1 = 15$ nonzero binary vectors of length 4 (without repetition), so the resulting binary linear code can correct all single-bit errors.

The corresponding generator matrix $G = \begin{bmatrix}I_{11}\\A\end{bmatrix}$ is a $15 \times 11$ matrix, so it takes an 11-bit message, and adds only $15 - 11 = 4$ check bits. This is much more efficient than the triple-repetition code of Example 19.1.6, which would have to add 22 check bits to detect every single-bit error in an 11-bit message.

*Remark 19.4.14.* Generalizing Example 19.4.13, a binary linear code is called a **Hamming code** if the columns of its parity-check matrix $P = [A\ I_r]$ are a list of all the $2^r - 1$ nonzero binary vectors of length $r$ (in some order, and without repetition). Every Hamming code can correct all single-bit errors. Because of their high efficiency, Hamming codes are often used in real-world applications. But they only correct single-bit errors, so other binary linear codes (which we will not discuss) need to be used in situations where it is likely that more than one bit is wrong.

**EXERCISES 19.4.15.**

1) Explain how to make a binary linear code of type $(29, 24)$ that corrects all single-bit errors.
2) Explain why it is impossible to find a binary linear code of type $(29, 25)$ that corrects all single-bit errors.
3) For each $k \leq 20$, find the smallest possible number $r$ of check bits in a binary linear code that will let you send $k$-bit messages and correct all single-bit errors. (That is, for each $k$, we want a code of type $(n, k)$ that corrects all single-bit errors, and we want $r = n - k$ to be as small as possible.)
4) What is the smallest possible number $r$ of check bits in a binary linear code that will let you send 100-bit messages and correct all single-bit errors?

## 19.5. Codes from designs

An error-correcting code can be constructed from any BIBD$(v, k, \lambda)$ for which $\lambda = 1$. More precisely, from each block of the design, create a binary string of length $v$, by placing a 1 in each of the positions that correspond to points in the design, and place 0s everywhere else. (The resulting code will not usually have a generator matrix, so it is not a binary linear code. In fact, it is encoding the blocks of the design rather than all binary strings of some length $r$, so unlike binary linear codes the total number of code words will not always be a power of 2.)

**EXAMPLE 19.5.1.** For the BIBD$(7, 3, 1)$ that has arisen in previous examples, with blocks

$$\{1, 2, 3\}, \{1, 4, 5\}, \{1, 6, 7\}, \{2, 4, 6\}, \{2, 5, 7\}, \{3, 4, 7\}, \{3, 5, 6\},$$

the corresponding code is

$$\mathcal{C} = \{1110000, 1001100, 1000011, 0101010, 0100101, 0011001, 0010110\}.$$

**PROPOSITION 19.5.2.** *A BIBD$(v, k, 1)$ can be used to produce a code that can correct up to $k - 2$ errors.*

**PROOF.** Let $B$ and $B'$ be blocks of the design, and let $b, b'$ be the corresponding binary strings of length $k$ as described at the start of this section. If the blocks have no points in common, then $d(b, b') = 2k$. If the blocks have 1 entry in common, then

$$d(b, b') = 2(k - 1)$$

(the strings differ in the $k - 1$ positions corresponding to points that are in $B$ but not in $B'$, and in the $k - 1$ positions corresponding to points that are in $B'$ but not in $B$). Since $\lambda = 1$, the blocks cannot have more than one point in common. So in any case,

$$d(b, b') \geq 2(k - 1).$$

Since $b$ and $b'$ were arbitrary output words of the code (because $B$ and $B'$ were arbitrary blocks), this means that $d(\mathcal{C}) \geq 2(k-1)$. This is greater than $2(k-2)$, so Theorem 19.2.12 tells us that the code can correct any $k - 2$ errors. □

**EXERCISES 19.5.3.**

1) Suppose that you use a BIBD to create a code whose words have length 10, that is 4-error-correcting. How many words will your code have?
2) How many errors can be corrected by a code that comes from a BIBD$(21, 4, 1)$?

3) Recall the 2-$(8, 4, 3)$ design given in Exercise 17.2.8.2. It is possible to show that this is also a 3-$(8, 4, 1)$ design; for the purposes of this problem, you may assume that this is true. If we convert these blocks to binary strings to form code words for a code, how many errors can this code correct?

---

## SUMMARY:

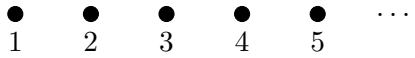
- how to make a code from a design
- Important definitions:
  - binary string
  - code, codeword
  - Hamming distance
  - minimum distance of a code
  - detect errors, correct errors
  - encode, decode
  - generator matrix
  - parity-check matrix
  - binary linear code of type $(n, k)$
  - Hamming code
- Notation:
  - $d(x, y)$
  - $d(\mathcal{C})$
  - $G = \begin{bmatrix} I_k \\ A \end{bmatrix}$
  - $P = [A\ I_r]$

---

# Appendix A

# Complex Numbers

When you were very young and first learning about numbers, you probably began with the numbers we use for counting: $1, 2, 3$, and so on. If you even had a concept of the number line at that time, it would have looked something like this:

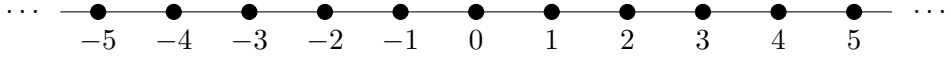



In fact, until you grasped the concept of infinity, your number line may also have had a definite ending.

At some point, you learned about the concept of 0. During elementary school, negative numbers were introduced to you and your number line expanded to include all of the integers:


but it still had lots of gaps.

Once you learned about fractions, you began to fill in those gaps between the integers, and even though there were still technically some gaps, you probably thought of the number line as continuous. Eventually you learned about the real numbers, and your number line solidified:




Each of these steps required a real leap of the imagination. In many societies, mathematical reasoning stopped before one of these developments, and they never developed words or symbols for some or all of 0, negative numbers, fractions, or real numbers. These concepts were outside of their frame of reference, and would probably not have made sense to someone from one of these societies without a fair bit of explanation and justification. When you were young, you might well have been confused by the idea that someone could have less than no cookies. Understanding what a negative number means usually involves the introduction of the concept of debt, which is far less natural than counting things that you see. Similarly, the concept of fractional items only becomes natural if you have items that you want to divide up and the result of that division is not an integer. Real numbers typically only arise through either algebraic manipulation, or geometry, and even in those situations rational numbers are close enough for most practical purposes.

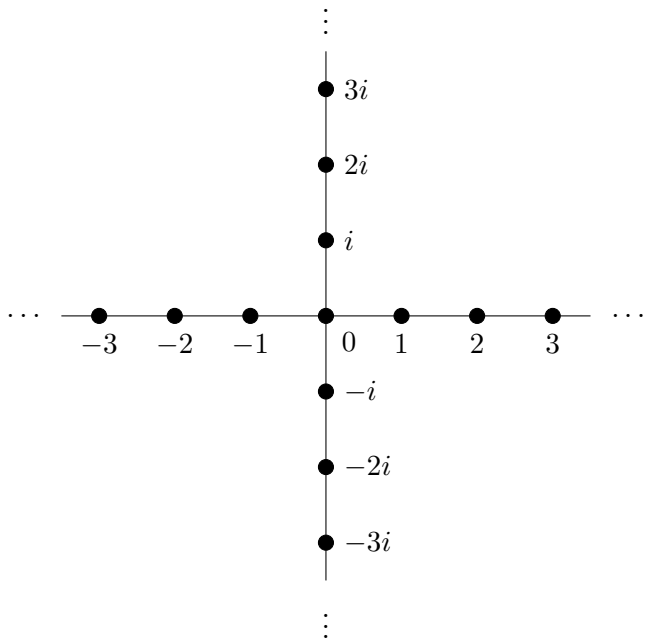
So maybe it's not so surprising that with complex numbers, we introduce a way to solve one of the few types of equation that have no solution in the real numbers: equations in which it is necessary to take the square root of a negative number. Is it really any more surprising to tell you that we *can* take the square root of a negative number after all, than it was to tell you that we could divide a single pie into seven equal pieces, obtaining a number that wasn't an integer? It does require us to once again expand our understanding of the number line, but we've done that before.

This, then, is the key concept of complex numbers: that $\sqrt{-1}$ exists. Since it would be cumbersome to write $\sqrt{-1}$ all the time, we give this quantity a symbol: $i$. Rather unfortunately, $i$ is short for "imaginary". While this is a natural term to use for the most basic number that isn't a "real" number, it does tend to reinforce many people's first impression on first learning about $i$: that these are numbers that don't really make any sense, or that have no practical value. In fact, complex numbers are extremely useful and no less natural than any of the other conceptual leaps we've been discussing.

Once we accept that there is no good reason for $\sqrt{-1}$ not to exist, then why shouldn't $\sqrt{-4}$ exist? In fact, it is an easy calculation that if $i = \sqrt{-1}$ so that $i^2 = -1$, then $(2i)^2 = 4i^2 = 4(-1) = -4$, so $\sqrt{-4} = 2i$. Similarly, the square root of any real negative number is some real multiple of $i$: if $x \in \mathbb{R}^+$ then $\sqrt{-x} = \sqrt{x}\sqrt{-1} = \sqrt{x}i$.

Actually, just as every equation of the form $x^2 = r$ has two real solutions when $r$ is positive, it also has two solutions when $r$ is negative. We see that $(-i)^2 = (-1)^2 i^2 = 1(-1) = -1$ and similarly, if $(ri)^2 = x$ then $(-ri)^2 = x$ also.

At this point, in addition to the real number line we have introduced the imaginary number line: every possible real multiple of $i$. Since $0 \cdot i = 0$ is on both of these number lines, it is natural to draw the two number lines perpendicular to each other, intersecting at 0:

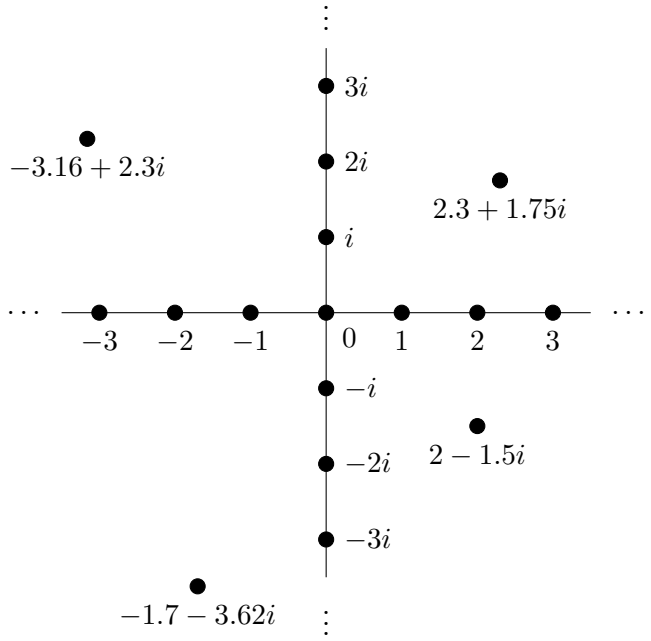



This image of the imaginary numbers led to a joke about an error message you might receive from your telephone service provider: "We're sorry; the number you have reached is imaginary. Please rotate your phone 90 degrees and try your call again."

It's pretty easy to see that if we take two numbers, each of which is on either the real or imaginary number line, and multiply them together, then we get another number that is on one of these lines. Also, if we add two real numbers the result is real. Similarly, if we add two imaginary numbers then the result is imaginary, since $ai + bi = (a + b)i$. But what if we add a real number to an imaginary number?

A number like $2 + 5i$ is not on either of our number lines. So again, we have to expand our concept of the numbers to include the whole of what we refer to as the *complex plane*. Any linear combination $a+bi$ of a real number with an imaginary number, is a complex number. We can view this number as lying in position $(a, b)$ of the plane, since its real part takes us $a$ steps horizontally along the real number line, and its imaginary part takes us $b$ steps vertically along

the imaginary number line. Thus, any point in the plane corresponds to a complex number, and vice versa. In the image below we've inserted a small number of these points.




In high school, you should have learned how to use the quadratic formula to find the roots of quadratic equations. This is a common context in which complex numbers arise naturally. You were probably taught that for an equation of the form

$$ax^2 + bx + c = 0$$

the roots have the form

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Very likely you learned that if $b^2 - 4ac = 0$ then there is only one (repeated) root (the graph of the quadratic equation just touches the $x$-axis at its maximum or minimum), and if $b^2 - 4ac < 0$ then there are no roots (the graph of the equation is either entirely above, or entirely below, the $x$-axis). This is true insofar as the real roots are concerned, but when $b^2 - 4ac < 0$ there are complex roots. When we allow complex numbers, an equation of degree $d$ in $x$ always has $d$ roots (some of which may be repeated).

Now that we understand what complex numbers are and where they come from, we need to quickly cover the rules of basic arithmetic involving them.

**Addition.** We've already talked a bit about adding two complex numbers. If you're familiar with linear algebra and think of complex numbers as vectors in the plane, then adding them is just like vector addition. It's pretty straightforward even if you don't have that background. We have

$$(a + bi) + (c + di) = (a + c) + (b + d)i.$$

We add the real parts of the summands to form the real part of the sum, and we add the complex parts of the summands to form the complex parts of the sum.

**EXAMPLE A.0.1.**

$$(1+\sqrt{2}i)+(-5+3i)=(1-5)+(3+\sqrt{2})i=-4+(3+\sqrt{2})i$$

**Subtraction.** Subtraction is really similar to addition. We have

$$(a+bi)-(c+di)=(a-c)+(b-d)i.$$

We take the difference of the real parts to form the real part of the difference, and we take the difference of the complex parts to form the complex parts of the difference.

**EXAMPLE A.0.2.**

$$(1+\sqrt{2}i)-(-5+3i)=(1-(-5))+(3-\sqrt{2})i=6+(3-\sqrt{2})i$$

**Multiplication.** Multiplication is like multiplying polynomials: treat $i$ as a variable, except that if we get an $i^2$ we use the rule $i^2=-1$ so that we always end up with something that looks like a real number plus an imaginary number:

$$(a+bi)(c+di)=ac+adi+bci+bdi^2=(ac-bd)+(ad+bc)i.$$

**EXAMPLE A.0.3.**

$$(1+\sqrt{2}i)(-5+3i)=-5+3i-5\sqrt{2}i+3\sqrt{2}i^2=(-5-3\sqrt{2})+(3-5\sqrt{2})i$$

Before explaining division, we need to introduce complex conjugates.

**Complex conjugation.** Given a complex number $a+bi$, we refer to

$$a-bi$$

as its **complex conjugate**. The value of this arises from the usefulness of differences of squares, and it will work in a similar manner to the technique of rationalising a denominator. If we take the product of a complex number with its complex conjugate, the result will be a difference of squares. Since the square of either a real number or a complex number is always real, this means that we get a real number! Here are the calculations:

$$(a+bi)(a-bi)=a^2-abi+abi-(bi)^2=a^2-b^2i^2=a^2-(-1)b^2=a^2+b^2.$$

**EXAMPLE A.0.4.** The complex conjugate of $1+\sqrt{2}i$ is $1-\sqrt{2}i$. If we multiply these two values we obtain

$$(1+\sqrt{2}i)(1-\sqrt{2}i)=(1-(\sqrt{2}i)^2)=1+2=3.$$

**Division.** If our denominator is a real nonzero number $r$ then it is easy to perform division:

$$(a+bi)/r=(a/r)+(b/r)i.$$

Otherwise, we use tricks (similar to those we might use if the denominator were irrational) to make the denominator easier to work with. In this case, we multiply both the numerator and

the denominator by the complex conjugate of the denominator. This makes our denominator real, which we understand how to work with. Here's how we do this:

$$\frac{a+bi}{c+di} = \frac{(a+bi)(c-di)}{(c+di)(c-di)} = \frac{(ac+bd)+(bc-ad)i}{c^2+d^2} = \frac{ac+bd}{c^2+d^2} + \frac{bc-ad}{c^2+d^2}i.$$

**EXAMPLE A.0.5.**

$$\frac{-5+3i}{1+\sqrt{2}i} = \frac{(-5+3i)(1-\sqrt{2}i)}{(1+\sqrt{2}i)(1-\sqrt{2}i)} = \frac{(-5+3\sqrt{2})+(3+5\sqrt{2})i}{1-(\sqrt{2}i)^2} = \left(-\frac{5}{3}+\sqrt{2}\right)+\left(1+\frac{5\sqrt{2}}{3}\right)i.$$

# Appendix B

# Biographical Briefs

In this appendix, I provide brief biographical sketches of the mathematicians whose names have arisen in this book.

Numerous mathematicians, with a wide diversity of backgrounds and identities, work in the field of combinatorics. Some (including many of those from the most marginalised backgrounds) have focused their attention on math education, generally providing significant support and inspiration to students. One example of this is that a number of the first Black people (and particularly the first Black women) to receive PhDs in mathematics in North America, went on to teach at Historically Black Colleges. Typically they would have heavy teaching loads, and little if anything in the way of research expectations. Some focus on writing books, or popular articles of mathematics, or on mathematical outreach to children. Such people may have a huge influence on the field, but are not often actively involved in research, so do not make discoveries or prove results that fit easily into a book such as this. Indigenous mathematicians likewise feel heavy pressures to teach, mentor, and serve their communities, that often limit the time they have for research.

It is important to acknowledge that underrepresentation is not accidental, and that systemic barriers as well as overt discrimination, harassment, and bullying have all contributed to it. Mathematicians continue to grapple with these issues.

Without leaving out anyone whose work seemed important to understanding the key topics of this book, I have tried to include results by people with diverse backgrounds and identities. It is often not apparent from someone's name whether or not they come from an underrepresented population, so I hope to make you aware of some of these mathematicians' identities through these sketches. There are likely to be other aspects to their identities that I remain unaware of. I hope that you may find some of these people relatable, and possibly inspirational.

Perhaps even more than mathematicians in other fields of research, combinatorists like to collaborate. We enjoy working together, and encourage our students to develop research skills through collaborations also. This has been a broader change to mathematical culture over time: mathematicians used to be much rarer and more isolated, so worked on their own a lot more than they typically do now. Particularly in the very recent context, it's therefore usual for someone's most significant piece of work to have been written in conjunction with others. This means that finding a theorem or object that has been named after a woman, for example, is challenging: her best work may have been completed with other researchers, and will bear their names as well. Please don't take this as a sign that women (or other underrepresented groups) haven't made important contributions!

Names are listed in alphabetical order, but this comes with some caveats. Following the style guides I found, medieval Arabic names are alphabetised under the first name. When the indefinite article "Al-" is part of the first name, it is ignored in the alphabetising. Chinese names have been alphabetised by the family name. The order in which Chinese family names

and given names are used has varied over time and through contexts, so is not consistent in this book. The medieval Zhu Shijie is referred to as such (his family name is Zhu); the modern Ming-Yao Xu publishes (when he writes in English) with his name in this order (his family name is Xu).

I have endeavoured to contact all of the living mathematicians included below, for approval and confirmation of the information I am including. Many of them responded very helpfully, for which I am extremely grateful. If you believe that something is inaccurate or should not be included, please do not hesitate to contact me so we can sort this out. Any errors are my own.

I am deeply indebted to Laci Babai who carefully read many of these sketches. Laci provided a great deal of clarification and additional information, and even provided suggestions for beautiful and carefully-constructed wording that I have adopted gratefully. To avoid filling the sketches with quotation marks and repeated references to Laci, this general acknowledgment stands in lieu of specific attribution for the many phrases and even paragraphs that he generously offered. In particular, without his knowledge of the Hungarian context and of the numerous Hungarian mathematicians listed below (most if not all of whom he knows or has known), these sketches would contain numerous misunderstandings and inaccuracies.

## B.1. List of Entries

F.
- Gino Fano (1871—1952)
- Leonardo Pisano ("Fibonacci") (c.1170—c.1250)
- Sir Ronald Aylmer Fisher (1890—1962)

G.
- Martin Gardner (1914—2010)
- Edgar Nelson Gilbert (1923—2013)

H.
- Wolfgang Haken (1928—)
- Halayudha (10th century CE)
- Philip Hall (1904—1982)
- Sir William Rowan Hamilton (1805—1865)
- Richard Wesley Hamming (1915—1998)
- Percy John Heawood (1861—1955)

I.

J.
- Jeannette Janssen (1963—)

K.
- Peter Keevash (1978—)
- Sir Alfred Bray Kempe (1849—1922)
- Reverend Thomas Penyngton Kirkman (1806—1895)
- Eszter Klein (1910—2005)
- Tamás Kővári (1930—2010)
- Kazimierz Kuratowski (1896—1980)

L.
- Clement Wing Hong Lam (1949—)
- Lu Jiaxi (1935—1983)

M.
- Gary McGuire (1967—)
- Wendy Joanne Myrvold (1961—)

N.

O.

P.
- Ernest Tilden Parker (1926—1991)
- Blaise Pascal (1623—1662)
- Peter Christian Julius Petersen (1839—1910)
- David Angus Pike (1968—)
- Cheryl Elisabeth Praeger (1948—)

Q.

R.
- Richard Rado (1906—1989)
- Frank Plumpton Ramsey (1903—1930)
- Dwijendra Kumar Ray-Chaudhuri (1933—)
- Alfréd Rényi (1921—1970)
- George Neil Robertson (1938—)
- Gordon F. Royle (1962—)

S.
- Al-Samaw'al ben Yahyā al-Maghribī (c.1130—c.1180)

- Issai Schur (1875—1941)
- Paul Seymour (1950—)
- Sharadchandra Shankar Shrikhande (1917—2020)
- Vera Sós (1930—)
- Jakob Steiner (1796—1863)
- Sushruta (c.800BCE—c.700BCE)
- Stan Swiercz (1955—)
- György Szekeres (1911—2005)

T.
- Peter Guthrie Tait (1831—1901)
- Larry Henry Thiel (1945—)
- Robin Thomas (1962—2020)
- Pál Turán (1910—1976)

U.
- John Cameron Urschel (1991—)

V.
- Varāhamihira (499—587)
- Vadim Georgievich Vizing (1937—2017)

W.
- Klaus Wagner (1910—2000)
- Jake Wellens (1992—)
- Richard Michael Wilson (1945—)
- Wesley Stoker Barker Woolhouse (1809—1893)

X.
- Ming-Yao Xu (1941—)

Y.

Z.
- Kazimierz Zarankiewicz (1902—1959)
- Zhu Shijie (1249—1314)

## B.2. Biographies

**Abraham ibn Ezra (c. 1093—1167).**
Abraham ibn Ezra was an 11th-century Jewish scholar whose commentaries on the bible have been highly influential, and who disseminated Arabic scholarly knowledge into Jewish communities. He was born in Tudela, in what is now the Spanish province of Navarre. In the Middle ages, the town of Tudela had one of the oldest and most important Jewish communities in that region. Abraham moved to Córdoba as a young man and spent about half of his adult life there. In the final 27 years of his life he travelled extensively (going as far as Baghdad), after fleeing from the attacks on hitherto tolerant Moorish Iberia by the fanatic Almohads around 1140. He seems to have made his living largely as a poet and author.

Abraham married and had five children, four of whom are believed to have died young. His youngest son Isaac, an influential poet, converted to Islam in 1140; this event was deeply troubling to Abraham, as reflected in some of his own poetry.

Abraham is primarily known for his commentaries on the Torah. He also wanted to spread the (largely Arabic) knowledge he gained in Spain to the many Jewish communities he visited and lived in during his travels. His own writing was in Hebrew, and he also translated other works into Hebrew. Along with his biblical commentaries, he wrote books on Hebrew grammar; mathematics; philosophy; and astrology, as well as writing poetry. The book in which he

discusses the problem described in Example 3.2.8 (mentioned again in passing in Example 4.2.7) is one of his books on astrology.

A lunar crater, Abenezra, was named after Abraham. Robert Browning's poem "Rabbi Ben Ezra" is a meditation on his life. It begins with the famous lines "Grow old along with me!/ The best is yet to be."

Sources: Stanford Encyclopedia of Philosophy, wikipedia, and Encyclopedia Britannica.

**Abū-l'Abbās Ahmad ibn al-Bannā' (1256—1321).**
Abū-l'Abbās Ahmad ibn al-Bannā' (al-Marrākushī) is also known as Abū'l-Abbas Ahmad ibn Muhammad ibn Uthman al-Azdi. He was a 13th century scholar and mathematician whose books include the earliest-known surviving examples of some mathematical formulas and notation that we use to this day. Abü lived in the region that is now Morocco. It is not completely clear whether he was born in Marrakesh, but he certainly spent most of his life in that region, so that "al-Marrākushī" is often added to his name. It was in Marrakesh that he studied mathematics, which he went on to teach at the university in Fez for most of his life.

He wrote at least 82 books, and translated Euclid's *Elements* into Arabic. It is not clear how much (if any) of the mathematical content of Abū's books is original and how much is due to earlier writers whose work has been lost; the writing seems to indicate mostly the latter. In his book *Raf al-Hijab* ("Lifting the Veil") Abū presents the general formula that we still use to calculate the binomial coefficient $\binom{n}{k}$. This was also the first known work to use algebraic notation, though there is again some uncertainty as to whether or not this originated with him. *Raf al-Hijab* was actually a commentary on a previous book of his that may have been too challenging for his readers, so Abū wrote the commentary to provide additional details. This is why he included the calculations required for a variety of operations, including binomial coefficients. Some of his work is described in Example 4.2.5.

A lunar crater, Al-Marrakushi, was named after Abū.

Sources: encyclopedia.com, wikipedia, and St. Andrews' math history web site.

**Atif Aliyan Abueida (1966—).**
Atif Aliyan Abueida is an American researcher specialising in combinatorics, who is originally from Palestine. He graduated with a B.Sc. in 1987 from the United Arab Emirates University, in Al Ain (UAE). After working for eight years as a high school teacher, he moved to the United States, where he completed his M.Sc. at East Tennessee State University in 1996. Abueida then moved to Auburn University in Alabama, where he worked on his Ph.D. under the supervision of Chris Rodger. His thesis was entitled "The Full Embedding Problem".

Immediately after defending his Ph.D. in 2000, Abueida moved to the University of Dayton in Ohio, where he has been working in the Department of Mathematics ever since. His areas of research interest include graph theory, design theory, and complex analysis. He has more than 25 publications in these areas, including the work with David Angus Pike (1968—) mentioned in Section 17.1. He has also been involved in outreach, running workshops for teachers in the Dayton area.

Abueida is fluent in English and Arabic.

Sources: University of Dayton, University of Dayton, and the Math Genealogy Project, as well as MathSciNet for a publication list. Updated and confirmed through personal communication.

**Ahmad ibn Mun'im al-'Abdarī (11??—1228).**
Ahmad ibn Mun'im al-'Abdarī was a 12th-century mathematician, doctor, and teacher, whose surviving works include what may be the earliest examples of combinatorial reasoning. Ahmad lived in the area that is now Morocco. Different sources use slightly different versions of Ahmad's

name. The most apparently reliable gives it as Ahmad ibn Ibrāhīm ibn 'Ali Ibn Mun'im al-'Abdarī. He was born near Valencia, Spain. He lived in Marrakesh (Morocco) for most of his life, and that is where he died. In addition to being a mathematical scholar and teacher, he learned medicine at the age of thirty, and was a practicing doctor as well as a scholar from that time.

Ahmad apparently wrote extensively about mathematics. Details remain known about only three of his texts, and only one of these, *Fiqh al-hisāb* ("Science of Calculation"), survives. This was written during the reign of the fourth Almohad caliph (1199—1213), who was a patron of education and science. In this book Ahmad considers a variety of problems involving permutations and combinations, with and without repetition, including the problems of silk threads and of words using the Arabic alphabet that is referred to in Example 4.2.7 and mentioned again in Section 4.3. The method of reasoning he uses to deduce an identity involving binomial coefficients may be the earliest known example of a combinatorial proof.

Abū-l'Abbās Ahmad ibn al-Bannā' (1256—1321) was a student of al-Qādhī ash-Sharīf, who in turn was a student of Ahmad's.

Sources: Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures, Combinatorics: Ancient and Modern, and muslimheritage.com.

**Michael Owen Albertson (1946—2009).**
Michael Owen Albertson was an American mathematician, particularly known for (with Karen Linda Collins (1959—)) introducing the distinguishing number of graphs. He was born in Philadelphia and raised in Abington, Pennsylvania. He began his undergraduate work at Michigan State University in 1964, and went on to complete a Ph.D at the University of Pennsylvania, where his thesis supervisor was Herbert Wilf.

Albertson defended his Ph.D. thesis (entitled "Irreducibility and Coloring Problems") in 1971, and shortly thereafter was hired to teach at Smith College in Northampton, Massachusetts. Albertson worked and taught at Smith, while raising a family, for more than 35 years until his untimely death in 2009. Albertson had 3 children. In 2004 Albertson was named the L. Clark Seelye Professor of Mathematics at Smith College. He was married to Debra Lynn Boutin (1957—) for the last 15 years of his life.

Albertson's main area of research was graph theory. Over the course of his career he published more than 70 research papers with 33 different collaborators, including both Boutin and Collins. His work with Collins on the distinguishing number is discussed in Section 12.5. He mentored and collaborated with many junior researchers and students; Collins first worked with him during her undergraduate studies.

Sources: memorial article by Joan Hutchinson, obituary, the Math Genealogy Project, and Smith College. Updated and confirmed through personal communication with Debra Boutin.

**Kenneth Ira Appel (1932—2013).**
Kenneth Ira Appel was an American algebraist with expertise in computers, famous for his 1976 computer-aided proof of the Four-Colour Theorem (with Wolfgang Haken (1928—)). Appel was born into a Jewish family in Brooklyn (New York), and raised in Queens. He graduated with a B.Sc. from Queens College in 1953. He then worked briefly as an actuary and spent two years in the army, where he served in the state of Georgia as well as overseas in Germany. After his time in the army, Appel studied at the University of Michigan under the supervision of Roger Lyndon. He was awarded a Master's in 1956. His Ph.D thesis (still at the University of Michigan) was entitled "Two Investigations on the Borderline of Logic and Algebra", and he defended it in 1959.

During his studies Appel spent summers programming computers for Douglas Aircraft, and after his doctorate he went to work for the federal government's *Institute for Defense Analyses* in Princeton, New Jersey, doing research in cryptography. He moved to Urbana, Illinois in

1961, where he joined the University of Illinois as a professor. It was there that he collaborated with Haken on the Four-Colour Theorem, discussed in Section 15.3. Haken brought the idea of the problem and the suggestion of using a computer to solve it, but had been convinced that this approach was not feasible. Appel had significant background in computing (having used computers in algebraic research as well as in cryptography), and proposed that they attempt this together.

Appel left the University of Illinois in 1993 to become chair of the mathematics department at the University of New Hampshire, which brought him closer to his grandchildren. He retired in 2003. During his career, Appel supervised at least five doctoral students of his own, one of whom (John Koch) assisted in some of the work on the Four-Colour Theorem.

Appel had a lifelong interest in politics. He served on the Urbana City Council; was treasurer of the Strafford County Democratic Party; and served on the Dover School Board for years until his death. His work on the school board also related to an ongoing interest and involvement in math education and outreach.

Appel was married with three children, two of whom became professors themselves (a daughter in biology and a son in computer science at Princeton). He and his wife met in graduate school, and were married after she obtained her Master's and he finished his Ph.D. Appel involved his children in checking calculations and computer output for the proof of the Four-Colour Theorem; they found hundreds of errors, many of which they were able to fix themselves.

Sources: New York Times obituary, St. Andrews' math history web site, the Math Genealogy Project, University of Illinois, and wikipedia.

**László Babai (1950—).**
László Babai is a Hungarian-American mathematician and computer scientist, best known for his coinvention of interactive proofs and for his quasipolynomial-time algorithm for the Graph Isomorphism problem. Born and educated in Budapest, Hungary, Babai (known as "Laci" to friends and colleagues) competed three times in the International Mathematical Olympiad for high school students. Representing Hungary, he won silver medals in 1966 and 1967 and a gold medal in 1968. He graduated from Loránd Eötvös University in Budapest in 1973. Mentored by Pál Turán (1910—1976) and Vera Sós (1930—), he earned his Ph.D. in 1975 from the Hungarian Academy of Sciences with his thesis entitled "Gráfok automorfizmuscsoportjai" ("Automorphism Groups of Graphs").

Babai worked at Eötvös University from 1973 until 1989. In 1984 he received a D.Sc. (doctor of science) from the Hungarian Academy of Sciences. This is a degree that is available in some countries to recognise notable and prolonged contributions to research. In 1984 Babai became a visiting professor at the Computer Science Department at the University of Chicago in the United States, while maintaining his affiliation with Eötvös University, embarking on a period of commuting between the two continents that lasted nearly a decade. He became a permanent professor of computer science at Chicago in 1987 and was additionally appointed a professor of mathematics in 1995. He held the George and Elizabeth Yovovich Professorship at the University of Chicago from 2010—2019, and since 2019 he has been holding the Bruce V. and Diana M. Rauner Distinguished Service Professorship. Meanwhile he continues to maintain strong ties with the mathematical community in Hungary. He was elected as a corresponding member of the Hungarian Academy of Sciences in 1990 and as a full member in 1994.

Babai enjoys mentoring and engaging young people in math and computer science at various stages of their careers, from high school to postdoc. He has helped to lead Research Experience for Undergraduates programs at the University of Chicago every summer from 2001 to 2016 and continues to mentor REU participants. He has supervised 27 Ph.D. students as of 2021. Five of his former mentees, including one from high school, have become invited speakers at the

International Congress of Mathematicians. He is one of the founders of the highly acclaimed *Budapest Semesters in Mathematics* study-abroad program for undergraduate students. At Chicago, he won a prestigious teaching award. Two pieces of advice that he frequently gives to students are: "It is better to learn several proofs of the same central theorem than to learn more theorems," and, encouraging students to be broad in their interests: "The only kind of math I never used is the math I never learned."

Babai has worked in combinatorics, group theory, the theory of algorithms, and complexity theory, with special attention to the interactions among these fields. The quasipolynomial graph isomorphism algorithm, which he discovered at the age of 65, is discussed in Section 11.4.

Babai has won numerous prizes and awards, including the Gödel prize (1993) and the Knuth Prize (2015). He was elected a Fellow of the American Academy of Arts and Sciences in 2015, and has been an invited speaker at the International Congress of Mathematicians three times (once as a plenary speaker).

Babai's native tongue is Hungarian. He is fluent in English and also conversant in German and Russian.

In the 2008 episode "First Contact" of the tv series *Stargate: Atlantis*, the character Dr. Radek Zelenka claims to have used Babai's work in combinatorics to trace a subspace transmission, but says that he cannot "dumb it down" for his superiors.

Sources: University of Chicago, American Academy of Arts and Sciences, the Math Genealogy Project, Gateworld, and wikipedia. Updated and confirmed through personal communication.

**Eric Temple Bell (1883—1960).**
Eric Temple Bell was a British-American analytic number theorist who studied diophantine equations. His work with generating functions is particularly noteworthy; he is also well-known for the popular books he wrote on math and the history of math (some of these are still in print). Bell was born in Peterhead, Scotland, but his family moved to California when he was just over a year old. The family returned to Bedford, England when his father died in 1896 (when Bell was 13), but he moved back to North America six years later, as he said "to escape being shoved into Woolwich [Royal Military Academy] or the India Civil Service".

Bell studied mathematics at Stanford University in California from 1902—1904, where he was able to take all of the available math courses (and nothing else) and graduate with his bachelor's degree in two years. He taught for three years at a private school in San Francisco, leaving after the great fire of 1906. In 1907 Bell entered the Master's program at the University of Washington, graduating in 1908. He then moved to San José, California, where he earned some money by writing a science fiction novel. During some of the breaks in his education, Bell also worked as a ranch hand, surveyor, and mule skinner. Over the course of his career, he wrote at least 16 science fiction novels under the pseudonym "John Taine". Many of these were written in a few weeks during the summers. They are now largely unknown and out of print.

Bell taught at Yreka High School in northern California for two years from 1909—1911. It was in Yreka that Bell met and married his wife, who also taught at the high school. After leaving Yreka, Bell moved to New York where he entered the Ph.D. program at Columbia University. He completed his Ph.D. in 1912, under the joint supervision of Frank Cole and Cassius Keyser, with a thesis entitled "The Cyclotomic Quinary Quintic".

Immediately after completing his doctoral work, Bell moved to Seattle, where he taught at the University of Washington from 1912—1926. His only child, a son named Taine, was born during this time. Bell's reputation in research grew significantly during these years, and he was offered professorships at a number of prominent universities. He ultimately accepted an offer from the California Institute of Technology, where he remained from 1926 until his retirement in 1959.

In the context of Bell's work on generating functions, he studied and wrote about the numbers that have become known as "Bell numbers" (see Section 9.3). In addition to his original research (which included more than 250 papers) and his science fiction, Bell published popular books with colourful accounts of mathematical problems and math history. The ongoing popularity of his *Men of Mathematics* (which includes one woman) may be due in part to his placing more weight on telling a good story than on strict accuracy. Critics describe Bell's accounts as inaccurate and fanciful, legends rather than histories (some use harsher terms).

Bell won the American Math Society (AMS)'s Bôcher Prize in 1921. He was appointed to the Council of the AMS in 1924, and served as its vice-president beginning in 1926. He was elected president of the Mathematical Association of America, a role he filled from 1931—1933.

Sources: St. Andrews' math history web site, encyclopedia.com, the Math Genealogy Project, *The Last Problem*, and wikipedia.

**Bhāskara II (1114—1185).**
Bhāskara was an Indian mathematician and astronomer from the 12th century whose works contain the first surviving systematic use of the decimal system and introduce many principles of calculus. He was born in Bijapur, India. He is known as Bhāskara II to distinguish him from another mathematician and astronomer of the same name, from the 7th century. He is also often known as Bhāskaracharya, meaning "Bhāskara the teacher". His father Maheśvara (a Brahman) was also a mathematician and astronomer/astrologer, and saw to his son's early training. As an adult, Bhāskara became the leader of an astronomical observatory in Ujjain. This was the centre of mathematical thought in India at that time.

At the age of 36, Bhāskara wrote the work for which he is best known, the *Siddhānta Siromani* ("Crown of treatises"). This is a four-part book that uses the decimal numbering system. The first part, Līlāvatī ("The Beautiful"; according to a story written by Fyzi, who translated this book into Persian in 1587, it is named for Bhāskara's daughter) is about calculations, including combinations and permutations. It is in this part that the problems referenced in Sections 3.1 (Example 3.1.6) and 4.2 appear. He did author other works as well. As was traditional in India at the time, his books were all written in verse.

Many extremely important results first appeared in various parts of the *Siddhānta Siromani*. Bhāskara explained a method for solving Pell's equation ($Nx^2+1 = y^2$) for a number of specific values of $N$. This is a problem that European mathematicians struggled with centuries later, though Archimedes had significant understanding of it much earlier than Bhāskara. He studied trigonometry, and gave several results including the rule $\sin(a+b) = \sin(a)\cos(b)+\cos(a)\sin(b)$. Though methods for finding solutions to quadratic equations geometrically were known before this time, the general quadratic formula first appeared in his work. Bhāskara also developed and explained many of the principles of differential calculus, in his study of the motion of planets and their instantaneous velocities. He formulated some concepts of integral calculus as well, for calculating the volume of a sphere (planet). This was more than 500 years before the work of Leibniz and Newton. His work included one of the well-known visual proofs of the Pythagorean Theorem. Despite all of these remarkable developments that appeared for the first time in his work, Bhāskara didn't get everything right; for example, he struggled with the concept of division by zero, claiming that $0a/0 = a$ for any $a$.

The Indian Space Research Organisation named one of their satellites after Bhāskara.

Sources: St. Andrews' math history web site, storyofmathematics.com, New World Encyclopedia, and wikipedia.

**Anthony Bonato (1971—).**
Anthony Bonato is an Italian-Canadian graph theorist who identifies as a gay man. In addition to his research in graph theory, he is known for his popular writing about math. He graduated with a B.Sc. from McMaster University in Hamilton, Ontario, in 1993. Bonato received a

Master's degree (in 1994) and Ph.D. from the University of Waterloo. He worked under the supervision of Ross Willard, and defended his thesis "Colourings, Generics, and Free Amalgams" in 1998.

After a post-doctoral position at Mount Allison University in New Brunswick, Bonato returned to Ontario and in 1999 took a tenure-track faculty position at Wilfrid Laurier University in Waterloo, Ontario. In 2008 he moved to a tenured faculty position at a university in Toronto then known as Ryerson University. (Because of the role of its namesake Egerton Ryerson in the genocidal residential school system into which Indigenous children were forced, pressure has been mounting for the university to change its name.) Bonato also holds adjunct appointments at Laurier and at Dalhousie University in Nova Scotia.

Bonato is open and vocal about his sexual orientation and advocates for the LGBTQ+ community in mathematics. He usually introduces himself as a "gay man and a mathematician", with the pronouns he/him/his. As Bonato has said, "the layers of our identity are critical to who we are and how our work is received." Bonato and his husband have been married since 2004. He organised the "LGBTQ+ Math Day" events to raise the profile of queer mathematicians and to encourage young people in math who identify as LGBTQ+. He has been actively involved in committees to improve equity, diversity and inclusion in mathematics.

Bonato has over 130 publications, including four books on mathematics, with a fifth book forthcoming in 2022. His research is highly regarded. He has served on Canada's science granting council (NSERC) selection committees, including chairing the pure math committee and has been an invited speaker at more than 30 international conferences.

Bonato is passionate about mentoring, collaborating, and communicating mathematics. He has collaborated with more than 100 different coauthors. One of his results is discussed in Section 11.5. He has supervised 48 graduate students and post-doctoral fellows, including 7 PhD students as of 2021. He has won awards for his research and graduate supervision. Bonato's book *Limitless Minds* is a collection of interviews with prominent mathematicians. He also writes a blog, The Intrepid Mathematician, aimed at communicating about mathematics and mathematicians to a broader audience, and has a very active presence on social media. In one tweet, Bonato wrote: "You're doing math right now without realizing it. Mathematicians are people who realize it."

Bonato's debut young adult, science-fiction novel *Patterns* will be published in 2022 and centres on a queer sixteen-year-old mathematician grappling with an alien invasion.

Sources: Ryerson University, The Intrepid Mathematician, LSE blogs, the Math Genealogy Project, and twitter. Updated and confirmed through personal communication.

**John Adrian Bondy (1944—).**
John Adrian Bondy is a graph theorist whose career began in England, grew to prominence in Canada, and who moved to France (where he still lives) in 1994. In addition to some important research results in graph theory, he is known for his (co-authored) introductory textbook on this subject. Bondy (who goes by "Adrian") was born in England in 1944. He obtained his D.Phil. from Oxford University in 1969, under the supervision of Dominic Welsh. His thesis was entitled "Some Uniqueness Theorems in Graph Theory."

Immediately after completing his doctorate, Bondy moved to Canada where he began to work at the University of Waterloo, in the Department of Combinatorics and Optimization. In 1976, his textbook *Graph Theory with Applications*, co-authored with colleague U.S.R. Murty, was published and quickly became the standard undergraduate textbook for graph theory. This was also the year in which Bondy's research with Václav Chvátal (1946—) about the closure of a graph (discussed in Section 13.2) was published. Bondy was one of the editors-in-chief of the prestigious *Journal of Combinatorial Theory, Series B* from 1985 until 2004.

In 1991, Bondy developed a relationship with a woman who lived and worked in Paris. In short succession, Bondy's father died, and he became a father himself. His partner and son remained in France. Bondy's increasing absences from Waterloo when he did not have teaching obligations on campus were a source of friction between himself and university administrators that increased when his request to be granted a sabbatical or to cut back to a half-time position were denied. In 1994, he negotiated a temporary leave with reduced pay in order to seek work in France. During this leave, Bondy applied for and was appointed to a position at Université Lyon I, which was still a long commute from his family in Paris. According to documents, he was subsequently dismissed from the University of Waterloo for a "substantial conflict of interest," apparently not having informed either Waterloo or Lyon that he was employed by both, and drawing salary from both. The dismissal was appealed, but was upheld at arbitration. This was a sad end to Bondy's distinguished career at Waterloo, and led to Pál Erdős (1913—1996) resigning his honorary degree, and to Chvátal resigning his adjunct appointment at Waterloo.

Bondy continued to work at Université Lyon I until his retirement in 2009. He is credited by the Math Genealogy Project for supervising 12 doctoral students (about half of them at Waterloo and half at Lyon), and published almost 100 papers during his career.

Since childhood, Bondy has been a keen amateur photographer. He began taking photographs more seriously in 1982 during a visit to Paris. Throughout the 1980s he exhibited his photographs frequently at the Kitchener-Waterloo Art Gallery and elsewhere. After his retirement in 2010 he founded *Mind's Eye*, a non-profit association that runs the "Galerie Adrian Bondy" in Paris. The goal of the association is to explore conceptual links between photography and mathematics. Bondy is a regular exhibitor at the gallery.

Sources: wikipedia, Journal of Combinatorial Theory, Waterloo Gazette (via Wayback Machine), the Math Genealogy Project, Letter from Chvátal, Letter from Erdős, and Mind's Eye. Updated and confirmed through personal communication.

**Raj Chandra Bose (1901—1987).**
Raj Chandra Bose was an Indian-American mathematician and statistician whose discoveries in coding theory are still being used in industrial and scientific applications. He is also known for his involvement in disproving a conjecture that had been made by Leonhard Euler (1707—1783). Bose was born in Hoshangabad, India, but raised in Rohtak. His father pushed him to excel in school, a task in which Bose's photographic memory was a significant asset. His mother died in 1918 of the influenza pandemic, and his father died within a year, leaving Bose with four younger siblings. He juggled this responsibility with his own desire for education, earning some money by tutoring while working on his bachelor's degree, which he completed in 1922. Bose managed to get a job teaching high school for the next year, but this interfered with his ability to attend classes and attain his Master's degree, and the following year he was only able to earn money by tutoring. Then the brother of one of his students offered to support him in Calcutta (now Kolkata) to study pure mathematics.

Bose moved to Calcutta in 1925, leaving his siblings in Delhi, where his brother had found a steady job. He drew the attention of Shyamadas Mukherjee, a geometer, who gave him a room to live in and found him tutoring work while acting as a mentor. Bose successfully completed his Master's degree in pure mathematics at the University of Calcutta's Rajabazar Science College in 1927.

Bose worked as a research assistant for two years. Jobs were hard to come by in 1930, and although Bose did manage to become a lecturer at Asutosh College in Calcutta, the work paid so poorly that he also needed to tutor. Bose was married in 1932. At the end of 1932 the director of the new Indian Statistical Institute, having seen his work in geometry, offered Bose a part-time position even though Bose had little background in statistics. Bose became one of the chief mathematicians at the institute, moving to full-time status in 1935. After a visit to

India by Sir Ronald Aylmer Fisher (1890—1962) in 1938—1939, Bose developed a significant interest in design theory. In 1940 Bose moved to the University of Calcutta, becoming head of the department of statistics there in 1945.

Bose wanted a position as a professor, but did not have a Ph.D. He submitted some of his published papers and was awarded a doctorate in 1947, examined by Fisher. After this he spent a couple of years visiting the United States, taking positions as a visiting professor before returning briefly to Calcutta in 1948. With job offers from Calcutta and a couple of American universities to choose from, Bose joined the University of North Carolina at Chapel Hill as a professor of statistics in 1949. In 1966 he was given the Kenan Chair there. He retired in 1971, but then took a chair at Colorado State University of Fort Collins. He retired again in 1980, was made a professor emeritus at Colorado State, and died in Colorado. Bose and his wife had two children.

Among Bose's more important discoveries were "BCH" codes. Together with his student Dwijendra Kumar Ray-Chaudhuri (1933—) he discovered these independently, at about the same time as their discovery by Alexis Hocquenghem in 1959. These codes are still used in applications such as compact disc players, DVDs, and solid-state drives; the "B" in their name is for Bose, and the "C" for Ray-Chaudhuri. Bose is also known for proving (with his student Sharadchandra Shankar Shrikhande (1917—2020) and Ernest Tilden Parker (1926—1991)) that pairs of orthogonal Latin squares of order $n$ exist whenever $n = 4k + 2$ with $k \geq 2$, disproving a conjecture of Euler's (as discussed in Section 16.2). This work earned them the nickname of "Euler's spoilers". The techniques that Bose and his co-authors developed for disproving Euler's conjecture played a significant role in the proof by Richard Michael Wilson (1945—) of Wilson's Theorem. Other significant discoveries include Bose-Mesner algebras, as well as the notions of partial geometry and strongly regular graphs. Bose published over 100 papers, and supervised at least 30 doctoral students in the United States, in addition to his influence on young mathematicians in India.

Among the honours Bose received were honorary degrees from the Indian Statistical Institute in 1974 and from Visva-Bharati University in 1979. In 1976 he was elected to the United States National Academy of Sciences.

Bose spoke a number of languages including English, and liked to recite poetry in Arabic, Bengali, Persian, Sanskrit, and Urdu. He never learned to drive; his wife did all the driving for them. He was an enthusiastic gardener. Bose had an excellent sense of humour, and liked to joke that, working modulo 2, "it is equally as good to give as to receive"!

Sources: wikipedia, St. Andrews' math history web site, the Math Genealogy Project, Indian Statistical Institute, Colorado State University, and Obituary in Journal of the Royal Statistical Society.

**Debra Lynn Boutin (1957—).**
Debra Lynn Boutin is an American whose research focuses on the area of algebraic graph theory. She was born in 1957, and joined the Navy after completing high school in Chicopee, Massachusetts in 1975. She served on active duty from 1975—1979, and remained in the Naval Reserve for 16 more years. She retired as a Chief Petty Officer in 1995. She also raised a daughter during these years. In 1985 Boutin enrolled in Springfield Technical Community College in Massachusetts. In 1988 she transferred to Smith College in Northampton, Massachusetts, where she completed her bachelor's degree in mathematics in 1991.

Boutin then entered the doctoral program at Cornell University in Ithaca, New York. She studied geometric group theory under the supervision of Karen Lee Vogtmann, and successfully defended her Ph.D. thesis, "Centralizers of Finite Subgroups of Automorphisms and Outer Automorphisms of Free Groups" in 1998. Shortly afterward, Boutin was hired at Hamilton College in Clinton, NY, where she holds the Samuel F. Pratt professorship. Her areas of research

interest include graph theory, geometric graph theory, and group theory. Boutin has published over 25 papers on these topics, and is well-known in the field. Her work on distinguishing cost is discussed in Section 12.5.

Boutin was married to Michael Owen Albertson (1946—2009) for the last 15 years of his life.

Sources: Hamilton College, Hamilton College, LinkedIn, and the Math Genealogy Project. Updated and confirmed through personal communication.

**John M. Boyer (1968—).**
John M. Boyer is a Canadian computer scientist who has spent his career working in industry. He lives and works in Victoria, British Columbia. He began working for the startup PureEdge Solutions on its first day of operations in 1993. While working for PureEdge, he also undertook graduate studies at the University of Victoria, beginning in 1995. Boyer completed his Ph.D. under the supervision of Wendy Joanne Myrvold (1961—) in 2001, with his thesis "Simplified $O(n)$ Algorithms for Planar Graph Embedding, Kuratowski Subgraph Isolation, and Related Problems".

The planarity testing algorithm by Boyer and Myrvold that is mentioned in Section 15.1 was published in 2004, based on ideas that appeared in Boyer's thesis. It is one of the two state-of-the-art algorithms currently in use for planarity testing. Boyer's implementations of the planarity algorithm and extensions for related problems are publicly available, and the planarity algorithm has been implemented in a variety of packages and applications.

In 2005 Boyer became an IBM employee as a result of IBM's acquisition of PureEdge. The products he had worked on for secure interactive data collection became IBM Forms, and Boyer became the Chief Architect of that division. During his time at IBM, Boyer also served as Chief Architect for several teams, including those working on social computing, machine learning, and data science platforms. Boyer has more than 25 published papers, has co-authored several computer industry standards, and holds over 40 patents. He was appointed an IBM Distinguished Engineer in 2010 and a Master Inventor at IBM in 2012.

Sources: wikipedia, ContactOut, LinkedIn, the Math Genealogy Project, and twitter. Updated and confirmed through personal communication.

**Rowland Leonard Brooks (1916—1993).**
Rowland Leonard Brooks was a British tax inspector who as an undergraduate student proved a result about graph colouring that appears in many textbooks. Brooks (who was known as "Leonard") was born in Caistor, Lincolnshire, England, in 1916. He studied at Trinity College of Cambridge University from 1935 to 1940. While there, he developed close friendships with three other students: Smith, Stone, and Tutte, and the four worked closely together on mathematical problems.

Brooks proved Brooks' Theorem as an undergraduate, and published it in the *Proceedings of the Cambridge Philosophical Society* in 1941. He and his friends also spent a lot of time as students trying to "square the square" (find a square with integral sides that could be decomposed into smaller unequal squares with integral sides). They were ultimately successful in finding the first known examples of this also, and developed related theory.

After leaving Cambridge, Brooks worked as a tax inspector in London and did not pursue mathematics further, though he continued to play with finding ways to "square squares" throughout his life. He died in Croydon, England. Brooks was a very private person and avoided making biographical information public.

Sources: wikipedia, squaring.net, talk by Bjarne Toft, and rootsweb.

**Eugène Charles Catalan (1814—1894).**

Eugène Charles Catalan was a French mathematician whose work included important results in number theory, geometry, and combinatorics. He was born in Bruges, which is now part of Belgium; at that time it was governed by France under Napoleon. Catalan considered himself French, even though he was only 1 when the Netherlands took control of Bruges in 1815. On his birth certificate, Catalan is registered as Eugène Charles Bardin; his mother was 17 and unmarried at the time of his birth. His parents were married in 1821 when Catalan was 7; his father acknowledged him and he took his father's surname of Catalan.

Catalan's father was described as a jeweller; he made his living in a variety of ways. Catalan briefly apprenticed as a jeweller at the age of 10. His family moved to Paris shortly thereafter, in about 1825. By this time his father was an architect, and Catalan entered school to learn this profession also.

At school, Catalan showed an aptitude for mathematics. Catalan passed the entrance examinations for the École Polytechnique in 1833. He had strong political opinions but did not take an active part in the many disturbances of the peace in Paris during this time. Nonetheless, along with all of his fellow students he was dismissed from the school and had to apologise for an "act of insubordination" before being readmitted. He graduated in 1835 and took a position teaching in Châlons-sur-Marne. He also published a number of research papers over the next few years.

Catalan wanted to return to Paris but was unsuccessful in applications for positions there. Liouville advised him to obtain additional qualifications. Catalan took Liouville's advice, and obtained a double baccalaureate in 1839 and a doctorate in 1841. His advisor was Joseph Liouville, and Catalan wrote two theses: "Attraction d'un ellipsoïde homogène sur un point extérieur ou sur un point intérieur" ("Attraction of a homogeneous ellipsoid on an exterior point or on an interior point") in mechanics, and "Sur le mouvement des étoiles doubles" ("On the movement of double stars") in astronomy.

Catalan took a more prominent role in politics and political unrest during this period, and this interfered with the advancement of his academic career. He took an active role in the revolution of 1848 that led to the Second Republic, and sat in France's Chamber of Deputies. When Louis-Napoléon Bonaparte assumed absolute power in 1851, Catalan refused to swear allegiance and lost the positions he still held. He continued to tutor and publish, but held no permanent positions until 1865, when he was appointed chair of mathematics at the University of Liège (Belgium). He held this position until his retirement in 1884, and remained in Liège until his death in 1894.

In addition to coming up with the Catalan numbers (see Section 9.2), Catalan published many important results in number theory, geometry, and combinatorics, and a surface that he discovered is also named for him. Catalan formulated a famous conjecture in number theory that was not proven until 2002.

Catalan was elected to many national and regional academies of science, which did not include the French Academy despite several attempts. He was awarded the Knight's Cross of the Légion d'Honneur, the Cross of the Knight of the Order of Léopold, and was made an Officer of the Order of Léopold in 1890. The Royal Academies for Science and the Arts of Belgium has named a prize after him that is awarded every five years for important progress in pure mathematics by a citizen of the European Union.

Sources: wikipedia, St. Andrews' math history web site, wikipedia, and the Math Genealogy Project.

**Maria Chudnovsky (1977—).**

Maria Chudnovsky is one of the foremost graph theorists of our time. She is particularly famous for her proof, in a monumental paper (with co-authors), of the Strong Perfect Graph Theorem. Chudnovsky was born in the USSR, and lived in Leningrad (now St. Petersburg, Russia) until

the age of 13, when her family moved to Israel. She finished school in Haifa, and completed both her bachelor's (in 1996) and Master's (in 1999) degrees at the Technion there. She was also completing her mandatory service in the Israel Defense Force from 1996 to 1999.

After her Master's, Chudnovsky moved to the United States to undertake graduate work at Princeton University in New Jersey. She worked under the supervision of Paul Seymour (1950—), and received an M.A. in 2002 and her Ph.D. in 2003. Her thesis was entitled "Berge Trigraphs and Their Applications", and as discussed in Section 14.3, it resulted in the proof of the Strong Perfect Graph Theorem.

Chudnovsky held a post-doctoral fellowship at the Clay Institute in Boston, Massachusetts from 2003—2008, also returning to Princeton as a Veblen Research Instructor (with a year at the Institute for Advanced Study) in 2003—2005, and then an Assistant Professor in 2005. She moved to Columbia University in 2006, and held the Liu Family Professorship in Industrial Engineering and Operations Research there in 2014. Chudnovsky returned to Princeton as a professor in 2015.

Chudnovsky was married in 2011, and has a son born in 2013. She has appeared in commercials for TurboTax and Comfortpedic (as herself, a brilliant mathematician). The TurboTax commercial can be found on YouTube.

As a young, prominent female mathematician, Chudnovsky has been in great demand for outreach and interviews as well as to speak at conferences. She is also a proud member of the Jewish scientific community. She has taken all of this seriously, and has served as a model to many. She has supervised 10 Ph.D. students and 6 post-docs as of 2021, as well as supervising research by undergraduate and Master's students. She has given the advice: "Don't let your self-doubt scare you too much. Just accept that everyone has their moments when they feel like a complete misfit. Just keep pushing."

Chudnovsky's research is in structural graph theory. In addition to her proof of the Strong Perfect Graph Theorem, some of her more important contributions include finding a polynomial-time algorithm to identify perfect graphs, and determining the structure of claw-free graphs. In 2004 Chudnovsky was named in the "Brilliant 10" by *Popular Science* magazine, and in 2012 she received a "genius award" from the MacArthur Foundation. She received the D.R. Fulkerson Prize in 2009, jointly with George Neil Robertson (1938—), Seymour, and Robin Thomas (1962—2020). She was an invited speaker at the International Congress of Mathematicians in 2014.

Sources: wikipedia, Princeton, Clay Institute, and the Intrepid Mathematician. Updated and confirmed through personal communication.

**Václav Chvátal (1946—).**

Václav Chvátal (known as "Vašek") is a mathematician from the former Czechoslovakia, whose career has been spent in North America and whose work has been influential across a number of branches of combinatorics. He was born in Prague, Czechoslovakia, where he studied mathematics at Charles University. He and his first wife fled Prague shortly after the Soviet invasion in 1968, and moved to Canada. By the time he enrolled as a Ph.D. student at the University of Waterloo in 1969, Chvátal had already published 6 papers, the first at the age of 19.

Chvátal took only a year to complete his doctorate under the supervision of Crispin Nash-Williams, with a thesis entitled "Hypergraphs and Ramseyan Theorems". He held a series of positions for relatively short periods over the next 8 years, bouncing back and forth between Montreal, Quebec, and Stanford, California. He started at McGill University in 1971; went to Stanford University for 1972; to the Université de Montréal from 1972—1974; back to Stanford from 1974—1977; and to the Université de Montréal from 1977—1978. He then took a position at McGill University again, where he stayed for longer, from 1978—1986.

In 1986 Chvátal moved to New Jersey to take a position at Rutgers University, where he worked from 1986—2004. He finally returned to Montreal in 2004, and held the Canada Research Chair in Combinatorial Optimization at Concordia University from 2004—2011, followed by the Canada Research Chair in Discrete Mathematics from 2011 until his retirement in 2014.

In addition to Chvátal's work on Hamilton cycles (which goes much further than the result mentioned in Section 13.2), he also proved significant results relating to hypergraphs, algorithmic complexity, linear programming, optimisation, and perfect graphs. In addition, Chvátal worked with David Applegate, Bob Bixby, and Bill Cook on the development of the record-breaking computer code Concorde for solving the Travelling Salesman Problem. He found the smallest triangle-free class-one graph in which every vertex has valency 4 (on 12 vertices), and it is named for him.

In a tribute to Claude Berge, whose book introduced him to graph theory and with whom he later developed a close friendship, Chvátal wrote that this experience "took me through the looking glass to enchanted worlds where I found myself." He has published more than 125 papers and 5 books. He supervised at least 12 doctoral students, and mentored many other young researchers. His talent in writing extends beyond math: in 1971 he wrote a prize-winning short story, *Déjà Vu*.

Sources: Concordia University, wikipedia, Vašek Chvátal: A Very Short Introduction, In Praise of Claude Berge, and the Math Genealogy Project. Updated and confirmed through personal communication.

**Gilles Civario (1972—).**
Gilles Civario is a high performance computing consultant from France, known in combinatorics for his contribution to the study of Sudokus. He was born in 1972, and began his career as a consultant for the CS Group Information Technology company near Paris, where he worked from 1999—2003. He then spent four years at Bull Information Technology, from 2004—2008, in the Grenoble region.

In 2008, Civario moved to Dublin, Ireland. He spent 8 years working as a Senior Software Architect for the Irish Centre for High-End Computing (ICHEC). This is a national service that provides computational resources and expertise for scientific research in Ireland. It was in this role that Civario carried out the computations necessary to determine that no 16-clue sudoku puzzle has a unique solution, as mentioned in Section 16.1. The computation took approximately 7,000,000 core hours and was being worked on for most of 2011.

Since leaving the ICHEC in 2016, Civario has returned to France, where he works for Dell Technologies as a High Performance Computing application specialist.

Sources: LinkedIn, "Sudoku meets Knights Corner", ICHEC, and the Irish Times. Updated and confirmed through personal communication.

**Karen Linda Collins (1959—).**
Karen Linda Collins is an American mathematician best known for her research in graph theory. She obtained her B.A. from Smith College in 1981, where she completed her honours thesis under the direction of Michael Owen Albertson (1946—2009). From there she went to the Massachusetts Institute of Technology (MIT). She worked at MIT under the supervision of Richard Stanley, and obtained her Ph.D. in 1986 with a thesis entitled "Distance Matrices of Graphs". She participated in the AT&T Bell Labs Research Program for Women during her graduate student summers, under the direction of Ron Graham and Fan Chung-Graham.

Collins started working at Wesleyan University in Connecticut immediately after completing her doctorate, in 1986, and says that she is thrilled to still be there. She enjoys teaching at all levels, and particularly likes to work with students on research projects. She served as chair of the department of Mathematics and Computer Science in 2007—2010, and is currently (in 2021) serving again. In 2018 she became the Edward Van Vleck Professor of Mathematics at

Wesleyan. Her husband, Mark Hovey, is a professor of math at Wesleyan, and currently (in 2021) an Associate Provost.

Collins has taken an active role in the combinatorics and graph theory community in the northeastern United States throughout her career, and is currently a co-organiser of the Discrete Math Days in the Northeast. She is best known for her work with Albertson on graph homomorphisms, and the distinguishing number of graphs (mentioned in Section 12.5) and her work with Ann Trenk of Wellesley College on the distinguishing chromatic number of graphs. She has published at least 30 articles and has supervised at least 6 Ph.D. students as well as 9 other graduate students as of 2021. She is the co-author with Trenk of a chapter on split graphs in the book "Topics in Algorithmic Graph Theory", which is part of the *Encyclopedia of Mathematics and its Applications*.

Sources: Wesleyan University, wikipedia, the Wesleyan Argus, and the Math Genealogy Project. Updated and confirmed through personal communication.

**Gabriel Andrew Dirac (1925—1984).**
Gabriel Andrew Dirac was a mathematician whose contributions to graph theory helped to establish this field of mathematical research. Born in Budapest, Hungary, and educated in the U.K., he spent most of his career in Denmark. At birth his name was Gábor Balázs. When he was 12, his mother married the physicist and Nobel Laureate Paul Dirac, and the family moved to England. Dirac and his sister were formally adopted, and took the surname of their stepfather.

Dirac began his mathematical studies at St. John's College, Cambridge in 1942. He interrupted his education in 1944 to work in the aircraft industry during the war. He obtained his Master's in 1949, and then went to the University of London for his Ph.D. He studied under Richard Rado (1906—1989) there, and received his doctorate in 1951 with a thesis entitled "On the Colouring of Graphs: Combinatorial topology of Linear Complexes".

Dirac had a somewhat peripatetic career, with appointments at universities in England (London), Canada (Toronto), Austria (Vienna), Germany (Hamburg and Ilmenau), Ireland (Dublin), and Wales (Swansea). His longest affiliation was with the University of Aarhus in Denmark, where he worked briefly in 1966. He returned there in 1970, and remained until his death in 1984 at the age of 59. His appointment in Dublin was to the Erasmus Smith professorship, from 1964—1966.

Dirac began to study graph theory when the field was still very young and not well-respected outside of Hungary (in Hungary there was an active group of researchers). Dirac's work brought broader recognition to the field. In addition to his important research on Hamilton cycles (mentioned in Section 13.2), Dirac made significant contributions to research into colouring and critical graphs. He also published results in number theory and geometry. In addition to his significant role in the development of graph theory, Dirac made notable contributions to the development of mathematical research in Denmark. Dirac was highly influential in the careers of many young researchers.

Sources: wikipedia, tribute by Carsten Thomassen, Aarhus University obituary, and the Math Genealogy Project.

**Pál Erdős (1913—1996).**
Pál (Paul) Erdős was one of the most influential mathematicians of the 20th century. It is hard to condense the life of Erdős into a brief biographical sketch. He was incredibly prolific, generous, collaborative, and also eccentric. Erdős was born in Budapest, Hungary to parents who were born Jewish but did not practice the religion. His two sisters both died of scarlet fever while his mother was in the hospital for his birth, leaving him an only child who grew up in the shadow of this tragedy. His parents were both teachers of mathematics, and Erdős

showed an early interest in math. During his young childhood, from 1914—1920, his father was absent, as a prisoner of war in Siberia.

Although anti-Jewish laws prevented most students of Jewish descent from entering university in Hungary at that time, Erdős won a national examination and was allowed to enrol in 1930 at the age of 17, at the university of science in Budapest (now named for physicist Loránd Eötvös). He obtained his doctorate there in 1934 (age 21), under Fourier analyst Lipót (Leopold) Fejér, with a thesis entitled "Über die Primzahlen gewisser arithmetischer Reihen" ("On the prime numbers in certain arithmetic progressions"). The situation in Hungary was becoming untenable for people who had Jewish heritage, and Erdős managed to find a postdoctoral fellowship at the University of Manchester, in England. He continued to visit Hungary when he could. Following world events in 1938, he moved to the United States. Most of his remaining close relatives were murdered during the Holocaust. His mother did survive, and he spent time with her often after he became able to visit Hungary without fear of being unable to leave. In fact, later in life at the age of 84, Erdős's mother joined him in his travels. She journeyed with him around the globe for the next 7 years until her death in Calgary, Canada in 1971.

The first mention of Erdős in this book occurs with the Erdős-Szekeres Theorem. This result appeared in the same paper as his work with Szekeres on the Happy Ending Problem (see Section 14.2). Erdős was also jointly responsible with Alfréd Rényi (1921—1970) for one of the random graph models discussed in Section 11.5. In related probabilistic work, the two also showed that almost every graph has no nontrivial automorphisms, as mentioned in Section 12.5.

Erdős held some temporary appointments through the next sixty years. For the most part, he lived as a professional itinerant scholar and collaborator. He travelled constantly, with only a briefcase early on, though later he acquired a suitcase. His focus was entirely on mathematics: many life skills such as cooking, driving, handling finances, and even tying his own shoes he acquired late, or never. He visited people, enthusiastically talked mathematics, and collaborated, right up to the day of his death in 1996. Erdős was welcomed around the world for his kindness, generosity, and humour.

Erdős had remarkable insight and a breadth of knowledge that resulted in his making significant contributions to a number of fields of mathematics, including combinatorics, number theory, set theory, classical analysis, and probability. He also built foundations of entirely new areas including Ramsey theory, transfinite combinatorics, probabilistic number theory, and probabilistic combinatorics. It is no accident that his name comes up repeatedly in this book. He cultivated his talent for asking interesting questions, and for finding people to solve them with. His lack of official positions meant that most of his mentoring was either unofficial, or came in the form of co-authorship. He had more than 500 coauthors, and published about 1500 research papers. It was this prolificacy that led to the idea of the "Erdős number" for mathematicians: how many degrees of separation are there between your collaborators, and Erdős? His co-authors have Erdős number 1; theirs have Erdős number 2, and so on. More than 12,000 people have an Erdős number of 2, and more than 80,000 people have an Erdős number of 5 (mine is 3).

Erdős had virtually no money. He lived very frugally aside from the cost of his travel itself (he generally stayed with mathematicians), and was extraordinarily generous. He often offered monetary prizes (anywhere from $25 to over $1000) for solutions to problems he posed, and gave money away whenever he saw a need. He said, "I never wanted material possessions. There is an old Greek saying that the wise man has nothing he cannot carry in his hands. If you have something beautiful, you have to look out for it, so I would rather give it away." Erdős had many unusual philosophies and perspectives. He also had his own terminology for a variety of things: for example, children were "epsilons", as were budding mathematicians. Each time he reunited with friends, he would ask, "Who are the new epsilons?" and would promptly invite

the gifted youngsters to lunch. Erdős met and mentored many mathematicians (Béla Bollobás, Lajos Pósa, Attila Máté, Noga Alon, and Imre Ruzsa to name just a few) while they were still in their teens or even younger.

Erdős believed passionately in the beauty of mathematics, and liked to say that "the SF has this transfinite Book that contains the best proofs of all mathematical theorems, proofs that are elegant and perfect…. You don't have to believe in God, but you should believe in the Book." The "SF" was Erdős's joking way of referring to God as the "Supreme Fascist", to express his frustration over how closely The Book is guarded, and how few glimpses a mortal is allowed. He believed that a mathematician's goal in life should be to have some of these glimpses.

Erdős was awarded the American Math Society's Cole Prize in 1951, Hungary's Kossuth Prize in 1958, and the Wolf Prize from Israel in 1983/4. He received many honorary degrees. An asteroid was named after him in 2021 ("Erdőspál"). A massive wallpainting entitled "Saints Dancing" wraps around the interior of the rotunda of St. Gregory of Nyssa Episcopal Church in San Francisco, showing about 90 saints. Erdős is one of these, dancing between Gandhi and Luther.

Sources: St. Andrews' math history web site, wikipedia, Math Association of America, New York Times, University of California San Diego, University of Chicago, Huffington Post, American Math Society, Education Resources Information Center, Purdue University, and the Erdős Number Project. Additional information and clarification by personal communication from Laci Babai.

**Euclid (c.325BCE—c.265BCE).**
Euclid was a mathematician from the 3rd century BCE, whose writings established mathematics as a deductive science. There is little to say about his life, because little is known. In contrast to other mathematicians who lived at that time and even earlier, there are no contemporary biographies, nor even much in the way of references to him. He seems to have lived most of his life in Alexandria, where he taught and wrote. There are a variety of stories about him, but significant doubt as to the authenticity of any of them. It has even been suggested, not without grounds, that "Euclid" may have been a pseudonym for a team of mathematicians. There is no evidence as to who might have been part of such a team if that were true, and it seems most likely that he was a real person. Whoever he was, he had many students and a school grew from his teachings.

In addition to his most famous work, the *Elements* (mentioned in Section 18.3), remnants of five works by Euclid survive to this day, and a number of other lost works are attributed to him. The *Elements* consists of 13 books, and includes rigorous proofs from geometry and number theory. Many of the results in the *Elements* were known prior to Euclid; practical geometry had received much attention in Greek, Babylonian, and Egyptian culture, and Euclid certainly draws heavily on work known to students of Plato's Academy.

The main contributions of Euclid are the clarity of the writing, and the rigorous proof methods that he insisted upon. Not only are these very well thought through, but this marked a conceptual change in the way in which mathematics is studied and understood as a science. Up to this time, historical mathematics (as studied in Egypt, India, China, and other ancient cultures) was an empirical science like other sciences, full of observations and experimentation. In Euclid's *Elements* for the first time, mathematics was treated as a deductive science. This is a distinction that still sets the study of mathematics apart from all other branches of intellectual endeavour. Euclid was very careful to explicitly lay out the definitions, axioms, and postulates on which his deductions relied, and to use only these in his proofs. He included as axioms things that might seem obvious, such as: "Things equal to the same thing are also equal to

each other". (There are some subtle and unmentioned assumptions in his work that were not noticed for centuries.)

The European Space Agency has been constructing a space telescope that has been named Euclid in his honour.

Sources: wikipedia, St. Andrews' math history web site, and worldhistory.org.

**Leonhard Euler (1707—1783).**

Leonhard Euler was one of the most influential and prolific mathematicians of all time. He was born in 1707 in Basel, Switzerland. His father, a minister, had studied mathematics along with Johann Bernoulli, under the tutelage of Jacob Bernoulli (and in fact living in the Bernoullis' house) as an undergraduate, and passed some of this learning along to Euler. Euler enrolled at the University of Basel in 1720 (not unusually young at the time). He earned a Master's degree in 1723, with a thesis comparing the philosophies of Newton and Descartes, and began to study theology according to his family's wishes.

By this time Euler had brought himself to the attention of Johann Bernoulli, who gave him advice on mathematical books to read, and made himself available on Saturday afternoons to answer Euler's questions. Between his own interest and Bernoulli's influence, Euler eventually stopped studying theology and devoted his attention to mathematics. His studies in Greek and Hebrew remained useful to him in this context. In 1726 he applied for a position in physics at the University of Basel, submitting a thesis on properties of sound ("Dissertatio physica de sono") with the support and advice of Bernoulli.

Although his application for the position in Basel was unsuccessful, Euler's connection with the Bernoullis stood him in good stead. Two of Bernoulli's sons were working at the Imperial Russian Academy of Sciences in St. Petersburg. When one of them died, the other recommended that the vacant position be offered to Euler. He accepted the position, and moved to St. Petersburg in 1727.

Euler thrived in St. Petersburg, where the Academy emphasised research and assigned limited teaching responsibilities. He learned Russian, and served the navy in addition to his mathematical work until 1730. He married while there; he and his wife Katharina had 13 children, of whom only 5 survived to adulthood. There was considerable political turmoil in Russia during these years, and in 1741 Euler moved his family to Germany, where he had been offered a post as a founding member of the Berlin Academy (the Academy was not fully and formally established until 1746). Euler spent 25 highly productive years there, before being lured back to St. Petersburg under very generous terms in 1766. (The conditions of his position in Berlin had also deteriorated.) He remained there until his death in 1783.

Euler lost the sight in his right eye after an illness in 1738 when he was 31. He began to lose his remaining sight in 1766 and lost it completely in 1771, due to a cataract in his left eye followed by surgery that led to an abscess. His phenomenal memory (he could recite Virgil's *Aenaeid*, and knew which line started and ended each page of the edition from which he had learned it) enabled him to continue his extraordinary mathematical productivity: for example, he produced more than 50 papers in the year 1775 alone. He was assisted by scribes in this work, including one of his three sons. These scribes worked out the details of many of his ideas.

Euler's mathematical contributions are immense. In addition to mentoring some highly influential students including Joseph Lagrange, Euler published more than 850 works, including numerous books that span many branches of mathematics and physics. Euler's research contributed to the areas of number theory, combinatorics, probability, infinite series, and partial differential equations, among others. He developed or popularised many familiar pieces of mathematical notation, including introducing the concept of a function and the notation $f(x)$; use of $\Sigma$ for summations; $\sin(x), \cos(x)$ and related trigonometric notation; $e$ for the base of

the natural logarithm (known as "Euler's number"); and $i$ for $\sqrt{-1}$. The identity $e^{i\pi} = -1$ is named after Euler, as its discoverer.

It would be far beyond the scope of this biographical sketch to even outline the most important of Euler's discoveries, but this book should make it clear that he had a fundamental influence on combinatorics. The contributions discussed in this book begin with his solution to the Königsberg bridge problem introduced in Section 11.1 and covered in detail in Section 13.1, and also include Euler's handshaking lemma, Euler's Formula as covered in Section 15.2, and initiating the study of Latin squares as discussed in Section 16.2. One of his most-read works is the compilation of his *Letters to a German Princess* written on a variety of topics, in the role Euler had been asked to fill as the princess' tutor.

Euler has been commemorated on stamps in Germany, Russia, and Switzerland, and has appeared on Swiss banknotes. In 1977, an asteroid that had been discovered in 1973 was named "2002 Euler" in his honour. The Mathematical Association of America supports an electronic archive with web pages devoted to each of Euler's works, in addition to writings about him.

Sources: St. Andrews' math history web site, wikipedia, wikipedia, biography.com, Story of Mathematics, and Purdue University.

**Gino Fano (1871—1952).**

Gino Fano was an Italian geometer who was instrumental in developing the field of finite geometry. He was born in Mantua, Italy, to wealthy Jewish parents. In that same year of 1871, patriotic troops captured Rome to complete the unification of Italy. Fano studied at the University of Turin under the supervision of Corrado Segre, from 1888 until he was awarded his doctorate in 1892. The contents of his dissertation were published in the paper, "Sopra le curve di dato ordine e dei massimi generi in uno spazio qualunque" ("On the curves of a given order and of maximum genus in any space").

At Segre's urging, Fano translated Felix Klein's "Erlangen Program" into Italian as an undergraduate, so it was natural that after graduating from Turin, Fano visited Klein to do what would probably now be called postdoctoral work with him in Göttingen from 1893 to 1894. In 1894, Fano moved to Rome, where he worked as the assistant of Guido Castelnuovo through 1898. Fano then spent a few years working in Messina, before accepting an appointment as a professor in Turin in 1901. He held this position until 1938 when at the age of 67 it was stripped from him because of his Jewish descent (along with all of his memberships in Italian scientific institutions and academies), by the Fascist regime.

Fano had married in 1911 and had two grown sons by this time. The family fled Italy; Fano and his wife spent the war years in Lausanne, Switzerland, where he taught in Italian refugee camps and maintained an association with the university. His sons both moved to the United States, where one (Robert) became a professor of computer science at MIT and the other (Ugo) an atomic physicist, reflecting the breadth of topics in which Fano engaged their interest from childhood. (Ugo Fano has reported being told about Bohr's theory of the atom, first introduced just 10 years previously, over dinner at the age of 12.) Fano returned to Italy with all status restored after the war in 1946. He visited his sons in the United States frequently for the remaining 6 years before his death in Verona, Italy, in 1952.

Fano published almost 150 works, including many textbooks. His work focused on projective and algebraic geometry, and he was a pioneer in finite geometry, as evidenced by his discovery of the Fano plane (see Example 18.4.2). Fano gave invited talks at the International Congress of Mathematicians in its inaugural meeting in 1897, and again in 1928. He was committed to public education; beginning in 1905 he was a teacher and organiser for the "Evening School for Women Workers of Turin", and in 1928 he was awarded a gold medal "Benemerito della Pubblica Istruzione" by the Italian government for this and related work.

Sources: St. Andrews' math history web site, wikipedia, encyclopedia.com, and Unione Italiana Matematica.

**Leonardo Pisano ("Fibonacci") (c.1170—c.1250).**
Leonardo Pisano ("of Pisa") is more often known as "Fibonacci", short for *filius Bonacci* (his father's surname was Bonacci). He was a 13th century mathematician who is largely credited with introducing the digit representations 0 through 9 along with decimal notation into Europe. Fibonacci was born in Pisa, Italy, and grew up in Bugia (now known as "Bejaia"), Algeria, where his father was a diplomat representing the merchants of Pisa. It was in Algeria that Fibonacci was educated in mathematics, giving him an understanding of the Hindu-Arabic number system and of Indian and Arabic mathematical knowledge, which were largely unknown in Europe at that time.

After returning to Pisa in approximately 1200, Fibonacci wrote a number of books, four of which have survived. These were part of his efforts, which also involved travelling and teaching in person, to educate Europeans about the mathematics he had learned. He was a particularly keen proponent of the Hindu-Arabic number system, which used place values and made arithmetic much simpler than the Roman numerals that were being used through most of Europe. With his personal teaching and his most significant book *Liber Abaci*, the "Book of Calculation", Fibonacci was highly influential in spreading Arabic numerals to Europe. This was very important to the development of banking and accounting throughout that continent. Fibonacci was a talented mathematician and his understanding and teaching went far beyond the number system itself, certainly into the realms of geometry and number theory.

One of the problems posed and solved in the *Liber Abaci* involves the speed at which a population of rabbits grows under certain idealised conditions. Calculating the answer, generation by generation, produces what has become known as the Fibonacci sequence (see Section 6.1).

An asteroid discovered in 1982 is named "6765 Fibonacci" in his honour.

Sources: St. Andrews' math history web site, wikipedia, Plus Magazine, and NPR.

**Sir Ronald Aylmer Fisher (1890—1962).**
Ronald Aylmer Fisher was a British statistician and geneticist, whose ground-breaking achievements in these fields have been tarnished by his vocal support for eugenics. He was born a twin in London, England (his twin was stillborn). Fisher studied at Cambridge University from 1909 until 1912, earning a First in Mathematics.

From 1913 until 1919 Fisher worked as a statistician in London. He also taught physics and math at a number of schools. He was not eligible to serve in the armed forces during World War I due to poor eyesight. In 1919 Fisher moved to Hertfordshire, and from then until 1933, he worked at the Rothamsted Experimental Station. This was an agricultural research centre, where he used his statistical skills to analyse vast quantities of data. He developed and published a great deal of theory in the course of his work, including in the area of design theory (for the design of experiments). Fisher taught at University College London from 1933—1939, and at Cambridge from 1940 until his retirement in 1956, holding endowed chairs at both places. After retirement, he emigrated to Adelaide, Australia, where he held a research fellowship and lived until his death in 1962.

Fisher held a strong and sustained belief in eugenics, at least from the time he was in university, when he assisted in the formation of a Cambidge University Eugenics Society. More precisely, he believed that "groups of mankind differ in their innate capacity for intellectual and emotional development" and spoke against the 1950 UNESCO statement "The Race Question" on this subject. Even after World War II in a testimony on behalf of a Nazi eugenicist Fisher expressed his belief that the Nazis had "sincerely wished to benefit the German racial stock". His legacy cannot help but be impacted by these unacceptable views, and some of the honours that were accorded to him have more recently been withdrawn. For this reason, this biography will also be briefer than his accomplishments would otherwise merit.

Despite his unacceptable beliefs, Fisher made huge contributions to the sciences of genetics and statistics, and to design theory. (Fisher's Inequality is stated and proved in Section 17.3.) He unified Darwin's theory of natural selection with Mendel's discoveries about genetic inheritance. He laid down fundamental aspects of the theory of statistics. He received numerous awards and honours, including being knighted in 1952, and being an invited speaker at the International Congress of Mathematicians on two occasions. A minor planet, "21451 Fisher", was named for him.

Sources: St. Andrews' math history web site, wikipedia, the Royal Society, and University of Adelaide.

**Martin Gardner (1914—2010).**
Martin Gardner was an American writer, known particularly for his popular columns about mathematics in the *Scientific American*. He was born in Tulsa, Oklahoma. Gardner enjoyed math and physics, and initially intended to study physics at university. Then he had trouble with calculus, and never took a course in math after high school. He graduated from the University of Chicago in 1936 with a degree in philosophy. He held a variety of jobs over the next 5 years, including reporting, media relations, and social work. During World War II, he served in the Navy for 4 years.

After the war Gardner returned to the University of Chicago briefly for graduate work, but did not complete a degree. From that time on he was a professional writer: sometimes an editor, sometimes a columnist, and sometimes working freelance. For most of these years, he lived in New York, though he moved to North Carolina in retirement. In 2004, several years after the death of his wife, Gardner moved to Norman, Oklahoma where one of his sons lived. He died there in 2010.

Gardner is best known for his column "Mathematical Games" in the *Scientific American*. He also gained a notable reputation as a magician, and as a skeptic and debunker of pseudoscience. He wrote not only about puzzles and games, but also traditional mathematical topics such as knot theory, transfinite numbers, and the four-colour problem (as discussed in Section 15.3), in addition to contemporary mathematical discoveries. In relation to his ability to write about even deep mathematical topics in an accessible way, Gardner downplayed his own understanding, saying, "If you are writing popularly about math, I think it's good not to know too much math." He wrote almost 300 columns over the years, which have been collected and republished in book form. He was very careful to give proper credit and attribution for the ideas he presented; it was through Gardner's columns that notables such as Penrose and the artist Escher first became well-known.

Gardner wrote books, columns, reviews, and articles about many other topics also, including magic, linguistics, and psychics. He was an authority on Lewis Carroll, and wrote annotated editions of several of Carroll's books as well as of a few other classics. Gardner also wrote two novels and a number of short stories. At the age of 95, he wrote an autobiography that appeared posthumously. In all, Gardner wrote or edited more than 100 books.

The asteroid "2587 Gardner" is named for him. The Mathematical Association of America has established an annual lecture that carries his name.

Sources: wikipedia, Scientific American, martin-gardner.org, BBC, and American Math Society.

**Edgar Nelson Gilbert (1923—2013).**
Edgar Nelson Gilbert was an American mathematician who spent his career doing research in the areas of combinatorics and probability for Bell Labs. He was born in Woodhaven, New York. In 1940 he began studying physics at Queen's College of the City University of New York, obtaining his B.Sc. in 1943. After a brief stint teaching math at the University of Illinois in Urbana-Champaign, he moved to Massachusetts to study at the Massachusetts Institute of

Technology (MIT). He obtained his Ph.D. in physics from MIT in 1948, under the supervision of Norman Levinson. His thesis was entitled "Asymptotic Solution of Relaxation Oscillation Problems". While at MIT he met and married his wife; the couple had three children.

After completing his doctorate, Gilbert moved to Whippany, New Jersey, where he took a job with Bell Laboratories. His arrival at Bell Labs was just 2 years after Richard Wesley Hamming (1915—1998) began working there, so they were longtime colleagues. Hamming described Gilbert as someone whose ideas he always found very stimulating. Gilbert worked at Bell for 48 years, from 1948 until his retirement in 1996.

During his career, Gilbert made a number of interesting discoveries, chiefly in the areas of combinatorics, optimisation, and probability. He is responsible for one of the models of random graphs discussed in Section 11.5. His combinatorial work included coding theory (stemming from Hamming's discovery of error-correcting codes) as well as graph theory. Gilbert published 85 research papers.

Sources: wikipedia, Daily Record obituary, the Math Genealogy Project, transcription of talk by Hamming, and prezi.com.

**Wolfgang Haken (1928—).**
Wolfgang Haken is a German-American topologist, best known in graph theory for his 1976 computer-aided proof, with Kenneth Ira Appel (1932—2013) of the Four-Colour Theorem. He was born Wolfgang Rudolf Gunther Haken in Berlin, Germany, but in 1976 dropped his middle names, changing his legal name to Wolfgang Haken. His father was a physicist who had obtained his doctorate under the supervision of Max Planck, and who worked in the German patent office. At the age of 4, Haken decided that counting should start with 0 rather than 1, and unsuccessfully urged his father to patent this idea. He grew up an only child (his brothers having died of scarlet fever), and his mother died when he was 11, just before the start of World War II.

At the age of 15, Haken was drafted into an anti-aircraft battery, which he served in for the remainder of the war. After working briefly as a farm hand and completing his high school equivalency, Haken entered the University of Kiel in 1946. He graduated with a degree in physics and math in 1948, and continued to study in Kiel, under the supervision of Karl-Heinrich Wiese. He obtained his doctorate in 1953, with a thesis entitled "Ein topologischer Satz über die Einbettung $(d-1)$-dimensionaler Mannigfaltigkeiten in $d$-dimensionale Mannigfaltigeiten" ("A topological theorem about the embedding of $d-1$-dimensional manifolds in $d$-dimensional manifolds"). In 1953, Haken also got married; his wife was a student at Kiel who also obtained a Ph.D. in math under the supervision of Wiese, in 1959.

From 1953 until 1962, Haken worked in Munich as an electrical engineer designing microwave devices for Siemans. During this time he continued to complete and publish mathematical research and obtained some impressive results on knot theory and topology that brought him to the attention of academics. He was invited to visit the University of Illinois in Urbana-Champaign (UIUC) for a year in 1962. From there he moved to Princeton (New Jersey), where he worked at the Institute for Advanced Studies for two years before returning to the UIUC as a tenured professor in 1965. He remained there until his retirement in 1998. Almost all of Haken's 6 children went to graduate school and many of them are involved in research or academics.

Haken has always been a keen outdoorsman. In 1956 he fell more than 30 feet while mountain climbing in the alps, a near-fatal accident. He was in a coma for several days, and one of his feet was damaged. He remained an avid hiker, and has been an active participant in the UIUC math department's regular Saturday hikes for decades.

Haken was an invited speaker at the International Congress of Mathematicians in 1978, and received the American Math Society's Fulkerson Prize jointly with Appel in 1979, for the Four-Colour Theorem, discussed in Section 15.3.

Sources: wikipedia, University of Illinois, the Math Genealogy Project, and Project Euclid.

**Halayudha (10th century CE).**
Halayudha was a 10th-century Indian mathematician, poet, and social reformer. Biographical information on Halayudha is contradictory and incomplete. Some sources write of a poet by this name, and others of a mathematician, with different writings attributed to each, while some assert that there was a single person who wrote these books. I will follow those who attribute everything to one individual, but the most definitive source I found on the poet has very detailed biographical information without any mention of the mathematical work.

Halayudha lived in India in the 900s (CE), originally in Manyakheta, where he wrote his poetry. He later moved to Ujjain, where he wrote a commentary "Mrta-sañjīvanī", on a mathematical book by the Indian mathematician Pingala. In this commentary Halayudha writes of the arithmetic triangle as representing Mount Meru, the holy mountain, and discusses its use in describing the number of combinations of long and short syllables that can be included in a line with a fixed number of syllables. This work is described in Section 4.2 and mentioned again in Section 4.3. It is not clear to what extent these ideas were present in the earlier work of Pingala. Halayudha was also famous for his efforts as a social reformer.

Sources: wikipedia, Dhanapāla and His Times, p. 228, Ton Smith, wisdomlib.org, Sanskrit discussion group, The Calcutta Review (pp. 168—183), and Mathematical Reasoning (p. 146).

**Philip Hall (1904—1982).**
Philip Hall was a British algebraist, whose fundamental discoveries in group theory set the stage for much of the more recent research in this field. He was born in London, England; his parents were not married and his father abandoned the family shortly after his birth. His mother used his father's surname for both herself and her son. Hall was too young to be drafted in World War I, but did serve in the Officer's Training Corps as a youth. Hall was raised to be very careful with money, which was not plentiful for his mother. He was extremely generous to students, young people, and charities later in life, when he had more funds.

Hall entered Cambridge University in 1922, and graduated in 1925. He did some work as a research assistant at University College, London, and in 1927 obtained a fellowship to return to Cambridge by writing a dissertation, "The Isomorphisms of Abelian Groups". After a series of fellowships he was appointed as a Lecturer in 1933. During the war years from 1941 to 1945, Hall worked at Bletchley Park, the centre of British efforts to break enemy codes and ciphers. His focus was on Italian and Japanese ciphers, and he learned to read Japanese characters as part of this work (he subsequently took great pleasure in Italian and Japanese poetry).

In 1945 Hall returned to Cambridge. He worked his way up through the ranks and was promoted to the Sadleirian Chair when it was vacated in 1953. He served as president of the London Mathematical Society from 1955—1957. Hall was significantly tied down by caring for his elderly mother in the late 1950s and early 1960s. His mother died in 1965 at the age of 93, and Hall retired two years later. In his retirement he engaged in a hobby of studying family connections between people who lived to be at least 90 (like his mother). Another retirement hobby of his was memorising a sonnet each day.

Hall's Theorem, published in 1935, and discussed in Section 16.3 notwithstanding, Hall's work was almost exclusively in the area of group theory, where he made many fundamental discoveries. (His motivation for Hall's Theorem also came from group theory.) Despite the coincidence of surnames, he should not be confused with Marshall Hall, Jr., who wrote one of the best-known reference books on group theory.

Hall's influence as a mentor was also significant. He supervised at least 34 doctoral students. One of the undergraduate students for whom he served as tutor was Alan Turing, with whom he later worked at Bletchley Park. Hall's opinion was highly influential when it came to making decisions about young people who were seeking fellowships, and he favoured the adventurous. One young person was turned down when Hall described him as "somewhat too reasonable" while another was elected after an admiring comment on his "almost repellant originality".

Hall was elected a Fellow of the Royal Society in 1951, and in 1961 was awarded the Sylvester Medal. He also received the London Math Society's Senior Berwick Prize in 1958, and the Larmor Prize and De Morgan Medal in 1965. Hall was invited to speak at the International Congress of Mathematicians in 1940 and again in 1950. He was unable to attend on either occasion (due to the war in 1940 and for family reasons in 1950).

Sources: wikipedia, St. Andrews' math history web site, the Math Genealogy Project, the Royal Society, and the London Math Society.

**Sir William Rowan Hamilton (1805—1865).**
William Rowan Hamilton was a 19th century Irish astronomer and mathematician, best known for his discovery of the quaternions. He was born in Dublin, Ireland. By the age of 3, Hamilton had been sent to his uncle in Trim (a curate who ran a local school) to be educated. His uncle, a linguist, trained Hamilton in languages, and by the age of 13 he had at least a basic understanding of an impressive number. These included French, Italian, Greek, Latin, Hebrew, Persian, Arabic, Hindustani, Sanskrit, Marathi, and Malay.

In 1823 Hamilton enrolled at Trinity College in Dublin. He studied math and the classics, and obtained a bachelor's degree in 1827 and a Master's in 1837. While still an undergraduate student he was appointed Andrews professor of Astronomy, which came with the title of Royal Astronomer of Ireland. From that time he lived at Dunsink Observatory until his death in 1865.

Although Hamilton's right to have his name associated with Hamilton cycles and paths (as discussed in Section 13.2) is questionable, he does deserve credit for many important discoveries in math and physics, including results in optics, dynamics, and the discovery of the quaternions. Quaternions exist in four dimensions, but can be identified with three-dimensional geometry in ways somewhat similar to the identification of complex numbers with the plane, and using this identification results in the fastest known methods for determining the outcome of successive rotations in three-dimensional space. Quaternions are therefore frequently used in computer graphics and in spaceship navigation, among other applications. In the context of this book, it should be noted that Hamilton also invented and studied "Icosian Calculus", which is really the group of symmetries of the dodecahedron. He used Hamilton cycles on the dodecahedron to help him understand the structure of this group.

Hamilton had a lifelong interest in poetry (first spurred by an early disappointment in love) and became friends with the poet Wordsworth. Wordsworth preferred the poetry of Hamilton's sister Eliza to Hamilton's own, and at one point wrote Hamilton, asking him pointedly "whether the poetical parts of your nature would not find a field more favourable to their nature in the regions of prose, not because those regions are humbler, but because they may be gracefully and profitably trod, with footsteps less careful and in measures less elaborate".

Hamilton was knighted in 1835, and was elected president of the Royal Irish Academy in 1837, a position he retained until 1846. He has two research institutes named for him, as well as a public lecture. Commemorative Irish stamps and a silver proof coin were also designed and issued in his honour.

Sources: wikipedia, St. Andrews' math history web site, Physics World, New World Encyclopedia, and the Irish Times.

**Richard Wesley Hamming (1915—1998).**
Richard Wesley Hamming was an American mathematician whose discovery of error-correcting codes had significant impact on the development of telecommunications and computer engineering. He was born in Chicago, Illinois (in the United States) in 1915. He was applying to universities during the Great Depression, and the only scholarship offer he received was to the University of Chicago. Hamming wanted to study engineering, but Chicago did not have an engineering school, so he completed a degree in science, graduating in 1937.

Hamming went to the University of Nebraska for his Master's degree, graduating in 1939, and returned to Illinois for his Ph.D. studies at the University of Illinois at Urbana-Champaign. His doctoral thesis was "Some Problems in the Boundary Value Theory of Linear Differential Equations", written under the supervision of Waldemar Trjitzinsky, and he completed it in 1942. He was also married in 1942.

After teaching for a couple of years at the University of Illinois, in 1944 Hamming took a position at the University of Louisville in Kentucky. He left this job in 1945 to work on the Manhattan Project (development of the atomic bomb) at the Los Alamos Laboratory in New Mexico. He had no idea of what he was getting into; a friend who was working there wrote him only that there was "something interesting going on down here. Come down and work." Hamming's position involved programming the IBM calculating machines to compute solutions to equations for the physicists on the project. His wife also joined the project as a human computer.

Hamming wrote a humorous account about having the gravity of the Manhattan Project brought home to him when he was asked to check calculations for the probability that the next test would ignite earth's entire atmosphere, but this seems not to have deterred him too much. The lesson he took from it was that "Mathematics is not merely an idle art form, it is an essential part of our society." He did say in an interview that the Los Alamos lab resembled "the mad scientist's laboratory", and his experiences there brought his keen interest in science fiction to an abrupt end. He remained with the project until 1946, when he took a position at Bell Telephone Labs. In fact, he stayed on at Los Alamos for 6 months after most others had left, feeling a responsibility to create a written record of the work that had been done there.

Hamming continued to work with calculating machines at Bell Labs. At the time, the machines used a parity check-bit for each block of bits in any bit string, so they could determine if an error occurred, but they simply terminated the active process if an error did arise. After experiencing the frustration of setting a computer to run calculations over a weekend only to discover on Monday that the process had terminated in an error, it occurred to Hamming that if the computer could determine which bit is incorrect, then it could correct the error and proceed. He introduced the theory of error-correcting codes (along with the Hamming distance metric) in a landmark 1950 paper. (See Chapter 19 and in particular Section 19.2 for more about Hamming distance and error-correcting codes.) This introduced a new field of study, as well as setting the stage for a quantum leap in the effectiveness of telecommunications and computer technology.

Hamming retired from Bell Labs in 1976, having evaded or avoided management responsibilities throughout his time there. He later saw this as a failure on his part. He held a number of visiting or adjunct appointments at universities during his time at Bell, including positions with Stanford and with Princeton. These positions allowed him to do some teaching in addition to his research. In 1976 Hamming moved to Monterey, California, where he worked in the computer science department at the Naval Postgraduate School. In 1997 he retired from this position and became a professor emeritus. He died soon thereafter, in 1998. Hamming was known to say: "The purpose of computing is insight, not numbers," and to students: "If you don't work on important problems, it is not likely that you will do important work."

Hamming acted as president for the Association for Computing Machinery (ACM) from 1958 to 1960. He earned many honours, beginning by winning the ACM's Turing Award in 1968; he became a Fellow of the ACM in 1994. He became a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) in 1968, and won the IEEE's Emanuel R. Piore Award in 1979 and its Computer Society Pioneer Award in 1980. In 1980 Hamming was elected a member of the National Academy of Engineering. He won the Harold Pender Award from the University of Pennsylvania in 1981. Fittingly, he was named the first recipient of the IEEE's "Richard W. Hamming Medal" in 1988; this is an annual award (with a prize of $10,000) given for "exceptional contributions to information sciences, systems, and technology". Hamming won the Rhein Foundation's Basic Research Award in 1996. He was also given the Navy's Distinguished Public Service Award posthumously.

In addition to his foundational contributions to coding theory, Hamming made significant advances in numerical analysis, numerical integration, and numerical filtering, and along with Ruth Weiss developed one of the earliest programming languages. He also worked on many other projects, and published more than 25 research papers. He became convinced that the ways in which we teach mathematics need to change, and wrote at least 8 textbooks using unconventional approaches.

Sources: wikipedia, St. Andrews' math history web site, IEEE, National Academies Press, and American Mathematical Monthly.

**Percy John Heawood (1861—1955).**
Percy John Heawood was a British mathematician, best known in graph theory for his contributions to the Four-Colour Conjecture. He was born in 1861 in Newport, England. He completed his studies at Oxford, beginning in 1880 and remaining there until 1887.

Heawood became a Lecturer in Mathematics at Durham University in 1887. He worked there throughout his career, and was appointed to the Chair of Mathematics in 1911. From 1926 until 1928 he was the Vice-Chancellor of the university. Heawood retired in 1939 at the age of 78.

One focus of Heawood's research throughout his life was the Four-Colour Conjecture. In 1890 he pointed out the error in Sir Alfred Bray Kempe (1849—1922)'s proof (which had been accepted for 11 years at that point) and adapted Kempe's proof to show that five colours suffice (the Five-Colour Theorem), as discussed in Section 15.3. He wrote 6 additional papers on the topic over the course of his career, with the last one in 1949. He also wrote papers on continued fractions, quadratic residues, and geometry. There is a graph on 14 vertices that is named after him. (This graph can be embedded on a torus so that 7 colours are required to colour its faces, and Heawood proved that seven colours suffice in general on a torus.)

Heawood lived to the age of 93, and was married for over 60 years (with 2 children). He had a singular and eccentric appearance. Gabriel Andrew Dirac (1925—1984) wrote about Heawood in the Journal of the London Math Society: "He had an immense moustache and a meagre, slightly stooping figure. He usually wore an Inverness cape of strange pattern and manifest antiquity, and carried an ancient handbag. His walk was delicate and hasty, and he was often accompanied by a dog, which was admitted to his lectures."

Durham Castle required extensive funds in 1928 to save it, as there was a problem with its foundation. The University of Durham tried to fundraise for the restoration of the castle, but only raised about one-fifth of the £150,000(GBP) that were ultimately required. Heawood doggedly continued working to fundraise for this cause for years, and ultimately succeeded. For these efforts he was awarded the Order of the British Empire in 1939.

Sources: wikipedia, St. Andrews' math history web site, Journal of the London Math Society, wikipedia, and the Times obituary.

**Jeannette Janssen (1963—).**
Jeannette Janssen is a Dutch-Canadian combinatorist, best known for her work in graph theory. She was born in the Netherlands in 1963. She enrolled at Eindhoven University of Technology in the Netherlands in 1982, and earned her first university degree there, graduating in 1988. She then spent two years lecturing at the Universidad de Guanajuato in Mexico. Janssen moved to the United States for her Ph.D., studying at Lehigh University in Bethlehem, Pennsylvania under the supervision of Edward Assmus, Jr. She completed her doctorate in 1993 with a dissertation entitled "Even and Odd Latin Squares".

After her Ph.D., Janssen took a postdoctoral position in Montreal, Canada, working jointly at the Université du Québec à Montréal and Concordia University. She also worked for 2 years at the London School of Economics in England, followed by a year at Acadia University in Nova Scotia, Canada, before moving to Dalhousie University in Halifax, Nova Scotia in 1998. Janssen has remained at Dalhousie University, where she became the first female chair of the math department in 2016. She directed the Atlantic Association for Research in the Mathematical Sciences from 2011 to 2016. Janssen was also chair of the Discrete Math Activity Group for the Society of Industrial and Applied Math for 2021—2022.

Janssen's research focuses on graph theory, in particular graphs as models of complex networks, and graph colouring. One of her most notable results can be expressed in graph theoretic terms, but also relates to extending Latin rectangles. Another of her results is discussed in Section 11.5. Janssen has published more than 60 research papers, and has supervised 5 Ph.D. students as of 2021.

Sources: Dalhousie University, wikipedia, LinkedIn, and the Math Genealogy Project. Updated and confirmed through personal communication.

**Peter Keevash (1978—).**
Peter Keevash is one of the foremost combinatorists of our time, best known for his resolution of the long-standing existence problem for combinatorial designs. He was born in London, England in 1978. His family subsequently moved to Brighton, and then to Leeds. Keevash showed an early interest in mathematics. He was a member of the team representing England at the International Mathematical Olympiad (a competition for high school students) in 1995, where he earned a bronze medal.

In 1995, Keevash enrolled at Trinity College of the University of Cambridge. He completed his bachelor's degree there in 1998. At this point he moved to the United States for his doctoral studies. In 2004 Keevash completed his Ph.D. at Princeton University in New Jersey, under the supervision of Benjamin Sudakov. His thesis was entitled "The Role of Approximate Structure in Extremal Combinatorics".

Keevash held a postdoctoral position at the California Institute of Technology in Pasadena, before moving back to England to work at Queen Mary, University of London, initially as a lecturer and then as a professor. In 2013 Keevash accepted a position at Oxford University.

Keevash has made major contributions to multiple areas of combinatorics, including extremal combinatorics, graph theory, Ramsey theory, and design theory. He proved the best known lower bound for the Ramsey number $R(3,k)$. However, he is best known for his 2014 result in design theory, specifically Keevash's Theorem mentioned in Section 18.2, establishing the existence of $t-(v,k,\lambda)$ designs whenever the necessary conditions are met and $v$ is sufficiently large. Keevash has more than 70 publications in all.

Keevash won the European Prize in Combinatorics in 2009. In 2018, he gave an invited talk at the International Congress of Mathematicians.

Sources: Oxford University, Oxford University, wikipedia, International Mathematics Olympiad, and the Math Genealogy Project. Updated and confirmed through personal communication.

**Sir Alfred Bray Kempe (1849—1922).**
Alfred Bray Kempe was a British barrister, who also carried out noteworthy mathematical research. He is best known for his attempted solution of the Four-Colour Conjecture. He was born in London, England, and studied at Trinity College, Cambridge University. After graduating with a B.A. in 1872, Kempe became a barrister in 1873. His father was a minister, and Kempe specialised in ecclesiastical law, ultimately becoming Chancellor (legal advisor) for several dioceses. He was highly respected in his chosen profession.

Despite his career in law, Kempe maintained an active interest in mathematics. He published papers sporadically from the 1870s through the 1890s, with his first paper appearing in the year he graduated from Cambridge. In 1881 he was elected a fellow of the Royal Society, and served as its treasurer and vice-president for 21 years. While he held this role, the Royal Society developed the National Physical Laboratory from its inception and ultimately handed it over to the government after 16 years. Kempe also served as president of the London Math Society (LMS) from 1892 to 1894 (many biographies list other years for this service, but these are the dates listed by the LMS).

It could be considered a bit sad that Kempe is best known for incorrect proofs. In addition to his "proof" of the Four-Colour Conjecture in 1879 that is mentioned in Section 15.3, in 1876 Kempe published a flawed proof of a result now known as the Kempe Universality Theorem, about a connection between linkages and algebraic curves. This theorem was ultimately proved in 2002, with somewhat different ideas, but a second proof in 2008 was based on Kempe's methods. His techniques of Kempe chains and unavoidable sets were used by Kenneth Ira Appel (1932—2013) and Wolfgang Haken (1928—) in their eventual proof of the Four-Colour Theorem. So for both of these results (as well as for the Five-Colour Theorem), a correct proof ultimately used several of the techniques that Kempe developed for the "proof"s that he published. It is also no mean accomplishment that on a popular topic like the Four-Colour Conjecture, it took researchers 11 years to notice the flaw in Kempe's proof. Kempe also discovered the Petersen graph a decade before Petersen did, but ironically did not notice its relationship to the Four-Colour Conjecture.

Kempe loved the mountains, and travelled to Switzerland frequently to walk and climb in the Alps. He was also devoted to music. He was married twice, and had three children by his second marriage. He received an honorary degree from the University of Durham in 1908, and was knighted in 1912.

Sources: wikipedia, St. Andrews' math history web site, Royal Society obituaries, MIT thesis proving the Universality Theorem with Kempe's ideas, London Math Society, and Nature magazine obituary.

**Reverend Thomas Penyngton Kirkman (1806—1895).**
Thomas Penyngton Kirkman was a 19th-century British minister, who also made extensive contributions to combinatorial research, particularly in the areas of graph theory and design theory. He was born in Bolton, England (and baptised "Thomas Pennington"). The local schoolmaster tried to persuade his family to send him to Cambridge, guaranteeing that he would win a scholarship, but his father (a cotton dealer) would not let him go. At the age of 14 he left school to work for his father. In 1829 he rebelled and left his father's office to attend Trinity College Dublin, working as a tutor to cover his expenses. His schooling had not included any math at all, but he picked up the background on his own. He completed his bachelor's degree in 1833 and moved back to England in 1835.

After university, Kirkman entered the church. He was ordained and became curate of Bury, and later of Lymm. In 1839 he was appointed rector at the parish of Croft with Southworth. He remained there until his retirement 52 years later, in 1892 at the age of 86. Kirkman married in 1841 and had 7 children. He earned extra income through tutoring until his wife came into

an inheritance that together with his income from the church sufficed to support the family. Upon his retirement, Kirkman moved a few miles away to Bowdon, where he died in 1895. His wife died just 10 days later.

Kirkman's ministerial duties were light and he had plenty of time for mathematics, particularly after he gave up tutoring. His first publication came in 1846, and he continued to work on math problems until his death. Kirkman was no mere dilettante; he made a number of significant contributions to combinatorics (only some of which are mentioned in this book), and also published work on geometry, group theory, and knot theory. Over the course of his life he published more than 60 substantial papers in addition to many minor ones. His work on Hamilton cycles is mentioned in Section 13.2, and his work in design theory is discussed in Section 18.1.

Kirkman was elected a fellow of the Royal Society in 1857. The institute of Combinatorics and its Applications awards a Kirkman medal annually, named in his honour.

Sources: wikipedia, St. Andrews' math history web site, Lecture by Alexander MacFarlane, and Bulletin of the London Math Society.

**Eszter Klein (1910—2005).**

Eszter Klein was a Hungarian-Australian mathematics educator. She was born in Budapest, Hungary, to a single mother who was devoted to Judaism (a devotion her daughter did not inherit). Klein is also known by the anglicised version "Esther" of her name, and by her married name of Szekeres, or the hyphenated Szekeres-Klein. During her high school years, Klein was a successful problem solver in the Hungarian High School Mathematical and Physical Monthly which published her photograph in the tableaux of the top 32 problem solvers for each of the years 1926—27 and 1927—28. In each of these years, the tableaux included 30 boys and 2 girls — Klein, and her classmate Márta Wachsberger from the Jewish Girls' High School.

In spite of laws in Hungary discriminating against people of Jewish heritage in university admissions, in 1928 Klein was admitted to the top university of science in the country (now named for physicist Loránd Eötvös (1848—1919)). At the university she became part of a brilliant group of young mathematicians and friends that included Tibor Gallai (who would go on to become a remarkable teacher and an eminent graph theorist, see the biographical sketch of Vera Sós (1930—)), Pál Erdős (1913—1996), György Szekeres (1911—2005), and Pál Turán (1910—1976). They met regularly to talk about math together. Klein, having first solved it in the case of quadrilaterals, posed the "Happy Ending Problem" (see Section 14.2) to this group, and it deepened their relationships. Erdős and Szekeres later published bounds for the general solution to the problem, and in 1937 Klein and Szekeres were married.

Szekeres had been studying chemical engineering, and spent 6 years working in Budapest as a chemist after graduation, to the late 1930s. During this period, Klein and Szekeres were not able to afford to live together, and saw each other only on weekends. By 1939 Hungary was an ally of Nazi Germany, and an increasingly menacing place for people of Jewish descent. Klein moved with Szekeres to Shanghai (which surprisingly and remarkably served as a safe haven for 20,000 Jewish refugees from Europe) to escape the Nazis. On their first arrival in China, exuberant at having escaped from Europe, they decided to start their family, and their oldest child, a son, was born in Shanghai in 1940. However, life in Shanghai was not easy. They lived through the Japanese occupation and the start of the Chinese Communist revolution. For years, they were without work, living on meagre refugee aid. But while back in Hungary Klein's mother was murdered in the Holocaust, the Szekeres-Klein family in Shanghai survived.

In 1948 Szekeres was offered a position at the University of Adelaide, and the family moved to Australia. For several years after moving to Adelaide, they shared an apartment with the family of Klein's high school friend Márta Wachsberger, now Marta Sved. Klein's second and

last child, a daughter, was born in 1954. Klein taught high school math and also tutored at the University of Adelaide, while raising the children.

In 1964 the family moved to Sydney, where Szekeres had been offered a professorship at the University of New South Wales. Klein became a lecturer in math at Macquarie University in Sydney. After their retirement, in 1984 Klein and Szekeres established a weekly problem-solving session as well as enrichment lessons for talented high school students, held at Mercy College. Klein supplied geometry problems and attended these sessions weekly until 2002, when it became too difficult for her. This program has expanded to about 30 groups across Australia and New Zealand.

In 2004 Klein and Szekeres moved back to Adelaide. Not long thereafter, Klein had to move into a nursing home, and Szekeres joined her there in late 2005. Seven weeks later, the two died together, within an hour of each other, after almost 70 years of marriage.

In 1990 Macquarie University gave Klein an honorary doctorate. In 1993, she won the B.H. Neumann Award from the Australian Mathematics Trust. This award is given for significant contribution to teaching mathematical problem-solving in Australia.

Sources: wikipedia, Quanta Magazine, Sydney Morning Herald, Women and Math blog, Australian Academy of Science, and Australian Math Society. Additional information and clarification by personal communication from Laci Babai.

**Tamás Kővári (1930—2010).**
Tamás Kővári was a Hungarian-British analyst. He was born in Budapest, Hungary. In later years he often used the anglicised version of his name ("Thomas"). He attended József Eötvös College in Baja, Hungary, beginning in 1948. He graduated from Loránd Eötvös University in Budapest and received his Ph.D. from the Hungarian Academy of Sciences. Kővári was also married in Hungary; his wife had a degree in physics and math from Eötvös University.

In October 1956 there was an uprising in Hungary against the Communist rule and Soviet occupation. It was quickly crushed by Soviet tanks. Before the borders were sealed, 200,000 people escaped the country. The refugees included Kővári and his wife. Eventually they reached London, England, where Kővári became a lecturer and later Reader at the University of London, from 1957 to his retirement in 1995. While working at the University of London, Kővári learned that an English qualification would be useful to him. He completed a Ph.D. under the supervision of Walter Hayman in 1961, with a thesis entitled, "On the Borel Exceptional Values of Lacunary Integral Functions".

Kővári and his wife had two children. The couple divorced in 1972, and Kővári produced very little new mathematics after this time. Kővári enjoyed music, particularly classical and disco. He also loved sweets, and was in a chocolate-tasting club.

It was in 1954, while still in Hungary, that Kővári's joint paper with Sós and Turán was published, giving an upper bound on the number of edges in a bipartite graph with bipartition sets of given size, before it necessarily contains a particular complete bipartite subgraph. This is probably Kővári's best-known work, and is stated more precisely in Section 14.2. He published more than 20 papers during his career; aside from this result in graph theory, all of his other research related to functions of complex variables.

Sources: London Math Society obituary, Stanford University, My Journey Home: Life after the Holocaust, and the Math Genealogy Project, as well as MathSciNet for a publication list. Additional information and clarification by personal communication from Laci Babai.

**Kazimierz Kuratowski (1896—1980).**
Kazimierz Kuratowski was a Polish topologist, best known for his fundamental contribution to graph theory, Kuratowski's Theorem characterising planar graphs. He was born in Warsaw, Poland. Poland at that time was partitioned among Russia, Prussia, and Austria, and Warsaw was under the control of the czar of the Russian empire. Russia had turned the University

of Warsaw into a Russian-language institution in 1869, and stopped permitting high school students to study in Polish. Although Polish high schools were allowed by the time Kuratowski was in high school, students attending such schools had to compete in examinations as external candidates for places in universities. As a result, many went abroad for university.

Kuratowski wanted to become an engineer. He attended the University of Glasgow in Scotland, which had a strong reputation in this area. Kuratowski began his studies there in 1913. After he returned home at the end of his first year, the outbreak of World War I made it impossible for him to return to Scotland to continue his studies. One year later in 1915, Russia withdrew from Warsaw. Germany took control, and changed the University of Warsaw back to being a Polish-language university. Kuratowski was one of the first students when it reopened in this form. At the end of World War I Poland became independent. Kuratowski graduated in 1919 and began doctoral work under Stefan Mazurkiewicz and Zygmunt Janiszewski, remaining at the University of Warsaw. Although Janiszewski died in 1920 at the age of 31, of the "Spanish flu" pandemic, Kuratowski was able to complete his Ph.D. in 1921.

Kuratowski moved to Lwów (now Lviv, Ukraine) in 1927, where he worked at the Technical University of Lwów. He married in 1929. He maintained a home near Warsaw that he visited each summer, and moved back to work at the University of Warsaw in 1934, having been offered a newly-established chair. While maintaining an active research career, from this time forward he also devoted significant effort to the development of mathematics in Poland.

In 1939, Nazi Germany invaded Poland. Many academics were killed or sent to concentration camps, and universities were closed. Kuratowski was reportedly of Jewish descent, and apparently had some narrow escapes from the Gestapo, but managed to survive in Warsaw. He also further risked his life to teach in one of the illegal universities that the Poles secretly established. Of this time, he wrote: "The importance of clandestine education consisted among others in keeping up the spirit of resistance, as well as optimism and confidence in the future, which was so necessary in the conditions of occupation. The conditions of a scientist's life at that time were truly tragic. Most painful were the human losses." He was reportedly completely imperturbable in later life, saying that after what he had already lived through during the war, nothing could ever upset him anymore.

After World War II education in Poland had to be rebuilt, and Kuratowski was at the centre of the mathematical part of this work. He served as president of the Polish Math Society for 8 years immediately after the war. He became director of the Mathematical Institute of the Polish Academy of Sciences in 1949, shortly after its establishment, and held this role for 19 years. He also served as a vice president of the Polish Academy of Sciences from 1957 to 1968, and of the International Mathematics Union from 1963 to 1966.

Kuratowski's research was primarily in topology and set theory, but he also discovered one of the central results in graph theory, Kuratowski's Theorem. The notation for complete graphs derives from his name (as mentioned in Section 11.3). He produced more than 150 publications, and his mentoring was broad and very influential. Kuratowski was elected to a number of national academies of science, and also received several honorary degrees. The Polish Mathematical Society has established a prize that is named for him.

Sources: wikipedia, St. Andrews' math history web site, the Math Genealogy Project, Prominent Poles, and obituary in the Polish Review.

**Clement Wing Hong Lam (1949—).**
Clement Wing Hong Lam is a Chinese-Canadian mathematician and computer scientist, best known for his computer-aided proof that there is no projective plane of order 10. He was born in Hong Kong, but left in 1968 to study at the California Institute of Technology, where he completed a B.Sc. in 1971 and a Ph.D. in 1974. His Ph.D. thesis, under the supervision of

Herbert Ryser, was entitled "Rational $g$-circulants satisfying the matrix equation $A^2 = dI+\lambda J$". He spent a year as a visiting assistant professor at the University of Calgary in Canada.

Lam moved to Montreal in 1975, where he took up a position in the computer science department at Concordia University. He remained there until retirement, and is now an emeritus professor there.

Lam determined that there is no projective plane of order 10 (equivalently, no affine plane of order 10, and no set of 9 mutually orthogonal latin squares of order 10), as mentioned in Section 16.2. In his expository paper "The Search for a Finite Projective Plane of Order 10", Lam wrote that his advisor Ryser had warned him against working on the projective plane of order 10 as his thesis topic as it might not lead to a successful thesis. So it was not until 1980 that Lam began working on this problem. It took almost 10 years of theory, algorithm development, letting the computers run, and patching issues before the result was complete in 1989. (Premature publicity came about on a couple of occasions before this.)

In 1992, Lam received the Lester Randolph Ford Award from the Mathematical Association of America for his expository article. He received the Euler Medal in 2006 from the Institute of Combinatorics and its Applications.

Sources: wikipedia, Concordia University, the Math Genealogy Project, and Mathematical Association of America. Updated and confirmed through personal communication.

**Lu Jiaxi (1935—1983).**

Lu Jiaxi was a Chinese teacher of physics who worked under very adverse conditions and received little recognition during his lifetime for his mathematical research despite proving impressive results in design theory. Lu was born in Shanghai, the only survivor of his parents' four children. After his father died when he was 14, he began working to help support himself. He taught himself high school material while working, as well as learning Russian, English, and Japanese in order to be able to read research papers.

Lu was admitted to university at Jilin Normal University (now Northeast Normal University) in 1957 to study physics. After graduating in 1961, he became a physics teacher but was also in charge of a school-run factory producing radio components. He had learned of the Kirkman Schoolgirl Problem in 1956 from a popular science book, spent a lot of time considering it while at university, and in 1961 wrote up his proof that solutions exist whenever the obvious necessary conditions are satisfied. Unfortunately, four years were wasted as he submitted his work to journals that were not suitable for high-level original research. Ultimately his work was inappropriately rejected from *Acta Mathematica Sinica* in 1965 as "not really new". The proof by Dwijendra Kumar Ray-Chaudhuri (1933—) and Richard Michael Wilson (1945—) was announced in 1968. Lu's proof has since become available in his collected works.

At this point the Cultural Revolution disrupted Lu's ability to conduct research. In 1972 he was married, and he and his wife started a family. Lu heard about the published solution to the Kirkman Schoolgirl Problem when he managed to resume acquiring copies of research papers to study, in the late 1970s. Although he found this discouraging, he moved on to study the problem of large sets of disjoint Steiner Triple Systems. He published a series of papers on this and on resolvable BIBDs in the early 1980s, mostly in international journals.

Lu remained relatively unknown in China despite increasing international recognition. John Adrian Bondy (1944—) had served as a referee for one of Lu's papers and asked to meet him when attending a conference in China in the summer of 1983, but the local mathematicians did not know who he meant. After this Lu began to be invited to conferences. The school where he worked was unsupportive of his research, refusing to assist him in attending conferences and assigning him extra duties since he apparently had time for other work.

Lu died suddenly of a heart attack on October 30, 1983, at the age of 48.

Sources: wikipedia, "Collected Works of Lu Jiaxi on Combinatorial Designs", "Triple Systems" by Charles Colbourn, p. 8.

**Gary McGuire (1967—).**
Gary McGuire is an Irish mathematician, best known for his computer-aided proof that at least 17 clues are required in a uniquely-solvable Sudoku puzzle. He was born in Dublin, Ireland. He obtained his bachelor's and master's degrees from University College Dublin, and then moved to the United States for his doctorate. McGuire completed his Ph.D. in 1995 at the California University of Technology under the supervision of Richard Michael Wilson (1945—). His thesis was entitled "Absolutely Irreducible Curves with Applications to Combinatorics and Coding Theory".

After completing his doctorate, McGuire held posts at the University of Virginia and the National University of Ireland, Maynooth before returning to University College Dublin, where he is now a Professor of Mathematics. McGuire has produced more than 80 publications, and has supervised 7 Ph.D. students as of 2021.

McGuire was able to show that at least 17 clues are required in a Sudoku that has a unique solution. It was previously known that there are uniquely-solvable puzzles with 17 clues. Although no 16-clue puzzles had been found, it was not certain that one might not exist. Together with his former student Bastian Tugemann who worked with him on this for several years, and with computational assistance from Gilles Civario (1972—) of the Irish Centre for High-End Computing, McGuire proved that no 16-clue puzzle has a unique solution, as mentioned in Section 16.1. This result was officially announced in 2012 and was covered in news stories worldwide, including prominent publications such as *Nature* and *Scientific American*.

Sources: the Irish Times, University College Dublin, the Math Genealogy Project, University College Dublin, and Nature. Updated and confirmed through personal communication.

**Wendy Joanne Myrvold (1961—).**
Wendy Joanne Myrvold is a Canadian computer scientist specialising in graph algorithms. She was born in Sarnia, Ontario, Canada. In her youth her family moved to Signal Mountain in Tennessee. In 1979 she went to McGill University in Montreal, Canada. After switching out of physiology and physics, she graduated with a degree in computer science in 1983. During the summers, Myrvold worked at an amusement park, maintained the software for memory testing the first PCs manufactured at IBM, and provided computational support for the chemistry labs at Imperial Oil and Syncrude.

Myrvold moved to Waterloo for her Master's degree. She began working in numerical analysis, but fell in love with graph theory during a course she took from John Adrian Bondy (1944—). She completed her master's degree in combinatorics and optimization in 1984. Her doctorate in computer science at Waterloo was under the supervision of Charles Colburn, and her thesis was entitled "The Ally and Adversary Reconstruction Problems". After completing her Ph.D. in 1988, Myrvold took a position in computer science at the University of Victoria. She was a professor there until her retirement in 2018, and is now a professor emeritus.

Myrvold has done research on graph theory, graph algorithms, network reliability, topological graph theory, Latin squares, and chemical graph theory. She has more than 60 publications, and had supervised 5 doctoral students as of 2021. The edge addition algorithm mentioned in Section 15.1 that she developed with John M. Boyer (1968—) for either determining that a graph is non-planar, or finding a planar embedding, is the state of the art for this problem.

Myrvold is married with 2 children. She fell in love with ice hockey at McGill and it became a lifelong passion. Her hockey highlights include playing varsity hockey at McGill, competing in the Western Shield and the BC Senior games, and playing with Wayne Gretzky at his Fantasy hockey camp. She also is an ardent skier and loves dance fitness classes.

Sources: wikipedia, University of Victoria, the Math Genealogy Project, and LinkedIn. Updated and confirmed through personal communication.

**Ernest Tilden Parker (1926—1991).**
Ernest Tilden Parker was an American mathematician best known for his work in disproving a conjecture by Leonhard Euler (1707—1783) about Latin squares. He was born in Detroit, Michigan (in the United States). Parker completed his Ph.D. at Ohio State University in 1957, under the supervision of Marshall Hall, Jr. His thesis was entitled "On Quadruply Transitive Groups".

After completing his doctorate, Parker took a job at Remington Rand Univac, which was an early computer company. For recreation, he worked on the problem of mutually orthogonal Latin squares, which fascinated him. H.F. MacNeish proved that there are at least $q - 1$ mutually orthogonal Latin squares of order $n$, where $q$ is the smallest factor in the prime power decomposition of $n$, and conjectured that there are no more than this. Since $21 = 3 \cdot 7$, by MacNeish's conjecture there should have been no more than $3 - 1 = 2$ mutually orthogonal Latin square of order 21, but in 1958 Parker found a set of 4 of them.

Although MacNeish's Conjecture would have implied Euler's conjecture, the failure of MacNeish's Conjecture did not imply anything about Euler's conjecture. Parker found this more of a curiosity than anything else, but his result caught the attention of Raj Chandra Bose (1901—1987) and Sharadchandra Shankar Shrikhande (1917—2020) and spurred them on. They managed to build on Parker's ideas to find a pair of orthogonal Latin squares of order 22. This in turn galvanised Parker, who was able to use a different method to find pairs of mutually orthogonal Latin squares for infinitely many orders of the form $4k + 2$, the smallest of which had order 10 (this was the smallest order for which the answer had been unknown). The three mathematicians then joined together to disprove Euler's conjecture in all cases, earning them the nickname of "Euler's spoilers". All of this work took place in 1958 and 1959, and the final announcement made the front page of the *New York Times* in April of 1959. (See also Section 16.2.) The techniques that Parker and his co-authors developed for disproving Euler's conjecture played a significant role in the proof by Richard Michael Wilson (1945—) of Wilson's Theorem.

Some reports say that Parker found his order 10 examples on a computer, but all of the publications use mathematical methods of construction, and Parker's sister stated that a computer was not used. Parker later became a professor at the University of Illinois at Urbana-Champaign. He published at least 50 papers during his career. The other result for which he is best known is disproving a conjecture by Pál Erdős (1913—1996) and Leo Moser about tournaments.

Parker died in Urbana-Champaign, Illinois, either on December 31, 1990, or in 1991 (accounts differ). A memorial fund providing fellowships for graduate students at the University of Illinois was established in his name.

Sources: wikipedia, University of Illinois, the Math Genealogy Project, Family Search, Bhāvanā mathematics magazine, and Mississippi State University.

**Blaise Pascal (1623—1662).**
Blaise Pascal was a 17th-century French mathematician, philosopher, and physicist, whose work has had lasting influence on all three fields. He was born in Clermont-Ferrand, France. His mother died when he was only 3 and his family moved to Paris in 1631 when Pascal was 8. Pascal suffered from ill health for most of his life, beginning at the age of 2, and his father educated him at home. A talented mathematician himself, Pascal's father initially refused to teach him any math, wanting to ground him in languages and classics. Pascal was drawn to the forbidden and began studying geometry on his own at the age of 12. After he explained to

his father some of the things he had discovered on his own, his father gave him a copy of the *Elements* of Euclid (c.325BCE—c.265BCE).

When Pascal was 15, his father also began to bring Pascal along with him to the regular meetings of a group organised by Mersenne for discussion of mathematics. At the age of 16, Pascal presented his own work to this group, "Essai pour les coniques" ("Essay on conics"), containing a remarkable result in projective geometry that is now known as Pascal's Theorem. This was published in 1640.

Pascal's father had to flee Paris in 1638 due to his opposition to some of Cardinal Richelieu's policies; he left his children in the care of a neighbour. A year later, he was forgiven and appointed a tax collector in Rouen, where the children joined him. At the age of 18 Pascal invented a mechanical calculator, in part to help his father with the endless addition and subtraction associated with his tax work. The machine was expensive and cumbersome, and not widely distributed, although Pascal continued to make design improvements over the next decade. The design was complicated by the fact that the machine calculated amounts of money, and the French currency did not use a decimal system (for example, there were 12 deniers in a sol). This was the second known mechanical calculator (the first having been manufactured in 1624). Pascal's father died in 1651, and Pascal's sister Jacqueline, who had been Pascal's primary caregiver, moved to a convent shortly thereafter.

Pascal, working with Fermat, laid the basis for probability theory, including the concept of expected value. In 1654 he wrote a treatise on the arithmetic triangle (as mentioned in Section 4.3), which included the first explicit statement of the principle of mathematical induction that we know of. Pascal also made significant contributions to physics, particularly in the areas of fluid mechanics and pressure. After having both a serious carriage accident and a religious experience in 1654, Pascal largely abandoned mathematical work in favour of religion and philosophy. He is also famous for his work in literature. This is an extraordinarily cursory summary of Pascal's many contributions to intellectual spheres.

Pascal's health worsened significantly when he was about 18. By the age of 24 he was only able to consume liquids. He died at the age of 39 at his sister Gilberte's home in Paris, and an autopsy found extensive damage in his abdomen and brain.

Pascal has been honoured in many ways. The University of Waterloo runs an annual contest for young teenagers that is named for him. Clermont-Ferrand in France is home to Université Blaise Pascal. There is also a school named after him in the Democratic Republic of the Congo. He has been the subject of several videos. The chameleon in the Disney film *Tangled* is named for Pascal. A unit of atmospheric pressure, and a programming language have also been named after him.

Sources: wikipedia, St. Andrews' math history web site, MIT, University of California at Berkeley, biography.com, and Stanford Encyclopedia of Philosophy.

**Peter Christian Julius Petersen (1839—1910).**
Peter Christian Julius Petersen (who went by "Julius") was an important 19th-century Danish mathematician and educator. He was born in 1839 in Sorø, Denmark. He left school at the age of 15 to work for his uncle as a grocer's apprentice in Kolding, because his family needed the money. About a year later his uncle died and left him some money, so Petersen was able to return to Sorø. He passed the high school exams and enrolled in Copenhagen's College of Technology in 1856. In 1860 he passed the first part of his civil engineering exams. Although he wanted to go on to study math at university, he had run out of money by this time.

From 1859 to 1871 Petersen taught at a private school to support himself. At the start of this time he was still finishing his engineering studies. After saving money from his job for a few years, in 1862 Petersen enrolled in the University of Copenhagen, having passed the entrance exams. He also got married in the same year, and continued to teach while studying, in order to

support his family (he and his wife had 3 children over time). Petersen wrote many textbooks over the course of his career, beginning during this period. A number of his textbooks were about geometry; his best-known book was translated from the original Danish into 8 other languages, and was still in print in French as recently as 1990. His style was considered elegant and concise, sometimes to a fault. He wrote textbooks for every level from the age of 14 through undergraduate studies, and his books were used almost universally in Danish schools in the late 19th and early 20th centuries.

Petersen obtained his Master's degree in 1866 and his doctorate in 1871 from the University of Copenhagen. His thesis was entitled "On equations which can be solved by square roots, with application to the solution of problems by ruler and compass". After completing his doctorate, he began teaching at the Polytechnical School in Copenhagen. He also taught at the Officers School of the Danish Army between 1881 and 1887. He took a post as professor of mathematics at the University of Copenhagen in 1887, and remained there until his retirement in 1909. Petersen died in Copenhagen.

Petersen had broad interests and expertise. He did research in algebra, number theory, geometry, analysis, and mechanics, and published individual articles even more broadly, including such areas as economics, cryptography, and physics. Petersen's work in geometry led him to an interest in graphs, and his 1891 paper "Die Theorie der regulären Graphs" is generally considered to be the first significant paper in graph theory as such. In it he proved (among other things) that every bridgeless cubic graph has a 1-factor. The Petersen graph (shown in Exercise 14.1.18.4 and Example 14.3.13 and discussed in Section 15.3) appears in a paper from 1898. He was sometimes limited by a dogged insistence on maintaining independence in his own thinking, which he cultivated by reading as little as possible of other mathematical work. As a result, he more than once discovered that things he proved were already known.

In addition to his research, Petersen was an important part of developing mathematical research in Denmark. His was one of the first theses to be written in Danish rather than Latin, and he continued to use Danish in much of his writing, to help promote math in Denmark. He was a founding member of the Danish Mathematical Society, which was established in 1873, and was elected to the Royal Danish Academy in 1879. He was one of the three members of the Commission of Education, appointed in 1887 and retaining the position for 13 years; this included the duty of setting problems for standardised exams taken at the end of high school. He was appointed to the Order of the Dannbrog in 1891. He took great pleasure in his work, and said: "When throughout life you have obtained honour and money for enjoying yourself, what more can you ask for!"

Sources: wikipedia, St. Andrews' math history web site, biographical article in *Cryptologia*, and Julius Petersen Graph Theory Centennial volume.

**David Angus Pike (1968—).**
David Angus Pike is a Canadian combinatorist (specialising in graph theory and design theory) who identifies as gay. He was born in 1968 and grew up in Ontario, Canada. He attended the University of Waterloo, where he completed a joint degree in math and computer science. As he was enrolled in the co-operative education programme, Pike held several jobs as a programmer, analyst, or trainer during his undergraduate studies, including stints at the Mutual Life Assurance Company, the Department of National Defence, and McNeil Consumer Products Company. Pike continued to work at McNeil during the summers while completing his Master's degree at Auburn University in Alabama, which he finished in 1994.

Pike remained at Auburn University for his Ph.D. in Discrete Math, which he completed under the supervision of Chris Rodger in 1996. His thesis was entitled "Hamilton Decompositions of Graphs". During his graduate studies, Pike came out as a gay man.

Pike was appointed to a tenure-track position at East Central University in Oklahoma in 1996. He worked there for 2 years before being offered a position in Canada, at Memorial University of Newfoundland (MUN), in the Department of Mathematics and Statistics. He has worked at MUN since 1998; since 2006 he has held a joint appointment in Computer Science. In June 2018 he was named a University Research Professor.

Pike has more than 50 publications, and has supervised about 20 graduate students and post-docs as of 2021, as well as mentoring many undergraduate students. His research includes the work with Atif Aliyan Abueida (1966—) mentioned in Section 17.1. He received a teaching award from MUN students in 2004, and was awarded the Hall Medal by the Institute of Combinatorics and its Applications (ICA) for his research in 2007.

Pike has provided significant service to the Canadian mathematical community also. He has served the Canadian science granting council (NSERC) for math, as both a member of the grant evaluation group, and the scholarships and fellowships selection committee. Pike has held many roles for the Canadian Math Society, of which he has been named a Fellow. These roles include vice-president, and he was elected president in 2021. He has also been vice-president of the ICA, and has served on both the scientific and equity advisory boards for the Banff International Research Station.

In addition to his mathematical work, Pike is a keen genealogist and has been the president of the Family History Society of Newfoundland and Labrador. He enjoys hiking and curling, and serves on the board of directors for his local LGBTQ+ curling league.

Sources: Memorial University of Newfoundland, Institute of Combinatorics and its Applications, the Math Genealogy Project, and LinkedIn. Updated and confirmed through personal communication.

**Cheryl Elisabeth Praeger (1948—).**
Cheryl Elisabeth Praeger is an Australian mathematician specialising in group theory, who has led the way for women in math in Australia. She was born in 1948 in Toowoomba, Australia. Her family moved several times during her childhood, whenever her father (who worked in a bank for most of her childhood) was transferred. A career guidance counsellor tried to discourage her from pursuing mathematics because she was a girl, but she persisted. Her parents had not been able to participate in higher education themselves, but encouraged Praeger to do so. She began her university studies at the University of Queensland, where she earned her bachelor's degree in 1969. Praeger then moved to England to study at Oxford University, where she received her Master's in 1972 and completed her D.Phil. in 1973 (though it was not awarded until 1974) under the supervision of Peter Neumann. Her thesis was entitled "On the Sylow Subgroups of Primitive Permutation Groups".

After completing her doctorate, Praeger was offered a three-year research fellowship at Australian National University, during which time she also worked for one semester at the University of Virginia in the USA. Praeger next moved to Western Australia in 1976, where she accepted a short-term position. This turned into a permanent position that she held until her retirement in 2017. She was the first female professor of math in Western Australia, and the second in all of Australia. A photograph used in a newspaper story about the appointment shows her on a bicycle, pulling a modified trailer that holds her two young children (both preschoolers). Praeger remains active as a professor emeritus, still conducting research, supervision, and outreach.

Praeger has had a noteworthy career as a researcher, teacher, and mentor. Her research is primarily in abstract algebra (more specifically, permutation groups), but her interests are broad and she has studied actions of permutation groups on many types of combinatorial structures, including codes, designs, and graphs. She has also worked on computational group theory. It would be impossible in a brief sketch to indicate all of the areas to which Prager has

contributed, as she has authored more than 400 publications with more than 180 co-authors. She served her university as department head, and as a dean and deputy dean at various times. Praeger has supervised more than 30 Ph.D. students as of 2021, in addition to many post-docs, undergraduate and Master's students, and young researchers. Even beyond her direct mentoring Praeger has served as an inspiration to many, particularly to young female mathematicians. The Praeger-Xu graphs that bear her name are described in Section 12.5.

Praeger's involvement in mathematics promotion and education has been enormous. She has been a member of the Curriculum Development Council of the (Australian government's) Commonwealth Schools Commission, the (Australian) Prime Minister's Science Council, and the Australian Federation of University Women. She became the first female president of the Australian Math Society in 1992. She chaired the committee that chose and trained Australia's team for the International Mathematical Olympiad (a high school competition) for about 20 years. She spent 8 years (2007—2014) on the executive of the International Mathematical Union; 4 years (2013—2016) as a vice-president of the International Commission on Mathematical Instruction; 4 years (2014—2018) as Foreign Secretary for the Australian Academy of Science, and also served the Association of Academies and Societies of Sciences in Asia on its executive committee and as chair of its Women in Science and Engineering Committee, for 6 years. Praeger also gives a lot of lectures on popular applications of mathematics at public events and in schools on topics such as the relationships between weaving and mathematics. She gives the advice: "There's nothing so wonderful as working at something you're passionate about. Go for it, grasp all the opportunities you can."

Along with this extraordinary record have come many honours. Praeger was elected a Fellow of the Australian Academy of Science in 1996, and named Western Australian Scientist of the Year in 2009. She was also made a Fellow of the American Math Society in 2012, and an honorary member of the London Math Society in 2014. Praeger was inducted into the Western Australian Science Hall of Fame in 2015. She has been awarded 6 honorary degrees, at universities in Europe and Asia in addition to Australia. Her many medals include the Centenary Medal (Australia); Moyal Medal (Macquarie University); Euler Medal (Institute of Combinatorics and its Applications); Thomas Ranken Lyle Medal (Australian Academy of Science); George Szekeres Medal (Australian Math Society); and the Ruby Payne-Scott Medal (Australian Academy of Science). She was made a Member of the Order of Australia in 1999, and then appointed a Companion of the Order in 2021. This is the highest civil honour available from the Australian government, and it was awarded for "eminent service to mathematics, and to tertiary education, as a leading academic and researcher, to international organisations, and as a champion of women in STEM careers".

Praeger was married in 1975 and raised two sons while pursuing her career. She and her husband figured out how to make things work, which wasn't always easy. She spent 6 weeks visiting Cambridge by herself for an important research project when her children were still very young, which was extremely unusual for a mother at that time. Her husband has a Ph.D. in statistics and runs his own consulting firm. Praeger also holds an Associate in Music, Australia (AMusA) in piano performance. An accident to one of her fingers when she was 11 made it impossible for her to pursue music as a career, but this is one of her keener interests along with hiking, sailing, and cycling.

Sources: wikipedia, St. Andrews' math history web site, University of Western Australia, STEM Women, Agnes Scott College, the Math Genealogy Project, Australian Government Honours, Australian Broadcasting, International Science Council, and Australian Academy of Science. Updated and confirmed through personal communication.

**Richard Rado (1906—1989).**

Richard Rado was a German-British mathematician who was involved in laying the foundations of Ramsey theory for infinite cardinals. He was born in Berlin, Germany, to Jewish parents. He was educated at the Universities of Berlin and Göttingen, completing his Ph.D. in Berlin in 1933 under the supervision of Issai Schur (1875—1941). His thesis was entitled "Studien Zur Kombinatorik" ("Studies of Combinatorics"). Rado also got married in 1933.

With Hitler coming into power in 1933, new laws made it impossible for Rado to work at a university in Germany. He obtained a scholarship to pursue further studies in Cambridge, and he and his wife moved to England. In 1935 Rado received a second Ph.D. from the University of Cambridge. His thesis was "Linear Transformations of Bounded Sequences", and his advisor was G.H. Hardy. While in Cambridge Rado met Pál Erdős (1913—1996) with whom he co-authored 18 papers over time.

Rado remained in Cambridge in a temporary position for one year after completing his Ph.D. He moved to the University of Sheffield in 1936, and then to King's College, London in 1947. His only child was born while he and his wife lived in Sheffield. Rado's last move was to take a chair at the University of Reading in 1954; he remained there until his retirement in 1971. Rado spent a year as a visiting professor at the University of Waterloo in Canada immediately after his retirement, in 1971—1972, and held a similar appointment at the University of Calgary from 1973—1974. He and his wife continued to live in Reading, dying in a nursing home in nearby Henley-on-Thames in 1989; the health of both was severely affected by a car accident in 1983.

Rado's most important mathematical work was in finite and transfinite combinatorics, including the early development of Ramsey theory for infinite cardinals (with Erdős), an area that decisively influenced subsequent developments in set theory. The Erdős-Ko-Rado Theorem became a cornerstone of the theory of extremal set systems. A number of objects and results bear his name, including the Rado graph, (see Section 11.5) which had been discovered earlier and independently (in different forms) by Ackermann, Erdős, and Alfréd Rényi (1921—1970). He also contributed to analysis, number theory, algebra, geometry, and measure theory.

Rado is only credited by the Math Genealogy Project with supervising 4 doctoral students, but his mentorship was broader and more influential than this would indicate. He served as secretary and then vice-president of the London Math Society (LMS) from 1953 to 1956, and was awarded the Senior Berwick Prize by the LMS in 1972. He founded the British Combinatorial Committee in 1977 and served as its chair from its inception until 1983. He was elected a Fellow of the Royal Society in 1978, and received an honorary doctorate from the Free University of Berlin in 1981, and from the University of Waterloo in 1986. At his retirement, Rado said, "There are almost as many types of mathematician as there are types of human being. Among them are technicians, there are artists, there are poets, there are dreamers…".

Rado was also a very talented pianist, but chose to pursue mathematics on the grounds that he believed that he could continue music as a hobby, while he couldn't do that with math. His wife was a professional-quality singer who also played the piano (they met when he was seeking someone to play piano duets with), and they enjoyed performing together in public and private recitals throughout their lives.

Sources: wikipedia, St. Andrews' math history web site, University of Reading, the Math Genealogy Project, Times obituary, and Royal Society.

**Frank Plumpton Ramsey (1903—1930).**
Frank Plumpton Ramsey was a British mathematician, philosopher, and economist. The work he produced during his short life had a lasting influence on all of these fields. Within Mathematics, his influence is strongly felt in mathematical logic, combinatorics, and set theory. Ramsey was born in Cambridge in 1903. His father, also a mathematician, was president of

Magdalene College there. After attending boarding school in Winchester, Ramsey returned to the University of Cambridge, where he received his bachelor's degree in mathematics in 1923.

Ramsey travelled to Vienna in 1924, seeking psychoanalysis for depression. He returned to England and became a fellow of King's College, Cambridge in October of that year. He was married a year later, and had two children. In 1926 he became a lecturer at King's College. Ramsey had been diagnosed with jaundice and became increasingly ill late in 1929. He underwent an operation in London early in 1930 to remove a suspected blockage. The operation revealed no blockage, but showed that his liver and kidneys were in very bad condition. Ramsey died in the hospital, shortly thereafter. The exact cause of his unexpected death is uncertain.

Ramsey's mathematical research was in the area of logic. In math, he is best known for the theorem that forms the basis of the subject now known as Ramsey theory (introduced in Section 1.3 and discussed in more detail in Section 14.2). To him it was a lemma being used for another purpose. Ironically, he was working toward a problem that is now known to be undecidable, and a more direct proof has since been found for the special case that he was able to prove using this lemma. So it has been said that "Ramsey's enduring fame in mathematics … rests on a theorem he didn't need, proved in the course of trying to do something we now know can't be done!" This whole paper was only 8 pages long.

Although he was employed as a mathematician, only 3 of Ramsey's publications appeared in mathematical journals. Ramsey also published important work in philosophy and economics; he studied with John Maynard Keynes at Cambridge and continued to work with him closely through his brief career. Ramsey was regarded as a genius; many of his important ideas were not fully appreciated for decades.

Sources: wikipedia, St. Andrews' math history web site, Stanford Encyclopedia of Philosophy, the New Yorker, Oxford University Press, and Times Literary Supplement.

**Dwijendra Kumar Ray-Chaudhuri (1933—).**
Dwijendra Kumar Ray-Chaudhuri (known as "Dijen") is an Indian-American combinatorist who has contributed to very important developments in both coding theory and design theory. Ray-Chaudhuri studied at the prestigious Rajabazar Science College of the University of Calcutta in India, where he received his M.Sc. in 1956. He went on to earn a Ph.D. in 1959 from the University of North Carolina at Chapel Hill. His advisor was Raj Chandra Bose (1901—1987), and his thesis was entitled "On the Application of the Geometry of Quadrics to the Construction of Partially Balanced Incomplete Block Designs and Error Correcting Binary Codes."

For about seven years after completing his dissertation, Ray-Chaudhuri served as a research associate or consultant at a number of universities and private businesses, including Bell Labs, IBM Research, Cornell Medical Center, and Sloan Kettering Institute. In 1966, he was hired at Ohio State University in Columbus, Ohio, where he spent the remainder of his career, and remains a professor emeritus. Ray-Chaudhuri served two terms as chair of his department, and was a visiting professor at a number of universities in Europe during stays there, as well as at the Tata Institute in Mumbai.

Ray-Chaudhuri is most famous for his work in error-correcting codes and in designs. With Bose, he discovered "BCH" codes independently, at about the same time as their discovery by Alexis Hocquenghem in 1959. These codes are still used in applications such as compact disc players, DVDs, and solid-state drives, and the "C" in their name is for Ray-Chaudhuri. With his Ph.D. student Richard Michael Wilson (1945—) in 1968, he determined exactly when Kirkman Triple Systems exist (see Theorem 18.1.11). During his career, Ray-Chaudhuri published over 80 papers, and supervised at least 33 doctoral students. He devoted significant effort and attention to developing research in combinatorics at Ohio State University, hiring and mentoring young researchers in addition to his own students.

Ray-Chaudhuri was a founding Fellow of the Institute of Combinatorics and its Applications (ICA), and was awarded the ICA's Euler Medal in 1999 for career achievements in combinatorics. He was an invited speaker at the International Congress of Mathematicians in 1970, and became a Fellow of the American Math Society in 2012. Ray-Chaudhuri is married with three children.

Sources: wikipedia, Ohio State University, wikipedia, and the Math Genealogy Project, as well as MathSciNet for a publication list.

**Alfréd Rényi (1921—1970).**

Alfréd Rényi (called "Buba" by his friends, including Pál Erdős (1913—1996)) was a Hungarian mathematician of Jewish descent, whose research centred around probability. In graph theory, he is particularly known for introducing the use of probabilistic methods in the study of graphs. Rényi was born in Budapest, Hungary. He became attracted to astronomy as a child, which led him to learn physics, and his interest in mathematics stemmed from this. Due to laws limiting Jewish students at Hungarian universities he was not able to attend university immediately upon finishing high school, so he spent some months working at a shipyard. Winning a prize in a math contest led to his admission to the science university in Budapest in 1940. (This university has been named Loránd Eötvös University since 1950, and will be referred to as Eötvös University in the rest of this biography.) He graduated with a bachelor's degree in 1944, and, as a Jewish male of military age, was sent to forced labour in a "labour batallion". While these batallions saw scenes of atrocities against their enlisted Jewish men throughout World War II, with the Nazi takeover of the Hungarian government on 16 October 1944, many of the labour batallions turned into death marches. Rényi managed to escape before this turn of events.

Rényi continued to live in hiding in Budapest for 6 months using false documents. During this period he rescued his parents, who were being confined in the Budapest ghetto, by the expedient of pretending to be a soldier and marching them out, wearing a uniform he somehow managed to obtain. This was an extraordinarily risky act that required quick thinking and bravery. While all of this was going on, he also managed to enrol at the University of Szeged, where he completed his doctorate in 1945 with Frigyes (Frederic) Riesz as his advisor. Riesz, one of the creators of functional analysis, had narrowly missed being transported to the Auschwitz gas chambers a year earlier. Rényi's thesis was entitled "Egy Stieltjes-féle inegrálról" ("On the Stieltjes transform"). He married in 1946; his wife was also a mathematician, Katalin Schulhof (she went by Kató Rényi after marriage). They had one child.

Rényi worked for a time as a statistician in Budapest. He then spent most of a year in Leningrad, USSR (now St. Petersburg, Russia), doing a postdoc with Yuri Linnik (ending in 1947). During this period he developed his own version of Linnik's number-theoretic "large sieve" method and used it to prove an approximate version of the famous Goldbach Conjecture. He briefly held an appointment at Eötvös University, until he was appointed Professor Extraordinary at the University of Debrecen in 1949. He held this role only until he became the founding director of the Mathematics Research Institute of the Hungarian Academy of Sciences in 1950. He continued to serve as its director until his death; it has since been renamed in his honour. In 1952 he became a professor in the Department of Probability and Statistics at Eötvös University. Rényi died of cancer in 1970 at the age of 48. His wife Kató had also died tragically young, about 6 months previously. In the summer of 1969, shortly after Kató's passing and unaware of his own deadly disease, Rényi delivered a plenary lecture on his recent joint work with his wife on "a theory of trees" at an international conference on combinatorics in Balatonfüred, Hungary.

Rényi's research was generally in the areas of number theory, graph theory, analysis, and probability. Among his more than 200 publications, he wrote 32 joint papers with Erdős.

These included their seminal 1960 paper on "The evolution of random graphs" that created the Erdős-Rényi model of random graph described in Section 11.5 and with it, an entirely new and highly influential field of study. Their joint work also included the result showing that almost all graphs are asymmetric, as mentioned in Section 12.5. Rényi also wrote a highly influential textbook in probability theory. He was a charismatic teacher. Through teaching, advising, and mentoring generations of mathematicians, he created a school of probabilists in Hungary, and this area remains one of the strongest suits of mathematics in that country. During his career he held visiting appointments at Stanford, Michigan State, the University of Michigan, Cambridge University, the University of Erlangen, and the University of North Carolina. Among the honours he received were the Kossuth Prize from the Hungarian government (twice).

Rényi used to say "A mathematician is a device for turning coffee into theorems," which Erdős loved to quote. Pál Turán (1910—1976) built on this joke by adding that weak coffee is only good for a lemma. Rényi was popular as a raconteur, played the piano, and enjoyed rowing, swimming, and skiing. He wrote a series of witty essays on mathematics in the form of imaginary dialogues or exchanges of letters between historical mathematicians.

Sources: wikipedia, St. Andrews' math history web site, Cambridge University obituary, the Math Genealogy Project, and Project Euclid obituary. Additional information and clarification by personal communication from Laci Babai.

**George Neil Robertson (1938—).**
George Neil Robertson (who goes by "Neil") is a Canadian-American mathematician, famous for having settled, in monumental series of works with coauthors, some of the most formidable open problems of graph theory. He was born in Killarney, Manitoba (in Canada). For several years during World War II the family moved through Manitoba, Alberta, and British Columbia, as Robertson's father served in Canada's "Veteran's Guard", but they returned to Killarney in 1944 and from 1945 on the family lived on a nearby farm. In 1956 Robertson enrolled at Brandon College (at that time a branch campus of the University of Manitoba), graduating in 1959 with a B.Sc.

Robertson's interest in mathematics solidified during his studies at Brandon College, and he went on to graduate school at the University of Manitoba. He taught first-year calculus to support himself during his Master's, which he completed in 1962. His Master's thesis was on systems of distinct representatives, and its quality earned him the attention of top advisors during his Ph.D. He spent one year in the Ph.D. program at Syracuse University in New York (in the United States) where he worked with Herbert Ryser. He then moved back to Canada to study at the University of Waterloo, where he was advised by William Tutte. Robertson completed his Ph.D. in 1969 with a thesis entitled "Graphs Minimal under Girth, Valency and Connectivity Constraints". For several summers during his doctoral studies he worked in Operations Research at an air force base near Montreal. Robertson also married during this period; he and his wife have 2 children. While still a student he discovered the Robertson graph, which has the fewest vertices among all graphs in which every vertex has 4 neighbours and the smallest cycle has length 5.

Although the degree was not conferred until 1969, Robertson's doctoral work was completed in 1968 and he spent a year at McGill University, working as a postdoc with William Brown. Immediately after completing his postdoc, Robertson was hired at Ohio State University, where he remained throughout his career. While working there, he also acted as a consultant for Bell from 1984 to 1996, and held a number of visiting positions, including at Princeton University and Victoria University of Wellington, New Zealand. He was awarded the title of Distinguished Professor of Mathematics and Physical Sciences at Ohio State in 2006, where he remains a professor emeritus since his retirement in 2008. He maintained a workload of one-third time for 6 years after retirement, not all of it at Ohio State. During his career, Robertson supervised

at least 23 Ph.D. students, and published more than 80 papers. Robin Thomas (1962—2020) worked with him as a postdoc.

Robertson's research is in graph theory. His most significant work has been in the areas of structural and topological graph theory. In a ground-breaking series of 22 papers dating from 1983 to 2004, he and Paul Seymour (1950—) proved that any family of graphs that is closed under minors can be characterised by a finite set of forbidden minors; this had been conjectured by Klaus Wagner (1910—2000), and is discussed in Section 15.2. Robertson was also an author (with Daniel Sanders, Seymour, and Thomas) of a new proof of the Four-Colour Theorem (which still required computers but was significantly simpler than the original proof), and (with Maria Chudnovsky (1977—), Seymour, and Thomas) of the proof of the Strong Perfect Graph Theorem. Robertson's work has been key to increasing the respect accorded to research in graph theory by mathematicians in other disciplines.

Robertson won the American Math Society's Fulkerson Prize on 3 occasions: in 1994, 2006, and 2009. He won the Ohio State University Distinguished Scholar Award in 1997, and in 2002 the Waterloo Alumni Achievement Award. In 2004 he also won the Pólya Prize from the Society for Industrial and Applied Math. Robertson was named a Fellow of the American Math Society in 2013, and was elected an Honorary Fellow of the Institute of Combinatorics and its Applications in 2018.

Sources: wikipedia, wikipedia, University of Waterloo, Institute of Combinatorics and its Applications,and the Math Genealogy Project, as well as MathSciNet for a publication list. Updated and confirmed through personal communication.

**Gordon F. Royle (1962—).**

Gordon F. Royle is an Australian mathematician and computer scientist, working primarily in the area of computational combinatorics. He was born in Australia, and received his Ph.D. from the University of Western Australia in 1987, under the joint supervision of Cheryl Elisabeth Praeger (1948—) and Brendan McKay. His thesis was entitled "Constructive Enumeration of Graphs".

Royle completed a postdoc at the University of Waterloo in Canada, where he worked with Chris Godsil. He also met his wife in Waterloo. They have 2 children. Royle and his wife returned to Perth, Australia, where he became a professor at the University of Western Australia, and where he still works.

Royle's research is primarily in computational combinatorics. He has expertise in algebra, computers and algorithms, and many branches of combinatorics. He has more than 60 publications. He is probably best known for the graduate textbook on algebraic graph theory that he co-authored with Godsil, and his work on the mathematics of Sudoku, including the web site mentioned in Section 16.1. Royle has supervised at least 2 doctoral students.

Sources: wikipedia, University of Western Australia, and the Math Genealogy Project. Updated and confirmed through personal communication.

**Al-Samaw'al ben Yahyā al-Maghribī (c.1130—c.1180).**

Al-Samaw'al ben Yahyā al-Maghribī was a 12th-century mathematician and doctor from the area that is now Iraq, whose writings include both mathematics and religion. He was born in Baghdad, capital of the Abbasid Caliphate (now capital of Iraq), to a Jewish family; his father was a Rabbi. Baghdad was the centre of scholarship in the Islamic world. Al-Samaw'al began to study mathematics alongside medicine as a youth. After surpassing his teachers in math, al-Samaw'al read all of the books he could find, and began to work out his own improvements to them.

When he was 19, al-Samaw'al wrote *al-Bahir fi'l-jabr* (the "Brilliant in Algebra"). In addition to including original material, it is valuable because it contains and builds on work by al-Karaji which has not survived. *Al-Bahir fi'l-jabr* includes 4 parts. In the first part he defines

and uses the basic arithmetic operations of addition, subtraction, multiplication, and division, on polynomials. He also describes methods for finding the roots of polynomials. The second part is largely about solving quadratic equations. It also includes the binomial theorem, though al-Samaw'al attributes this to al-Karaji. He also uses an inductive type of argument, though it is not clear or well-developed and it would be a stretch to say that he fully understood the principle of induction. The third part is about calculating with irrational numbers, and does not contain much original material. The fourth and final part includes the problem described in Example 3.2.9, in which (to be more precise) he asks for the values of 10 unknowns, given the 210 equations that show the sums of any 6 of the variables.

After writing this treatise, al-Samaw'al travelled extensively through countries in the region of Iraq. He practised medicine during his travels. In 1163, during his travels, he converted to Islam but kept this a secret for 4 years because he was concerned about his father's reaction. In fact, when his father received the letter in which al-Samaw'al told him of his conversion, he immediately set out to see his son. He died along the way.

Including his first book, al-Samaw'al is said to have written about 85 works; most of these have not survived. *Al-Bahir fi'l-jabr* contains his only surviving mathematical writing of value. His surviving works also include a book explaining the problems he saw with Judaism and Christianity.

Sources: wikipedia, St. Andrews' math history web site, Encyclopedia.com, and Columbia University.

**Issai Schur (1875—1941).**

Issai Schur was a Russian-German mathematician of Jewish descent, best known for his work in algebra, but whose important discoveries have impacted a number of fields including Ramsey theory. He was born in Mogilev, which at the time was part of the Russian Empire (it is now in Belarus). When he was 13, Schur moved to Libau (now Liepāja, Latvia) to live with his sister. The local Jewish community spoke German, and Schur attended a German-language school.

In 1894 Schur moved to Germany to attend the University of Berlin. He completed all of his studies there, finishing in 1901. His Ph.D. thesis, "Über eine Klasse von Matrizen, die sich einer gegebenen Matrix zuordnen lassen" ("On a class of matrices which can be assigned to a given matrix"), was written under the joint supervision of Ferdinand Frobenius and Lazarus Fuchs.

Schur worked at the University of Berlin from 1903 to 1911, and then moved to the University of Bonn, where he worked until 1916. In 1916, Schur returned to the University of Berlin. He was promoted to professor in 1919. In 1922, Schur was elected to the Prussian Academy of Sciences. After the Nazis rose to power, anti-Jewish laws passed in Germany in 1933 required that Schur be suspended and barred from the university. Due to an outcry and a technicality (he was respected and popular in the community, and had held an appointment prior to World War I, which made him qualify for an exemption under the law) Schur was given a temporary dispensation to lecture in 1933—1935, and was in fact the last Jewish professor to lose his position. Schur was invited to move to various universities in other countries during this period, but he declined these offers. He continued to think of himself as a German and was baffled by the Nazis' inability to see anything but his Jewish heritage. However, he was increasingly isolated and disrespected, and was ultimately fired in 1935. By the end of 1938 he was pressured into resigning from the Prussian Academy of Sciences.

In 1938 Schur was summoned to an interview with the Gestapo. In his increasing isolation and depression, he had told his wife that he would commit suicide if summoned by the Gestapo, so she managed to send him out of town and go in his place. She was asked why they had not left Germany. A Reich Flight Tax amounting to a quarter of their assets had to be paid if they were to be allowed to leave. They were unable to raise enough money, because a substantial

portion of their assets were tied up in a mortgage on a house in Lithuania that at that time could neither be sold nor used as collateral. The tax was ultimately paid; the source of the money is not known, but the Schurs left Germany in 1939. After a brief stay with their daughter in Bern, Switzerland, the Schurs moved to Palestine. Schur had to sell his collection of books to the Institute for Advanced Study in Princeton in order to cover their living expenses in Tel Aviv. Schur died there of a heart attack on his 66th birthday.

Schur's research was mostly in the field of algebra; more specifically, group theory and representation theory. Among his almost 90 papers he also published important results in number theory, analysis, and Ramsey theory. There are numerous structures and results across these fields that bear his name, including the theorem whose statement is given in Exercise 14.2.10. Schur was extremely popular as a lecturer and mentor, and built a school of researchers who worked with him and learned from him. These included the doctoral students he supervised (there were at least 22 of these, and some sources credit him with 35), one of whom was Richard Rado (1906—1989).

Sources: wikipedia, St. Andrews' math history web site, the Math Genealogy Project, Encyclopedia.com, and London Math Society.

**Paul Seymour (1950—).**

Paul Seymour is a British mathematician and a towering figure in contemporary combinatorics. He was born in Plymouth, England. He attended Exeter College at the University of Oxford throughout his university education, earning his bachelor's degree in 1971 and his Master's of Science in 1972. Working under the supervision of Aubrey Ingleton, Seymour completed his doctorate in 1975 with a thesis entitled "Matroids, Hypergraphs and the Max-Flow Min-Cut Theorem". Seymour was also awarded an M.A. from Oxford in 1975.

Between 1974 and 1980 Seymour held research positions: first for 2 years at the University College of Swansea; then for 4 years at Merton College, Oxford University. During his time at Oxford, Seymour also spent a year at the University of Waterloo, Canada, as a visiting researcher. This is where he met George Neil Robertson (1938—). In 1980, he moved to the United States, and took a position at Ohio State University, partly to collaborate with Robertson. Seymour continued to work there until 1983. In 1984 he was hired as a senior scientist at Bellcore (Bell Communications Research) in New Jersey, where he continued to work until 1996. He maintained his academic connections throughout this period, with adjunct positions first at Rutgers University and then at the University of Waterloo, followed by a visiting position at Princeton University. In 1996 he accepted a position as a professor at Princeton, where he has remained ever since. In 2016, he was appointed to the Albert Baldwin Dod Professorship there. Seymour married in 1979 and has two children.

Seymour's influence on combinatorics has been immense. He began his research career working on matroids, where he made some significant contributions before moving into graph theory. He works primarily in the areas of structural and topological graph theory. His research has been instrumental in significantly increasing the respect with which other mathematicians view graph theory. In a ground-breaking series of 22 papers dating from 1983 to 2004, he and Robertson proved that any family of graphs that is closed under minors can be characterised by a finite set of forbidden minors; this had been conjectured by Klaus Wagner (1910—2000), and is discussed in Section 15.2. Seymour was also an author of a new proof of the Four-Colour Theorem (this still required computers but was significantly simpler than the original proof), and of the proof of the Strong Perfect Graph Theorem. Seymour has over 250 publications. He has supervised 11 doctoral students in addition to undergraduate and Master's students as of 2021. He is joint editor-in-chief of the *Journal of Graph Theory*. Seymour also holds two patents.

Seymour has an impressive list of honours, as befits his illustrious career. He won the American Math Society's Fulkerson Prize on 4 occasions (on his own in 1979, and jointly in 1994, 2006, and 2009). He also won the Pólya Prize from the Society for Industrial and Applied Math in 1983, and again (jointly with Robertson) in 2004. Seymour was awarded a Sloan Fellowship in 1983, and the Ostrowski Prize in 2004. He holds honorary doctorates from the University of Waterloo (2008) and the Technical University of Denmark (2013), and was given the Commemorative Medal of Comenius University (Slovakia) in 2019.

Sources: wikipedia, Princeton University, the Math Genealogy Project, and University of Waterloo. Updated and confirmed through personal communication.

**Sharadchandra Shankar Shrikhande (1917—2020).**
Sharadchandra Shankar Shrikhande was an Indian mathematician who was nicknamed one of "Euler's Spoilers" for his work in disproving a conjecture of Leonhard Euler (1707—1783). Shrikhande was born in Sagar, India in 1917. He completed his bachelor's degree at the Government College of Science in Nagpur, after which he spent a year as a Research Fellow at the Indian Statistical Institute in Calcutta, where he met Raj Chandra Bose (1901—1987). Shrikhande briefly taught at the college in Jabalpur and then the College of Science in Nagpur, but regularly returned to Calcutta to visit Bose.

Shrikhande moved to the United States in 1947 for his doctorate, which he completed in 1950. He worked on this under the supervision of Bose at the University of North Carolina in Chapel Hill (Bose moved there from India in 1949). Shrikhande's thesis was entitled "Construction of Partially Balanced Designs and Related Problems".

Shrikhande moved back to India after completing his doctorate, and held a position at the Science College in Nagpur from about 1953 to 1958. He spent some time back in the United States from 1959 to 1960; it was on this visit that he collaborated with Bose and Ernest Tilden Parker (1926—1991) to disprove Euler's 1782 conjecture that there are no pairs of orthogonal Latin squares of order $n$ when $n = 4k + 2$ (see Section 16.2); this earned them the nickname of "Euler's Spoilers". Shrikhande took a position as a professor at Banaras Hindu University on his return to India in 1960. In 1963 he became the founding head of the math department at the University of Mumbai and the founding director of the Center of Advanced Study in Mathematics of Mumbai; he retired from these positions in 1978. Shrikhande then took on the directorship of Mehta Research Institute (which has since been renamed after Harischchandra) in Allahabad. He stayed there until about 1988, when his wife died. After this reports vary, but it seems that he lived with his sons in the United States and New Delhi for a couple of decades. In 2009 he moved to an Ashram in Vijayawada run by one of his grandchildren, where he lived for his final years of life. He died there in 2020 at the age of 102.

In addition to Shrikhande's status as one of "Euler's Spoilers", he also has a graph named after him. He published more than 75 research papers, and was a mentor to many, although there is no formal record of most of these relationships. The techniques that he and his co-authors developed for disproving Euler's conjecture played a significant role in the proof by Richard Michael Wilson (1945—) of Wilson's Theorem. He was a Fellow of the Indian National Science Academy and the Indian Academy of Science. He was also a Fellow of the Institute of Mathematical Statistics in the United States, and was named an Honorary Fellow of the Institute for Combinatorics and its Applications in 1991.

Shrikhande was married with three children. One of his sons was a professor of mathematics at Central Michigan University until his own retirement in 2015.

Sources: wikipedia, Institute of Mathematical Statistics, the Math Genealogy Project, Resonance, Institute for Combinatorics and its Applications,and News 18 (India).

**Vera Sós (1930—).**

Vera Sós [pronounced "Shosh"] is a Hungarian number theorist and combinatorist, who has been influential through both her impressive research results and her mentorship. She was born to Jewish parents in Budapest, Hungary. Anti-semitic laws meant that her father lost his job as a teacher during her childhood. After surviving the Holocaust, Sós was able to resume her studies in the Jewish Girls' High School in 1945. Her math teacher was renowned graph theorist Tibor Gallai, a close friend of Pál Erdős (1913—1996). Her classmates included another future math professor, Edit Gyarmati (1930—2014), attesting to the extraordinary qualities of Gallai as a teacher. Sós subsequently studied at the science university in Budapest, which was renamed Loránd Eötvös University while she was there in 1950. Sós studied math and physics, graduating in 1952, and began to teach at the university in 1950 while still a student. Although she was surrounded by men through most of her career, Sós never felt that her gender was an issue in her work.

Sós completed her Ph.D. in 1957, with Fourier analyst Lipót (Leopold) Fejér as her advisor. The title of her thesis (written in Hungarian) translates as "A geometric treatment of continued fractions and its application to the theory of diophantine approximation"; the thesis includes her famous "three-gaps-theorem" in diophantine approximation. She worked at Eötvös University in the Department of Analysis until 1987. At that time, she moved to work at the Alfréd Rényi Institute of Mathematics. She continued to give lectures at the university until about 2006.

While studying at university, Sós met Pál Turán (1910—1976), who was a professor. They married in 1952, and had two children. Sós's name often appears (for example in publications) as Vera T. Sós; the "T" is for Turán. She met Alfréd Rényi (1921—1970) and Erdős while still in high school, as her teacher (Gallai) fostered her abilities not only by assigning her beautiful, challenging problems, but also by introducing her to prominent mathematicians. Recognising Sós's talent, Rényi soon became her mentor, inspiring and challenging her in weekly sessions while she was still in high school and starting a research collaboration with her during university. Sós began working more closely with Erdős in the mid-1960s, and became one of his most frequent collaborators; the two published 35 joint papers. The period of Sós's graduate education was a turbulent time in Hungary, as the Hungarian revolution against Soviet rule took place (and was crushed) in 1956. Sós had a 3-year-old son at this time. Schools and universities closed, and Sós avoided leaving the house to remain safe during fighting on the streets. She said in an interview that this gave her more time to work on mathematics, and downplayed the challenges it presented in comparison with those of mathematicians like her husband who found ways to work on mathematical research while in forced labour camps during World War II.

Sós's research is primarily in the areas of number theory and combinatorics, and her more than 100 publications include results that now bear her name in both of these fields. She initiated several new areas of research, including Ramsey-Turán theory and the extremal theory of structured intersections. In the 2000s she joined a group of researchers led by László Lovász in exploring graph limit theory; their joint paper appeared in 2012, when she was 82, in the *Annals of Math.* Sós's mentorship and influence have been profound. In 1965, she jointly founded a weekly combinatorics seminar at the Mathematical Institute of the Hungarian Academy of Sciences that has been running ever since. It serves as a forum for new results in combinatorics, and often features the research of young mathematicians including undergraduate students. Starting in the 1960s, she was also a key organiser of a long series of international conferences on combinatorics held in Hungary; these allowed mathematicians from the Soviet bloc (including Hungary) to interact with others worldwide, at a time when such opportunities were very limited. One of her joint works with her husband is mentioned in Section 14.2.

Sós won the Tibor Szele Prize for research mentorship from the János Bolyai Math Society in 1974, and the Academy Award of the Hungarian Academy of Sciences in 1983. She was elected to the Hungarian Academy of Sciences as a corresponding member in 1985, and as an

ordinary member in 1990. She won Hungary's Benedikt Otto Prize in 1994, and was elected to the Austrian Academy of Sciences in 1995. Sós also won Hungary's Széchenyi Prize in 1997, the Cross of Merit in 2002, and the "My Country" award in 2007. She became a member of the Academia Europaea (a pan-European Academy of scholarly inquiry) in 2013. Sós received an honorary degree from the Hebrew University of Jerusalem in 2018.

Sources: wikipedia, St. Andrews' math history web site, the Math Genealogy Project, Gil Kalai's blog, and Simons Foundation. Additional information and clarification by personal communication from Laci Babai.

**Jakob Steiner (1796—1863).**
Jakob Steiner was a 19th-century Swiss-German mathematician whose research focus was geometry, but who had a profound impact on the early development of design theory. He was born in the village of Utzenstorf in Switzerland. He grew up on a farm, helping his parents, and did not learn to read and write until he was 14. He defied his parents by leaving home to go to school when he was 18, in 1814. After studying at a school in Yverdom for about 4 years, he moved to Heidelberg in 1818, where he continued his studies at the university, while earning his living by tutoring.

In 1821 Steiner moved to Berlin, Germany, where he continued to tutor. In order to be allowed to teach he attempted to pass qualifying exams. Due to his performance in some of the humanities exams Steiner was only granted a restricted license. Nonetheless, he was appointed to teach mathematics at a local school. He was soon dismissed as he refused to use the methods followed by the director of the school or to use the textbook the director had written. Steiner went back to tutoring while studying at the University of Berlin from 1822 to 1824. In 1825 he found another position as a teacher, but again came into conflict with the director of the school. Steiner was carrying out research and publishing during this period, and despite his disagreements with the director he was promoted to the level of senior teacher in 1829.

In 1832 Steiner published his first book. He was awarded an honorary degree from Königsberg University in 1833 in recognition of the research he had published. He was elected to the Prussian Academy of Sciences in 1834, and also accepted a new extraordinary chair in geometry at the University of Berlin. He retained this position for the rest of his life. During a stay in Paris in 1855 he was elected to the Académie des Sciences. Steiner was also elected to the Royal Society but died before the vote was ratified, so never became a member.

Steiner's research was almost entirely in geometry. He made many fundamental discoveries in this field, and a number of theorems and structures are named for him. His work in combinatorics (on Steiner systems) was published in 1853. Although Reverend Thomas Penyngton Kirkman (1806—1895) preceded him in finding Steiner triple systems (see Section 18.1), Steiner was unaware of Kirkman's work. The fundamental design theory structure of Steiner systems (see Section 18.2) also bear his name. Steiner published extensively, including more than 60 papers in Crelle's journal alone (one of the first pure math journals that was not the proceedings of an academy; it was established by a friend of Steiner's). He also produced a number of books.

Steiner apparently had an abrasive personality. He never married; he had many challenging relationships with colleagues and students, and was described as being crude and blunt. He gave nicknames to everyone, and these were never flattering. During his last decade, ill health kept him in Switzerland for most of each year, and he only visited Berlin to give his lectures each winter. He eventually became completely confined to his bed and was unable to continue teaching. He died in Switzerland. He bequeathed one third of his estate to the Berlin Academy for the establishment of a prize named after him.

Sources: wikipedia, St. Andrews' math history web site, University of Evansville, Your Dictionary, The Mathematical Intelligencer, and encyclopedia.com.

**Sushruta (c.800BCE—c.700BCE).**
Sushruta was a renowned Indian doctor from the 8th century BCE whose influence on medicine (and surgery in particular) in India may have been as profound as that of Hippocrates on western medicine, and for similar reasons. The work attributed to Sushruta in Section 4.1 (mentioned again in passing in Section 4.2) appears in a book called the *Sushruta Samhita*, meaning the "Compendium of Sushruta". Scholars have estimated the date of this work at anywhere from 1000BCE to 500CE. It is likely that the core of the book dates from the early end of this range, and that successive people (perhaps even doctors named Sushruta) added to it over the years. We cannot be certain of the original date of any given piece of the work. The book lists its author as Sushruta, and locates him in Varanasi, India.

There was a very eminent Indian doctor named Sushruta who lived sometime between approximately 800BCE and approximately 700BCE. He is referred to as the "father of Indian surgery" and the "father of plastic surgery" as the topics of surgery and plastic surgery are also covered in the *Sushruta Samhita*. The original part of the manuscript is most likely attributable to this doctor.

The *Sushruta Samhita* is a very important work in the history of medicine. The surviving text includes 184 chapters describing illnesses, injuries, surgeries, medicines, and treatments. Sushruta taught others, referred to as "Saushrutas", who were required to undertake 6 years of study and to take an oath (not the Hippocratic Oath, but with similar intent) to do no harm, before being allowed to practice hands-on surgery under supervision.

Sources: wikipedia, Journal of Indian Philosophy, and wikipedia.

**Stan Swiercz (1955—).**
Stan Swiercz is a Canadian computer programmer and software engineer who contributed to the computer-assisted proof that there is no finite projective plane of order 10. He studied electrical engineering at McGill University in Montreal, Canada, but decided that he preferred software to hardware. In 1978, Swiercz was hired by the Department of Computer Science at Concordia University in Montreal as a programmer. He has worked there ever since.

Swiercz is the Manager of Software Applications for what is now the Gina Cody School of Engineering and Computer Science at Concordia University. He is responsible for installing and maintaining required software. It was in this capacity that he worked with Clement Wing Hong Lam (1949—) and Larry Henry Thiel (1945—) on writing, optimising, and running the code to prove the nonexistence of a finite projective plane of order 10, as mentioned in Section 16.2. At the time, Thiel was Swiercz's boss, and Swiercz worked on this under his supervision.

Swiercz is married with two children.

Sources: Concordia University, and Mathematical Association of America. Updated and confirmed through personal communication.

**György Szekeres (1911—2005).**
György Szekeres was a Hungarian-Australian mathematician and educator of Jewish descent, who played a key role in establishing the foundations of Ramsey theory. He was born in Budapest, Hungary. He later adopted the anglicised version of his given name, "George", which is what he is most commonly known by. Laws in Hungary at the time when Szekeres finished high school limited the enrolment of Jewish students in university. Nonetheless, Szekeres secured a place at the Technological University in Budapest. He had long been interested in math, but studied chemical engineering at university, to learn skills that would be useful in his family's leather business.

During university, Szekeres met regularly with a group of students including Tibor Gallai (a notable teacher and mathematician mentioned in the biographical sketch of Sós), Eszter Klein (1910—2005), Pál Turán (1910—1976), and Pál Erdős (1913—1996), who gathered to talk about math. Even though they did not all attend the same university in Budapest, they

knew of each other from high school, when they had all worked on solving problems for the famed "Hungarian High School Mathematical Monthly". This was a national publication that inspired generations of Hungarian mathematicians in their love of mathematics and problem-solving. During one of this group's meetings in about 1933, Klein proposed the Happy Ending Problem (see Section 14.2). Erdős and Szekeres worked on this problem and were able to obtain bounds on the solution; the Erdős-Szekeres Theorem appeared in this same paper. Klein and Szekeres subsequently married in 1937.

After graduating from university in 1933, Szekeres spent about 6 years working as an analytical chemist in a leather factory in Simontornya (his family's business collapsed while he was in university). He and Klein were only able to spend weekends together during this period. Because of their Jewish heritage, Szekeres and Klein were not safe in Hungary, which was an ally of Nazi Germany. In 1939 Szekeres and his wife (as well as one of his brothers) fled to Shanghai, China, which did not require much in the way of documentation for immigrants. Their son was born there in 1940. Life was not easy in China during World War II. Szekeres found work in a leather factory, but it closed in 1940. There were times when they had to flee for their lives from Japanese bombing. Food was sometimes hard to come by, and Szekeres traded a bicycle for a bag of rice on one occasion. Later Szekeres did obtain work as a clerk in an American air force base.

Szekeres' formal qualifications were his bachelor's degree in chemistry, and he had taken virtually no mathematics at university (in an interview he reported having taken none; other sources claim that he took one calculus course). Nonetheless, he had a number of noteworthy publications in math research, beginning in 1935, and he had friends in Australia who held him and his work in high regard. In 1948, Szekeres was hired by the University of Adelaide as a lecturer in math, and the family moved to Australia. Szekeres and Klein had a daughter born there in 1954. Szekeres accepted a professorship at the University of New South Wales in Sydney in 1963, and remained there for the rest of his career. He retired in 1975 and became an emeritus professor, remaining active in research for many years.

Szekeres was a talented musician who played both violin and viola. He started learning violin at the age of 6, and was a member of the Ku-ring-gai Philharmonic Orchestra in Sydney as well as the North Sydney Symphony Orchestra. In addition to maintaining his research in retirement, music and hiking were his principal interests. When Szekeres lost his driver's license the somewhat remote family home near Sydney became too hard to live in, and he and Klein returned to Adelaide, where their children lived. They died within an hour of each other in 2005, in hospital beds side-by-side, after almost 70 years of marriage.

While in Australia, Szekeres was active in mathematical outreach as well as in research. He devised problems for math competitions, and with his wife organised weekly enrichment and problem-solving sessions for high school students that grew and spread, and are still running. He also started a journal for high school students interested in math. Combinatorics was the focus of his research, which included a variety of areas of mathematics. In particular, he was very interested in general relativity, and is credited for developing the mathematical theory that forms the basis for our understanding of black holes. One of the central ideas in this work is the "Kruskal-Szekeres coordinate system" (discovered independently by these two researchers) that is an important element in Carl Sagan's science fiction novel *Contact*. Szekeres' other best-known contribution to math was in the early development of Ramsey theory, which grew out of his work with Erdős on the Happy Ending Problem. Szekeres had an early interest in the use of computers for mathematical research; he taught himself programming in about 1960, and used computers regularly in his research from that time forward. He produced about 90 papers in all, and supervised at least 16 doctoral students.

Szekeres was elected to the Australian Academy of Science' in 1963, and was awarded its Thomas Ranken Lyle Medal in 1968. He was also elected to the Hungarian Academy of Science.

He was a founding member of the Australian Math Society in 1956, and served as its president from 1972 to 1974. The University of New South Wales granted him an honorary doctorate in 1976, shortly after his retirement. In 2001, Szekeres received the Australian Centenary Medal "for service to Australian society and science". In 2002, he was appointed a Member of the Order of Australia with a similar citation. The Australian Math Society established a medal in 2001 that is named after him.

Sources: wikipedia, St. Andrews' math history web site, Australian Academy of Science, Journal of the Australian Math Society, the Math Genealogy Project, and obituary in the Sydney Morning Herald. Additional information and clarification by personal communication from Laci Babai.

**Peter Guthrie Tait (1831—1901).**
Peter Guthrie Tait was a 19th-century Scottish mathematician and physicist. He is probably best known in physics; his most significant work in mathematics was his research on the Four-Colour Conjecture. He was born in Dalkeith, Scotland. His father died when Tait was 6 years old, and the family moved to Edinburgh where they lived with one of his uncles. This uncle was enthusiastic about science, and likely sparked Tait's own interest.

In 1847 Tait enrolled at the University of Edinburgh. He stayed there for only one year before moving to Peterhouse College at Cambridge University. He completed his bachelor's degree in 1852. He remained at Cambridge as a fellow and lecturer until 1854, when he moved to Ireland to become a professor at Queen's College in Belfast.

In 1860 Tait returned to Scotland, becoming a professor of natural philosophy at the University of Edinburgh, a chair that he held almost until he died. It was only after his return to Scotland that Tait began to publish research papers. Some of his early publications were based on things he had studied in Ireland, where he met Sir William Rowan Hamilton (1805—1865). He had also published a book while still in Ireland. Tait knew and worked closely with Lord Kelvin; much of Tait's research was closer to physics than to pure mathematics. He also did research in knot theory and topology, and of course in combinatorics. It was in 1880 that Tait published his proof that the Four-Colour Conjecture is equivalent to the nonexistence of a planar snark (discussed in Section 15.3), which at the time he thought was a proof of the Four-Colour Conjecture. He wrote 16 textbooks, one of which was a joint work with Lord Kelvin that became a standard reference. He also wrote 133 research papers and 232 other publications.

Tait was married in 1857, and had 7 children. He died in Edinburgh in 1901. Tait was a keen golfer and also studied the mathematics and physics of golf. He had some significant disputes with other researchers, which seem to have been based on clinging to his own preferences, whether scientific (holding quaternions as superior to vectors) or patriotic (pro-British). Lord Kelvin quoted Tait as having said that "nothing but science is worth living for,". Kelvin added that Tait seemed not to live by this maxim as he loved to read and had a terrific memory, which allowed him to be ready with apposite quotations for many situations.

Tait was elected a Fellow of the Royal Society of Edinburgh shortly after moving back to Scotland, and served as its general secretary from 1879 until his death in 1901. He won both the Gunning Victoria Jubilee Prize and the Keith Prize (twice) from the Royal Society of Edinburgh, and in 1866 received the Royal Medal of the Royal Society of London. He was awarded honorary degrees by the University of Glasgow and the University of Ireland. Tait was elected an honorary member of the scientific academies of Denmark, Holland, Ireland, and Sweden. A road in Edinburgh and a chair in the department of physics at the University of Edinburgh have been named for him.

Sources: wikipedia, St. Andrews' math history web site, Clerk Maxwell Foundation, and encyclopedia.com.

**Larry Henry Thiel (1945—).**
Larry Henry Thiel is an American-Canadian computer programmer and analyst who contributed to the computer-assisted proof that there is no finite projective plane of order 10. He was born in Trenton, New Jersey (in the United States). He received a bachelor's degree in mathematics from Michigan State University in 1967. He immigrated to Canada in 1968 to work at the University of Alberta as a programmer/analyst.

Thiel moved to Montreal in 1973 to work as a programmer/analyst in the brand new department of Computer Science at Sir George Williams University (which in 1974 merged with Loyola College to form Concordia University). Over time, he was promoted and ended up managing the technical staff and computing labs for the department. In this role, Thiel was often called on to write and/or optimise code being used by researchers. He retired in 2000.

Thiel is a co-author on at least 15 papers to which he was asked to contribute with his expertise in programming. Most notably, he was involved in proving the non-existence of a projective plane of order 10, with Clement Wing Hong Lam (1949—) and Stan Swiercz (1955—), as mentioned in Section 16.2. In his expository paper about this result, Lam says that Thiel "has a reputation of making any computer program run faster."

Thiel has 4 children, including a son who now teaches computer science at Concordia, and with whom he sometimes collaborates.

Sources: Mathematical Association of America, Concordia University, Theorem of the Day, MathSciNet, and Concordia University. Updated and confirmed through personal communication.

**Robin Thomas (1962—2020).**
Robin Thomas was a Czech-American mathematician, renowned for his work in proving (with co-authors) tremendous results in graph theory including the Strong Perfect Graph Theorem. He was born in Prague, Czechoslovakia, where he completed his education with a doctorate from Charles University in 1985, under the supervision of Jaroslav Nešetřil. In 1987, while Czechoslovakia was still under communist control, Thomas emigrated to the United States for a postdoc at Ohio State University in Columbus, where he began to work with George Neil Robertson (1938—). In 1989 Thomas began working at Georgia Institute of Technology in Atlanta. He became a Regents' Professor there in 2010, and remained there until his death.

Thomas was diagnosed with Amyotrophic Lateral Sclerosis (ALS, also known as Lou Gehrig's disease) in about 2008. He was confined to a wheelchair for most of the last decade of his life. He died in 2020 at the age of 57. Thomas was married with 3 children. His wife is a professor at Georgia Tech whose research is in stochastic systems and computer simulation. He advised students to "Follow your passion, value your education, and work hard. Don't give up in the face of hardship, and have fun."

Thomas' research interests were broad. He published in algebra, geometry, topology, and theoretical computer science as well as in combinatorics. His best-known achievements, though, lie in his outstanding work in structural graph theory, particularly his result on the Hadwiger Conjecture with Robertson and Paul Seymour (1950—), and their work on the Strong Perfect Graph Theorem, joined by Maria Chudnovsky (1977—). During his lifetime Thomas published more than 100 research papers, and supervised at least 16 doctoral students (several others were in progress when he died). His influence as a mentor was far broader, though; Thomas served as the head of the Algorithms, Combinatorics and Optimization Ph.D. program and kept a watchful eye on the progress of all of the students in that program. He maintained close ties with Charles University in what had become the Czech Republic, and frequently invited promising researchers to visit or complete a postdoc with him at Georgia Tech.

Thomas won the Fulkerson Prize twice: in 1994, and in 2009. He gave an invited talk at the International Congress of Mathematicians in 2006. In 2011 Thomas received the Karel

Janeček Foundation Neuron Prize for Lifetime Achievement in Mathematics. He was named a fellow of the American Math Society in 2012, and of the Society for Industrial and Applied Mathematics in 2018. In 2016 he received the Class of 1934 Distinguished Professor Award at Georgia Tech, and gave the commencement address to graduate students. After his death, a graduate fellowship was established at Georgia Tech in his name.

Sources: wikipedia, Georgia Institute of Technology, Charles University, Charles University, and the Math Genealogy Project. Updated and confirmed through personal communication with Sigrun Andradöttir.

**Pál Turán (1910—1976).**

Pál Turán (who often used the anglicised version of his given name, "Paul") was a Hungarian mathematician of Jewish descent. His research focus was in analytic number theory and although he is best known for his results in that field, in combinatorics he was responsible for establishing the research area of extremal graph theory. Turán was born in Budapest, Hungary. Until 1919, his surname was Rosenfeld. Anti-Jewish laws in Hungary limited the number of students of Jewish descent who could enter university, but Turán was able to enrol in the science university in Budapest (now Loránd Eötvös University) in 1928, thanks to placing highly in a national competition. While at university, Turán was part of an active group of young mathematical problem-solvers that included Tibor Gallai (who would go on to become a remarkable teacher and an eminent graph theorist, see the biographical sketch of Vera Sós (1930—)), Pál Erdős (1913—1996), Eszter Klein (1910—2005), and György Szekeres (1911—2005). Turán's collaborations with Erdős continued through the rest of his life.

Turán graduated with a teaching degree in 1933, and received a Ph.D. in 1935 with Fourier analyst Lipót (Leopold) Fejér as his advisor. His thesis was entitled "Az egész számok prímosztóinak számáról" ("On the number of prime divisors of integers"). Anti-semitic laws also prevented Turán from obtaining a position after his graduation, either at a university or as a school teacher. He survived by tutoring, finally obtaining a teaching job in 1938 at the high school of the Rabbinical Seminary. Turán married in 1939 and had a son from this first marriage.

From 1940 to 1944 (during the war) Turán spent several long periods in the forced labour batallions to which men of Jewish descent were conscripted. The situation was horrific; men in the labour batallions were subject to daily abuse. Turán's primary interest throughout his career was in analytic number theory. This was not a subject that he could easily work on without writing down large formulas as he went. During his time in labour camps, he discovered that he could perform graph theory research in his head. Many of the problems that Turán came up with in graph theory, as well as the solutions for some of them, came out of this terrible time. He said, "I got my best ideas while pulling wires, because then I could be alone and nobody noticed that I was thinking." The full force of the Holocaust hit Hungary during the last year of the war. Turán survived, but his three siblings (a sister and two brothers) were murdered.

Remarkable stories came out of Turán's time in the labour camps. Some of these are told beautifully in his own words in the final source listed below, in an effort to illustrate "the enchantment and help [graph theory] gave me in the most difficult times of my life". I summarise them here. Turán's initial task at labour camp involved carrying railway ties. The officer overseeing this work happened to hear one of Turán's companions call him by name; this officer was an engineer with an interest in math. In an amazing coincidence, as a civilian the officer worked as a proofreader in a publishing house of the Hungarian Academy of Sciences, and had been involved in the production of some of Turán's papers. This officer reassigned Turán to a lighter duty: directing crews to the proper piles of logs. This lighter work freed Turán's mind and he was able to consider (and ultimately solve) a problem that had occurred

to him earlier in the year, as he mulled over the mathematical ramifications of some items in a letter he had received from his friend Szekeres, then a refugee in Shanghai.

In 1944 Turán was in a labour camp outside Budapest, working at a brick factory. Small wheeled trucks carried the bricks from the kilns along tracks to the storage yards; every kiln was connected by rail to every yard. Where the rails crossed, the trucks often jumped the tracks and spilled their loads, which caused problems for the workers. In this context, Turán began to wonder: in a complete bipartite graph (as this system represented), what is the minimum number of edge-crossings? Turán considered this and solved some related extremal problems, but not the general case of this one. Even later in 1944, while expecting at any day to be sent to the gas chambers, Turán kept up his spirits by trying to find bounds on a problem very similar to the Ramsey numbers: given $n$, what is the largest value $M(n)$ such that any 2-edge-colouring of $K_n$ (using red and blue, say) must contain either a red or a blue $K_{M(n)}$? One of his results in Ramsey theory is mentioned in Section 14.2.

After the war, in 1945 Turán was given a position at the science university in Budapest. He was promoted to the Chair of Algebra and Number Theory there in 1949. In 1952 he remarried, to Vera Sós (1930—); they had two children, one of whom also became a mathematician. In 1955 Turán was named head of the complex function theory department in the Mathematical Institute of the Hungarian Academy of Sciences.

Turán died at the age of 66. He had been ill from leukemia for about 6 years, but Sós concealed the diagnosis from him as she felt that the knowledge would only depress and limit him. Turán once referred to his research work as "building my pyramid": developing his claim to immortality.

Turán's most significant research is likely his work in analytic number theory. He developed the power sum method, which has seen many uses over the years. He worked on many problems related to the Riemann Hypothesis; Erdős jokingly refers to Turán as a "heretic" for not believing in the Riemann Hypothesis. In graph theory he was the founder of what is now known as extremal graph theory: finding largest or smallest graphs that have certain properties, or bounding various characteristics of such graphs. This is now a major area of research, and it came out of a result Turán proved in his mind while performing forced labour. Turán produced more than 200 publications, and mentored generations of Hungarian mathematicians.

Turán was elected to the Hungarian Academy of Sciences in 1948, initially as an associate member, and then as a regular member in 1952. He was awarded Hungary's Kossuth Prize in 1948 and again in 1952. He served the János Bolyai Math Society in a variety of ways, including by acting as its president from 1963 to 1966. In 1975 he was awarded the Szele Prize for research mentorship by the János Bolyai Math Society. He was elected a fellow of the American Math Society, the Austrian Math Society, and the Polish Math Society.

Sources: wikipedia, St. Andrews' math history web site, Acta Arithmetica, Yivo Encyclopedia of Jews in Eastern Europe, the Math Genealogy Project, encyclopedia.com, Simons Foundation, and Journal of Graph Theory. Additional information and clarification by personal communication from Laci Babai.

**John Cameron Urschel (1991—).**
John Cameron Urschel is a Black Canadian-American who was a professional American football player, and is a mathematician. He was born in Winnipeg, Manitoba (in Canada). His parents divorced when he was very young, and he moved with his mother to Buffalo, New York (in the United States), which is where he grew up. When Urschel was a child, his first-grade teacher contacted his mother to say that he might need to repeat first grade, as he was not fitting in or demonstrating understanding. His mother, suspecting a possible racial element to this assessment, challenged the school to test him; Urschel performed extremely well, and the school suggested that he pass immediately into third grade instead.

In middle school, Urschel began to play sports, primarily as a way of fitting in, but later also as something to do with his father, who had played football at the University of Alberta. His father, who had moved to Buffalo to live nearby, also took graduate courses in math and engineering at the University of Buffalo to keep his mind active, and talked about them with Urschel. In 2004 Urschel's father gave him a math book in which he had inscribed: "To live a happy life, one has to be able to see the beauty that is around us. That sounds easy, but it is surprisingly difficult to do. It requires mental training. Studying mathematics is an ideal form of mental training. Mathematics strips away the dirt of the world to leave the beauty and purity of mathematical reasoning."

Urschel studied mathematics at Pennsylvania State University, where he also played football as an offensive lineman. He earned his bachelor's degree and his Master's at Penn State, graduating in 2014. Upon graduation, Urschel was drafted by the Baltimore Ravens of the National Football League. He played three seasons for the Ravens before announcing his retirement in 2017 at the age of 26.

Even before his retirement, Urschel enrolled in the Ph.D. program in math at the Massachusetts Institute of Technology (MIT). He first applied during his rookie year, in 2015. His application was very atypical, but MIT decided to admit him to begin his studies in 2016. This was the major factor in Urschel's decision to retire from football. He quickly found that trying to complete a doctorate while playing professional football was too demanding, as the MIT program required that he maintain full-time student status. There were other factors involved in his decision also: his fiancée was pregnant, and new evidence about long-term effects of concussions kept emerging. Urschel completed his doctorate in 2021, with a thesis entitled "Topics in Applied Linear Algebra", under the supervision of Michel Goemans. Urschel published a number of results before the end of his Ph.D., including his work with Jake Wellens (1992—) on the difficulty of testing whether or not a graph is $k$-planar that is mentioned in Section 15.1.

Urschel is married, with one child. His wife is a writer who assisted in the writing of his memoir, *Mind and Matter: A Life in Math and Football*. Urschel has been actively involved in outreach, particularly as a mentor and role model. He gives many public talks and interviews, in attempts to increase the representation of African American youth in mathematics. His status as a professional football player draws in an audience that might not otherwise be receptive to his messages about mathematics. As Urschel points out, "It's very hard to dream of being in a career if you can't relate to anyone who's actually in that field." Of his own motivation he says, "Mathematics speaks to this side of me where I'm really curious and want to know why." Urschel also enjoys playing chess.

Sources: wikipedia, American Math Society, Quanta Magazine, ESPN, MIT, Chicago Tribune, Sports Illustrated, and MIT.

**Varāhamihira (499—587).**
Varāhamihira was a 6th-century Indian mathematician, astronomer, and astrologer. He is also referred to as Varāha or Mihira. He was born in India; the exact place of his birth is unknown. One of his books says that he was educated at Kapitthaka, but it is not certain what this is now called, or where it is. What is known is that he worked at Ujjain; it was already an important mathematical centre, and he increased its prominence. There is a story that he used astrology to predict the death of the prince, and was given the emblem of the boar (varāha) in recognition of his wisdom, which led to his being known as Varāhamihira instead of Mihira.

One of Varāhamihira's most notable works is the *Brhatsambitā* (the "Great Compilation"), which covers topics including architecture, astronomy, weather, mathematics, perfumes, and more. The book has 106 chapters. It is in this work that he discussed the problem described in Example 3.2.6. He wrote that at least some of his work was a summary of previous results

in works that have not survived. The mathematical content of the *Brhatsambitā* includes versions of the arithmetic triangle, formulas for computing binomial coefficients, and trigonometric identities.

Varāhamihira also wrote some other books on astronomy and astrology; his most famous work is the *Pancasiddhantika* ("Five Astronomical Canons"), which describes five books (by previous sages) that are no longer in existence. He knew Greek, and was aware of Greek scientific understanding as well as Indian and Babylonian knowledge.

Sources: wikipedia, St. Andrews' math history web site, the Free Press, and Sanskriti magazine.

**Vadim Georgievich Vizing (1937—2017).**
Vadim Georgievich Vizing was a pioneering graph theorist in the Soviet Union, best known for Vizing's Theorem. This is the decisive result on the chromatic index of graphs that now appears in many textbooks, including the one by Bondy and Murty referenced in the biographical sketch of John Adrian Bondy (1944—). He was born in Kiev, USSR (now Kyiv, Ukraine). As his mother was partly German, in 1947 (after World War II) the family was forced by the Stalin government to move to Siberia. Beginning in 1954, Vizing studied math at Tomsk State University in Siberia, graduating in 1959. He then moved to Moscow to work on his Ph.D. at the Steklov Institute of Mathematics. He was working in the area of function approximation, which he did not enjoy.

After being refused permission to switch topics, Vizing left Moscow in 1962 without completing his graduate work. He returned to Novosibirsk in Siberia, and worked at the Siberian branch of the Soviet Academy of Sciences in nearby Akademgorodok (one of the premier research centres in the Soviet Union) from 1962 to 1968, studying graph theory. He earned his doctorate there in 1966. A.A. Zykov acted as a mentor to him.

It was during his doctoral studies that Vizing published the result for which he is best known: that the edges of a graph can be coloured with at most $\Delta + 1$ colours, where $\Delta$ is the maximum number of neighbours of any vertex. This is described in Section 14.1. The paper was written in Russian; the theorem is now considered essential content in graph theory courses. Vizing also introduced the much-studied idea of list colouring, in which the permitted colours for a vertex have to come from a predetermined list (each vertex has its own list). He posed the conjecture that the vertices and edges of a graph can be coloured with at most $\Delta + 2$ colours so that no pair of adjacent or incident objects has the same colour. This problem remains open to this day.

Vizing wanted to move back to Ukraine where it would be less cold. He was not given permission to return to Kiev. He worked in a number of small towns, and in 1974 was offered a position at the Academy for Food Technology in Odessa, where he remained for the rest of his career. In 1976 Vizing switched his research interests to scheduling problems, and only returned to the study of graph theory in 1995. He had retired by 2000, while continuing to pursue research. Vizing published more than 50 papers.

Vizing was unable to travel much outside of the Soviet Union during his career, as the Soviet government did not allow him to accept the invitations he received prior to the Perestroika reforms of the late 1980s. Vizing's pension was the equivalent of just $70USD/month; remarkably, he said that the additional $45USD/month he received from a European Union research grant enabled him to travel and meet colleagues. (Despite the effects of inflation, these numbers are as unimaginably low as they seem.) A world-class mathematician, the highest recognition Vizing received was the "Great Silver Medal of the Institute of Mathematics of the Siberian Department of the Russian Academy of Sciences".

Sources: wikipedia, European Math Society, the Math Genealogy Project, and MathSciNet. Additional information and clarification by personal communication from Laci Babai.

**Klaus Wagner (1910—2000).**
Klaus Wagner was a German graph theorist best known for his pioneering work on topological graph theory and graph minors. He was born in Cologne, Germany. The family moved (within Cologne) in 1923, and Wagner continued to live in this house for the rest of his life, taking particular interest in maintaining the garden. Wagner enrolled at the University of Cologne in 1930, studying math, physics, and chemistry. By the end of 1934 he had already submitted his doctoral thesis, entitled "Über zwei Sätze der Topologie: Jordanscher Kurvensatz und Vierfarbensatz" ("On two theorems in topology: Jordan's curve theorem, and the four-colour theorem"), under the supervision of Karl Dörge. Dörge had been a student of Issai Schur (1875—1941). Wagner's Ph.D. was awarded in 1935.

In 1937, Wagner published Wagner's Theorem, characterising planar graphs as graphs that have no $K_{3,3}$ or $K_5$ minor. Much of his best-known research focussed around similar questions of planarity, minors, and colouring. He also made the conjecture that when proved became the Robertson-Seymour Theorem (a true magnum opus of the careers of George Neil Robertson (1938—) and Paul Seymour (1950—)), showing that every family of graphs that is closed under minors can be characterised by a finite collection of forbidden minors (this result is discussed further in Section 15.2). Wagner did spend a period of his career focussed more on other topics, including topology, calculus, and analysis. He returned to graph theory and particularly topological graph theory in the late 1950s. He organised many conferences on graph theory at Oberwolfach, and wrote an introductory book about graph theory that appeared in 1970.

After completing his doctorate, Wagner worked for a time as a meteorologist for the German air force at the airports in Cologne and Berlin. The Allies took over this work after the war ended, leaving Wagner out of work, so he returned to the University of Cologne. In Germany, one way to become a full professor is to complete a "habilitation"; essentially completing an additional thesis and examination at a more advanced level, typically while working in a junior position. Wagner went through this process, completing his habilitation in 1949.

Wagner was married in 1950, and had one son. He and his wife also adopted the daughter of close friends who had died in a car accident. Wagner taught for many years at the University of Cologne, becoming a professor in 1956. He moved in 1970 to the University of Duisberg, about an hour away from Cologne. He remained there until he retired in 1978, and continued active in mathematics after his retirement. He gave his last lecture at the age of 84.

Over the course of his career, Wagner is credited by the Math Genealogy Project with supervising 25 doctoral students; he wrote more than 60 research papers, and published 7 books. He was made an honorary professor at the University of Cologne in 1971, and received an honorary doctorate from the University of Duisburg in 1997.

Sources: wikipedia, Festcolloquium for Wagner, the Math Genealogy Project, and Memorial article in Results in Mathematics.

**Jake Wellens (1992—).**
Jake Wellens is an American combinatorist still in the early stages of his career. He was born in Philadelphia, Pennsylvania (in the United States) in 1992. During high school he enjoyed competing in math competitions. He moved to California for university, attending the California Institute of Technology beginning in 2011 and graduating in 2015.

Wellens moved back east to Massachusetts, enrolling in the Ph.D. program at the Massachusetts Institute of Technology (MIT). He worked under the supervision of Henry Cohn, and graduated in 2020 during the global pandemic. His thesis was entitled "Assorted results in boolean function complexity, uniform sampling and clique partitions of graphs". It was at MIT that he met John Cameron Urschel (1991—) and they collaborated on several projects, including their work on the difficulty of testing whether or not a graph is $k$-planar mentioned in Section 15.1.

Sources: MIT, Main Line Times, and CalTech. Updated and confirmed through personal communication.

**Richard Michael Wilson (1945—).**
Richard Michael Wilson is an outstanding American design theorist best known for his impressive proofs establishing the existence of certain kinds of designs. He was born in Gary, Indiana (in the United States). He completed his bachelor's degree at Indiana University, graduating in 1966. He then moved to Columbus, Ohio, where he received a Master's degree from Ohio State University in 1968. He remained at Ohio State University to complete his doctorate in 1969 under the supervision of Dwijendra Kumar Ray-Chaudhuri (1933—). His thesis was entitled "An Existence Theory for Pairwise Balanced Designs".

After completing his Ph.D., Wilson stayed in Columbus, immediately taking a position at Ohio State University. He worked his way up through the ranks there, becoming a professor in 1974. In 1980 he moved to Pasadena, California, where he accepted a job at the California Institute of Technology (CalTech). Wilson remained there until his retirement in 2014, and is now a professor emeritus at CalTech.

Among Wilson's many enormous contributions to combinatorics, here are just a few that are particularly relevant to the material in this book. During his graduate studies and in collaboration with Ray-Chaudhuri, Wilson determined exactly when Kirkman Triple Systems exist (see Theorem 18.1.11). Subsequently, in 1974 Wilson proved Wilson's Theorem, that BIBDs exist whenever $v$ is sufficiently large and the natural divisibility conditions are satisfied, as discussed in Section 17.2. Wilson also produced the best known lower bounds for the number of mutually orthogonal Latin squares of order $n$. Wilson published more than 60 papers during his career. He had supervised at least 29 doctoral students as of 2021, including Gary McGuire (1967—). He is also a co-author of a graduate-level introductory textbook in combinatorics.

In 1975, Wilson won the Pólya Prize from the Society for Industrial and Applied Math. He held a research fellowship from 1975 to 1977, and was the Sherman Fairchild Distinguished Scholar at CalTech in the winter of 1976. He has held visiting appointments at the University of London, the University of Illinois at Chicago, and the University of Minnesota, and an adjunct appointment at the University of Waterloo from 1982 to 1987.

In the late 1970s Wilson developed an interest in historical flutes. He collects, studies, and performs on original European flutes from the 18th and 19th centuries, as well as on replicas from these periods or earlier. He jointly wrote a book that has been published, on flute performance in London in about 1830. His interest has broadened to traditional flutes from around the world, which he also collects and performs on, sometimes at poetry-readings (his wife is a poet).

Sources: wikipedia, Wilson's c.v., the Math Genealogy Project, and Designs, Codes, and Cryptography.

**Wesley Stoker Barker Woolhouse (1809—1893).**
Wesley Stoker Barker Woolhouse was a 19th century British actuary and editor, whose writings included mathematics, music, and actuarial science. He was born in North Shields, England, in 1809. When he was 13 he won a prize for math offered by *The Ladies' Diary* that was open to all ages. Woolhouse spent most of his career as an actuary, but also had interests in music, steam engines, and other diverse fields. He was active in publishing through much of his life, including both books and periodicals.

Woolhouse published a number of papers, problems, and solutions to problems over the next years, making his name known in mathematical circles. He was made deputy superintendent of the *Nautical Almanac*'s office from 1830 to 1837, where he published tables and papers on astronomy. After a difference of opinion with his superior, he left the *Almanac*, and in 1839 he took a position as actuary of the International Loan fund. At one point in his career he

was asked by Lord Ashley to calculate the distance walked by factory workers who had to retie threads when they broke, for the purposes of motivating a reform bill limiting the working day to 10 hours. Using probability, Woolhouse estimated the distance at 30 miles per day.

In 1841 the *Ladies' Diary* and the *Gentleman's Diary* were succeeded by the *Lady's and Gentleman's Diary*, and Woolhouse became the editor in 1844. He held this post until 1865; this annual periodical itself only lasted until 1871, so he was the editor for most of its existence. It was in this publication that (in his first year as editor) Woolhouse included the problem (see Section 18.1) that inspired Reverend Thomas Penyngton Kirkman (1806—1895) in his work on designs, and Kirkman's schoolgirl problem also appeared in the *Diary* in 1850.

During his lifetime, Woolhouse wrote and published books on subjects as diverse as geometry, musical theory, mortality in the Indian Army (this was an actuarial investigation), calculus, and coins and calendars. He co-founded the English Institute of Actuaries in 1848, and played a significant role in the development of actuarial science and the actuarial profession in England. Woolhouse also enjoyed needlework, and amused himself by calculating the exact length of thread he would need for various projects.

Sources: wikipedia, Journal of the Institute of Actuaries, wikipedia, and Monthly Notices of the Royal Astronomical Society.

**Ming-Yao Xu (1941—).**
Ming-Yao Xu is a contemporary Chinese mathematician who is responsible for the formation of two significant schools of mathematical research in China. He was born in Tianjin, China, during World War II (part of the period known in China as the War of Resistance Against Japan). During the war years the family traveled through the south of China evading Japanese attacks, settling in Peking in 1945. The Chinese Communist Revolution was still ongoing, and from 1945 to 1949 the Kuomintang (Nationalist Party) controlled Peking; the Chinese Communist Party captured the city in 1949.

His wartime experiences made Xu a strong supporter of the Communist Party. He joined the Communist Youth League in 1956 and was elected as a leader of the League at his school.

Xu enrolled in Peking University in 1959. His university years fell between periods of significant political movements in China, but life was not calm. Xu expressed his own opinions on politics and came under criticism from his classmates twice. Besides these political controversies, chaotic "teaching reform" efforts at the university during the time Xu was there made it difficult to follow prerequisites through anything like a standard curriculum. As a result it took Xu almost 6 years to complete his bachelor's degree; he graduated in 1965. Despite the chaos, Xu studied hard. His teacher Shisun Ding became a close mentor and led him to study group theory. Ding was later president of Peking University (1984—1989).

After completing university, Xu was assigned to work in the Tangshan District. His time in Tangshan coincided very closely with the 10 years of the tumultuous Cultural Revolution. For almost all of this period, Xu was officially assigned to teach at the Tangshan Number 5 Middle School, but the schools worked abnormally during this period and were even closed for the first few years. During these years, Xu was often assigned to manual labour including as a purchaser for the school cafeteria; in the Tangshan Steel Plant; and in the Kailuan Coal Mine. In 1976, Xu escaped the worst of the deadly Tangshan earthquake (7.8 on the Richter scale) as he happened to be visiting Beijing. On his return he found that his bed had been split in two and the floor had collapsed. Toward the end of this time China's restrictions on publishing mathematical research lightened; Xu's first paper (based on his undergraduate thesis) was published in 1976.

At this time there was a shortage of people in higher education in China. Peking University wanted Xu to return to teach, but the Tangshan Education Bureau refused to release him. Eventually Xu got around this by passing the postgraduate entrance examination. Xu returned to Peking University in 1978 at the age of 37 as a graduate student. He worked under the

supervision of Xiaofu Duan (also known as Hsio-Fu Tuan) and Efang Wang, and finished his graduate studies in 1980, immediately beginning to teach.

During his 24 years at Peking University, Xu built around himself a school of Chinese researchers working on groups, and on group interactions with graphs and with maps. He also developed expertise in computational group theory. Xu's efforts included writing a new graduate textbook on group theory; organising a seminar; and building relationships with international experts. Chinese researchers had been largely isolated from the international research community for many years, so building international relations was a challenge. Xu organised international conferences, but also travelled extensively in order to build these relationships, which he then used to develop opportunities for his students. He made long visits to Australia in 1985, 1987, and 1989; during the first of these he visited Cheryl Elisabeth Praeger (1948—) and they discovered the Praeger-Xu graphs described in Section 12.5. Xu also visited Canada in 1991 and 1992; Slovenia in 1995; and South Korea in 1999. In 1993 Xu began to supervise his own doctoral students for the first time; in the 11 years between then and his retirement from Peking University, Xu supervised 13 Ph.D. students.

In 2003, just as Xu was preparing to retire from Peking University, he accepted an opportunity to become the first Distinguished Professor at Shanxi Normal University. Here too he developed a school of research. This time the focuses were on $p$-groups and computational group theory. Again, one of his first steps was to write a graduate-level textbook. After establishing this school of research, Xu retired again in 2010.

Xu enjoys music, and takes pleasure in reading literature and history. In his leisure time he has independently conducted an extensive study of modern Chinese history, with a particular focus on the political, cultural, and ideological history of China between 1957 and 1976. He has published about 10 scholarly articles in journals that focus on these topics. Xu is married; his daughter Jing Xu is an associate professor of math at Capital Normal University, having completed her doctorate under the supervision of Praeger.

Xu has written more than 90 mathematical research papers, principally in the areas of group theory and group actions on graphs. Several of his papers introduced important concepts that were subsequently studied extensively by other researchers. Groups and group actions on graphs are now major areas of mathematical research in China, and most Chinese researchers in these fields owe something to Xu, whether it is for his textbooks, the international relationships he established, or mentorship (direct or indirect). Xu has won the Science and Technology Progress Award from the Ministry of Education in China.

Sources: personal communications, Amazon, and MathSciNet. Most of the content of this biography comes from material shared by Shaofei Du, Xu's first Ph.D. student, from an article in preparation for the occasion of Xu's 80th birthday.

**Kazimierz Zarankiewicz (1902—1959).**
Kazimierz Zarankiewicz was a Polish mathematician, working in topology and other areas of mathematics. In graph theory, his name is best known for a problem he shared with Pál Turán (1910—1976). He was born in Częstochowa, then in the Austro-Hungarian empire (now in Poland). He enrolled at the University of Warsaw in 1919. By then, Warsaw was the capital of the newly-independent Poland. Zarankiewicz graduated with a Ph.D. in topology in 1923, although his thesis was not published until 1927. Topology was a major strength of researchers at the University of Warsaw at that time.

Poland followed the German tradition of the "habilitation", in which young researchers complete an additional (more advanced) thesis and examination, typically while working at a junior level. Zarankiewicz completed his habilitation in 1929 while working as an assistant at the University of Warsaw, and was promoted. He left in 1930 to visit Vienna and Berlin, and was unable to obtain a position at the University of Warsaw on his return. He taught instead at

the Warsaw University of Technology, the Agricultural College, and at the University of Tomsk in the years from 1931 to 1939.

In 1939, Nazi Germany invaded Poland. After the invasion, many academics were killed or sent to concentration camps, and universities were closed. During the war, Zarankiewicz continued to teach at the underground university that had been established, defying Nazi edict. In 1944 he was arrested for this, and sent to a concentration camp. He survived the camp, returned to Warsaw in 1945, and in 1946 took a position at the Warsaw University of Technology, where he was promoted to professor in 1948. Zarankiewicz also visited the United States in 1948, teaching at a number of universities including Harvard.

Zarankiewicz published at least 45 papers. His research was in topology, complex functions, number theory, and graph theory. The problem for which he is best known in graph theory (mentioned in Section 14.2) was originally stated as a problem about matrices, but can be reformulated in terms of bipartite graphs. Zarankiewicz traded problems with Turán: Turán was involved in finding an upper bound for the Zarankiewicz problem, and Zarankiewicz found and proved an upper bound on Turán's problem about the number of crossings in a complete bipartite graph, and conjectured that his upper bound is the correct answer. (This is one of the problems Turán formulated while in a labour batallion, and is mentioned in his biographical sketch.)

Zarankiewicz was also a noteworthy mentor, supervising Poland's Mathematical Olympiad (a competition for high school students) from 1949 to 1957, and remaining on as a member of the board. He also served as an editor for a math journal aimed at high school teachers. From 1948 to 1951 he was president of the Warsaw section of the Polish Mathematical Society, and he was a founder of the Polish Astronautical Society in 1956. Zarankiewicz died at the age of 57 in London, England, where he was attending the Congress of the International Astronautical Federation, of which he was the vice-president. There is a street in Warsaw named for him.

Sources: wikipedia, St. Andrews' math history web site, and encyclopedia.com.

**Zhu Shijie (1249—1314).**
Zhu Shijie was a 13th-century Chinese mathematician who wrote, travelled, and taught during the Yuan dynasty. He was born near what is now Beijing, China, in 1249. The Yuan dynasty came into power during his lifetime in 1279, when Kublai Khan united China. Zhu wrote two books that have survived: the *Suan hsüeh Ch'i-mong* ("Introduction to Computational Studies") in 1299, and the *Siyuan Yuchian* ("Jade Mirror of the Four Unknowns") in 1303. In the preface to the *Siyuan Yuchian*, Zhu says that he has travelled around China for more than 20 years, teaching mathematics. Widespread travel had only become safe once China was unified.

The *Suan hsüeh Ch'i-mong* is written as a textbook on elementary mathematics. It discusses basic computations, areas, and volumes. It also includes polynomials. It was lost in China for a long time, but had reached Korea and Japan, where it survived, and was translated back into Chinese in 1839.

The *Siyuan Yuchian* is Zhu's most important work. It includes algebra in four unknowns, finds cube roots and square roots, and uses matrix reduction to solve systems of linear equations. Most importantly for our purposes, it includes the arithmetic triangle (which Zhu calls "the table of the ancient method of powers") and binomial coefficients, as discussed in Section 4.3. This book too was lost for a time. A copy was eventually found, but the version had passed through many hands and it is not always clear when or where material and errors may have been introduced.

Sources: wikipedia, St. Andrews' math history web site, and Mathigon.

# Appendix C

# Solutions to selected exercises

For the reader's convenience, solutions are given with full work shown as well as a final numerical solution. Typically the final numerical solution would not be expected, but makes it easier to verify an answer that has been reached using a different method.

## Solutions for Chapter 2

**Solutions to Exercise 2.1.7:**

2.1.7.1 There are 4 choices for colour; 2 choices for air conditioning; 5 choices for stereo; and 3 choices for floor mats. So in total, there are $4 \cdot 2 \cdot 5 \cdot 3 = 120$ different combinations of options. Of these, 3 combinations are available at the dealership, so the probability that one of these cars has the options Ace wants is $3/120 = 1/40$.

2.1.7.2 In Candyce's book, the reader will have 3 choices at the first decision point, and 2 choices at each of the following three decision points. Thus, there are a total of $3 \cdot 2 \cdot 2 \cdot 2 = 3 \cdot 2^3 = 24$ possible storylines. Candyce must write 24 endings.

**Solutions to Exercise 2.2.6:**

2.2.6.1 Let's call the black markers Black A, Black B, and Black C. The four possible options are: I leave behind the blue marker; I leave behind Black A; I leave behind Black B; I leave behind Black C. In each of the final three cases, I take the blue marker. Therefore, the probability that I take the blue marker is $(1 + 1 + 1)/4 = 3/4$.

2.2.6.2 If Ocean is thinking of a letter, there are 26 things they could be thinking of. If they are thinking of a digit, there are 10 things they could be thinking of. In total, there are $10+26 = 36$ things they could be thinking of.

**Solutions to Exercise 2.3.4:**

2.3.4.1 We divide the possible passwords into three cases, depending on whether the digit is in the first, second, or third position. In each case, we have 10 choices for the digit, 26 choices for the first lowercase letter, and 26 choices for the second lowercase letter. Thus, in each case we have $10 \cdot 26^2$ possible passwords. In total, there are $10 \cdot 26^2 + 10 \cdot 26^2 + 10 \cdot 26^2 = 30 \cdot 26^2 = 20,280$ different passwords with these constraints.

2.3.4.2 We divide the possible passwords into two cases, depending on whether there are 8 or 9 characters. If there are 8 characters, then the product rule tells us that there are $10^8$ passwords

consisting entirely of digits (10 choices for the digit in each of the 8 positions). Similarly, if there are 9 characters, then there are $10^9$ passwords consisting entirely of digits. In total, there are $10^8 + 10^9 = 1,100,000,000$ passwords with these constraints.

**Solutions to Exercise 2.4.1:**

It is sometimes possible to turn a use of the sum rule into a use of the product rule, or vice versa, so you might get different answers that could be correct. The answers below represent one natural way of solving each problem.

2.4.1.1 Use both rules. There are two cases that should be added together: the number of numbers that have two digits, and the number of numbers that have four digits. For each of these cases, use the product rule to determine how many numbers have this property. (The answer is $9 \cdot 10 + 9 \cdot 10^3 = 9,090$. Note that in order for a number to have exactly $k$ digits, the leading digit cannot be zero.)

2.4.1.2 Use only the product rule. There are 6 outcomes from the red die, and for each of these, there are 6 outcomes from the yellow die, for a total of $6 \cdot 6 = 36$ outcomes.

# Solutions for Chapter 3

**Solutions to Exercise 3.1.9:**

3.1.9.1 The band may choose any one of the 6 people to play lead guitar. They may then choose any one of the remaining 5 people to play bass. Therefore, the band can be completed in $6 \cdot 5 = 30$ ways. You might also observe that the number of ways to complete the band is the number of 2-permutations (for the two open spots) of 6 people, which is $6 \cdot \ldots \cdot (6 - 2 + 1) = 6 \cdot 5 = 30$.

3.1.9.2 Divide this into two cases, depending on which of the two parts Garth got. If they got the first part, then there were 8 other people competing for 4 other roles, so the number of ways of completing the cast in this case is the number of 4-permutations of the 8 people. Repeating this argument for the second case (if Garth got the other part) and adding the two numbers, we see that in total there are $2 \cdot 8 \cdot 7 \cdot 6 \cdot 5$ (since $8 - 4 + 1 = 5$) ways of completing the cast. This works out to $3,360$.

**Solutions to Exercise 3.2.10:**

3.2.10.2 At the end of this trick, the only sets of cards that they could not possibly end up with are sets that contain nothing but spades. There are $\binom{13}{3}$ sets that include only spades (choose any 3 of the 13 spades), and $\binom{52}{3}$ possible sets of 3 cards from the deck as a whole, so the number of sets of three cards that are not all spades is $\binom{52}{3} - \binom{13}{3} = 22,100 - 286 = 21,814$.

3.2.10.3 The leading digit cannot be a zero, so if there are to be exactly two zeroes, we have 4 possible positions in which they can be placed. Thus, there are $\binom{4}{2}$ ways of choosing where to place the two zeroes. In each of the remaining three positions, we can place any of the digits 1 through 9, so there are $9^3$ choices for the remaining digits. Thus, there are $\binom{4}{2}9^3 = 4,374$ 5-digit numbers that contain exactly two zeroes.

**Solutions to Exercise 3.3.7:**

3.3.7.2 Using the Binomial Theorem, we see that

$$(a+b)^5(c+d)^6 = \left(\sum_{r=0}^{5}\binom{5}{r}a^r b^{5-r}\right)\left(\sum_{s=0}^{6}\binom{6}{s}c^s d^{6-s}\right).$$

To find the coefficient of $a^2b^3c^2d^4$, we must take $r=2$ and $s=2$. This gives us the term $\binom{5}{2}a^2b^3\binom{6}{2}c^2d^4 = \binom{5}{2}\binom{6}{2}a^2b^3c^2d^4$. Thus, the coefficient of $a^2b^3c^2d^4$ is $\binom{5}{2}\binom{6}{2} = 10\cdot 15 = 150$.

3.3.7.4 Using the Binomial Theorem, we see that

$$(a+b)^5 + (a+b^2)^4 = \sum_{r=0}^{5}\binom{5}{r}a^r b^{5-r} + \sum_{s=0}^{4}\binom{4}{s}a^s (b^2)^{4-s}.$$

The coefficient of $a^3b^2$ in the first summand arises when $r=3$; in the second summand, it arises when $s=3$. This gives us the term $\binom{5}{3}a^3b^2 + \binom{4}{3}a^3(b^2)^1$. Thus, the coefficient of $a^3b^2$ is $\binom{5}{3} + \binom{4}{3} = 10+4 = 14$.

# Solutions for Chapter 4

**Solutions to Exercise 4.1.3:**

4.1.3.1 When counting the number of subsets of an $n$-set, we saw that there is a bijection between that number and the number of binary strings of length $n$: identify each element of the set with a position in the string, and put a 0 in that position if the element is not in the subset, and a 1 if it is.

Analogously, we can find a bijection between the number of these structures and the number of ternary strings of length $n$ (strings containing 0, 1, or 2 in each position). Identify each element of the set with a position in the string, put a 0 in that position if the element is not in the structure, a 1 if it occurs once, and a 2 if it occurs twice. Thus, we can form $3^n$ structures from the set $\{1,\ldots,n\}$: there are 3 choices for each of the $n$ entries in the ternary string.

4.1.3.3 We identify each of the ten Olympic contenders with a crib, and each of the three dolls with one of the three medals. If the doll corresponding to the gold medal goes into crib $i$, this corresponds to the competitor corresponding to crib $i$ winning the gold medal. Similarly, if the doll corresponding to the silver medal goes into crib $j$, this is equivalent to the contender corresponding to crib $j$ winning the silver medal; and the doll corresponding to bronze going into crib $k$ is equivalent to the contender corresponding to crib $k$ winning the bronze medal.

**Solutions to Exercise 4.2.10:**

4.2.10.2 **COMBINATORIAL PROOF.** *The problem:* We use the problem given to us in the hint, so will be counting the number of ways to start with $n$ dogs, determine $r$ who will enter a competition and $k$ of those who will be finalists.

*Counting method 1:* From the $n$ dogs, we first choose the $r$ who will enter the competition. This can be done in $\binom{n}{r}$ ways. For each of these ways, we can choose $k$ of the $r$ competitors to become finalists in $\binom{r}{k}$ ways. Thus, there are a total of $\binom{n}{r}\binom{r}{k}$ ways to choose the dogs.

*Counting method 2:* From the $n$ dogs, choose $k$ who will be the finalists. This can be done in $\binom{n}{k}$ ways. For each of these ways, we can look at the remaining $n-k$ dogs and choose $r-k$ to be the competitors who will not be finalists, in $\binom{n-k}{r-k}$ ways. Thus, there are a total of $\binom{n}{k}\binom{n-k}{r-k}$ ways to choose the dogs.

*Conclusion:* Since both of these solutions count the answer to the same problem, the answers must be equal, so we have $\binom{n}{r}\binom{r}{k} = \binom{n}{k}\binom{n-k}{r-k}$. ■

4.2.10.3 **COMBINATORIAL PROOF.** *The problem:* We will count the number of ways to choose a random sample of $n$ people from a class of $n$ men and $n$ women.

*Counting method 1:* From the $2n$ total people, choose $n$ of them for the random sample. This can be done in $\binom{2n}{n}$ ways.

*Counting method 2:* Let $r$ represent the number of men who will be in the sample. Notice that $r$ may have any value from 0 up to $n$. We divide the problem into these $n+1$ cases, and take the sum of all of the answers. In each case, we can choose the $r$ men for the sample from the $n$ men, in $\binom{n}{r}$ ways. For each of these ways, from the $n$ women, we choose $r$ who will not be part of the sample (so the remaining $n-r$ will be in the sample, for a total of $r+n-r=n$ people in the sample). There are $\binom{n}{r}$ ways to do this. Thus the total number of ways of choosing $r$ men and $n-r$ women for the sample is $\binom{n}{r}^2$. Adding up the solutions for all of the cases, we obtain a final answer of $\sum_{r=0}^{n}\binom{n}{r}^2$.

*Conclusion:* Since both of these solutions count the answer to the same problem, the answers must be equal, so we have $\sum_{r=0}^{n}\binom{n}{r}^2 = \binom{2n}{n}$. ■

**Solutions to Exercise 4.2.11:**

4.2.11.1 We could use this expression to count the number of ways of choosing one leader and some number (which could be zero) of other team members for a project, from a group of $n$ people. There are $n$ ways to choose the leader, and for each of these, there are $2^{n-1}$ ways of choosing a subset of the other $n-1$ people to be team members.

4.2.11.3 We could use this expression to count the number of ways of starting with a collection of $n$ books, choosing $r$ books to put out on bookshelves and some other number (which could be zero) of other books to keep but not display. We break this down into cases depending on the total number $k$ of books that are kept (including the displayed books), which could be anywhere from $r$ up to $n$. We will take the sum of all of the answers. In each case, there

are $\binom{n}{k}$ ways of choosing the $k$ books to keep, and for each such choice, there are $\binom{k}{r}$ ways of choosing the $r$ books to display from the books that are kept. Thus, there are a total of $\sum_{k=r}^{n} \binom{n}{k}\binom{k}{r}$ solutions to this problem.

# Solutions for Chapter 5

**Solutions to Exercise 5.1.5:**

5.1.5.2 There are $6^3$ ways for the teacher gifts to be chosen (each child can choose any one of the six types of prizes to give to his teacher). There are $\left(\binom{6}{3}\right)$ ways for Kim to choose his other three prizes; $\left(\binom{6}{2}\right)$ ways for Jordan to choose his other two prizes, and $\left(\binom{6}{5}\right)$ ways for Finn to choose his other five prizes. Thus, the total number of ways for the prizes (including teacher gifts) to be chosen is

$$\begin{aligned} 6^3 \left(\binom{6}{3}\right)\left(\binom{6}{2}\right)\left(\binom{6}{5}\right) &= 6^3 \binom{6+3-1}{3}\binom{6+2-1}{2}\binom{6+5-1}{5} \\ &= 6^3 \binom{8}{3}\binom{7}{2}\binom{10}{5} \\ &= 6^3 \cdot 56 \cdot 21 \cdot 252 = 64,012,032. \end{aligned}$$

5.1.5.3 Since the judges must choose at least one project from each age group, this is equivalent to a problem in which they are choosing only six projects to advance, with no restrictions on how they choose them. They can choose six projects from three categories in $\left(\binom{3}{6}\right) = \binom{3+6-1}{6} = \binom{8}{6} = 28$ ways.

**Solutions to Exercise 5.1.6:**

5.1.6.1 **COMBINATORIAL PROOF.** *The problem:* We will count the number of ways of choosing $k$ items from a menu that has $n$ different entries (including mac and cheese), in two ways.

*Counting method 1:* By definition, the answer to this is $\left(\binom{n}{k}\right)$.

*Counting method 2:* We divide our count into two cases, according to whether or not we choose any orders of mac and cheese. If we do not choose any mac and cheese, then we must choose our $k$ items from the other $n-1$ entries on the menu. We can do this in $\left(\binom{n-1}{k}\right)$ ways. If we do choose at least one order of mac and cheese, then we must choose the other $k-1$ items from amongst the $n$ entries on the menu (with mac and cheese still being an option for additional choices). We can do this in $\left(\binom{n}{k-1}\right)$ ways. By the sum rule, the total number of ways of making our selection is $\left(\binom{n-1}{k}\right) + \left(\binom{n}{k-1}\right)$.

*Conclusion:* Since both of these methods are counting the same thing, the answers must be equal, so $\left(\binom{n}{k}\right) = \left(\binom{n-1}{k}\right) + \left(\binom{n}{k-1}\right)$. ■

**Solutions to Exercise 5.2.5:**

5.2.5.1 There are 14 words in the list. The word "the" appears three times; the words "on" and "child" appear twice each; the other seven words each appear once. Thus, the number of "poems" (orderings of the set) is

$$\binom{14}{3,2,2,1,1,1,1,1,1,1} = \frac{14!}{3!2!2!} = 3,632,428,800.$$

# Solutions for Chapter 6

**Solutions to Exercise 6.1.6:**

6.1.6.1 Various formulas are possible. Most simply, the sequence can be described by the recurrence relation $s_1 = 4$, $s_i = 2s_{i-1} + 1$ for $i \geq 2$. With this description, the next term is $s_5 = 2(39) + 1 = 79$.

6.1.6.3 Adjusting the recurrence relation from Example 6.1.5, we obtain the new relation

$$r_n = r_{n-1} - 20 + .01(r_{n-1} - 20).$$

This simplifies to $r_n = 1.01(r_{n-1} - 20)$. We still have $r_0 = 2000$. We now have

$$r_1 = 1.01(r_0 - 20) = 1.01(1980) = 1999.80.$$

Stavroula is (marginally) losing money from the beginning. This situation will only get worse as her starting balance each year dwindles.

**Solutions to Exercise 6.2.6:**

6.2.6.1 **PROOF.** *Base case:* $n = 1$. We have $b_1 = 5$ and $5 + 4(1 - 1) = 5$, so $b_n = 5 + 4(n - 1)$ when $n = 1$.

*Inductive step:* We begin with the inductive hypothesis. Let $k \geq 1$ be arbitrary, and suppose that the equality holds for $n = k$; that is, assume that $b_k = 5 + 4(k - 1)$.

Now we want to deduce that

$$b_{k+1} = 5 + 4(k + 1 - 1) = 5 + 4k.$$

Using the recursive relation, we have $b_{k+1} = b_k + 4$ since $k + 1 \geq 2$. Using the inductive hypothesis, we have $b_k = 5 + 4(k - 1)$. Putting these together gives

$$b_{k+1} = 5 + 4(k - 1) + 4 = 5 + 4k - 4 + 4 = 5 + 4k = 5 + 4(k + 1 - 1),$$

as desired. This completes the proof of the inductive step.

By the Principle of Mathematical Induction, $b_n = 5 + 4(n - 1)$ for every $n \geq 1$. ■

6.2.6.3 **PROOF.** *Base case:* $n = 0$. We have $0! = 1$ (by definition) and $n = 0$, so $n! = 1 \geq 0 = n$. Thus, $n! \geq n$ when $n = 0$.

*Inductive step:* We begin with the inductive hypothesis. Let $k \geq 0$ be arbitrary, and suppose that the inequality holds for $n = k$; that is, assume that $k! \geq k$.

Now we want to deduce that $(k + 1)! \geq k + 1$. Using the definition of factorial, we have $(k + 1)! = (k + 1)k!$ since $k + 1 \geq 0 + 1 = 1$. Using the inductive hypothesis, we have $k! \geq k$. Putting these together gives

$$(k + 1)! = (k + 1)k! \geq (k + 1)k.$$

If $k \geq 1$, then

$$(k+1)k \geq (k+1)1 = k+1$$

and we are done. If $k = 0$, then $(k+1)! = 1! = 1 = k+1$ and again the inequality is satisfied. This completes the proof of the inductive step.

By the Principle of Mathematical Induction, $n! \geq n$ for every $n \geq 0$. ■

**Solutions to Exercise 6.3.5:**

6.3.5.2 **PROOF.** Base cases: We will have four base cases: $n = 12$, $n = 13$, $n = 14$, and $n = 15$.

For $n = 12$, I can get \$12 onto my gift card by buying three increments of \$4, since $4 + 4 + 4 = 12$.

For $n = 13$, I can get \$13 onto my gift card by buying two increments of \$4 and one of \$5, since $4 + 4 + 5 = 13$.

For $n = 14$, I can get \$14 onto my gift card by buying two increments of \$5 and one of \$4, since $4 + 5 + 5 = 14$.

For $n = 15$, I can get \$15 onto my gift card by buying three increments of \$5, since $5 + 5 + 5 = 15$.

Inductive step: We begin with the (strong) inductive hypothesis. Let $k \geq 15$ be arbitrary, and assume that for every integer $i$ with $k - 3 \leq i \leq k$, I can put $\$i$ onto my gift card.

Now I want to deduce that I can put $\$(k+1)$ onto my gift card. Using the inductive hypothesis in the case $i = k - 3$, I see that add can put $\$(k-3)$ onto my gift card by buying increments of \$4 or \$5. Now if I buy one additional increment of \$4, I have put a total of $\$(k-3+4) = \$(k+1)$ onto my gift card, as desired. This completes the proof of the inductive step.

By the (strong) Principle of Mathematical Induction, I can put any amount of dollars that is at least \$12 onto my gift card. ■

# Solutions for Chapter 7

**Solutions to Exercise 7.1.3:**

7.1.3.2 Since the $n$th term of the sequence is $2^n$, the generating function is $\sum_{n=0}^{\infty} 2^n x^n$.

7.1.3.3 The generating function is $1 + 5x + 10x^2 + 15x^3 + 10x^4 + 5x^5 + x^6$.

**Solutions to Exercise 7.2.7:**

7.2.7.1 We have

$$\binom{-5}{7} = (-1)^7 \binom{5+7-1}{7} = -\binom{11}{7} = -330.$$

7.2.7.2 By the Generalised Binomial Theorem, the coefficient of $y^4$ in $(1+y)^{-2}$ is $\binom{-2}{4}$, so (replacing $y$ with $-x$) the coefficient of $x^4$ in $(1-x)^{-2}$ is

$$(-1)^4 \cdot \binom{-2}{4} = (1) \cdot \frac{(-2)(-3)(-4)(-5)}{4!} = 5.$$

**Solutions to Exercise 7.3.5:**

7.3.5.1 **PROOF.** *Base case:* $n = 1$. The left-hand side of the equation in this case is $1 + x$. The right-hand side is $\dfrac{1 - x^2}{1 - x}$. Since $1 - x^2 = (1 - x)(1 + x)$, we can rewrite the right-hand side as $\dfrac{(1 - x)(1 + x)}{1 - x}$. Cancelling the $1 - x$ from the top and bottom gives $1 + x$, so the two sides are equal. Since a generating function is a formal object, $x$ is acting as a placeholder, and we do not need to worry about the possibility that $1 - x = 0$ that would prevent us from cancelling these factors.

*Inductive step:* Let $k \geq 1$ be arbitrary, and suppose that

$$1 + \cdots + x^k = \frac{1 - x^{k+1}}{1 - x}.$$

Now we must deduce that

$$1 + \cdots + x^{k+1} = \frac{1 - x^{k+2}}{1 - x}.$$

We have

$$1 + \cdots + x^{k+1} = (1 + \cdots + x^k) + x^{k+1}.$$

Applying our inductive hypothesis, this is $\dfrac{1 - x^{k+1}}{1 - x} + x^{k+1}$. Adding this up over a common denominator of $1 - x$ gives

$$\frac{1 - x^{k+1} + x^{k+1} - x^{k+2}}{1 - x} = \frac{1 - x^{k+2}}{1 - x},$$

as desired.

By the Principle of Mathematical Induction,

$$1 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x}$$

for every $n \geq 1$. ■

7.3.5.4 The generating function for this problem is

$$(x + x^2 + x^3 + x^4 + x^5 + x^6)^5.$$

We can rewrite this as

$$x^5(1 + x + x^2 + x^3 + x^4 + x^5)^5.$$

Finding the coefficient of $x^{11}$ in this expression is equivalent to finding the coefficient of $x^6$ in

$$(1 + x + x^2 + x^3 + x^4 + x^5)^5 = \left(\frac{1 - x^6}{1 - x}\right)^5.$$

Using the Binomial Theorem and substituting $y = -x^6$, we see that

$$\begin{aligned}(1 - x^6)^5 &= (-x^6)^0 + \binom{5}{1}(-x^6)^1 + \binom{5}{2}(-x^6)^2 + \binom{5}{3}(-x^6)^3 + \binom{5}{4}(-x^6)^4 + (-x^6)^5 \\ &= 1 - 5x^6 + 10x^{12} - 10x^{18} + 5x^{24} - x^{30}.\end{aligned}$$

The function we're interested in is the product of this with $(1-x)^{-5}$, and we are looking for the coefficient of $x^6$. The only ways of getting an $x^6$ term from this product are by taking the $x^0$ term above and multiplying it by the $x^6$ term from $(1-x)^{-5}$, or by taking the $x^6$ term above and multiplying it by the $x^0$ term from $(1-x)^{-5}$.

Using the Generalised Binomial Theorem (and substituting $y=-x$), the coefficient of $x^0$ in $(1-x)^{-5}$ is

$$(-1)^0\binom{-5}{0} = (-1)^0(-1)^0\binom{5+0-1}{0} = 1.$$

Similarly, the coefficient of $x^6$ in $(1-x)^{-5}$ is

$$(-1)^6\binom{-5}{6} = (-1)^6(-1)^6\binom{5+6-1}{6} = \binom{10}{6} = 210.$$

Thus, the number of ways in which Trent can roll a total of 11 on his five dice is the coefficient of $x^{11}$ in our generating function, which is $\binom{10}{6} - 5 = 205$. The probability of this happening is 205 divided by the total number of outcomes of his roll, which is $6^5 = 7776$, so 205/7776, or about 2.5%.

# Solutions for Chapter 8

**Solutions to Exercise 8.1.2:**

8.1.2.1 First we rewrite the generating function as a sum of two parts:

$$\frac{1}{(1+2x)(2-x)} = \frac{A}{1+2x} + \frac{B}{2-x} = \frac{A(2-x)+B(1+2x)}{(1+2x)(2-x)}.$$

Now the numerator gives $2A+B+(2B-A)x = 1$ as polynomials. Hence we must have $2B-A=0$ and $2A+B=1$. Combining these gives $B=1/5$ and $A=2/5$. Thus the given generating function is equal to

$$\frac{2}{5}(1+2x)^{-1} + \frac{1}{5}(2-x)^{-1} = \frac{2}{5}(1+2x)^{-1} + \frac{1}{10}(1-\frac{1}{2}x)^{-1}.$$

Using the Generalised Binomial Theorem, the coefficient of $x^r$ in the first of these summands is $\frac{2}{5}(-1)^r 2^r$, while the coefficient of $x^r$ in the second summand is $\frac{1}{10}\left(\frac{1}{2}\right)^r$. Thus, the coefficient of $x^r$ is $\frac{2}{5}(-1)^r 2^r + \frac{1}{10}\left(\frac{1}{2}\right)^r$.

8.1.2.3 First we rewrite the generating function as a sum of three parts:

$$\begin{aligned}\frac{1+2x}{(1-2x)(2+x)(1+x)} &= \frac{A}{1-2x} + \frac{B}{2+x} + \frac{C}{1+x} \\ &= \frac{A(2+x)(1+x)+B(1-2x)(1+x)+C(1-2x)(2+x)}{(1-2x)(2+x)(1+x)}.\end{aligned}$$

Now the numerator gives

$$\begin{aligned}&A(2+3x+x^2)+B(1-x-2x^2)+C(2-3x-2x^2) \\ =&2A+B+2C+(3A-B-3C)x+(A-2B-2C)x^2 \qquad\qquad = 1+2x\end{aligned}$$

as polynomials, so we have $2A + B + 2C = 1$, $3A - B - 3C = 2$, and $A - 2B - 2C = 0$. Solving this gives $C = -1/3$, $B = 3/5$, and $A = 8/15$. Thus (taking a factor of 2 out of the denominator of the second piece) the given generating function is equal to

$$\frac{8}{15}(1-2x)^{-1} + \frac{3}{10}(1+\frac{1}{2}x)^{-1} - \frac{1}{3}(1+x)^{-1}.$$

Using the Generalised Binomial Theorem, the coefficient of $x^r$ in the first of these summands is $\frac{8}{15}2^r$; the coefficient of $x^r$ in the second summand is $\frac{3}{10}(-1)^r\left(\frac{1}{2}\right)^r$; and the coefficient of $x^r$ in the third summand is $-\frac{1}{3}(-1)^r$. We conclude that the coefficient of $x^r$ in this generating function is

$$\frac{8}{15}2^r + \frac{3}{10}(-1)^r\left(\frac{1}{2}\right)^r - \frac{1}{3}(-1)^r.$$

**Solutions to Exercise 8.2.3:**

8.2.3.2 To use the method of partial fractions, we first factor the denominator:

$$2x^2 + x - 1 = (2x-1)(x+1).$$

Now, write

$$\begin{aligned} f(x) &= \frac{2+x}{2x^2+x-1} = \frac{2+x}{(2x-1)(x+1)} = \frac{A}{2x-1} + \frac{B}{x+1} \\ &= \frac{A(x+1)+B(2x-1)}{(2x-1)(x+1)} = \frac{(A-B)+(A+2B)x}{2x^2+x-1}. \end{aligned}$$

Equating the coefficients in the numerators yields the two equations

$$A - B = 2, A + 2B = 1.$$

Subtracting the second equation from the first tells us that $-3B = 1$, so $B = -1/3$. Then the first equation tells us that $A = 2 - (1/3) = 5/3$. So we have

$$f(x) = \frac{5/3}{2x-1} - \frac{1/3}{x+1} = -\frac{5/3}{1-2x} - \frac{1/3}{1+x}.$$

The coefficient of $x^r$ in $1/(1-x)$ is 1, so

- the coefficient of $x^r$ in $1/(1-2x)$ is $2^r$ (by replacing $x$ with $2x$), and
- the coefficient of $x^r$ in $1/(1+x)$ is $(-1)^r$ (by replacing $x$ with $-x$)

Therefore, the coefficient of $x^r$ in the generating function $f(x)$ is

$$-\frac{5}{3}(2^r) - \frac{1}{3}(-1)^r.$$

**Solutions to Exercise 8.3.3:**

8.3.3.1 Let $C(x) = \sum_{n=0}^{\infty} c_n x^n$ be the generating function of $\{c_n\}$. Then

$$\begin{aligned} C(x) &= c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + \cdots \\ xC(x) &= \quad c_0 x + c_1 x^2 + c_2 x^3 + c_3 x^4 + \cdots \\ x^2 C(x) &= \quad\quad c_0 x^2 + c_1 x^3 + c_2 x^4 + \cdots \end{aligned}$$

so

$$\begin{aligned} (1 - x - 2x^2)C(x) &= C(x) - xC(x) - 2x^2 C(x) \\ &= c_0 + (c_1 - c_0)x + \sum_{n=2}^{\infty} (c_n - c_{n-1} - 2c_{n-2})x^n \\ &= c_0 + (c_1 - c_0)x, \end{aligned}$$

since (by the recurrence relation) we have $c_n - c_{n-1} - 2c_{n-2} = 0$ for $n \geq 2$. Therefore

$$C(x) = \frac{c_0 + (c_1 - c_0)x}{1 - x - 2x^2}.$$

Since $c_0 = 2$ and $c_1 = 0$, this means

$$C(x) = \frac{2 + (0-2)x}{1 - x - 2x^2} = \frac{2 - 2x}{(1+x)(1-2x)}.$$

We now use partial fractions. Write

$$\frac{2-2x}{(1+x)(1-2x)} = \frac{A}{1+x} + \frac{B}{1-2x} = \frac{A(1-2x) + B(1+x)}{(1+x)(1-2x)} = \frac{(A+B) + (B-2A)x}{(1+x)(1-2x)}.$$

Equating the coefficients in the numerators yields the equations

$$2 = A + B, \qquad -2 = B - 2A.$$

Subtracting the second equation from the first tells us that

$$2 - (-2) = (A + B) - (B - 2A) = 3A,$$

so $A = 4/3$. Then the second equation tells us that

$$B = 2A - 2 = 2(4/3) - 2 = 2/3.$$

So

$$C(x) = \frac{2-2x}{(1+x)(1-2x)} = \frac{A}{1+x} + \frac{B}{1-2x} = \frac{4/3}{1+x} + \frac{2/3}{1-2x} = \frac{4}{3}\left(\frac{1}{1+x}\right) + \frac{2}{3}\left(\frac{1}{1-2x}\right).$$

From the generalized binomial theorem, we know:

- The coefficient of $x^n$ in $\dfrac{1}{1+x} = (1+x)^{-1}$ is
$$\binom{-1}{n} = (-1)^n \binom{1+n-1}{n} = (-1)^n \binom{n}{n} = (-1)^n.$$

- The coefficient of $x^n$ in $\dfrac{1}{1-2x} = (1-2x)^{-1}$ is
$$\binom{-1}{n}(-2)^n = (-1)^n\binom{1+n-1}{n}(-2)^n = (-1)^{2n}\binom{n}{n}2^n = 2^n.$$

Therefore $c_n$, the coefficient of $x^n$ in $C(x)$, is $\dfrac{4}{3}(-1)^n + \dfrac{2}{3}\cdot 2^n$.

8.3.3.3 Let $E(x) = \displaystyle\sum_{n=0}^{\infty} e_n x^n$ be the generating function of $\{e_n\}$. Then

$$\begin{aligned} E(x) &= e_0 + e_1x + e_2x^2 + e_3x^3 + e_4x^4 + \cdots \\ xE(x) &= \quad e_0x + e_1x^2 + e_2x^3 + e_3x^4 + \cdots \\ \frac{1}{1-x} &= 1 \quad + \quad x + \quad x^2 + \quad x^3 + \quad x^4 + \cdots \end{aligned}$$

so

$$\begin{aligned} (1-3x)E(x) + \frac{2}{1-x} &= E(x) - 3xE(x) + 2\cdot\frac{1}{1-x} \\ &= (e_0+2) + \sum_{n=1}^{\infty}(e_n - 3e_{n-1} + 2)x^n \\ &= (e_0+2), \end{aligned}$$

since (by the recurrence relation) we have $e_n - 3e_{n-1} + 2 = 0$ for $n \geq 1$. Therefore

$$E(x) = \frac{(e_0+2) - \dfrac{2}{1-x}}{1-3x} = \frac{(e_0+2)(1-x)-2}{(1-3x)(1-x)}.$$

Since $e_0 = 2$, this means

$$E(x) = \frac{(2+2)(1-x)-2}{(1-3x)(1-x)} = \frac{2-4x}{(1-3x)(1-x)}.$$

We now use partial fractions. Write

$$\frac{2-4x}{(1-3x)(1-x)} = \frac{A}{1-3x} + \frac{B}{1-x} = \frac{A(1-x)+B(1-3x)}{(1-3x)(1-x)} = \frac{(A+B)-(A+3B)x}{(1-3x)(1-x)}.$$

Equating the coefficients in the numerators yields the equations

$$2 = A + B, \qquad -4 = -(A+3B).$$

Adding the two equations tells us that $2-4 = (A+B)-(A+3B) = -2B$, so $B = 1$. Then the first equation tells us that $A = 2 - B = 2 - 1 = 1$. So

$$E(x) = \frac{2-4x}{(1-3x)(1-x)} = \frac{A}{1-3x} + \frac{B}{1-x} = \frac{1}{1-3x} + \frac{1}{1-x}.$$

From the generalized binomial theorem, we know that the coefficient of $x^n$ in $\dfrac{1}{1-3x}$ is $3^n$, and the coefficient of $x^n$ in $\dfrac{1}{1-x}$ is 1. Therefore $e_n$, the coefficient of $x^n$ in $E(x)$, is $3^n + 1$.

# Solutions for Chapter 9

**Solutions to Exercise 9.1.3:**

9.1.3.2 To prove Proposition 9.1.2 requires strong induction, since the recursive relation calls on two previous terms. Thus, two base cases are required.

9.1.3.3 The formula from Proposition 9.1.2 gives

$$\begin{aligned} D_5 &= 5!\left(\frac{(-1)^0}{0!}+\frac{(-1)^1}{1!}+\frac{(-1)^2}{2!}+\frac{(-1)^3}{3!}+\frac{(-1)^4}{4!}+\frac{(-1)^5}{5!}\right) \\ &= 120\left(1-1+\frac{1}{2}-\frac{1}{6}+\frac{1}{24}-\frac{1}{120}\right) \\ &= 60-20+5-1=44. \end{aligned}$$

9.1.3.4 The initial conditions are $D_1 = 0$ and $D_2 = 1$. The recursive relation $D_n = (n-1)(D_{n-1} + D_{n-2})$ for $n \geq 3$ gives $D_3 = 2(D_2 + D_1) = 2(1+0) = 2$. Now

$$D_4 = 3(D_3 + D_2) = 3(2+1) = 9,$$

and

$$D_5 = 4(D_4 + D_3) = 4(9+2) = 4(11) = 44.$$

**Solutions to Exercise 9.2.3:**

9.2.3.1 **PROOF.** *Base case:* $n = 0$. We have $c_0 = 1$ (by definition) and $1 > 0$, so $c_0 > 0$.

*Inductive step:* We begin with the inductive hypothesis. We will require strong induction. Let $k \geq 0$ be arbitrary, and suppose that the inequality holds for every $j$ with $0 \leq j \leq k$; that is, assume that for every such $j$, $c_j > 0$.

Now we want to deduce that $c_{k+1} > 0$. Using the recursive relation, we have

$$c_{k+1} = \sum_{i=0}^{k} c_i c_{(k+1)-i-1} = \sum_{i=0}^{k} c_i c_{k-i}.$$

Using the inductive hypothesis, we have $c_j > 0$ for every $j$ such that $0 \leq j \leq k$. Putting these together gives that $c_{k+1}$ is a sum of $k+1$ terms where each term has the form $c_i c_{k-i}$ with $0 \leq i \leq k$. Since $0 \leq k-i \leq k$, we see that $c_i > 0$ and $c_{k-i} > 0$ so that $c_i c_{k-i} > 0$. Hence

$$c_{k+1} = \sum_{i=0}^{k} c_i c_{k-i} > 0.$$

This completes the proof of the inductive step.

By the Principle of Mathematical Induction, $c_n > 0$ for every $n \geq 0$. ■

**Solutions to Exercise 9.3.6:**

9.3.6.1

$$\begin{aligned} B_4 = \sum_{k=1}^{4} \binom{3}{k-1} B_{4-k} &= \binom{3}{0}B_3 + \binom{3}{1}B_2 + \binom{3}{2}B_1 + \binom{3}{3}B_0 \\ &= 5 + 3(2) + 3(1) + 1(1) = 15. \end{aligned}$$

9.3.6.3 If $b_i = (i+1)!/2$ then the expanded exponential generating function for this sequence is

$$\sum_{i=0}^{\infty} \frac{b_i x^i}{i!} = \sum_{i=0}^{\infty} \frac{(i+1)! x^i}{2i!} = \sum_{i=0}^{\infty} (i+1)x^i/2.$$

This is

$$\frac{1}{2}\sum_{i=0}^{\infty}(i+1)x^i = \frac{1}{2}\left(\sum_{i=0}^{\infty}(i+1)x^i\right) = \frac{1}{2(1-x)^2}.$$

# Solutions for Chapter 10

**Solutions to Exercise 10.1.12:**

10.1.12.1 Since there are 8 rows on a chessboard, and $17 > 2(8)$, the Generalised Pigeonhole Principle says that there must be at least $2 + 1 = 3$ rooks that are all in the same row of the board. Choose such a row, and call it Row A. Note that Row A contains at most 8 rooks.

There are at least $17 - 8 = 9$ rooks that are not in Row A. Since there are 7 other rows on the chessboard, and $9 > 1(7)$, the Pigeonhole Principle says that there must be at least $1 + 1 = 2$ rooks that are in the same row, from amongst the other rows of the board. Choose such a row, and call it Row B. Note that Row B also contains at most 8 rooks.

There is at least $17 - 8 - 8 = 1$ rook remaining, so there must be a rook somewhere on the board that is in neither Row A nor Row B. Choose such a rook, Rook 1, and call the row that it is in Row C. Since there are at least 2 rooks in Row B, at least one of them must not be in the same column as Rook 1. Choose such a rook, Rook 2. Since there are at least 3 rooks in Row A, at least one of them must not be in the same column as either Rook 1 or Rook 2. Choose such a rook, and call it Rook 3. Now Rooks 1, 2, and 3 do not threaten each other, so fulfil the requirements of the problem.

10.1.12.3 We use the even more generalised pigeonhole principle with $n - 1 = 15$ for the adults, and $n_2 = 23$ for the children (and $m = 2$ categories: adults and children). The principle tells us that as long as at least

$$n_1 + n_2 - m + 1 = 15 + 23 - 2 + 1 = 37$$

people are approached, the artist will have enough people to carry their art in the parade.

**Solutions to Exercise 10.2.8:**

10.2.8.2 Of the basic pieces of information that we need to complete a Venn diagram, one has not been given to us: the number of Kevin's apps that are free and require internet access. Fortunately, we can use inclusion-exclusion to work this out from the others. We use $F$ to represent the set of free apps; $G$ to represent the games, and $I$ to represent the apps that require internet. Then we have been told:

$$|F| = 78, |I| = 124, |G| = 101, |F \cap G| = 58, |G \cap I| = 62, |F \cap G \cap I| = 48, |F \cup G \cup I| = 165.$$

Using inclusion-exclusion, we have

$$165 = 78 + 124 + 101 - 58 - 62 - |F \cap I| + 48,$$

so

$$|F \cap I| = 78 + 124 + 101 - 58 - 62 + 48 - 165 = 66.$$

The value we have been asked for is

$$|F \cap I \cap \overline{G}| = |F \cap I| - |F \cap I \cap G| = 66 - 48 = 18.$$

10.2.8.3 The number of integers between 1 and 60 that are divisible by 2 is $60/2 = 30$. Call the set of these integers $A$. The number of integers between 1 and 60 that are divisible by 3 is $60/3 = 20$. Call the set of these integers $B$. The number of integers between 1 and 60 that is divisible by 5 is $60/5 = 12$. Call the set of these integers $C$. Then $|A \cap B|$ is the number of integers between 1 and 60 that are divisible by 2 and 3; that is, the number that are divisible by 6. This is $60/6 = 10$. Similarly, $|A \cap C|$ is the number of integers between 1 and 60 that are divisible by 2 and 5; that is, the number that are divisible by 10. This is $60/10 = 6$. Also, $|B \cap C|$ is the number of integers between 1 and 60 that are divisible by 3 and 5; that is, the number that are divisible by 15. This is $60/15 = 4$. Finally, $|A \cap B \cap C|$ is the number of integers between 1 and 60 that are divisible by 2, 3, and 5; that is, the number that are divisible by 30. This is $60/30 = 2$.

We have been asked for $|A \cup B \cup C|$. Using inclusion-exclusion, we see that the answer is $30 + 20 + 12 - 10 - 6 - 4 + 2 = 44$.

# Solutions for Chapter 11

**Solutions to Exercise 11.2.11:**

11.2.11.1

- The only edge incident with $a$ is $e_1$, so the valency of $a$ is 1.
- The only edge incident with $b$ is $e_2$, so the valency of $b$ is 1.
- The edges incident with $c$ are $e_1$, $e_3$, and $e_4$, so the valency of $c$ is 3.
- The edges incident with $e$ are $e_4$, $e_5$, and $e_6$, and $e$ appears twice as an endvertex of $e_6$, so altogether $e$ appears 4 times as the endvertex of some edge. Thus, the valency of $e$ is 4.
- The edges incident with $e$ are $e_4$, $e_5$, and $e_6$, so the valency of $e$ is 3.

Since $e_6 = \{e, e\}$ is a loop, the graph is *not* simple. There is no isolated vertex, because no vertex has valency 0. The only neighbour of $a$ is $c$, and the only edge incident with $a$ is $e_1$.

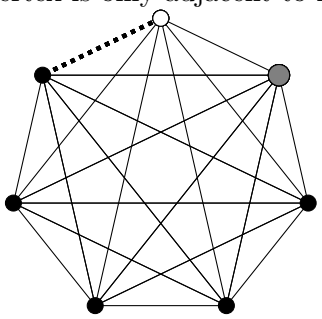
11.2.11.3

- The edges incident with $a$ are $e_1$ and $e_2$, so the valency of $a$ is 2.
- The edges incident with $b$ are $e_1$ and $e_3$, so the valency of $b$ is 2.
- The edges incident with $c$ are $e_2$ and $e_3$, so the valency of $c$ is 2.
- No edges are incident with $d$, so the valency of $d$ is 0.

There are no loops or multiple edges, so the graph is simple. The graph does have an isolated vertex, namely, $d$ (because the valency of $d$ is 0). The neighbours of $a$ are $b$ and $c$, and the edges incident with $a$ are $e_1$ and $e_2$, as was already mentioned above.

**Solutions to Exercise 11.3.13:**

11.3.13.3 In the following picture, the dotted line represents the edge we will delete. If we then delete the white vertex, the graph that remains is complete. If instead we then delete the large grey vertex (which is the next one clockwise from the white vertex), the remaining graph will not be complete, since the white vertex is only adjacent to four of the five black vertices.



11.3.13.4 **PROOF.** Let $G$ be a graph with $e$ edges.

*Base case:* $e = 0$. Since $G$ has no edges, every vertex has valency 0. So the number of vertices of odd valency is 0, which is even.

*Inductive step:* We begin with the inductive hypothesis. Fix $e \geq 0$, and assume that every graph with $e$ edges has an even number of vertices of odd valency.

Now we want to deduce that every graph with $e + 1$ edges has an even number of vertices of odd valency. Let $H$ be an arbitrary graph with $e + 1$ edges. Choose one edge $f$ (there is one since $e + 1 \geq 1$), and call its endvertices $u$ and $v$. Let $H' = H \setminus \{f\}$. Notice that $H'$ has $e + 1 - 1 = e$ edges, so our induction hypothesis applies to $H'$. Therefore, $H'$ has an even number of vertices of odd valency. Call this number $2m$, where $m \in \mathbb{Z}$.

Observe that the valency of every vertex of $H$ is the same as its valency in $H'$ if the vertex is not $u$ or $v$, and is one greater than its valency in $H'$ if the vertex is either $u$ or $v$. Consider the three possible cases: $u$ and $v$ both have even valency in $H'$; $u$ and $v$ both have odd valency in $H'$; or exactly one of $u$ and $v$ has even valency in $H'$.

If $u$ and $v$ both have even valency in $H'$, then they both have odd valency in $H$, so the number of vertices of odd valency in $H$ must be $2m + 2$, which is even.

If $u$ and $v$ both have odd valency in $H'$, then they both have even valency in $H$, so the number of vertices of odd valency in $H$ must be $2m - 2$, which is even.

If exactly one of $u$ and $v$ has even valency in $H'$, then exactly one of $u$ and $v$ will have even valency in $H$ (the other one, since the valency of each of $u$ and $v$ goes up by 1). So the number of vertices of odd valency in $H$ must be $2m$ (even though one of the specific vertices of odd valency has changed between $u$ and $v$), which is even.

In all cases, $H$ has an even number of vertices of odd valency. This completes the proof of the inductive step.

By the Principle of Mathematical Induction, every graph with at least 0 edges has an even number of vertices of odd valency. ■

**Solutions to Exercise 11.4.7:**

11.4.7.1 The graphs are not isomorphic, because the only vertex of valency 1 in $G$ (namely, $c$) is adjacent to a vertex of valency 3 (namely $d$), but the only vertex of valency 1 in $H$ (namely, $z$) is adjacent only to a vertex of valency 2 (namely, $y$).

Here is a more complete proof. Suppose $\varphi\colon G \to H$ is an isomorphism. (This will lead to a contradiction.) We must have

$$d_H\big(\varphi(c)\big) = d_G(c) = 1.$$

(This principle was pointed out in the proof of Proposition 11.4.4.3.) The only vertex of valency 1 in $H$ is $z$, so this implies that $\varphi(c) = z$.

Now, since $d \sim c$, we must have $\varphi(d) \sim \varphi(c)$. Since $\varphi(c) = z$, and the only neighbour of $z$ is $y$, this implies $\varphi(d) = y$. So

$$d_H(y) = d_H\big(\varphi(d)\big) = d_G(d).$$

However, $d_H(y) = 2$ and $d_G(d) = 3$, so $d_H(y) \neq d_G(d)$. This is a contradiction.

11.4.7.3 There is no vertex of valency 0 in $G_1$, but $A$ is a vertex of valency 0 in $G_2$. Therefore $G_1$ and $G_2$ do not have the same degree sequence, so they are not isomorphic.

**Solutions to Exercise 11.4.8:**

11.4.8.2 Of the five vertex labels, we can choose any two to join with an edge. Thus, the number of labeled graphs on five vertices with one edge is $\binom{5}{2} = 10$.

11.4.8.3 Notice that there are $\binom{5}{2} = 10$ total edges possible in a graph on 5 vertices. Thus, the number of labeled graphs on 5 vertices with 3 edges is the number of ways of choosing 3 of these 10 labeled edges. So there are $\binom{10}{3} = 120$ labeled graphs on 5 vertices that have 3 edges.

Similarly, there are $\binom{10}{4} = 210$ labeled graphs on 5 vertices that have 4 edges. Thus, in total there are $120 + 210 = 330$ labeled graphs on 5 vertices that have 3 or 4 edges.

# Solutions for Chapter 12

**Solutions to Exercise 12.1.6:**

12.1.6.1 **PROOF.** We prove, by induction on $n$, that if $n \geq 1$, and $G$ is any digraph with $n$ vertices that has no loops or multiarcs, then

$$|A(G)| = \sum_{v \in V(G)} d_G^+(v) = \sum_{v \in V(G)} d_G^-(v).$$

*Base case:* $n = 1$. Let $G$ be a digraph with no loops or multiarcs, and with only one vertex $v_1$. Then there are no arcs in $G$, so $|A(G)| = 0 = d_G^+(v_1) = d_G^-(v_1)$. So the desired conclusion is true when $n = 1$.

*Inductive step:* Assume that $n \geq 1$, the formula holds for every digraph with $n$ vertices that has no loops or multiarcs, and $G$ is a digraph with $n+1$ vertices that has no loops or multiarcs.

Pick an arbitrary vertex $u$ of $G$. Let

- $N^+$ be the set of outneighbours of $u$, and $N^-$ the set of inneighbours of $u$,
- $s = |N^+| = d_G^+(u)$ be the number of arcs that begin at $u$,
- $t = |N^-| = d_G^-(u)$ be the number of arcs that end at $u$, and
- $G'$ be the digraph obtained from $G$ by deleting $u$ and its $s + t$ incident arcs.

Note that:

- $V(G') = V(G) \setminus \{u\}$, so $G'$ has $n$ vertices.
- $|A(G')| = |A(G)| - s - t$.

- For $v \in V(G') \setminus N^-$, we have $d^+_{G'}(v) = d^+_G(v)$ (because the outneighbours of $v$ in $G'$ are exactly the same as the outneighbours of $v$ in $G$).
- For $v \in N^-$, we have $d^+_{G'}(v) = d^+_G(v) - 1$ (because $u$ is counted as an outneighbour of $v$ in $G$, but it is not in $G'$ so it cannot be counted as an outneighbour in $G'$).
- Similar statements hold with $N^-$ replaced by $N^+$.

Hence

$$\begin{aligned}
\sum_{v \in V(G)} d^+_G(v) &= \sum_{v \in V(G) \setminus (N^- \cup \{u\})} d^+_G(v) + \sum_{v \in N^-} d^+_G(v) + \sum_{v \in \{u\}} d^+_G(v) \\
&= \sum_{v \in V(G) \setminus (N^- \cup \{u\})} d^+_{G'}(v) + \sum_{v \in N^-} \left(d^+_{G'}(v) + 1\right) + d^+_G(u) \\
&= \left( \sum_{v \in V(G) \setminus N^- \cup \{u\}} d^+_{G'}(v) + \sum_{v \in N^-} d^+_{G'}(v) \right) + |N^-| + d^+_G(u) \\
&= \sum_{v \in V(G')} d^+_{G'}(v) + t + s \\
&= |A(G')| + s + t \qquad \text{(induction hypothesis)} \\
&= |A(G)|.
\end{aligned}$$

Similarly, we can argue that $\sum_{v \in V(G)} d^- G(v) = |A(G)|$. This completes the inductive step and the proof. ■

12.1.6.3 Beginning at the top and working clockwise, label the vertices of the digraph $a$, $b$, $c$, $d$, and $e$. Then:

- $a$ has outvalency 2 and invalency 1;
- $b$ has outvalency 2 and invalency 2;
- $c$ has outvalency 1 and invalency 2;
- $d$ has outvalency 2 and invalency 2;
- $e$ has outvalency 1 and invalency 1.

**Solutions to Exercise 12.2.6:**

12.2.6.2 This graph is connected. To see this, note that $(a, j, g, v, e, d, i, h, c, b)$ is a walk that passes through all of the vertices of $G$, so it is possible to walk from $a$ to any other vertex. Therefore, the connected component that contains $a$ is $V(G) = \{a, b, c, d, e, f, g, h, i, j\}$.

There are several walks of length 5 from $a$ to $f$. One example is $(a, g, a, j, g, f)$.

12.2.6.3 This graph is *not* connected. To see this, note that there are no edges from any vertex in $\{a, d, e, f, g, j\}$ to any vertex in $\{b, c, h, i\}$. Indeed the connected component that contains $a$ is $\{a, d, e, f, g, j\}$. (The walk $(a, d, e, f, g, j)$ passes through all of these vertices, but none of these vertices are adjacent to any vertex that is not in the subset.)

There are several walks of length 3 from $a$ to $d$. One example is $(a, d, a, d)$.

**Solutions to Exercise 12.3.10:**

12.3.10.1

a. There are many paths of length 3. One example is $(a, b, c, h)$.

b. $(b, c, f, b)$ is a cycle of length 3.

c. $(a, b, c, b)$ is neither a path nor a cycle. It is not a path because the vertices are not all distinct. (Namely, the vertex $b$ occurs twice.) It is not a cycle, because the first vertex (namely, $a$) is not the same as the final vertex (namely, $b$).

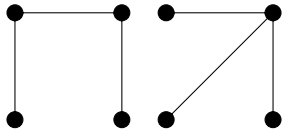
12.3.10.3 **PROOF.** Let $(u = u_1, u_2, \ldots, v = u_k, u)$ be a cycle of $G$ in which $u$ and $v$ appear as consecutive vertices. Let $G' = G \setminus \{uv\}$.

Let $x$ and $y$ be arbitrary vertices of $G$. Since $G$ is connected, there is a walk $(x = x_1, x_2, \ldots, x_m = y)$ from $x$ to $y$ in $G$. If this walk does not contain the edge $uv$ then it is also a walk in $G'$. If it does contain the edge $uv$, then we can find some $i$ with $1 \leq i \leq m-1$ such that either $x_i = u$ and $x_{i+1} = v$, or vice versa. For every such $i$, replace the pair $(x_i, x_{i+1})$ in the walk by either $(u = u_1, u_2, \ldots, v = u_k)$ or $(v = u_k, u_{k-1}, \ldots, u = u_1)$ (as appropriate, depending on whether $x_i = u$ or $x_i = v$). The result is a walk from $x$ to $y$ that does not use the edge $uv$, so is in $G'$. Since $x$ and $y$ were arbitrary vertices of $G'$, for any two vertices $x$ and $y$ of $G'$ there is an $x - y$ walk, so by definition, $G'$ is connected. ■

**Solutions to Exercise 12.4.6:**

12.4.6.1 **PROOF.** Let $T$ be a tree, and let $v$ be a leaf of $T$. Consider $T \setminus \{v\}$. Certainly it cannot have any cycles, since $T$ has no cycles. Let $x$ and $y$ be arbitrary vertices of $T \setminus \{v\}$. Since $T$ is connected, there is an $x - y$ walk in $T$, so by Proposition 12.3.4, there is an $x - y$ path in $T$. Since $v$ is a leaf of $T$, if an $x - y$ walk uses the vertex $v$ then the neighbour of $v$ would have to come both before and after $v$ in the walk, since $v$ has only one neighbour, so such a walk would not be a path. Thus, the $x - y$ path cannot use the vertex $v$, so it is still a path in $T \setminus \{v\}$. Since $x$ and $y$ were arbitrary vertices, this shows that $T \setminus \{v\}$ is connected. This completes the proof. ■

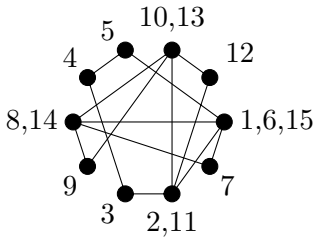
12.4.6.3 The two nonisomorphic unlabeled trees on 4 vertices are:



# Solutions for Chapter 13

**Solutions to Exercise 13.1.6:**

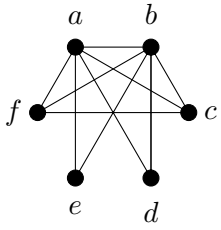
13.1.6.1.a This graph has Euler tours, because it is connected and all vertices have even valency. One Euler tour is $(d, f, g, j, a, d, e, i, h, b, f, c, b, i, d)$. The following figure numbers the vertices $1, 2, 3, \ldots$ in the order they are visited.



**Solutions to Exercise 13.2.12:**

13.2.12.2

a. In the closure, we can join $a$ to $b$; we can join $a$ to $c$; and we can join $b$ to $f$. This completes the closure, shown below. It is not easy to see from this whether or not the graph has a Hamilton cycle. In fact, it does not.




b. The closure of this graph is $K_6$. We can easily see from this that the graph does have a Hamilton cycle. (To see that the closure is $K_6$, observe that every vertex of the graph has valency at least 2. Thus, the two vertices of valency 4 can be joined to each of their non-neighbours. After doing so, every vertex has valency at least 3, so every vertex can be joined to every other vertex.)

13.2.12.3 Let $G$ be the graph that has been shown here. Using the notation of Theorem 13.2.2, let $S = \{a, f\}$. Then $|S| = 2$, but $G \setminus S$ has 3 connected components: $\{b, e\}$, $\{c, h\}$, and $\{d, g\}$. Since $3 > 2$, $G$ cannot have a Hamilton cycle.

# Solutions for Chapter 14

**Solutions to Exercise 14.1.18:**

14.1.18.1 **PROOF.** Trees are bipartite (they have no cycles at all, so certainly do not have any cycles of odd length), so Theorem 14.1.17 tells us that they are class one. ■

14.1.18.2 **PROOF.** The proof is by contradiction: suppose $n$ is odd, and the cycle

$$C_n = (v_0, v_1, \ldots, v_n = v_0)$$

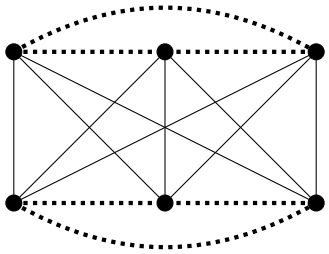
is class one. Since every vertex of $C_n$ has valency two, this means that the graph has a proper edge-colouring that uses only 2 colours. Let us call the colours red and blue.

Assume, without loss of generality, that the edge $v_0v_1$ is red. The edge $v_1v_2$ cannot be the same colour as $v_0v_1$ (because they are both incident to $v_1$), so $v_1v_2$ must be blue. The edge $v_2v_3$ cannot be the same colour as $v_1v_2$ (because they are both incident to $v_2$), so $v_2v_3$ must be red. Continuing in this way, we see (by induction on $k$) that $v_kv_{k+1}$ is red whenever $k$ is even, and it is blue whenever $k$ is odd. (That is, the two colours must alternate red, blue, red, blue, red, blue,… as we go around the cycle.)

In particular, since $n$ is odd, we know that $n - 1$ is even, so this means that the edge $v_{n-1}v_n$ is red. However, we have $v_n = v_0$ so the edges $v_{n-1}v_n$ and $v_0v_1$ are both incident to the vertex $v_0$), so they cannot be the same colour. The contradicts the fact that both edges are red. ■

**Solutions to Exercise 14.2.5:**

14.2.5.1 The following 2-colouring of the edges of $K_6$ has no solid triangle and no dotted $K_4$ (because the solid edges form $K_{3,3}$, which has no cycles of odd length, and each connected component of the dotted graph is only $K_3$):



14.2.5.3 **PROOF.** We prove this by induction on $k+\ell$.

*Base case:* $k+\ell = 2$ (so $k = \ell = 1$). Then

$$R(k,\ell) = R(1,1) = 1 < 4 = 2^{1+1} = 2^{k+\ell}.$$

So the inequality is valid in the base case.

*Inductive step:* Assume $k+\ell \geq 2$, and that $R(k',\ell') \leq 2^{k'+\ell'}$, whenever $k'+\ell' < k+\ell$. Since $R(k,\ell) = R(\ell,k)$, we may assume $k \leq \ell$ (by interchanging $k$ and $\ell$, if necessary). If $k = 1$, then

$$R(k,\ell) = R(1,\ell) = 1 = 2^0 < 2^{k+\ell}.$$

Therefore, we may assume $2 \leq k \leq \ell$. Since $(k-1)+\ell < k+\ell$ and $k+(\ell-1) < k+\ell$, the induction hypothesis tells us that

$$R(k-1,\ell) \leq 2^{(k-1)+\ell} \text{ and } R(k,\ell-1) \leq 2^{k+(\ell-1)}.$$

Therefore, we see from Proposition 14.2.4 that

$$\begin{aligned} R(k,\ell) &\leq R(k-1,\ell) + R(k,\ell-1) \\ &\leq 2^{(k-1)+\ell} + 2^{k+(\ell-1)} && \text{(induction hypothesis)} \\ &= 2^{k+\ell-1} + 2^{k+\ell-1} \\ &= 2 \cdot 2^{k+\ell-1} \\ &= 2^{k+\ell}. \end{aligned}$$

This completes the proof. ■

14.2.5.4 Since $R(k,\ell) \leq R(k',\ell')$ whenever $k \leq k'$ and $\ell \leq \ell'$, we have

$$40 \leq R(3,10) \leq R(3,11).$$

Also, since $R(k,\ell) \leq R(k-1,\ell) + R(k,\ell-1)$ and $R(2,\ell) = \ell$, we have

$$R(3,11) \leq R(3-1,11) + R(3,11-1) = R(2,11) + R(3,10) \leq 11 + 42 = 53.$$

So $40 \leq R(3,11) \leq 53$.

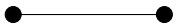
**Solutions to Exercise 14.2.7:**

14.2.7.2 **PROOF.** We assume that $N-1 > (c+1)(n-1)$. Consider an arbitrary colouring of the edges of $K_N$ with $c+1$ colours. Fix a vertex $v$. Since $v$ has $N-1 > (n-1)(c+1)$ incident edges, the generalised pigeonhole principle tells us that there must be some set of at least $n$ edges incident with $v$ that are all coloured with the same colour, say colour $i$. Look at the induced subgraph of $K_N$ on the $n$ other endpoints of these edges. If any edge $xy$ of this induced subgraph is coloured with colour $i$, then all of the edges of the triangle $\{v,x,y\}$ have been coloured with colour $i$, so $K_N$ has a monochromatic triangle.

If on the other hand no edge of the induced subgraph has been coloured with colour $i$, then the induced subgraph is a $K_n$ whose edges have been coloured with the remaining $c$ colours. By hypothesis, every such colouring has a monochromatic triangle. This completes the proof. ■

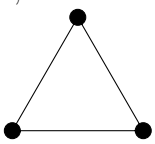
**Solutions to Exercise 14.2.9:**

14.2.9.1 We are looking for the smallest value of $n$ such that every edge-colouring of $K_n$ with dotted, dashed, and solid lines has either a dashed $K_2$ or a dotted $K_2$ or a solid triangle. The following colouring shows that $R(2,2,3) > 2$:



However, $R(2,2,3) = 3$. This is because if any edges are dotted or dashed, then there is a dotted or dashed $K_2$; if no edges are dotted or dashed, then every edge is solid, so there is a solid $K_3$.

14.2.9.2 We will show that $R(2,4) = 4$. We are looking for the smallest value of $n$ such that every edge-colouring of $K_n$ with dashed or solid lines has either a dashed $K_2$ or a solid $K_4$. The following colouring shows that $R(2,4) > 3$;



However, in $K_4$ if any edge is dashed, then there is a dashed $K_2$, while if no edges are dashed, then there is a solid $K_4$.

**Solutions to Exercise 14.3.14:**

14.3.14.2 **PROOF.** We will proceed by induction on $n$.

*Base case:* $n = 1$. Then $K_1$ is the graph with one vertex and no edges, and $\chi(K_1) = 1$. Thus, when $n = 1$ we have $\chi(K_n) = n$.

*Induction step:* We begin with the induction hypothesis. Let $k \geq 1$ be arbitrary, and assume that $\chi(K_k) = k$, so we can properly colour $K_k$ using $k$ colours, and $k$ colours are required to do so.

Now consider the graph $K_{k+1}$. Let $v$ be an arbitrary vertex of this graph. By our induction hypothesis, $\chi(K_{k+1} \setminus \{v\}) = k$. Thus, any proper colouring of $K_{k+1}$ must use at least $k$ colours on the vertices other than $v$. It is not possible to colour $v$ with any of these $k$ colours (since $v$ is adjacent to all of the other vertices, so has a neighbour that is coloured with each of these $k$ colours). Therefore, $\chi(K_{k+1}) \geq k+1$. In fact, since $v$ is the only vertex not yet coloured by these $k$ colours, it is clear that $k+1$ colours suffice to colour the graph: we colour $v$ with a new colour, which is the $(k+1)$st colour. This will certainly be a proper colouring of $K_{k+1}$. Thus, $\chi(K_{k+1}) = k+1$, completing the induction step.

By the Principle of Mathematical Induction, $\chi(K_n) = n$ for every $n \geq 1$. ■

14.3.14.4 The fact that $G$ contains a subgraph isomorphic to $K_i$ implies that $\chi(G) \geq i$. The fact that $\Delta(G) \leq j$ implies that

$$\chi(G) \leq \Delta(G) + 1 \leq j + 1.$$

So $4 \leq i \leq \chi(G) \leq j + 1 \leq 7$. If we also know that $G$ is connected and is neither a complete graph nor a cycle of odd length, then $\chi(G) \leq \Delta(G) \leq j$, so $4 \leq i \leq \chi(G) \leq j \leq 6$ in this case.

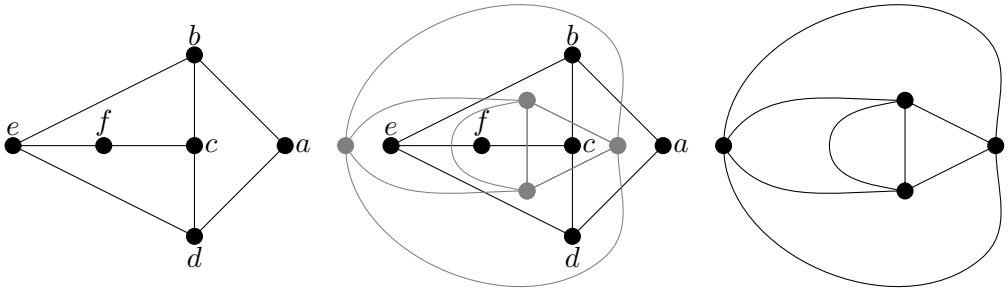
# Solutions for Chapter 15

**Solutions to Exercise 15.1.16:**

15.1.16.1 **PROOF.** Let $G$ be a graph with a nonplanar subgraph $H$. Suppose (towards a contradiction) that $G$ is planar. Find a planar embedding of $G$, and delete vertices and/or edges (as appropriate) to reach the subgraph $H$. No edges that do not share an endvertex have points in common in the embedding of $G$, and edges that do share an endvertex have no other points in common. This property is not changed by deleting vertices and/or edges, so our result is a planar embedding of $H$. But this is impossible, since $H$ is nonplanar. The contradiction shows that $G$ must be planar.

Since $K_5$ is a subgraph of $K_n$ for every $n \geq 6$ and $K_5$ is nonplanar by Theorem 15.1.3, this shows that $K_n$ is nonplanar for every $n \geq 6$. ■

15.1.16.3 We show a planar embedding of the graph, the planar embedding with the dual graph shown in grey, and the dual graph.




**Solutions to Exercise 15.2.10:**

15.2.10.1 We will prove that for any planar embedding of a disconnected planar graph with exactly two connected components, $|V| - |E| + |F| = 3$.

**PROOF.** We will prove this formula by induction on the number of faces of the embedding. Let $G$ be a planar embedding of a graph with exactly two connected components.

*Base case:* If $|F| = 1$ then $G$ cannot have any cycles (otherwise the interior and exterior of the cycle would be 2 distinct faces). So $G$ must consist of two connected graphs that have no cycles, i.e., two trees, $T_1$ and $T_2$. By Theorem 12.4.5 we know that we must have $|E(T_1)| = |V(T_1)| - 1$ and $E(T_2) = V(T_2) - 1$, so

$$|V| - |E| + |F| = |V(T_1)| + |V(T_2)| - (|V(T_1)| - 1) - (|V(T_2) - 1) + 1 = 3.$$

*Inductive step:* We begin by stating our inductive hypothesis. Let $k \geq 1$ be arbitrary, and assume that for any planar embedding of a graph that has exactly two connected components, such that the embedding has $k$ faces, $|V| - |E| + |F| = 3$.

Let $G$ be a planar embedding of a graph that has exactly two connected components, such that the component has $k+1 \geq 2$ faces. Since forests have only one face, $G$ must have a cycle in at least one of its components. Choose any edge $e$ that is in a cycle of $G$, and let $H = G \setminus \{e\}$. Clearly, we have

$$|E(H)| = |E(G)| - 1$$

and $|V(H)| = |V(G)|$. Also,

$$|F(H)| = |F(G)| - 1 = k$$

since the edge $e$ being part of a cycle must separate two faces of $G$, which are united into one face of $H$. Furthermore, since $e$ was in a cycle and $G$ has two connected components, by an argument similar to that given in Proposition 12.3.9 $H$ has two connected components, and $H$ has a planar embedding induced by the planar embedding of $G$. Therefore our inductive hypothesis applies to $H$, so

$$\begin{aligned} 3 &= |V(H)| - |E(H)| + |F(H)| \\ &= |V(G)| - (|E(G) - 1) + (|F(G)| - 1) \\ &= |V(G)| - |E(G)| + |F(G)| \end{aligned}$$

This completes the inductive step.

By the Principle of Mathematical Induction, $|V| - |E| + |F| = 3$ for any planar embedding of graph that has exactly two connected components. ■

15.2.10.4 The value for $|V| - |E| + |F|$ on a torus is 0. For example, consider the graph on 5 vertices consisting of two cycles of length 3 that meet at a vertex. Draw this graph on a torus so that one cycle goes through the hole in the middle, and one cycle goes around the outside edge of the torus. This embedding has one face, since the first cycle cuts the torus into something resembling a cylinder, and the second cuts the cylinder into a rectangle. There are 5 vertices and 6 edges, so $|V| - |E| + |F| = 5 - 6 + 1 = 0$, as claimed.

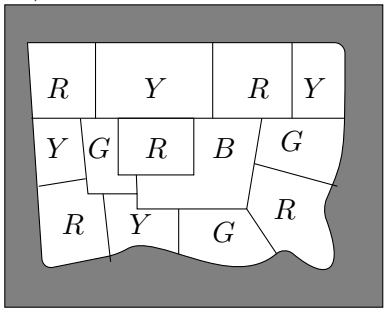
**Solutions to Exercise 15.3.8:**

15.3.8.1 First, notice that since $G$ is cubic, every vertex has valency 3, which is odd. Therefore, by Corollary 11.3.12, $G$ must have an even number of vertices.

This means that the number of vertices in the Hamilton cycle, and the number of edges in the Hamilton cycle (which are equal) are both even. Thus, we can colour the edges of the Hamilton cycle with 2 colours, say blue and red, alternating between the two colours all the way around the cycle.

Since the graph is cubic, each vertex is now incident to exactly one edge that has not yet been coloured. Therefore, we can colour all of the remaining edges with a single colour — green, say. Thus, we have properly 3-edge-coloured $G$. Since $\Delta(G) = 3$, this means that $G$ is a class one graph.

15.3.8.2 We use the letters $R$, $G$, $B$, and $Y$ to represent the four colours. The exterior face (which appears grey in the picture) will be assigned the colour $B$.



# Solutions for Chapter 16

**Solutions to Exercise 16.1.3:**

16.1.3.1 **PROOF.** Let $L$ be an $n \times n$ Latin square whose entries come from a set $N$ of cardinality $n$, and let $L'$ be the result of exchanging row $i$ with row $j$.

Let $k \in \{1, \ldots, n\}$ be arbitrary, and consider column $k$ of $L'$. Its entries are exactly the same as the entries of column $k$ of $L$, except that the $i$th entry has been exchanged with the $j$th entry. Since every element of $N$ appears exactly once in column $k$ of $L$, it also appears exactly once in column $k$ of $L'$ (although possibly in a different position). Since $k$ was arbitrary, every element of $N$ appears exactly once in each column of $L'$.

Now consider row $k$ of $L'$. If $k \neq i, j$, then this row is exactly the same as row $k$ of $L$. Since every element of $N$ appears exactly once in row $k$ of $L$, it also appears exactly once (and in the same position even) in row $k$ of $L'$. If $k = i$ or $k = j$, then row $k$ of $L'$ is the same as some other row (the $j$th or $i$th row, respectively) of $L$. Since every element of $N$ appears exactly once in that row of $L$, it also appears exactly once in row $k$ of $L'$.

Thus, $L'$ satisfies the definition of a Latin square. ■

16.1.3.2 There are three different ways to complete the square:

$$\begin{array}{cccc} 1 & 3 & 4 & 2 \\ 2 & 1 & 3 & 4 \\ 4 & 2 & 1 & 3 \\ 3 & 4 & 2 & 1 \end{array} \qquad \begin{array}{cccc} 1 & 3 & 4 & 2 \\ 3 & 1 & 2 & 4 \\ 2 & 4 & 1 & 3 \\ 4 & 2 & 3 & 1 \end{array} \qquad \begin{array}{cccc} 1 & 3 & 4 & 2 \\ 3 & 1 & 2 & 4 \\ 4 & 2 & 1 & 3 \\ 2 & 4 & 3 & 1 \end{array}$$

So the completion is not unique.

**Solutions to Exercise 16.2.9:**

16.2.9.2 As explained in the first paragraph of the proof of Theorem 16.2.4, we may assume the first row is $1, 2, 3, 4$.

Now, we use the idea that is explained in the second paragraph of the proof of Theorem 16.2.4. For any position in a row after the first row, the entry in our new Latin square cannot be the same as the entry in this position of either of the two given squares (because, for any $j$, the ordered pair $(j, j)$ has already appeared in the top row of the given square and our new square), and it also cannot be the same as the entry in the first row of the column. This eliminates three possibilities for the entry in this position, so there is only one possibility left. Putting this remaining entry into each position yields the following Latin square, which must be the one that was requested:

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{array}$$

16.2.9.3

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | 1 | 2 | 7 | 8 | 5 | 6 |
| 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 |
| 7 | 8 | 5 | 6 | 3 | 4 | 1 | 2 |
| 4 | 3 | 2 | 1 | 8 | 7 | 6 | 5 |
| 2 | 1 | 4 | 3 | 6 | 5 | 8 | 7 |
| 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| 6 | 5 | 8 | 7 | 2 | 1 | 4 | 3 |

**Solutions to Exercise 16.3.6:**

16.3.6.2 This collection has no system of distinct representatives, because the five sets $A_2$, $A_3$, $A_4$, $A_5$, and $A_6$ have union $A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6 = \{v, w, x, y\}$, which has cardinality 4.

16.3.6.4 This collection does have a system of distinct representatives: $x$, $y$, and $z$, for $A_1$, $A_2$, and $A_3$ (respectively).

**Solutions to Exercise 16.3.12:**

16.3.12.1 Adam, Ella, Justin, and Bryant are the only friends that have visited either England, Scotland, Ireland, France, or Italy. Therefore, Onyx has only 4 friends to choose from to answer the questions for these 5 countries. Since the number of friends is less than the number of countries, Onyx cannot choose a different friend for each of these countries.

16.3.12.2 No, this cannot be completed to a $4 \times 4$ Latin square:

- The third entry in the third row must be 1, because 2, 3, and 4 already appear in either the third row or the third column.
- The last entry in the third row must also be 1, because 2, 3, and 4 already appear in either the third row or the last column.

So we cannot complete the third row: we are forced to have 1 appear twice in this row, but that is not allowed.

Hall's Theorem does *not* apply to this situation, because, as we have seen, the third column and fourth column must both choose their entry for the third row from the set $\{1\}$, which has less than two elements. (Theorem 16.3.8 does not apply because the partial Latin square does not consist only of complete rows — it has a row that has only been partially filled in.)

# Solutions for Chapter 17

**Solutions to Exercise 17.1.11:**

17.1.11.1 Numerically, this property is easy to verify from the proof of Theorem 17.1.9, which tells us that if $b$ is the number of blocks, then $b = \dfrac{\lambda v(v-1)}{k(k-1)}$, so (dividing both sides by 2 and multiplying through by $k(k-1)$) we have $b\binom{k}{2} = \lambda\binom{v}{2}$.

The correspondence is due to the colouring explained in Theorem 17.1.8.

17.1.11.2 We are given that $v = 16$, $k = 6$, and $\lambda = 3$.

From the formula $r(k-1) = \lambda(v-1)$, we see that

$$r = \frac{\lambda(v-1)}{k-1} = \frac{3(16-1)}{6-1} = 9,$$

which means that each point is in 9 blocks.

Now, from the formula $bk = vr$, we have

$$b = \frac{vr}{k} = \frac{16 \cdot 9}{6} = 24.$$

This means that the design has 24 blocks.

**Solutions to Exercise 17.2.8:**

17.2.8.1 **PROOF.** Clearly the complement of a design will still have $v$ varieties. (It will also have $b$ blocks, since each of its blocks comes from one block of the original design.)

From a block $B$ of size $k$, the corresponding block of the complementary design will have the $v - k$ elements of $V \setminus B$.

We need to count how many times any given pair of varieties appear together in a block of the complementary design. Two varieties appear together in a block of the complementary design if and only if neither of them was in the corresponding block of the original design. Each of the two varieties appeared in $r$ blocks of the original design, and they appear together in $\lambda$ blocks of the original design. We can now use inclusion-exclusion to count the number of blocks in which at least one of the two varieties appears: $r + r - \lambda = 2r - \lambda$. Thus, the number of blocks in which neither of them appears is $b-(2r-\lambda) = b-2r+\lambda$, as claimed. Since this count in no way depended on the choice of our two varieties, the complement is indeed a design, as every pair of varieties appear together in some block $b - 2r + \lambda$ times. ■

17.2.8.3 The set $\{1, 3, 7\}$ gives the differences $\pm 2$, $\pm 4$, and $\pm 6$, while the set $\{1, 6, 13\}$ gives the differences $\pm 5$, $\pm 7$, and $\pm 12$. So we need to find two sets that contain the differences $\pm 1$, $\pm 3$, $\pm 8$, $\pm 9$, $\pm 10$, and $\pm 11$. The sets $\{1, 2, 11\}$ and $\{1, 4, 12\}$ work.

**Solutions to Exercise 17.3.2:**

17.3.2.1 Many examples are possible (but they may be hard to find). For example, let

$$v = 16,\ k = 6, \text{ and } \lambda = 1\ .$$

Then

$$\lambda \frac{v-1}{k-1} = 1 \cdot \frac{16-1}{6-1} = 3$$

and

$$\lambda \frac{v(v-1)}{k(k-1)} = 1 \cdot \frac{16(16-1)}{6(6-1)} = \frac{16 \cdot 15}{6 \cdot 5} = 8$$

are integers, so the conditions in Theorem 17.1.9 are satisfied.

From the formula $r(k-1) = \lambda(v-1)$, we see that

$$r = \frac{\lambda(v-1)}{k-1} = \frac{1 \cdot (16-1)}{6-1} = 3.$$

Then, from the formula $bk = vr$, we have

$$b = \frac{vr}{k} = \frac{16 \cdot 3}{6} = 8.$$

Therefore $b = 8 < 16 = v$, so Fisher's Inequality is not satisfied.

Since Fisher's Inequality is not satisfied, there is no BIBD with these parameters.

17.3.2.2 It is shown just before the proof of Fisher's Inequality that Fisher's Inequality is equivalent to $\lambda(v-1) \geq k(k-1)$. Since $\lambda = 1$ and $k = 20$, this means

$$1 \cdot (v-1) \geq 20(20-1) = 380,$$

so $v \geq 380 + 1 = 381$. Therefore, $v$ must be at least 381 to satisfy Fisher's Inequality.

Since

$$\lambda \frac{v-1}{k-1} = 1 \cdot \frac{381-1}{20-1} = \frac{380}{19} = 20$$

and

$$\lambda \frac{v(v-1)}{k(k-1)} = 1 \cdot \frac{381(381-1)}{20(20-1)} = \frac{381(380)}{380} = 381$$

are integers, the conditions in Theorem 17.1.9 are also satisfied. So 381 is the smallest value for $v$ that satisfies all three conditions.

# Solutions for Chapter 18

**Solutions to Exercise 18.1.12:**

18.1.12.2 Since $v = 39 = 6 \cdot 6 + 3 \equiv 3 \pmod 6$, the proof of Theorem 18.1.8 tells us that we should use a Latin square constructed in Lemma 18.1.4. Since $v/3 = 39/3 = 13$, the Latin square is of order $n = 13$, so the first sentence of the proof of Lemma 18.1.4 tells us that the first row of the Latin square is

$$1 \quad \frac{13+3}{2} \quad 2 \quad \frac{13+5}{2} \quad 3 \quad \frac{13+7}{2} \quad 4 \quad \cdots \quad 13 \quad \frac{13+1}{2}.$$

In other words, the first row is

$$1 \quad 8 \quad 2 \quad 9 \quad 3 \quad 10 \quad 4 \quad 11 \quad 5 \quad 12 \quad 6 \quad 13 \quad 7.$$

Then the second sentence of the proof of Lemma 18.1.4 tells us that the rest of the rows are obtained by shifting to the left. So the Latin square is

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 8 | 2 | 9 | 3 | 10 | 4 | 11 | 5 | 12 | 6 | 13 | 7 |
| 8 | 2 | 9 | 3 | 10 | 4 | 11 | 5 | 12 | 6 | 13 | 7 | 1 |
| 2 | 9 | 3 | 10 | 4 | 11 | 5 | 12 | 6 | 13 | 7 | 1 | 8 |
| 9 | 3 | 10 | 4 | 11 | 5 | 12 | 6 | 13 | 7 | 1 | 8 | 2 |
| 3 | 10 | 4 | 11 | 5 | 12 | 6 | 13 | 7 | 1 | 8 | 2 | 9 |
| 10 | 4 | 11 | 5 | 12 | 6 | 13 | 7 | 1 | 8 | 2 | 9 | 3 |
| 4 | 11 | 5 | 12 | 6 | 13 | 7 | 1 | 8 | 2 | 9 | 3 | 10 |
| 11 | 5 | 12 | 6 | 13 | 7 | 1 | 8 | 2 | 9 | 3 | 10 | 4 |
| 5 | 12 | 6 | 13 | 7 | 1 | 8 | 2 | 9 | 3 | 10 | 4 | 11 |
| 12 | 6 | 13 | 7 | 1 | 8 | 2 | 9 | 3 | 10 | 4 | 11 | 5 |
| 6 | 13 | 7 | 1 | 8 | 2 | 9 | 3 | 10 | 4 | 11 | 5 | 12 |
| 13 | 7 | 1 | 8 | 2 | 9 | 3 | 10 | 4 | 11 | 5 | 12 | 6 |
| 7 | 1 | 8 | 2 | 9 | 3 | 10 | 4 | 11 | 5 | 12 | 6 | 13 |

18.1.12.4 No, it is not a Kirkman system.

We will call $\{u_1, \ldots, u_5\}$ the $u$-girls; $\{v_1, \ldots, v_5\}$ the $v$-girls; and $\{w_1, \ldots, w_5\}$ the $w$-girls. Of the 35 blocks that we obtained through the construction, 5 have one $u$-girl, one $v$-girl, and

one $w$-girl; the other 30 have either two $u$-girls with a $v$-girl, two $v$-girls with a $w$-girl, or two $w$-girls with a $u$-girl.

A Kirkman system requires us to divide the blocks into 7 groups of 5 blocks such that each girl appears exactly once in each group of blocks. Since there should be 7 groups of 5 blocks, but there are only 5 blocks that have a $u$-girl, a $v$-girl, and a $w$-girl, there must be at least one group of blocks (in fact, at least two) that has no block consisting of a $u$-girl, a $v$-girl, and a $w$-girl.

Consider such a group of 5 blocks. We must have all 5 of the $u$-girls. If no block contained more than one $u$-girl, then in order to get all 5 $u$-girls we would have to choose only blocks that have two $w$-girls and a $u$-girl. However, this would mean that we had 10 $w$-girls and no $v$-girls, which is not allowed. So we must choose at least one block that has two $u$-girls and a $v$-girl. Repeating the same argument with $v$ or $w$ taking the place of $u$, we see that we must also choose at least one block that has two $v$-girls and a $w$-girl, and at least one block that has two $w$-girls and a $u$-girl. Since we are only choosing 5 blocks but there are these three classes of blocks, there must be some class of blocks of which we only choose one.

Without loss of generality, suppose that we only choose one of the blocks that has two $u$-girls and a $w$-girl. In order to have all 5 of the $u$-girls, we must choose three blocks that have two $w$-girls and a $u$-girl. But this means that we have six $w$-girls, which is not allowed.

Therefore, there is no way to partition the blocks of this design into seven groups of five blocks so that every girl appears exactly once in each group.

**Solutions to Exercise 18.2.7:**

18.2.7.2 We must have $b\binom{k}{t} = \lambda\binom{v}{t} = \binom{15}{t}$.

Since we are not including any trivial $t-(v,t,1)$ design, we have $t \geq 2$, $3 \leq k \leq 14$, and $t < k$.

Now

$$\frac{15!}{t!(15-t)!} = b\frac{k!}{t!(k-t)!},$$

which means that

$$\frac{15 \cdot 14 \cdots (16-t)}{k(k-1)\cdots(k+1-t)}$$

is an integer.

Furthermore, we have $\binom{k-1}{t-1}$ divides $\binom{14}{t-1}$, so that $\frac{(k-1)!}{(k-t)!}$ divides $\frac{14!}{(15-t)!}$. In other words,

$$\frac{14!(k-t)!}{(15-t)!(k-1)!} = \frac{14 \cdot 13 \cdots (16-t)}{(k-1)(k-2)\cdots(k+1-t)}$$

is an integer. If we call this integer $y$, combining this with the previous paragraph tells us that $k$ is a divisor of $15y$. We can also further work with the algebra to obtain

$$y = \frac{14 \cdot 13 \cdots k}{(15-t)(14-t)\cdots(k+1-t)}.$$

When $k = 14$, this gives $y = 14/(15-t)$. Since $k$ divides $15y$ and $k$ is coprime to 15, we must have $k$ divides $y$. But then $y/14 = 1/(15-t)$ is an integer, implying $t = 14$. This contradicts $t < k$. Thus $k = 14$ cannot arise.

When $k = 13$ this gives $y = \dfrac{14 \cdot 13}{(15-t)(14-t)}$. Since $k$ divides $15y$ and $k$ is coprime to 15, we must have $k$ divides $y$. But then $\dfrac{y}{13} = \dfrac{14}{(15-t)(14-t)}$ is an integer. Since $t < k = 13$, we have $14 - t \geq 2$, but no two consecutive integers each of which is at least 2 are both divisors of 14, a contradiction. Thus $k = 13$ cannot arise.

When $k = 12$, this gives

$$y = \frac{14 \cdot 13 \cdot 12}{(15-t)(14-t)(13-t)}.$$

Now $k$ dividing $15y$ implies that

$$\frac{15 \cdot 14 \cdot 13}{(15-t)(14-t)(13-t)}$$

is an integer. Since the numerator is not a multiple of $2^2$, the denominator cannot be either, leaving only the possibilities $t = 4, 8$. Since the numerator is not a multiple of $3^2$, the denominator cannot be either, which eliminates $t = 4$. When $t = 8$, the numerator of $y$ is not a multiple of 5, but the denominator is, so this is also impossible. Thus $k = 12$ cannot arise.

When $k = 11$ this gives

$$y = \frac{14 \cdot 13 \cdot 12 \cdot 11}{(15-t)(14-t)(13-t)(12-t)}.$$

Since $k$ divides $15y$ and $k$ is coprime to 15, we must have $k$ divides $y$. But then

$$\frac{y}{11} = \frac{14 \cdot 13 \cdot 12}{(15-t)(14-t)(13-t)(12-t)}$$

is an integer. Since the numerator is not a multiple of 5, the four consecutive numbers that are the factors of the denominator must be 6 through 9 (since $t \geq 2$, they cannot be 11 through 14, and since $t < 11$ they cannot be 1 through 4). Thus, we must have $t = 6$. But then the numerator is not divisible by $3^2$, while the denominator is divisible by $3^3$, contradicting $y/11$ being an integer. Thus $k = 11$ is not possible.

When $k = 10$, this gives

$$y = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10}{(15-t)(14-t)(13-t)(12-t)(11-t)}.$$

Now $k$ dividing $15y$ implies that

$$\frac{15 \cdot 14 \cdot 13 \cdot 12 \cdot 11}{(15-t)(14-t)(13-t)(12-t)(11-t)}$$

is an integer. Since the numerator is not a multiple of $2^4$, the denominator cannot be either. In particular, 8 cannot be one of the factors that appears in the denominator (since some other even factor would appear with it), nor can 2, 4, and 6 all be factors that appear in the denominator. Also, the numerator is not divisible by $3^3$, so we cannot have $11 - t = 9$. This leaves $t = 8$ as the only possibility. However, the numerator of $y$ is not divisible by $3^2$, so $t = 8$ is also not possible. Thus $k = 10$ is not possible.

When $k = 9$, we see that

$$y = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9}{(15-t)(14-t)(13-t)(12-t)(11-t)(10-t)}.$$

So $k$ dividing $15y$ gives

$$\frac{15 \cdot 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10}{(15-t)(14-t)(13-t)(12-t)(11-t)(10-t)}$$

being an integer. Since the numerator is not divisible by $2^5$, the denominator cannot be either. In particular, 8 cannot appear as one of the factors in the denominator (or two other numbers divisible by 2 would also appear as factors), so the only possibility is $t = 8$. However, if we take $k = 9$, $t = 8$, and $i = 2$, the necessary condition is $\binom{7}{6} = 7$ divides $\binom{13}{6}$, which is not true. Thus, $k = 9$ is not possible.

When $k = 8$, we calculate

$$y = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{(15-t)(14-t)(13-t)(12-t)(11-t)(10-t)(9-t)}.$$

Since $k$ divides $15y$ and $k$ is coprime to 15, we must have $k$ divides $y$. But then

$$\frac{y}{8} = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9}{(15-t)(14-t)(13-t)(12-t)(11-t)(10-t)(9-t)}.$$

Since the consecutive factors in the denominator include 8 and at least two other even numbers, this implies that the numerator should also be a multiple of $2^5$, but it is not. Thus $k = 8$ is not possible.

When $k = 7$ we see that

$$y = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7}{(15-t)(14-t)(13-t)(12-t)(11-t)(10-t)(9-t)(8-t)}.$$

Since the consecutive factors in the denominator include 6 and 9, if they also include another multiple of 3 then the numerator must be divisible by $3^4$, but it is not. This leaves the possibility that $t = 4$ so the factors in the denominator are 4 through 11, but this is divisible by $5^2$, which the numerator is not. Thus, $k = 7$ is not possible.

If $k = 6$ then

$$y = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{(15-t)(14-t)(13-t)(12-t)(11-t)(10-t)(9-t)(8-t)(7-t)}.$$

So $k$ dividing $15y$ gives

$$\frac{15 \cdot 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7}{(15-t)(14-t)(13-t)(12-t)(11-t)(10-t)(9-t)(8-t)(7-t)}$$

being an integer. The numerator is not divisible by $2^8$, so the denominator cannot be either; in particular it cannot include as factors all of the even integers from 4 through 10 as well as one other. This leaves the possibilities $t = 2$ and $t = 4$. If $t = 2$ then $y = 14/5$ which is not an integer, and similarly if $t = 4$ then we have $y = \dfrac{14 \cdot 13 \cdot 12}{5 \cdot 4 \cdot 3}$ which is not an integer. So $k = 6$ is not possible.

If $k = 5$ then

$$y = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{(15-t)(14-t)(13-t)(12-t)(11-t)(10-t)(9-t)(8-t)(7-t)(6-t)}.$$

The numerator is not divisible by $2^9$, so the denominator cannot be either; in particular, the denominator cannot include all of 4, 6, 8, 10, and 12 as factors in the product. This leaves only

the possibility $t = 4$. If $k = 5$ and $t = 4$ then $i = 0$ gives $\binom{5}{4} = 5$ divides $\binom{15}{4} = 1365$ which is true; $i = 1$ gives $\binom{4}{3} = 4$ divides $\binom{14}{3} = 364$ which is true; $i = 2$ gives $\binom{3}{2} = 3$ divides $\binom{13}{2} = 78$, which is true; $i = 3$ gives $\binom{2}{1} = 2$ divides $\binom{12}{1} = 12$, which is true. Thus a $4-(15, 5, 1)$ design could exist, but this is the only possibility with $k = 5$.

When $k = 4$ we have

$$y = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4}{(15-t)(14-t)(13-t)(12-t)(11-t)(10-t)(9-t)(8-t)(7-t)(6-t)(5-t)}.$$

The denominator includes 3, 6, 9, and 12, so is divisible by $3^5$, but the numerator is not. Thus, $k = 4$ is not possible.

If $k = 3$ then $2 \leq t < k$ implies $t = 2$. We know these parameters are possible, as these are the parameters of a Steiner triple system.

Thus, the only possible values of $k$ and $t \geq 2$ for which nontrivial $t$-designs might exist with $v = 15$ and $\lambda = 1$ are $k = 5$ and $t = 4$: a $4-(15, 5, 1)$ design, or $k = 3$ and $t = 2$: a $(15, 3, 1)$ Steiner triple system.

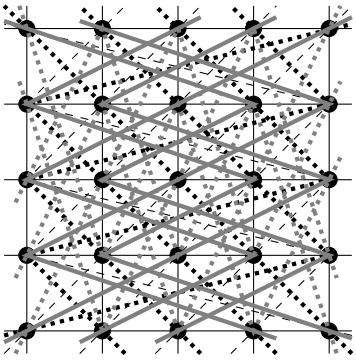
18.2.7.3 If a $3-(16, 6, 1)$ design exists then we have $\binom{6-i}{3-i}$ divides $\binom{16-i}{3-i}$ for $0 \leq i \leq 2$. With $i = 0$ this gives $\binom{6}{3} = 20$ divides $\binom{16}{3} = 560$, which is true. With $i = 1$, we have $\binom{5}{2} = 10$ divides $\binom{15}{2} = 105$, which is not true. Therefore such a design is not possible.

**Solutions to Exercise 18.3.9:**

18.3.9.1 **PROOF.** Let $L$, $M$, and $N$ be lines of an affine plane such that $L$ is parallel to both $M$ and $N$; that is, no point lies on both $L$ and $M$ or on both $L$ and $N$. Let $p$ be an arbitrary point on $M$. Since $L$ and $M$ are parallel, $p$ does not lie on $L$. By the parallel postulate, there is a unique line through $p$ that is parallel to $L$; this line is $M$. Therefore $N$ cannot contain $p$. Since our choice of $p$ on $M$ was arbitrary, no point of $M$ can also lie on $N$, so $M$ and $N$ are parallel. ■

18.3.9.3 A finite affine plane of order 19 has $19^2 = 361$ points, and $19(19 + 1) = 380$ lines.

18.3.9.5 Without colours it is difficult to effectively draw this plane so that the parallel classes can be clearly seen. We will use solid vertical lines; solid horizontal lines; dashed lines; dotted lines; solid grey lines; and dotted grey lines to represent the six parallel classes of lines, but some of these may be difficult to distinguish. Note that the lines that are neither vertical nor horizontal will "turn corners" or zig-zag to join their sets of 5 points.



We obtain the following 4 MOLS of order 5 from this affine plane, by using the vertical and horizontal lines to create the coordinates. To make things easier to see, we will have the positions in the Latin squares correspond visually to the positions in the 5 by 5 array of points that we have drawn, so the top-left entry in the Latin squares will come from the top-left point of the array, etc. We will number the lines in each parallel class so as to ensure that the entries in the top row of each square are 1, 2, 3, 4, and 5, in that order. The first square corresponds to the dashed lines; the second to the dotted lines; the third to the solid grey lines, and the fourth to the dashed grey lines.

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 2 | 3 | 4 | 5 | 1 | 5 | 1 | 2 | 3 | 4 | 3 | 4 | 5 | 1 | 2 | 4 | 5 | 1 | 2 | 3 |
| 3 | 4 | 5 | 1 | 2 | 4 | 5 | 1 | 2 | 3 | 5 | 1 | 2 | 3 | 4 | 2 | 3 | 4 | 5 | 1 |
| 4 | 5 | 1 | 2 | 3 | 3 | 4 | 5 | 1 | 2 | 2 | 3 | 4 | 5 | 1 | 5 | 1 | 2 | 3 | 4 |
| 5 | 1 | 2 | 3 | 4 | 2 | 3 | 4 | 5 | 1 | 4 | 5 | 1 | 2 | 3 | 3 | 4 | 5 | 1 | 2 |

**Solutions to Exercise 18.4.4:**

18.4.4.1 No, not every design with $\lambda = 1$ is is a projective plane. The condition $\lambda = 1$ ensures that every two points have a unique line that is incident with both of them. However, there is no requirement in a design that every two blocks have a nonempty intersection. (If every two blocks do have a nonempty intersection, then the condition $\lambda = 1$ does force the intersection to have exactly one point.) The condition that there exist four points such that no three are incident with a single line can also fail, but only in trivial or complete situations.

18.4.4.3 From Theorem 16.2.7, we know that there are $p - 1$ MOLS of order $p$ whenever $p$ is prime (in fact, whenever $p$ is a prime power, though we did not prove this). This implies that there is a projective plane that has $p + 1$ points on each line whenever $p$ is prime (or a prime power).

18.4.4.4 Since there are not 5 MOLS of order 6 (as we saw in Euler's problem), there is no projective plane that has 7 points on each line.

# Solutions for Chapter 19

**Solution to Exercise 19.1.5:** The only string with an odd number of 1s is 10101, so it is not an allowable message, but all of the others are allowed.

**Solutions to Exercise 19.2.5:**

19.2.5.1 The only such word is "math."

19.2.5.3 There are many possibilities, such as "bats," "gash," and "many."

19.2.5.5 We have answered this in each solution given above.

**Solutions to Exercise 19.2.6:**

19.2.6.2 **PROOF.** Let $x$ and $y$ be words of the same length. We have $d(x, y) = 0$ if and only if $x$ and $y$ differ in no positions. This means that $x$ must have the same entry as $y$ in every position, which means $x = y$. ■

19.2.6.4 **PROOF.** Let $x$, $y$, and $z$ be words of the same length. Suppose that $d(x, z) = k$, so that $x$ and $z$ differ in $k$ positions. Suppose that $d(x, y) = i$, so $y$ differs from $x$ in $i$ positions. If $i \geq k$ then since $d(y, z) \geq 0$ by Exercise 19.2.6.1, we have $d(x, z) \leq d(x, y) + d(y, z)$. Otherwise, there must be some list of at least $k - i$ positions in which $x$ differs from $z$ but does not differ from $y$. In each of these positions, since $y$ has the same entry as $x$, $y$ must have a different entry than $z$. Therefore $d(y, z) \geq k - i$. Now $d(x, y) + d(y, z) \geq i + k - i = k = d(x, z)$, completing the proof. ■

**Solutions to Exercise 19.2.10:**

19.2.10.3 The minimum distance is 2. To see this, first note that $d(01011, \underline{10}011) = 2$, so the minimum distance is no more than 2. Since each of the nonzero codewords has exactly three 1s, its distance from 00000 is 3, and its distance to any other nonzero codeword is greater than 1, because changing a single bit will change the number of 1s. So the minimum distance is at least 2.

**Solutions to Exercise 19.2.13:**

19.2.13.2

a. The code can detect 5 errors, but not 6 (because the number of errors detected must be less than the minimum distance).

b. The code can correct 2 errors (because $2 \times 2 < 6$, but $2 \times 3 \not< 6$).

**Solutions to Exercise 19.3.6:**

19.3.6.1 Since

$$G = \begin{bmatrix} I_k \\ A \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix},$$

we have

$$G \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad G \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad G \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

This means that 0101 encodes as 010111, 0010 encodes as 001000, and 0010 encodes as 111001.

**Solution to Exercise 19.4.5:** Since $G = \begin{bmatrix} I_k \\ A \end{bmatrix}$, and the given matrix $G$ has 4 columns, we must have $k = 4$, so $I_k = I_4$ has 4 rows. Therefore, $A$ is all but the first 4 rows of $G$, which means

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}.$$

Since $A$ is a $3 \times 4$, matrix, we have $r = 3$, so the parity-check matrix is

$$P = [A\ I_r] = [A\ I_3] = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

**Solutions to Exercise 19.4.12:**

19.4.12.2

a. Yes, all six columns of the parity-check matrix are different from each other (and none of them are all 0), so Theorem 19.4.9 tells us that the code can correct all single-bit errors.

b. Let $P$ be the given parity-check matrix. Then:

- $P \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. This is the th column of $P$, so changing the th bit corrects the error. The received word 001001 decodes as 0010$\underline{1}$1.
- $P \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. This is 0, so there is no error. The received word 110011 decodes as 110011.
- $P \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. This is the nd column of $P$, so changing the nd bit corrects the error. The received word 000110 decodes as 0$\underline{1}$0110.

**Solutions to Exercise 19.4.15:**

19.4.15.1 There are $2^5 - 1 = 31$ different nonzero 5-bit strings. Of these 31 strings, 5 of them have only one 1. Thus, there are $31 - 5 = 26$ different nonzero strings with at least two 1s. Therefore, we can make a $5 \times 24$ matrix $A$, such that the columns of $A$ are 24 different binary column vectors with at least two 1s in each column (because there are 26 different possible columns to choose from, and we need only 24 of them). The columns of the resulting parity-check matrix $P = [A\ I_5]$ are all nonzero and distinct, so Theorem 19.4.9 tells us that the resulting binary linear code can correct every single-digit error.

Furthermore, since $P$ is $5 \times 24$, we know that $r = 5$ and $k = 24$. Since $r = n - k$, this implies $n = k + r = 24 + 5 = 29$. So the code is of type $(n, k) = (29, 24)$, as desired.

19.4.15.3 Suppose $P$ is the the parity-check matrix of a binary linear code of type $(n, k)$ that corrects all single-bit errors, and let $r = n - k$. Then Theorem 19.4.9 tells us that the columns of $P$ must be distinct (and nonzero). However, $P$ is $r \times n$, and $n = k + r$, so it has $k + r$ columns of length $r$, and there are only $2^r - 1$ different possible nonzero columns of length $r$. Therefore, we must have $k + r \leq 2^r - 1$. Conversely, if this inequality is satisfied, then we can construct a $k \times (k + r)$ parity-check matrix whose columns are all distinct and nonzero. Thus, the smallest possible number of check bits is the smallest value of $r$ that satisfies the inequality $k + r \leq 2^r - 1$.

Thus:

- $r = 2$ check bits suffice for $k = 1$, because $k + r = 1 + 2 = 3 = 2^2 - 1 = 2^r - 1$. (But $r = 1$ check bit does not suffice, because $k + r \geq 1 + 1 = 2 > 2^1 - 1 = 2^r - 1$.)
- $r = 3$ check bits suffice for $k = 2, 3, 4$, because $k + r \leq 4 + 3 = 7 = 2^3 - 1 = 2^r - 1$. (But $r = 2$ check bits do not suffice, because $k + r \geq 2 + 2 = 4 > 2^2 - 1 = 2^r - 1$.)
- $r = 4$ check bits suffice for $5 \leq k \leq 11$, because $k + r \leq 11 + 4 = 15 = 2^4 - 1 = 2^r - 1$. (But $r = 3$ check bits do not suffice, because $k + r \geq 5 + 3 = 8 > 2^3 - 1 = 2^r - 1$.)
- $r = 5$ check bits suffice for $12 \leq k \leq 20$, because $k + r \leq 20 + 5 = 25 < 31 = 2^5 - 1 = 2^r - 1$. (But $r = 4$ check bits do not suffice, because $k + r \geq 12 + 4 = 16 > 2^4 - 1 = 2^r - 1$.)

**Solutions to Exercise 19.5.3:**

19.5.3.1 Using Proposition 19.5.2, we have $v = 10$, $k - 2 = 4$ so that $k = 6$, and $\lambda = 1$. The question is asking us for $b$. Using $bk(k - 1) = \lambda v(v - 1)$ gives $30b = 90$ so $b = 3$. Such a code will have only 3 words.

# Appendix D

# List of Notation

| Symbol | Description | Page |
|---|---|---|
| $n!$ | $n$ factorial | 20 |
| $\binom{n}{r}$ | $n$ choose $r$ | 23 |
| $\left(\!\binom{n}{r}\!\right)$ | $n$ multichoose $r$ | 43 |
| $\binom{n}{r_1,\ldots,r_m}$ | $n$ choose $r_1$ and ... and $r_m$ | 45 |
| (IH) | Inductive Hypothesis | 52 |
| $V$ | set of vertices of a graph | 108 |
| $E$ | set of edges of a graph | 108 |
| $E(G)$ | edge set of the graph $G$ | 108 |
| $V(G)$ | vertex set of the graph $G$ | 108 |
| $u \sim v$ | $u$ is adjacent to $v$ | 109 |
| $uv$ | the edge between vertices $u$ and $v$ | 109 |
| $\mathrm{val}(v)$ | valency of $v$ | 109 |
| $\deg(v)$ | valency (degree) of $v$ | 109 |
| $d(v)$ | valency (degree) of $v$ | 109 |
| $d_G(v)$ | valency (degree) of $v$ in $G$ | 109 |
| $G \setminus \{v\}$ | $G$ with vertex $v$ deleted | 111 |
| $G \setminus S$ | $G$ with the set $S$ of vertices deleted | 111 |
| $G \setminus \{e\}$ | $G$ with edge $e$ deleted | 111 |
| $K_n$ | complete graph on $n$ vertices | 112 |
| $G^c$ | complement of $G$ | 113 |
| $\varphi\colon G_1 \to G_2$ | $\varphi$ is a map from the vertices of $G_1$ to the vertices of $G_2$ (in this course, always an isomorphism) | 115 |
| $G_1 \cong G_2$ | $G_1$ is isomorphic to $G_2$ | 115 |
| $P_n$ | path of length $n$ | 127 |
| $C_n$ | cycle of length $n$ | 128 |
| $\delta$ | minimum valency | 145 |
| $\Delta$ | maximum valency | 145 |
| $\delta(G)$ | minimum valency of $G$ | 145 |
| $\Delta(G)$ | maximum valency of $G$ | 145 |
| $\chi'(G)$ | chromatic index of $G$ | 150 |
| $\chi'$ | chromatic index | 150 |
| $K_{m,n}$ | complete bipartite graph | 152 |
| $c(v)$ | number of colours used on edges incident with $v$ | 153 |
| $R(n_1, \ldots, n_c)$ | Ramsey number | 157 |

(Continued on next page)

| Symbol | Description | Page |
|---|---|---|
| $\chi(G)$ | chromatic number of $G$ | 161 |
| $\chi$ | chromatic number | 161 |
| $F(G)$ | set of faces of $G$ | 169 |
| $F$ | set of faces | 169 |
| $G^*$ | planar dual of $G$ | 169 |
| MOLS | mutually orthogonal latin squares | 188 |
| $a \equiv b \pmod{n}$ | $a$ is equivalent to $b$ modulo $n$ | 190 |
| $\mathcal{B}$ | collection of blocks in a design | 199 |
| $v$ | number of varieties in a design | 199 |
| $b$ | number of blocks in a design | 199 |
| $r$ | replication number (number of times each variety appears in the blocks) of a design | 199 |
| $k$ | cardinality of the blocks of a design | 199 |
| $\lambda$ | number of times each pair (or $t$-set) of varieties appear together in a block of a ($t$-)design | 199 |
| $(b, v, r, k, \lambda)$-design | BIBD with parameters $b, v, r, k, \lambda$ | 199 |
| BIBD | balanced incomplete block design | 200 |
| BIBD$(v, k, \lambda)$ | BIBD with parameters $v, k, \lambda$ | 200 |
| STS$(v)$ | Steiner triple system on $v$ varieties | 211 |

# Index