

MA241 Combinatorics

Keith Ball

Books

Bender and Williamson, Foundations of Combinatorics with Applications.

Harris, Hirst and Mossinghoff, Combinatorics and Graph Theory.

Bollobás, Graph Theory: An Introductory Course.

Ball, Strange Curves, Counting Rabbits,...

Cameron, Combinatorics: Topics, Techniques, Algorithms.

Chapter 0. Introduction

Combinatorics is not an easy subject to define: the syllabus doesn't do so. Combinatorial problems tend to deal with finite structures and frequently involve counting something. Instead of defining it I will give an example of the kind of arguments we shall use.

It is a famous fact that if you select 23 people at random, there is a roughly 50:50 chance that some pair of them share a birthday: that two of the 23 people celebrate their birthdays on the same day.

Let us confirm this. We shall make a simplifying assumption that there are 365 days in every year.

It is easier to calculate the probability that all the people have **different** birthdays, and then subtract the result from 1.

Consider the first person, Alice. Her birthday can fall on any day of the year. Now look at the second, Bob. If he is not to share his birthday with Alice, there are only 364 of the 365 dates available for his birthday. The chance that he was born on one of those is $\frac{364}{365}$.

Now take the third person, Carol. If she is to avoid the birthdays of both Alice and Bob, she has only 363 possible days. So the chance that she falls into one of them is $\frac{363}{365}$.

Hence the chance that these three people are born on different days of the year is

$$\frac{365}{365} \times \frac{364}{365} \times \frac{363}{365}.$$

Continuing in this way for 23 people we get the probability that all 23 are born on different days to be

$$\frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \dots \times \frac{365 - 22}{365}.$$

With a calculator or computer you can check that this number is about 1/2. This is the chance that there is no matching pair. When you subtract from 1 to get the chance that there **is** a pair you also get about 1/2.

This argument doesn't help us to understand the phenomenon. It proves the statement but gives no real insight into why we need so few people.

To make it easier to see what's going on let's ask what the probability would be for k people instead of 23 and with 365 replaced by n . The answer is

$$\frac{n-1}{n} \times \frac{n-2}{n} \dots \times \frac{n-(k-1)}{n}$$

which is the same as

$$\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right).$$

$$\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right).$$

Each factor is pretty close to 1 so we still might be a bit surprised that the product is only $1/2$. We can't really tell because the product looks like a horrible function of k and n . Can we estimate the function so as to be able to see how large it is?

It is hard to estimate a product but (usually) much easier to estimate a sum. So we take logs.

We want to estimate

$$\log \left(1 - \frac{1}{n}\right) + \log \left(1 - \frac{2}{n}\right) + \cdots + \log \left(1 - \frac{k-1}{n}\right)$$

or

$$\sum_{j=1}^{k-1} \log \left(1 - \frac{j}{n} \right). \quad (1)$$

So far this doesn't help much because logs are complicated. But we now bring in calculus. We know that $\log(1+x)$ has a Taylor expansion for small x

$$\log(1+x) \approx x - \frac{x^2}{2} + \frac{x^3}{3} - \dots.$$

So if j is quite a bit smaller than n

$$\log \left(1 - \frac{j}{n} \right) \approx -\frac{j}{n}.$$

Ignoring the fact that we are adding up approximations, so the errors might add up to something a bit big, the expression (1) is approximately

$$\sum_{j=1}^{k-1} \log \left(1 - \frac{j}{n} \right) \approx \sum_{j=1}^{k-1} \left(-\frac{j}{n} \right) = -\frac{k(k-1)}{2n}.$$

This is an approximation to the logarithm of the probability, so the probability itself should be

$$\exp \left(-\frac{k(k-1)}{2n} \right).$$

Now if $k = 23$ and $n = 365$ we get $e^{-0.69315\dots} = 0.5000\dots$

Now we have a much clearer understanding of why we need k to be much smaller than n . For the exponential

$$\exp\left(-\frac{k(k-1)}{2n}\right)$$

to be about $1/2$ we need

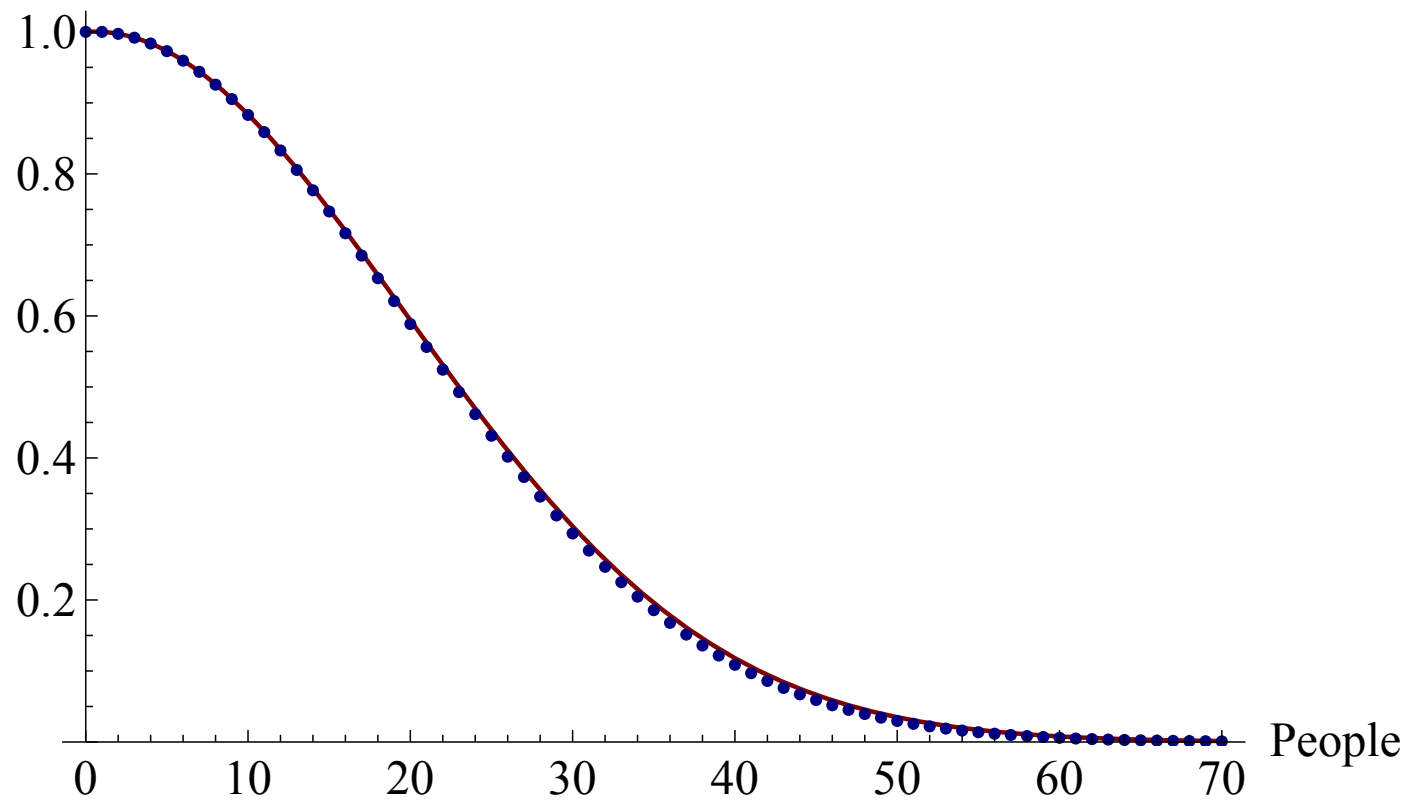
$$\frac{k(k-1)}{2n} \approx \log 2$$

and this means that $k(k-1)$ should be about the same size as n . So k is only about \sqrt{n} .

It happens that the approximation we made is pretty good for most values of k . Below is a graph showing the true probabilities of avoiding a match and a graph of our approximation

$$k \mapsto \exp\left(-\frac{k(k-1)}{2n}\right).$$

Chance of no match



How many pairs

We found the chance of getting no match. Let's ask **how many** matches we expect with k people?

One way to calculate this is to find the probability p_1 of exactly one match, the probability p_2 of exactly two matches and so on and then form the sum

$$p_1 + 2p_2 + 3p_3 + \dots$$

This would be madness. There is a much easier way: “probabilities are difficult but expectations are easy”.

For example, if you toss a fair coin 40 times the probability of k heads is

$$\binom{40}{k} \frac{1}{2^{40}}$$

for each k between 0 and 40.

The expected number of heads is therefore

$$\sum_{k=0}^{40} \binom{40}{k} \frac{1}{2^{40}} k.$$

But if you toss a coin 40 times the expected number of heads is obviously 20. Each toss contributes on average, half a head.

The expected number of heads is the chance of getting a head on each go multiplied by the number of goes.

Back to birthdays.

A “go” is a pair of people who might share.

The chance that a given pair of people share their birthday is $1/365$ (or $1/n$ in our algebraic version).

How many pairs of people are there? Each of the k people can be paired with each of the other $k - 1$ so the product $k(k - 1)$ counts each pair exactly twice. The total number of pairs is $k(k - 1)/2$. This number is the binomial coefficient “ k choose 2”.

So the expected number of pairs is

$$\frac{k(k-1)}{2} \times \frac{1}{n} = \frac{k(k-1)}{2n}.$$

This is the same expression that appeared in (our estimate for) the probability of no matches.

The expected number of pairs is K and the probability of no pair is roughly e^{-K} .

This reminds you of the Poisson distribution. Indeed, the number of matches has roughly a Poisson distribution if k is not too large.

The birthday example illustrates several points.

- We are counting something: the number of ways of distributing k names among n boxes so that each name lands in a different box.
- The problem involves a discrete structure: finitely many names and finitely many boxes: but we use analysis (calculus) to help us understand it.
- We needed to calculate the number of pairs that can be chosen from k people: something that appears in the Binomial Theorem which we shall return to.

As the example shows, combinatorics has links to probability theory. In particular it has close ties with statistical mechanics: the study of random models of particle systems.

It also has many links to computer science and in particular the theory of algorithms: “How can you compute such and such quickly?”

In this course we shall concentrate on two main parts of the subject: enumerative combinatorics and graph theory. Both of these appear in other applications.

Enumerative combinatorics

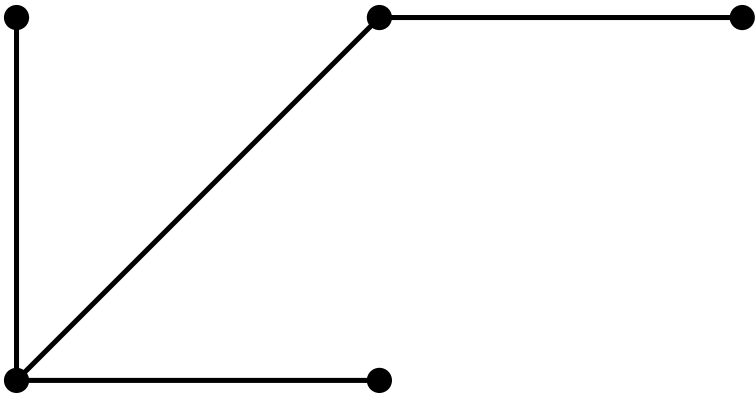
We examine ways to count things. How many ways can you reorder the numbers $1, 2, \dots, n$ so that no number appears in the same place as it started? (How many ways can you place the numbers into n numbered boxes, one in each box, so that no number lands in its own box?)

There is an obvious simpler problem to which you already know the answer: How many ways are there to order the n numbers?

Answer:

Graph theory

Informally a graph is a collection of points (or vertices) together with some of the lines joining pairs of them (edges).



Graphs have been used to model communication networks, the human brain, water percolating through porous rock and many other structures.

One of the most famous problems in graph theory is the four-colour problem. Is it possible to colour every planar map with 4 colours so that countries with a common border are always differently coloured?

Volume I. Enumerative combinatorics

Chapter 1. Basic counting and the Binomial Theorem

By convention, a **set** $\{x, y, z\}$ contains certain elements with no ordering on them and no repetition. A **sequence** (or list or vector) is ordered

$$(x_1, x_2, \dots, x_k)$$

and repetitions are allowed unless we specify otherwise.

Example The number of sequences of length k whose terms are selected from an n -element set such as $\{1, 2, 3, \dots, n\}$ is

$$n^k.$$

To see this we observe that there are n choices for the first entry, n for the second and so on.

In the case that $k = 2$ this is easy to draw. The possible 2-term sequences are

$$\begin{array}{cccc} (1, 1) & (1, 2) & \dots & (1, n) \\ (2, 1) & (2, 2) & \dots & (2, n) \\ \vdots & & & \\ (n, 1) & (n, 2) & \dots & (n, n) \end{array}$$

This is related to the idea of independence in probability theory.

Example The number of subsets of a set of size m is 2^m .

To see this observe that each subset is determined by which of the m elements are in and which are out. To build a subset we go through the m elements one by one and each time we choose: in or out. There are 2 choices each time so there are 2^m possible subsets.

The number of subsets of a set of size m is 2^m because we can pair off the subsets with sequences of length m whose terms are selected from the set $\{Y, N\}$.

If the set is $\{1, 2, 3\}$ then the pairing is

\emptyset	(N, N, N)
$\{1\}$	(Y, N, N)
$\{2\}$	(N, Y, N)
$\{3\}$	(N, N, Y)
$\{1, 2\}$	(Y, Y, N)
$\{2, 3\}$	(N, Y, Y)
$\{1, 3\}$	(Y, N, Y)
$\{1, 2, 3\}$	(Y, Y, Y)

A modification of the product idea can be used to calculate the number of **permutations** of the set $\{1, 2, \dots, n\}$. A permutation of the set is a sequence of length n in which each of the numbers appears exactly once. This time we have n choices for the first entry, but only $n - 1$ for the second and so on. So there are $n!$ permutations altogether.

More generally, the number of sequences of length k that can be formed from n elements **without repetition** is

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

In the case of sequences without repetition we do not have a simple Cartesian product structure on the sequences. For example, if $k = 2$ we get a pair of triangles

	(1, 2)	(1, 3)	(1, 4)
(2, 1)		(2, 3)	(2, 4)
(3, 1)	(3, 2)		(3, 4)
(4, 1)	(4, 2)	(4, 3)	

For larger k we would get a brick with various slices removed.

The choice of the first term in the sequence affects what you may choose later but it doesn't affect **how many** choices you make at each go.

Example

- How many 3 digit numbers are there? We know it is $999 - 99 = 900$. But we could argue: we have 9 choices for the first digit, then 10 for the second and 10 for the third.
- How many 3 digit numbers have all their digits different? This time we have 9 choices for the first digit. For the second we can use the digit **0** but not the one we already used: so there are 9 choices. For the third digit there are 8 choices so the answer is $9 \times 9 \times 8 = 648$.
- How many 3 digit numbers contain the string **11**?

How many 3 digit numbers contain the string **11**? We have 9 choices for the first digit but what happens next depends heavily upon whether we choose **1** or not.

It is easier to break into cases:

	1	1	*
not 1 or 0	1	1	1

$$10 + 8 = 18.$$

How many subsets of size k are there in a set of size n ? How many ways can we choose k objects from among n ?

We imagine choosing the elements one at a time as if we were writing down a sequence. As before we get the number of **sequences** of k distinct elements to be:

$$n(n - 1) \dots (n - k + 1).$$

In writing down the sets as (ordered) sequences we have counted each set not once but $k!$ times. So the actual number of sets is

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

The sequences of length k can be arranged in a grid: for example if $n = 4$ and $k = 3$.

(1, 2, 3)	(1, 2, 4)	(1, 3, 4)	(2, 3, 4)
(2, 1, 3)	(2, 1, 4)	(3, 1, 4)	(3, 2, 4)
(2, 3, 1)	(2, 4, 1)	(3, 4, 1)	(3, 4, 2)
(3, 2, 1)	(4, 2, 1)	(4, 3, 1)	(4, 3, 2)
(3, 1, 2)	(4, 1, 2)	(4, 1, 3)	(4, 2, 3)
(1, 3, 2)	(1, 4, 2)	(1, 4, 3)	(2, 4, 3)
<hr/>			
{1, 2, 3}	{1, 2, 4}	{1, 3, 4}	{2, 3, 4}

The number we calculated is the familiar **Binomial Coefficient**

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

Lemma (Choosing subsets). *The number of ways to choose a subset of k objects from among n is*

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

Since the total number of subsets of a set of size n is 2^n we have proved

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

This can be recognised as a special case of the Binomial Theorem.

Example A standard deck of cards consists of 52 cards divided into 4 “suits”: ♠, ♥, ♦ and ♣. Each suit consists of 13 cards A,2,3,4,5,6,7,8,9,10,J,Q,K.

- How many 5-card poker hands are there?

$$\binom{52}{5} \approx 2.6 \text{ million.}$$

- How many hands contain 4 of a kind? There are 13 ways to choose which kind. After those 4 cards have been selected there are 48 choices for the spare. So the answer is $13 \times 48 = 624$.

The Binomial Theorem

Theorem (Binomial). *If x and y are numbers and n is a non-negative integer then*

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$$

where $\binom{n}{k}$ is the number of ways of choosing k objects from among n and is given by

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

Proof We can expand the power $(x+y)^n$ by multiplying out the brackets

$$(x + y)(x + y) \dots (x + y).$$

$$(x + y)(x + y) \dots (x + y)$$

Each term in the expansion will be of the form $x^{n-k}y^k$ for some k between 0 and n , which we obtain whenever we choose y from k of the brackets and x from the remaining $n - k$. The number of ways of doing this is the number of ways of selecting k brackets from among n : namely

$$\binom{n}{k}.$$

□

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$$

We proved this directly by expanding and counting the subsets. Thus, I chose to **define** the binomial coefficient as the number of ways of choosing subsets and then prove two facts: the factorial formula and the expansion of $(x + y)^n$. But usually a better way to think of it is that all 3 objects are the same: the number of subsets, the coefficients in the expansion, and the factorial expression.

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$$

As mentioned above, if we set $x = y = 1$ we recover

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

The binomial coefficients may be arranged into what is known as Pascal's Triangle: it appears in Chinese texts 300 years before Pascal and in the Chandah Sutra of Pingal from 200BC.

The top 0^{th} row contains $\binom{0}{0} = 1$.

				1			
			1		1		
		1		2		1	
	1		3		3		1
	1	4		6		4	1
1		5	10		10	5	1

Each number is the sum of the two numbers immediately above it.

The fact that the binomial coefficients appear in Pascal's triangle can be summarised in the following lemma:

Lemma (The inductive property of Binomial Coefficients). *For each n and $1 \leq k \leq n - 1$*

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

To illustrate the idea that all 3 definitions of the binomial coefficients are equal we prove the formula in 3 ways.

Proof From the factorial formula

$$\begin{aligned}\binom{n-1}{k-1} + \binom{n-1}{k} &= \frac{(n-1)!}{(k-1)!(n-k)!} + \frac{(n-1)!}{k!(n-k-1)!} \\ &= \frac{(n-1)!}{(k-1)!(n-k-1)!} \left(\frac{1}{n-k} + \frac{1}{k} \right) \\ &= \frac{(n-1)!}{(k-1)!(n-k-1)!} \left(\frac{n}{(n-k)k} \right) = \binom{n}{k}.\end{aligned}$$

□

Proof From the binomial expansion, for example

$$\begin{aligned}(x + y)^4 &= (x + y)(x^3 + 3x^2y + 3xy^2 + y^3) \\ &= x^4 + 3x^3y + 3x^2y^2 + xy^3 \\ &\quad + x^3y + 3x^2y^2 + 3xy^3 + y^4 \\ &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4.\end{aligned}$$

□

Finally we can prove it with a combinatorial “story” .

Proof We are trying to write down the subsets of size k of $\{1, 2, \dots, n\}$. Each one either contains the symbol n and $k - 1$ of the other $n - 1$ symbols; or it fails to contain n but contains k of the others. So we can break the collection of k -subsets into two disjoint collections: $\binom{n-1}{k-1}$ that contain n and $\binom{n-1}{k}$ that don't. \square

This illustrates another obvious counting principle: if our family can be written as the union of disjoint subfamilies then the number of elements is the sum of the numbers in the subfamilies. This is related to disjointness in probability theory.

It would have been a little contrived but I could have approached the binomial theorem the opposite way around. Starting with the problem of counting subsets I could have proved the inductive formula using the combinatorial story and then proved the binomial expansion and the factorial formula by induction.

We shall see in the rest of this volume (enumerative combinatorics) that this approach usually works better. In more difficult problems it is not easy to see the formula immediately and so we approach it step by step.

The shape of the binomial coefficients

It is easy to see that the n^{th} row of Pascal's triangle is symmetric: for each n and k

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

We can get a much more detailed picture using probability theory.

If you toss a fair coin n times then the chance of getting k heads is

$$\binom{n}{k} \frac{1}{2^n}$$

because each sequence of heads and tails appears with probability $1/2^n$ and $\binom{n}{k}$ of them have k heads.

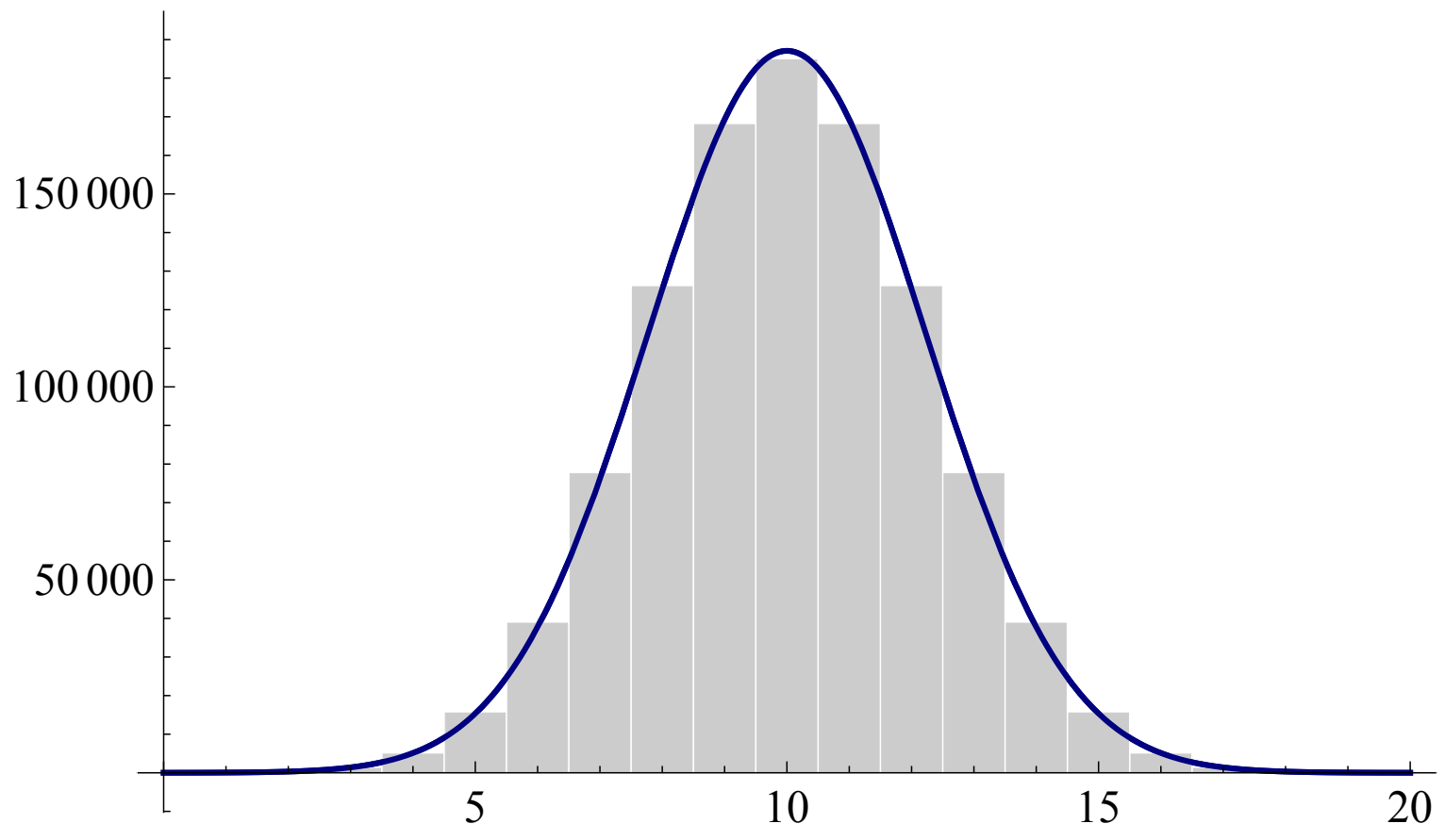
By the **Central Limit Theorem** these probabilities have a roughly **normal** or **Gaussian** distribution.

As a consequence the binomial coefficients themselves can be approximated by the curve

$$y = \frac{2^n}{\sqrt{\pi n/2}} \exp\left(-\frac{(k - n/2)^2}{n/2}\right).$$

The picture shows a bar chart of the binomial coefficients for $n = 20$ and the bell-shaped curve approximating them.

In the rabbits book you can find a direct derivation of this without the use of the Central Limit Theorem.



The bell-shaped curve gives good estimates for coefficients in the middle but not at the ends. If k is not too large then an obvious estimate which we shall use later is

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

The estimate can be simplified still further (without much loss of accuracy) by means of Stirling's formula which we shall discuss briefly in Chapter 2.5066....

Summary

Lemma (Choosing subsets). *The number of ways to choose a subset of k objects from among n is*

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

Theorem (Binomial). *If x and y are numbers and n is a non-negative integer then*

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$$

where $\binom{n}{k}$ is the number of ways of choosing k objects from among n and is given by

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

Lemma (The inductive property of Binomial Coefficients). For each n and $1 \leq k \leq n - 1$

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

$$\begin{array}{cccccc} & & & & & 1 \\ & & & & & & 1 & 1 \\ & & & & & 1 & 2 & 1 \\ & & & & 1 & 3 & 3 & 1 \\ & & 1 & 4 & 6 & 4 & 1 \end{array}$$

Chapter 2. Applications of the Binomial Theorem

A Mean Value Theorem

If we set $x = 1$ and $y = -1$ in the Binomial Theorem we get

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

as long as $n \geq 1$.

For example

$$1 - 4 + 6 - 4 + 1 = 0.$$

The sequence of signed binomial coefficients

$$(1, -4, 6, -4, 1)$$

is perpendicular to

$$(1, 1, 1, 1, 1).$$

This formula can be generalised. The signed sequence is also perpendicular to other powers:

$$(0, 1, 2, 3, 4)$$

$$(0, 1, 4, 9, 16)$$

$$(0, 1, 8, 27, 64).$$

Theorem (Orthogonality for Binomial Coefficients). *If $r < n$ are non-negative integers then*

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

Equivalently, for any polynomial p of degree at most $n - 1$

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

Any polynomial of degree at most $n - 1$ can be written as a linear combination of the powers up to $n - 1$.

For any polynomial p of degree at most $n - 1$

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

Proof It suffices to check this for a basis for the space of polynomials of degree at most $n - 1$:

$$p_0(x) = 1$$

$$p_1(x) = x$$

$$p_2(x) = x(x - 1)$$

$$\vdots$$

$$p_{n-1}(x) = x(x - 1) \dots (x - n + 2)$$

By the Binomial Theorem we have that for each t

$$\sum_{k=0}^n \binom{n}{k} (-1)^k t^k = (1-t)^n \quad (2)$$

and we saw that if you substitute $t = 1$ you get since $p_0(k) = 1$

$$\sum_{k=0}^n \binom{n}{k} (-1)^k p_0(k) = 0.$$

Differentiating (2) with respect to t we get

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

Now substituting $t = 1$ we get since $p_1(k) = k$

$$\sum_{k=0}^n \binom{n}{k} (-1)^k p_1(k) = 0$$

as long as $n \geq 2$.

Differentiating (2) twice with respect to t we get

$$\sum_{k=0}^n \binom{n}{k} (-1)^k k(k-1)t^{k-2} = n(n-1)(1-t)^{n-2}.$$

So now we get since $p_2(k) = k(k-1)$

$$\sum_{k=0}^n \binom{n}{k} (-1)^k p_2(k) = 0$$

as long as $n \geq 3$. We can continue in this way for p_3 and so on up to p_{n-1} . □

Theorem (Orthogonality for Binomial Coefficients). *If $r < n$ are non-negative integers then*

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

It is often useful to know what happens “next”: when $r = n$.

Lemma.

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

Lemma.

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

Proof The sum will be unchanged if we replace the polynomial $x \mapsto x^n$ by the polynomial $p_n : x \mapsto x(x-1)\dots(x-n+1)$ which differs from x^n by a polynomial of degree $n-1$. The polynomial p_n vanishes at $0, 1, 2, \dots, n-1$ and so for this polynomial we get

$$\sum_{k=0}^n \binom{n}{k} (-1)^k p_n(k) = \binom{n}{n} (-1)^n p_n(n) = (-1)^n n!.$$

□

The theorem and the lemma can be souped up to give a more general statement which is intuitively much easier to understand.

Theorem (Mean Value Theorem for Divided Differences). *If f is n times differentiable on an open interval including $[0, n]$ then*

$$\sum_{k=0}^n \binom{n}{k} (-1)^k f(k) = (-1)^n f^{(n)}(t)$$

for some t between 0 and n .

In other words $f(0) - f(1)$ is like a derivative.

If we do it twice

$$f(0) - 2f(1) + f(2) = (f(0) - f(1)) - (f(1) - f(2))$$

we get something like a second derivative, and so on.

Theorem (Mean Value Theorem for Divided Differences). *If f is n times differentiable on an open interval including $[0, n]$ then*

$$\sum_{k=0}^n \binom{n}{k} (-1)^k f(k) = (-1)^n f^{(n)}(t)$$

for some t between 0 and n .

It is easy to deduce the orthogonality property and the lemma from this Mean Value Theorem.

If $f(x) = x^r$ for $0 \leq r \leq n - 1$ then $f^{(n)} = 0$ so we get orthogonality.

If $f(x) = x^n$ then $f^{(n)}(x) = n!$ so we get the lemma.

The harder direction will be HW.

Theorem (Orthogonality for Binomial Coefficients). *If n is a positive integer and p is a polynomial of degree $n - 1$ then*

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

The mean value formulation suggests a different approach to the orthogonality.

Let p be a polynomial of degree m . What can we say about

$$p^{[n]}(x) = \sum_{k=0}^n \binom{n}{k} (-1)^k p(x + k)?$$

Note that

$$p^{[1]}(x) = p(x) - p(x + 1)$$

has degree $m - 1$ because the leading order terms cancel.

Now

$$\begin{aligned} p^{[2]}(x) &= p(x) - 2p(x + 1) + p(x + 2) \\ &= (p(x) - p(x + 1)) - (p(x + 1) - p(x + 2)) \\ &= p^{[1]}(x) - p^{[1]}(x + 1). \end{aligned}$$

So $p^{[2]}$ has degree $m - 2$.

In the same way

$$\begin{aligned} p^{[3]}(x) &= p(x) - 3p(x+1) + 3p(x+2) - p(x+3) \\ &= (p(x) - 2p(x+1) + p(x+2)) \\ &\quad - (p(x+1) - 2p(x+2) + p(x+3)) \\ &= p^{[2]}(x) - p^{[2]}(x+1). \end{aligned}$$

and so $p^{[3]}$ has degree $m - 3$.

If we continue in this way using the inductive property of the binomial coefficients we conclude that $p^{[m]}$ has degree 0: it is constant. Hence $p^{[n]}$ is zero if $n > m$.

Multisets

A **multiset** is an **unordered** collection in which repetition is allowed:
so

$$\{\{1, 2, 3\}\} = \{\{1, 3, 2\}\}$$

but

$$\{\{1, 2, 3\}\} \neq \{\{1, 2, 3, 3\}\}$$

A set is determined by what are its elements: which things are in and which are not.

A multiset is determined by what are its elements and how often each one is present.

In the spirit of the binomial coefficients we may ask: how many multisets are there of size d with elements from a set of size m ?

In algebra and algebraic geometry it is often useful to know the dimension of the space of polynomials or the space of homogeneous polynomials. In one variable this is not too hard: the space of polynomials of degree at most d has dimension $d + 1$. The space is spanned by the monomials

$$1, x, x^2, \dots, x^d.$$

For each d the space of **homogeneous** polynomials of degree d is 1-dimensional.

For two variables the monomials are

$$1, x, y, x^2, xy, y^2, x^3, \dots$$

and the dimension of the space of homogeneous polynomials of degree d is $d + 1$.

In fact the homogeneous polynomials of degree d in 2 variables can be identified with the polynomials in one variable of degree at most d :

$$\begin{array}{cccc} x^3 & x^2y & xy^2 & y^3 \\ 1 & y & y^2 & y^3 \end{array}$$

The monomials of degree d in m variables x_1, x_2, \dots, x_m can be identified with the multisets of size d whose elements come from $1, 2, \dots, m$.

For example if $d = 10$

$$x_1^3 x_2^2 x_3^5$$

corresponds to the multiset

$$\{\{1, 1, 1, 2, 2, 3, 3, 3, 3, 3\}\}.$$

So a second formulation of the problem is this: what is the dimension of the space of homogeneous polynomials of degree d in m variables?

A third formulation of the same problem is as follows. How many ways are there to put d identical oranges into m labelled boxes? The multiset

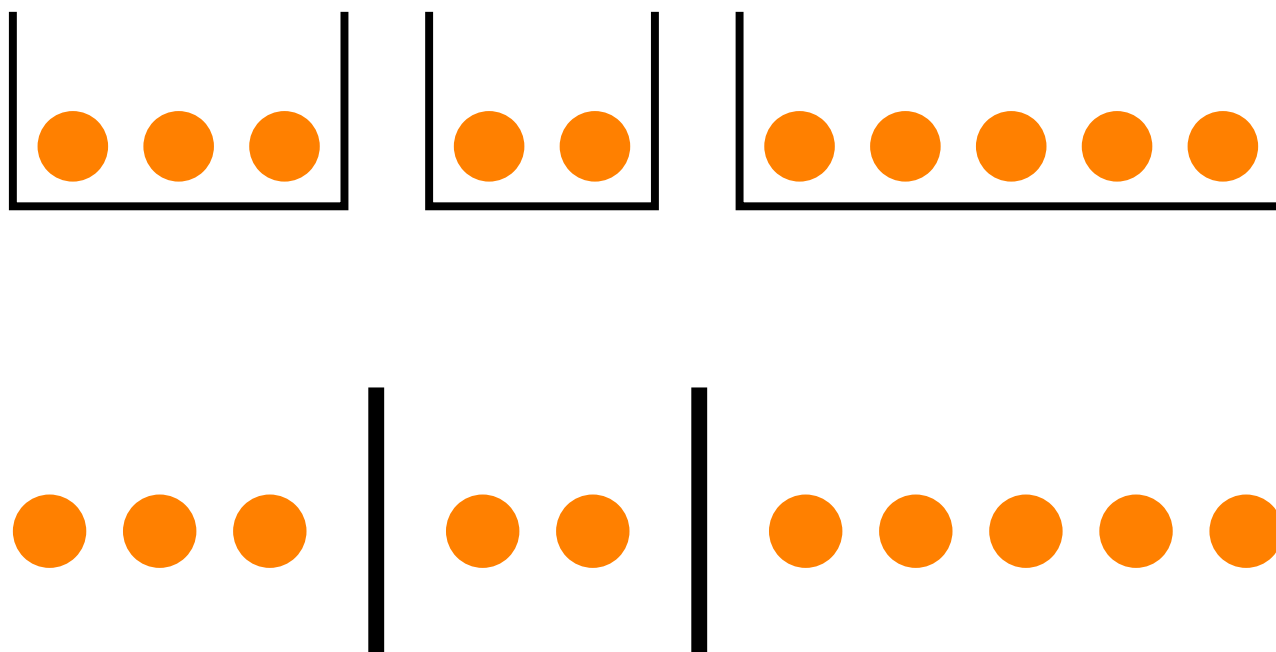
$$\{\{1, 1, 1, 2, 2, 3, 3, 3, 3, 3\}\}$$

corresponds to putting 3 oranges in box 1, 2 in box 2 and 5 in box 3.

Theorem (Multiset formula). *The number of multisets of size d with elements from a set of size m is*

$$\binom{d + m - 1}{m - 1} = \binom{d + m - 1}{d}.$$

Proof Imagine that instead of separating the oranges by putting them into boxes, we divide them using $m - 1$ pencils. We replace the gaps between neighbouring boxes by pencils. Each arrangement of oranges in boxes corresponds to a sequence of oranges and pencils.



We want the number of ways of arranging d oranges and $m - 1$ pencils. We have $d + m - 1$ items and we want to choose which of the $d + m - 1$ slots is occupied by pencils. There are

$$\binom{d + m - 1}{m - 1}$$

ways to do it.

□

Corollary (Dimension of spaces of polynomials). *The space of homogeneous polynomials of degree d in m variables has dimension*

$$\binom{d + m - 1}{m - 1}.$$

The space of polynomials of degree at most d in m variables has dimension

$$\binom{d + m}{m}.$$

Proof We already did the first one. The second one follows because each polynomial in m variables of degree at most d can be paired with a homogeneous polynomial of degree d in $m + 1$ variables:

$$1 + x + y + x^2 \quad \mapsto \quad w^2 + wx + wy + x^2.$$

□

Example

The space of polynomials of degree at most 2 in 3 variables has dimension $\binom{5}{2} = 10$.

The monomials are

1

x

x^2

y

z

xy

xz

y^2

yz

z^2

Multinomial Theorem

We may rewrite the binomial expansion in a more symmetric way

$$(x + y)^n = \sum_{k=0}^n \frac{n!}{k!(n-k)!} x^{n-k} y^k = \sum_{i,j \geq 0, i+j=n} \frac{n!}{i!j!} x^i y^j.$$

k and $n - k$ are just non-negative integers adding up to n .

There is a generalisation of the BT to more than two numbers:

$$(x + y + z)^n = \sum_{i,j,k \geq 0, i+j+k=n} \frac{n!}{i!j!k!} x^i y^j z^k.$$

Similarly for 4 numbers and so on.

Permutations and the Inclusion-Exclusion formula

How many numbers between 1 and 120 are divisible by 3? Answer 40.

How many are divisible by 5? Answer 24.

How many are divisible by 3 or 5? The answer is not 64 because we have counted some of them twice: all the ones divisible by 15 of which there are 8. So the answer is $40+24-8=56$.

How many are divisible by 3, 4 or 5? There are 40 divisible by 3, 30 by 4 and 24 by 5. There are 10 divisible by 3 and 4, 6 by 4 and 5 and 8 by 3 and 5. So the answer is

$$40 + 30 + 24 - 10 - 6 - 8 = 70?$$

No, because the two numbers 60 and 120 that are divisible by 3, 4 and 5 have been added in 3 times and then removed 3 times. So we need to add them back in again.

$$40 + 30 + 24 - 10 - 6 - 8 + 2 = 72.$$

A group of n absent-minded professors attend a lecture and each leaves his or her coat outside the door. At the end of the lecture the professors file out and each picks a coat at random from the rack. What is the chance that they all get the wrong coats?

The question is asking “How many permutations of $1, 2, \dots, n$ move every symbol?”

On the face of it this is quite tricky. We know there are $n!$ permutations altogether but what about those that fix nothing?

The key to understanding such a problem is to realise that there is **something** we can easily compute.

The number of permutations that fix 1 is $(n - 1)!$ because we are just permuting the other $n - 1$ symbols. The number that fix the symbol 2 is also $(n - 1)!$ and so on.

The number that fix both 1 and 2 is $(n - 2)!$ and so on. **Just as it was easy to calculate how many numbers were divisible by 3 and 5.**

Thus if we let A_i be the set of permutations that fix symbol i we can compute the size of each intersection of these sets

$$A_1, A_7, A_2 \cap A_3, A_1 \cap A_2 \cap A_9, \dots$$

The problem is to find out how many permutations don't belong to any A_i : to find the size of the complement of the union. There is a formula that resembles the Binomial Theorem which does it for us.

The problem is to find out how many permutations don't belong to any A_i : to find the size of the complement of the union. There is a formula that resembles the Binomial Theorem which does it for us.

Theorem (Inclusion-Exclusion formula). *Let A_1, A_2, \dots, A_n be subsets of a set Ω . Then*

$$\begin{aligned} & |\Omega - (A_1 \cup A_2 \cup \dots \cup A_n)| \\ &= |\Omega| - \sum_i |A_i| + \sum_{i < j} |A_i \cap A_j| - \sum_{i < j < k} |A_i \cap A_j \cap A_k| + \dots \end{aligned}$$

$$\begin{aligned}
& |\Omega - (A_1 \cup A_2 \cup \dots \cup A_n)| \\
&= |\Omega| - \sum_i |A_i| + \sum_{i < j} |A_i \cap A_j| - \sum_{i < j < k} |A_i \cap A_j \cap A_k| + \dots .
\end{aligned}$$

Proof For each element of Ω let us count how many terms on the right it contributes to. Suppose it belongs to exactly q of the sets. Then it contributes once to $|\Omega|$. It contributes q times to $\sum_i |A_i|$, $\binom{q}{2}$ times to the next term and so on. So the total contribution is

$$1 - \binom{q}{1} + \binom{q}{2} - \binom{q}{3} + \dots + (-1)^q \binom{q}{q}$$

and we already saw that this sum is zero except if $q = 0$. In this case the sum is 1. So each element that belongs to **none** of the A_i contributes once and no other elements contribute. \square

Returning to the professors problem, we have that each A_i contains $(n - 1)!$ elements each $A_i \cap A_j$ contains $(n - 2)!$ and so on. So we get

$$\begin{aligned} |\Omega - (A_1 \cup A_2 \cup \dots \cup A_n)| &= n! - \binom{n}{1}(n - 1)! + \binom{n}{2}(n - 2)! - \dots \\ &= \sum_{k=0}^n \binom{n}{k} (-1)^k (n - k)! \\ &= n! \sum_{k=0}^n (-1)^k \frac{1}{k!}. \end{aligned}$$

Examples If $n = 2$ the only permutation that works is the transposition. Sure enough

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

If $n = 3$ then the 3-cycles are the only ones that work.

$$6(1 - 1 + 1/2 - 1/6) = 2.$$

The **proportion** of permutations that fix nothing is the number of permutations divided by $n!$ so it is

$$\sum_{k=0}^n (-1)^k \frac{1}{k!}.$$

When n is large this partial sum is very close to e^{-1} . The chance that all professors get the wrong coats is about e^{-1} .

What is the expected number of professors who get the correct coat?

“Expectations are easy.”

How many “goes” are there: n professors.

What is the chance that a particular professor will get the right coat?

Each has $1/n$ chance.

So the expected number is 1 and the probability of no correct coats is close to e^{-1} .

This is another example of roughly Poisson behaviour **but I'm cheating a bit.**

Summary

Theorem (Orthogonality for Binomial Coefficients). *If $r < n$ are non-negative integers then*

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

Lemma.

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

Theorem (Mean Value Theorem for Divided Differences). *If f is n times differentiable on an open interval including $[0, n]$ then*

$$\sum_{k=0}^n \binom{n}{k} (-1)^k f(k) = (-1)^n f^{(n)}(t)$$

for some t between 0 and n .

Theorem (Multiset formula). *The number of multisets of size d with elements from a set of size m is*

$$\binom{d + m - 1}{m - 1} = \binom{d + m - 1}{d}.$$

Corollary (Dimension of spaces of polynomials). *The space of homogeneous polynomials of degree d in m variables has dimension*

$$\binom{d + m - 1}{m - 1}.$$

The space of polynomials of degree at most d in m variables has dimension

$$\binom{d + m}{m}.$$

Multinomial Theorem

$$(x + y + z)^n = \sum_{i,j,k \geq 0, i+j+k=n} \frac{n!}{i!j!k!} x^i y^j z^k.$$

Theorem (Inclusion-Exclusion formula). Let A_1, A_2, \dots, A_n be subsets of a set Ω . Then

$$\begin{aligned} & |\Omega - (A_1 \cup A_2 \cup \dots \cup A_n)| \\ &= |\Omega| - \sum_i |A_i| + \sum_{i < j} |A_i \cap A_j| - \sum_{i < j < k} |A_i \cap A_j \cap A_k| + \dots \end{aligned}$$

Chapter 2.5066.... Stirling's formula

How large is $n!$? This may seem like an odd question. $3! = 6$: what more is there to say?

Example Express 2345^{2345} approximately in standard notation.

$$\log_{10} (2345^{2345}) = 2345 \log_{10} 2345 \approx 7902.985$$

So the number itself is approximately

$$10^{0.985} 10^{7902} \approx 9.66 \times 10^{7902}.$$

Example Express 2345! approximately in standard notation.

You can't easily feed factorials into standard functions like logs.

Theorem (Stirling).

$$n! \asymp \sqrt{2\pi} e^{-n} n^{n+1/2}$$

as $n \rightarrow \infty$.

The symbol \asymp means that the ratio of the two sides approaches 1.

It is not too hard to show that

$$r_n = \frac{n!}{e^{-n} n^{n+1/2}}$$

converges (to something) as $n \rightarrow \infty$.

To get the $\sqrt{2\pi}$ is trickier. Most arguments at some point use the Gaussian integral

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}.$$

One approach is given in the rabbits book.

Chapter 3. The Fibonacci numbers and linear difference equations

The Fibonacci sequence is probably the best known sequence in maths.

$$1, 1, 2, 3, 5, 8, 13, \dots$$

in which each term is the sum of the two previous ones. We label the sequence u_1, u_2 and so on. The recurrence relation that defines the sequence is

$$u_{n+1} = u_n + u_{n-1}.$$

This relation immediately shows that the terms are integers. It doesn't enable us to see easily how large they are.

It is convenient to extend the sequence backwards at least by defining $u_0 = 0$ which preserves the recurrence.

Binet is credited with finding a closed formula although it was almost certainly known earlier.

Theorem (Binet). *If $n \geq 0$ the Fibonacci number u_n is*

$$u_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right) = \frac{\phi^n - \psi^n}{\phi - \psi}$$

where $\phi = \frac{1 + \sqrt{5}}{2}$ and $\psi = \frac{1 - \sqrt{5}}{2}$ are the solutions of the equation

$$x^2 = x + 1.$$

$$u_n = \frac{\phi^n - \psi^n}{\phi - \psi}.$$

Proof It is easy to see that if $n = 0$ or $n = 1$ then we get the right value. So it suffices to check that for every n the expression $w_n = A\phi^n + B\psi^n$ satisfies the recurrence relation (whatever the values of A and B) because our expression is of this form. But

$$w_{n+1} - w_n - w_{n-1} = A(\phi^{n+1} - \phi^n - \phi^{n-1}) + B(\psi^{n+1} - \psi^n - \psi^{n-1})$$

and this is

$$A\phi^{n-1}(\phi^2 - \phi - 1) + B\psi^{n-1}(\psi^2 - \psi - 1) = 0.$$

□

Linear difference equations

As you know this parallels the theory of linear **differential** equations, at least in the case of constant coefficients.

Example A sequence is defined by

$$v_0 = 0$$

$$v_1 = 2$$

$$v_{n+1} = 4v_n - 3v_{n-1} \quad \text{for } n \geq 1.$$

Can we find an analogue of the Binet formula?

We test $v_n = r^n$. For this to work we need

$$r^{n+1} - 4r^n + 3r^{n-1} = 0$$

and unless $r = 0$ this says that

$$r^2 - 4r + 3 = 0$$

which implies that $r = 1$ or $r = 3$.

We now know that any sequence of the form $A + B3^n$ will satisfy the recurrence.

Can we choose A and B to satisfy the initial conditions $v_0 = 0$ and $v_1 = 2$?

$$A + B = 0$$

$$A + 3B = 2$$

Yes, $B = 1$ and $A = -1$. So $v_n = 3^n - 1$.

Example A sequence is defined by

$$v_0 = 0$$

$$v_1 = 1$$

$$v_{n+1} = 4v_n - 4v_{n-1} \quad \text{for } n \geq 1.$$

The auxiliary equation is

$$r^2 - 4r + 4 = 0$$

and this has a repeated root, $r = 2$. So we appear to have only one solution $v_n = 2^n$

Imagine that we chose an equation with two roots very close together $2 + d$ and 2 . Then

$$\frac{(2 + d)^n - 2^n}{d}$$

would be a solution.

As $d \rightarrow 0$ we get the derivative $n2^{n-1}$.

You can check that this is indeed a solution of the original recurrence:

$$v_{n+1} = 4v_n - 4v_{n-1}.$$

$$v_0 = 0$$

$$v_1 = 1$$

$$v_{n+1} = 4v_n - 4v_{n-1} \quad \text{for } n \geq 1.$$

The general solution is

$$v_n = A 2^n + B n 2^{n-1}$$

and to satisfy the initial conditions we need $A = 0$ and $B = 1$.

So the solution is

$$v_n = n 2^{n-1}.$$

The Continued fraction for the Golden Ratio

The Fibonacci numbers are given by

$$u_n = \frac{\phi^n - \psi^n}{\phi - \psi}$$

So the ratio of two successive terms is

$$\frac{u_{n+1}}{u_n} = \frac{\phi^{n+1} - \psi^{n+1}}{\phi^n - \psi^n}.$$

$\phi = 1.61803\dots$ while $\psi = -0.61803\dots$. So when n is large the ratio is very close to ϕ . This number is known as the Golden Ratio.

The sequence of approximations is

$$r_1 = \frac{1}{1}, \quad r_2 = \frac{2}{1}, \quad r_3 = \frac{3}{2}, \quad r_4 = \frac{5}{3}, \quad \dots$$

Observe that

$$r_n = \frac{u_{n+1}}{u_n} = \frac{u_n + u_{n-1}}{u_n} = 1 + \frac{u_{n-1}}{u_n} = 1 + \frac{1}{r_{n-1}}.$$

So we can build these approximations by truncating the continued fraction

$$1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \dots}}}}$$

It can be shown that every real number larger than 1 has a continued fraction expansion

$$a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \dots}}}}$$

where the a_i are positive integers. The Golden Ratio is the one for which all these integers are equal to 1 and this translates into a statement that ϕ is the most difficult number to approximate by fractions. This is the real reason that the Fibonacci numbers are important.

The Fibonacci matrices

Let Q be the matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Theorem (Fibonacci matrix theorem). *The powers of Q generate the Fibonacci numbers as follows:*

$$Q^n = \begin{pmatrix} u_{n+1} & u_n \\ u_n & u_{n-1} \end{pmatrix}.$$

This is closely related to the continued fraction. See the rabbits book.

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} u_{n+1} & u_n \\ u_n & u_{n-1} \end{pmatrix}.$$

Proof The formula is clear if $n = 1$ since $u_2 = u_1 = 1$ and $u_0 = 0$. The result will follow by induction if we check that

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u_{n+1} & u_n \\ u_n & u_{n-1} \end{pmatrix} = \begin{pmatrix} u_{n+2} & u_{n+1} \\ u_{n+1} & u_n \end{pmatrix}.$$

Multiplying out the left side gives

$$\begin{pmatrix} u_{n+1} + u_n & u_n + u_{n-1} \\ u_{n+1} & u_n \end{pmatrix}$$

and this is the right side. □

Corollary. For each n ,

$$u_{n+1}u_{n-1} - u_n^2 = (-1)^n.$$

Proof The left side is the determinant of Q^n while $\det Q = -1$.

$$\det(AB) = \det A \det B.$$

□

Now we move onto something that would be a fair bit tougher without matrices.

Theorem (Divisibility of Fibonacci numbers). If $m|n$ then $u_m|u_n$.

For example, $u_7 = 13$ and $u_{14} = 377 = 13 \times 29$.

Proof If $n = km$ then

$$\begin{pmatrix} u_{n+1} & u_n \\ u_n & u_{n-1} \end{pmatrix} = \begin{pmatrix} u_{m+1} & u_m \\ u_m & u_{m-1} \end{pmatrix}^k.$$

We can use induction on k if we check that whenever A, B, C are integers and we multiply

$$\begin{pmatrix} u_{m+1} & u_m \\ u_m & u_{m-1} \end{pmatrix} \begin{pmatrix} A & Bu_m \\ Bu_m & C \end{pmatrix}$$

we retain the property that the off-diagonal entries are divisible by u_m .

This is easily checked.

Alternatively, work with numbers modulo u_m . The matrix

$$\begin{pmatrix} u_{m+1} & u_m \\ u_m & u_{m-1} \end{pmatrix}$$

is congruent to a diagonal matrix so its powers are too. □

Aside for algebraicists: We are really building a representation of the field $\mathbb{Q}(\sqrt{5})$ on the 2×2 matrices over \mathbb{Q} :

$$a + b\sqrt{5} \mapsto \begin{pmatrix} a + b & 2b \\ 2b & a - b \end{pmatrix}.$$

The matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

is symmetric so its eigenvalues are real numbers.

The characteristic polynomial is

$$\det \begin{pmatrix} 1 - \lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - \lambda - 1.$$

The characteristic polynomial of the Fibonacci matrix is

$$\lambda^2 - \lambda - 1$$

so the eigenvalues of the matrix are ϕ and ψ .

We can diagonalise the matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = U^{-1} \begin{pmatrix} \phi & 0 \\ 0 & \psi \end{pmatrix} U$$

for some orthogonal matrix U .

Hence

$$\begin{pmatrix} u_{n+1} & u_n \\ u_n & u_{n-1} \end{pmatrix} = U^{-1} \begin{pmatrix} \phi^n & 0 \\ 0 & \psi^n \end{pmatrix} U$$

from which the Binet formula could be read off.

Theorem (Divisibility of Fibonacci numbers). *If $m|n$ then $u_m|u_n$.*

Suppose h is the highest common factor of m and n . Then $h|m$ and $h|n$ so $u_h|u_m$ and $u_h|u_n$. An obvious question: might it be true that u_h is the h.c.f. of u_m and u_n ? **Exotic!**

Recall that this means that if q divides u_m and u_n then it divides u_h .

By the Euclidean algorithm we know that there are integers a and b with $h = am + bn$. **The key fact we know about h .**

So

$$Q^h = (Q^m)^a (Q^n)^b.$$

$$\begin{pmatrix} u_{h+1} & u_h \\ u_h & u_{h-1} \end{pmatrix} = \begin{pmatrix} u_{m+1} & u_m \\ u_m & u_{m-1} \end{pmatrix}^a \begin{pmatrix} u_{n+1} & u_n \\ u_n & u_{n-1} \end{pmatrix}^b.$$

This appears to solve the problem since if q divides u_m and u_n then modulo q both matrices on the right are diagonal and so q divides u_h .

There is a slight problem because one of a and b will be negative. Suppose it's b . What we need is that

$$\begin{pmatrix} u_{n+1} & u_n \\ u_n & u_{n-1} \end{pmatrix}^{-1}$$

is a matrix with integer entries whose off diagonal entry is divisible by u_n .

Remember that the determinant of

$$\begin{pmatrix} u_{n+1} & u_n \\ u_n & u_{n-1} \end{pmatrix}$$

is $(-1)^n$ so its inverse is

$$(-1)^n \begin{pmatrix} u_{n-1} & -u_n \\ -u_n & u_{n+1} \end{pmatrix}$$

which is as we want.

Theorem (Highest common factor of Fibonacci numbers). *The highest common factor of u_m and u_n is the Fibonacci number u_h where $h = \text{hcf}(m, n)$.*

Summary

Theorem (Binet). *If $n \geq 0$ the Fibonacci number u_n is*

$$u_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right) = \frac{\phi^n - \psi^n}{\phi - \psi}$$

Theorem (Fibonacci matrix theorem). *The powers of Q generate the Fibonacci numbers as follows:*

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} u_{n+1} & u_n \\ u_n & u_{n-1} \end{pmatrix}.$$

Theorem (Divisibility of Fibonacci numbers). *If $m|n$ then $u_m|u_n$.*

Theorem (Highest common factor of Fibonacci numbers). *The highest common factor of u_m and u_n is the Fibonacci number u_h where $h = \text{hcf}(m, n)$.*

Chapter 4. Generating Functions and the Catalan Numbers

Given a sequence of numbers p_0, p_1, p_2, \dots it is often useful to encode or represent the sequence by a transform

$$g(x) = \sum_0^{\infty} p_k x^k.$$

If the sum has a positive radius of convergence we call the resulting function the **generating function** for the sequence.

The simplest examples are powers: if $p_k = 2^k$ then

$$g(x) = \sum_0^{\infty} (2x)^k = \frac{1}{1 - 2x}.$$

If two power series $f(x) = \sum a_n x^n$ and $g(x) = \sum b_n x^n$ have radius of convergence at least R then for $|x| < R$ we can differentiate

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$$

and multiply

$$\begin{aligned} f(x)g(x) &= (a_0 + a_1x + a_2x^2 + \dots)(b_0 + b_1x + b_2x^2 + \dots) \\ &= a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + \dots \end{aligned}$$

We have already seen one example of a generating function.

If $p_k = \binom{n}{k}$ then by the Binomial Theorem

$$g(x) = \sum_0^n p_k x^k = (1 + x)^n.$$

Since the Fibonacci numbers are given as a sum of two powers it is clear that we can compute their generating function as a rational function, but there is an easier way to do it.

$$\begin{aligned}(1 - x - x^2) \sum_{k=0}^{\infty} u_k x^k &= (1 - x - x^2)(0 + x + x^2 + 2x^3 + 3x^4 + 5x^5 + \dots) \\ &= 0 + x + 0x^2 + (2 - 1 - 1)x^3 + (3 - 2 - 1)x^4 + (5 - 3 - 2)x^5 + \dots \\ &= x.\end{aligned}$$

So the generating function is

$$\frac{x}{1 - x - x^2}.$$

The generating function for the multiset formula parallels the Binomial Theorem. At level m there are now arbitrarily large multisets using m symbols so the sum is an infinite one. The generating function is

$$g(x) = \sum_{d=0}^{\infty} \binom{d + m - 1}{m - 1} x^d$$

and I claim that this is

$$\frac{1}{(1 - x)^m}.$$

This is a standard argument from analysis.

What about a story?

Recall that the multiset formula tells us the number of monomials in m variables with total degree d .

By the geometric series formula

$$\frac{1}{1 - x_1} \frac{1}{1 - x_2} \cdots \frac{1}{1 - x_m}$$

$$= (1 + x_1 + x_1^2 + \cdots)(1 + x_2 + x_2^2 + \cdots) \cdots (1 + x_m + x_m^2 + \cdots).$$

$$\frac{1}{1-x_1} \frac{1}{1-x_2} \cdots \frac{1}{1-x_m}$$

$$= (1 + x_1 + x_1^2 + \cdots)(1 + x_2 + x_2^2 + \cdots) \cdots (1 + x_m + x_m^2 + \cdots).$$

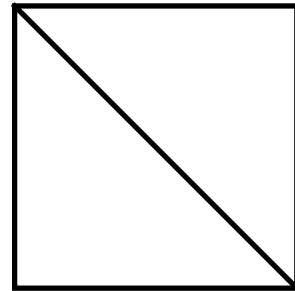
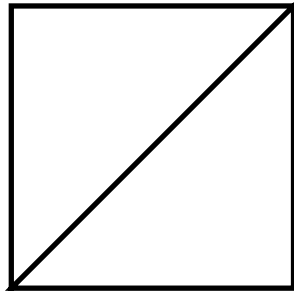
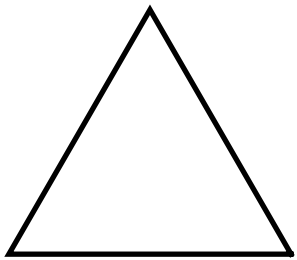
When you multiply out this product, collecting terms of the same total degree, you get all possible monomials in the m variables

$$1 + x_1 + x_2 + \cdots + x_m + x_1^2 + x_1x_2 + \cdots .$$

If you now set each x_i equal to x , the coefficient of x^d will be exactly the number of monomials of degree d .

Euler's dissection problem

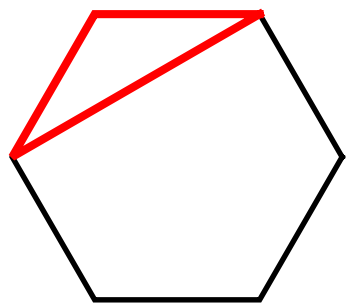
Find C_n : the number of ways to dissect a (regular) $(n + 2)$ -gon into n triangles (using $n - 1$ lines joining pairs of vertices).



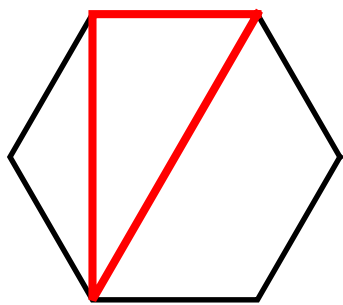
For $n = 1$ the triangle has only one way, $C_1=1$. We can cut the square with either diagonal so $C_2=2$. These numbers are called the **Catalan** numbers.

It is pretty difficult to calculate C_n directly. However there is a nice argument which expresses C_n in terms of earlier values of C .

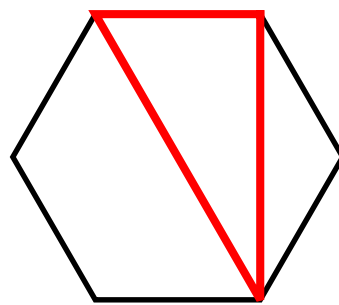
Consider a fixed edge of the $(n + 2)$ -gon. In each decomposition it belongs to some triangle, whose other vertex is one of the remaining n . Once this has been chosen the decomposition is obtained by decomposing each half of what's left. If $n = 4$:



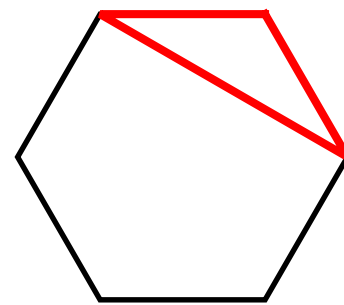
2 and $(n + 1)$



3 and n



.....



$(n + 1)$ and 2

This could be a 2-gon and an $(n + 1)$ -gon or a 3-gon and an n -gon and so on.

In the 3-gon/ n -gon case the number of ways to do it is C_1C_{n-2} , in the 4-gon/ $(n - 1)$ -gon case it is C_2C_{n-3} and so on.

At the ends we have a slightly different situation since we don't have to do anything to a 2-gon. In this case we are just cutting up an $(n + 1)$ -gon and the number of ways to do it is C_{n-1} . So we have proved

$$C_n = C_{n-1} + C_1C_{n-2} + C_2C_{n-3} + \cdots + C_{n-2}C_1 + C_{n-1}.$$

If we adopt the convention that $C_0 = 1$ we can write this in a neater form.

Theorem (Catalan Recurrence). *The Catalan numbers satisfy*

$$C_n = C_0C_{n-1} + C_1C_{n-2} + C_2C_{n-3} + \cdots + C_{n-2}C_1 + C_{n-1}C_0$$

for each n .

This looks like the formula for multiplying power series.

This inductive relationship allows us to calculate the first few C_n relatively quickly. It is still not so easy to see a general formula for the numbers.

$$C_0 = 1$$

$$C_1 = C_0^2 = 1$$

$$C_2 = 2C_0C_1 = 2$$

$$C_3 = 2C_0C_2 + C_1^2 = 5$$

$$C_4 = 2C_0C_3 + 2C_1C_2 = 14$$

Let's consider the generating function

$$g(x) = C_0 + C_1x + C_2x^2 + C_3x^3 + \dots .$$

Suppose that it has positive radius of convergence and consider $g(x)^2$.

$$\begin{aligned} g(x)^2 &= C_0^2 + (C_0C_1 + C_1C_0)x + (C_0C_2 + C_1C_1 + C_2C_0)x^2 + \dots \\ &= C_1 + C_2x + C_3x^2 + \dots = \frac{g(x) - C_0}{x} = \frac{g(x) - 1}{x}. \end{aligned}$$

If we write $g(x) = g$ then

$$xg^2 - g + 1 = 0.$$

$$g(x) = C_0 + C_1x + C_2x^2 + C_3x^3 + \dots .$$

If we write $g(x) = g$ then

$$xg^2 - g + 1 = 0.$$

We can solve to get

$$g(x) = \frac{1 \pm \sqrt{1 - 4x}}{2x}.$$

The function

$$\frac{1 + \sqrt{1 - 4x}}{2x}$$

is not given by a power series near 0 since it behaves like $1/x$.

The function

$$\frac{1 - \sqrt{1 - 4x}}{2x}$$

does have a power series with radius of convergence $1/4$ which starts

$$1 + x + \dots$$

The coefficients of this function **do** satisfy the recurrence so they are the C_n .

We can now calculate the numbers because we have a Binomial Theorem for fractional powers. We know how to differentiate $(1 - 4x)^{1/2}$.

We get

$$C_n = \frac{1}{n+1} \binom{2n}{n}.$$

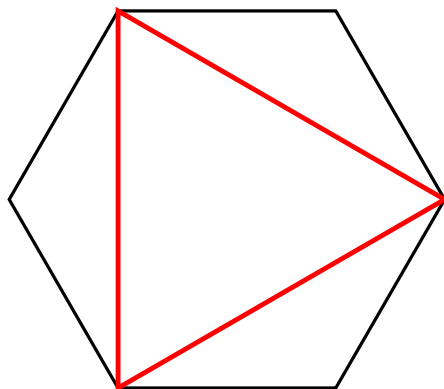
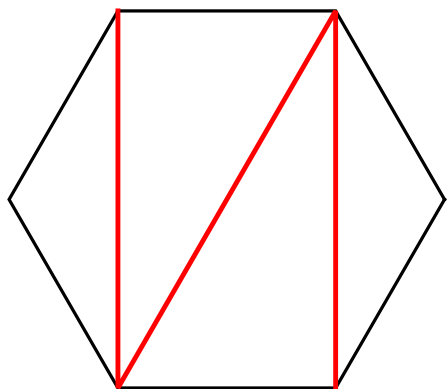
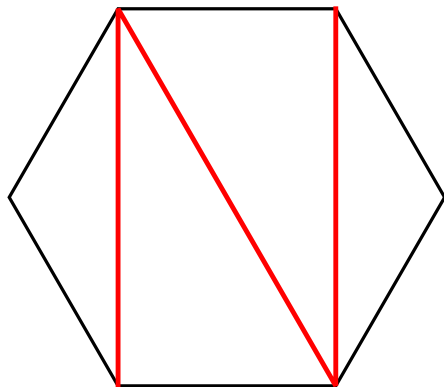
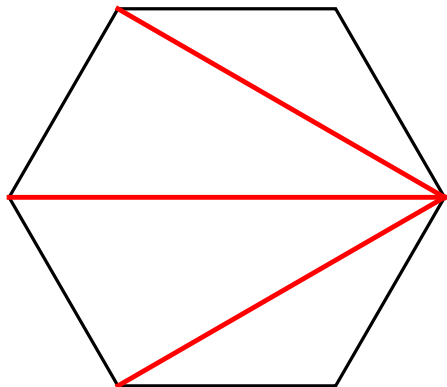
This will be HW.

Theorem (Euler's Dissection Problem). *The number of ways to dissect a regular $(n+2)$ -gon into n triangles using $n-1$ diagonals is the Catalan number*

$$C_n = \frac{1}{n+1} \binom{2n}{n}.$$

It isn't clear from the formula that the Catalan numbers are integers. It follows from the fact that they count dissections.

$$C_4 = 14.$$



Summary

The generating function for the Fibonacci numbers is

$$0 + x + x^2 + 2x^3 + 3x^4 + 5x^5 + \dots = \frac{x}{1 - x - x^2}.$$

The generating function for the multiset formula is

$$\sum_{d=0}^{\infty} \binom{d + m - 1}{m - 1} x^d = \frac{1}{(1 - x)^m}.$$

Theorem (Catalan Recurrence). *The Catalan numbers satisfy*

$$C_n = C_0 C_{n-1} + C_1 C_{n-2} + C_2 C_{n-3} + \dots + C_{n-2} C_1 + C_{n-1} C_0$$

for each n .

The generating function for the Catalan numbers

$$g(x) = C_0 + C_1x + C_2x^2 + C_3x^3 + \dots$$

is given by

$$\frac{1 - \sqrt{1 - 4x}}{2x}.$$

Theorem (Euler's Dissection Problem). *The number of ways to dissect a regular $(n + 2)$ -gon into n triangles using $n - 1$ diagonals is the Catalan number*

$$C_n = \frac{1}{n + 1} \binom{2n}{n}.$$

Chapter 4.8. The Poisson distribution

A random variable X is said to have a Poisson distribution with mean μ if it takes only the values 0, 1, 2 and so on, with probabilities

0	1	2	3	4
$e^{-\mu}$	$e^{-\mu} \mu$	$e^{-\mu} \frac{\mu^2}{2!}$	$e^{-\mu} \frac{\mu^3}{3!}$...

The expectation is

$$\begin{aligned} \sum_{k=0}^{\infty} k e^{-\mu} \frac{\mu^k}{k!} &= \sum_{k=1}^{\infty} e^{-\mu} \frac{\mu^k}{(k-1)!} \\ &= e^{-\mu} \mu \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} \\ &= e^{-\mu} \mu \sum_{j=0}^{\infty} \frac{\mu^j}{j!} = \mu. \end{aligned}$$

The Poisson phenomenon

Let A_1, A_2, \dots, A_n be independent events with probabilities p_1, p_2, \dots, p_n respectively. Let $K = \sum_i p_i$.

Then the expected number of events that occur is K .

The probability that no events occur is

$$(1 - p_1)(1 - p_2) \dots (1 - p_n).$$

If the p_i are all small then $1 - p_i \approx e^{-p_i}$ and so

$$(1 - p_1)(1 - p_2) \dots (1 - p_n) \approx \exp\left(-\sum_i p_i\right) = e^{-K}.$$

Chapter 5. Permutations, Partitions and the Stirling numbers

Partitions, the Stirling numbers of the second kind and Bell numbers

How many ways are there to partition $\{1, 2, \dots, n\}$ into k (non-empty) subsets?

Example $k = 1$. This is easy: one way.

Example $k = 2$. This is still fairly easy: you pick a non-empty subset and use that set and its complement as long as the complement is also non-empty. There are 2^n subsets and thus 2^{n-1} pairs. But we have to remove the pair that consists of \emptyset and $\{1, 2, \dots, n\}$. So the answer is $2^{n-1} - 1$.

Let $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ be the number of partitions of a set of n symbols into k non-empty subsets. These are called the Stirling numbers of the second kind.

It is not so easy to write down a formula for the Stirling numbers. We shall start with a recurrence

Theorem (Recurrence for Stirling II). *The Stirling numbers $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ satisfy the following recurrence*

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = k \left\{ \begin{matrix} n-1 \\ k \end{matrix} \right\} + \left\{ \begin{matrix} n-1 \\ k-1 \end{matrix} \right\}$$

for $n, k \geq 1$.

Proof We shall break up the partitions according to whether the singleton $\{n\}$ is one of the subsets.

Each partition which includes the singleton is obtained by partitioning the remaining $n - 1$ numbers into $k - 1$ non-empty sets and then adding the singleton as the k^{th} subset.

Each partition in which $\{n\}$ is **not** one of the subsets can be obtained in an unique way as follows: partition the remaining $n - 1$ numbers into k nonempty subsets and then throw n into one of those k subsets. Each of the lower level partitions is used k times.

Consequently

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = k \left\{ \begin{matrix} n - 1 \\ k \end{matrix} \right\} + \left\{ \begin{matrix} n - 1 \\ k - 1 \end{matrix} \right\}.$$

□

This makes it easy to calculate values level by level.

$$\binom{n}{k} = k \binom{n-1}{k} + \binom{n-1}{k-1}.$$

	$k = 0$	1	2	3	4	5
$n = 1$	0	1				
2	0	1	1			
3	0	1	3	1		
4	0	1	7	6	1	
5	0	1	15	25	10	1

From the recurrence we can produce a formula for the numbers.

Theorem (The Stirling numbers of the second kind). For each $n \geq 1$ and $k \geq 0$

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{j=0}^k \binom{k}{j} (-1)^j (k-j)^n = \frac{1}{k!} \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} j^n.$$

Notice that the expression on the right is 0 if $k > n$ by the orthogonality property of the binomial coefficients.

Proof The statement is easy to check if $n = 1$. Assume inductively that it holds with n replaced by $n - 1$.

Look at the expression with exponent n :

$$\frac{1}{k!} \sum_{j=0}^k \binom{k}{j} (-1)^j (k-j)^n = \sum_{j=0}^k \frac{1}{j!(k-j)!} (-1)^j (k-j)^n.$$

We want to show that

$$\sum_{j=0}^k \frac{1}{j!(k-j)!} (-1)^j (k-j)^n = \left\{ \begin{matrix} n \\ k \end{matrix} \right\}.$$

We can split the expression into two using one factor of $k-j$:

$$\begin{aligned} & \sum_{j=0}^k \frac{1}{j!(k-j)!} (-1)^j (k-j)^n \\ &= k \sum_{j=0}^k \frac{1}{j!(k-j)!} (-1)^j (k-j)^{n-1} - \sum_{j=0}^k \frac{j}{j!(k-j)!} (-1)^j (k-j)^{n-1} \end{aligned}$$

By the inductive hypothesis the first term is

$$k \left\{ \begin{matrix} n-1 \\ k \end{matrix} \right\}.$$

So to complete the inductive step we need to check that the second term (with its negative sign) is

$$\left\{ \begin{matrix} n - 1 \\ k - 1 \end{matrix} \right\}.$$

Now in view of the factor j in the numerator we may start the sum from $j = 1$ so we want to show that

$$- \sum_{j=1}^k \frac{j}{j!(k-j)!} (-1)^j (k-j)^{n-1} = \left\{ \begin{matrix} n - 1 \\ k - 1 \end{matrix} \right\}.$$

But

$$\begin{aligned} & - \sum_{j=1}^k \frac{j}{j!(k-j)!} (-1)^j (k-j)^{n-1} \\ &= \sum_{j=1}^k \frac{1}{(j-1)!(k-j)!} (-1)^{j-1} (k-j)^{n-1} \\ &= \sum_{j=1}^k \frac{(-1)^{j-1}}{(j-1)!(k-1-(j-1))!} (k-1-(j-1))^{n-1} \\ &= \sum_{r=0}^{k-1} \frac{1}{r!(k-1-r)!} (-1)^r (k-1-r)^{n-1} \\ &= \left\{ \begin{matrix} n-1 \\ k-1 \end{matrix} \right\}. \end{aligned}$$

□

The Bell numbers

It is natural to ask what can be said about the **total** number of partitions of $\{1, 2, \dots, n\}$.

The n^{th} Bell number B_n is the total number of partitions of a set of size n :

$$B_n = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\}.$$

These numbers were actually studied in medieval Japan 500 years before E. T. Bell.

We have a formula for the Stirling numbers so we can produce a formula for the Bell numbers but it involves a double sum and is a bit hard to analyse.

By interchanging the order of summation we can produce a much nicer formula.

Theorem (The Bell numbers). *The Bell numbers are given by*

$$B_n = e^{-1} \sum_{k=0}^{\infty} \frac{k^n}{k!}.$$

If X is Poisson random variable with mean 1 then the expected value of X^n is

$$\sum_{k=0}^{\infty} k^n p_k = e^{-1} \sum_{k=0}^{\infty} \frac{k^n}{k!}.$$

So the Bell numbers are the moments of a Poisson r.v. with mean 1.

Proof Observe that because of the symmetry of the binomial coefficients we may switch round the formula for the Stirling numbers

$$j \mapsto k - j$$

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{j=0}^k \binom{k}{j} (-1)^j (k-j)^n = \frac{1}{k!} \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} j^n.$$

Now

$$B_n = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\}$$

but we can continue the sum to infinity since the numbers are 0 if $k > n$.

$$B_n = \sum_{k=0}^{\infty} \left\{ \begin{matrix} n \\ k \end{matrix} \right\}.$$

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} j^n \quad (3)$$

and

$$B_n = \sum_{k=0}^{\infty} \left\{ \begin{matrix} n \\ k \end{matrix} \right\}.$$

Notice that the expression in (3) is 0 if $k > n$ by the orthogonality property of the binomial coefficients.

Thus

$$B_n = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} j^n = \sum_{k=0}^{\infty} \sum_{j=0}^k \frac{1}{j!(k-j)!} (-1)^{k-j} j^n.$$

$$B_n = \sum_{k=0}^{\infty} \sum_{j=0}^k \frac{1}{j!(k-j)!} (-1)^{k-j} j^n.$$

Since the sum is absolutely convergent we may interchange the order to get

$$B_n = \sum_{j=0}^{\infty} \frac{j^n}{j!} \sum_{k=j}^{\infty} \frac{(-1)^{k-j}}{(k-j)!} = \sum_{j=0}^{\infty} \frac{j^n}{j!} \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} = e^{-1} \sum_{j=0}^{\infty} \frac{j^n}{j!}.$$

□

Finally, we can obtain a generating function for the numbers $B_n/n!$: sometimes called an exponential generating function for the B_n .

Theorem (Exponential generating function for Bell).

$$\sum_{n=0}^{\infty} \frac{B_n}{n!} x^n = \exp(e^x - 1).$$

Proof

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n &= e^{-1} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{j=0}^{\infty} \frac{j^n}{j!} x^n \\ &= e^{-1} \sum_{j=0}^{\infty} \frac{1}{j!} \sum_{n=0}^{\infty} \frac{j^n}{n!} x^n && \text{Interchanging summation order} \\ &= e^{-1} \sum_{j=0}^{\infty} \frac{1}{j!} e^{jx} \\ &= e^{-1} \exp(e^x). \end{aligned}$$

From this we can calculate a few values:

n	0	1	2	3	4	5	6
B_n	1	1	2	5	15	52	203

We actually adopted the convention that $B_0 = 1$ without really noticing, because we used the 0^{th} moment of the Poisson.

Permutations and the Stirling numbers of the first kind

Recall that each permutation of n symbols can be written uniquely (except for trivial cycles) as a product of disjoint cycles.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 5 & 6 & 3 & 7 & 1 & 2 \end{pmatrix}$$

We track a symbol under repeated application

$$1 \rightarrow 4 \rightarrow 3 \rightarrow 6 \rightarrow 1$$

$$2 \rightarrow 5 \rightarrow 7 \rightarrow 2$$

so the permutation is the product

$$(1436)(257)$$

The cycle type of the permutation is important since it specifies the conjugacy class.

How many permutations on n symbols use k cycles?

We have to make a choice about cycles of length 1. We include them. So the identity is

$$(1)(2)(3)\dots(n).$$

Similarly

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 4 & 5 & 6 & 3 & 7 & 1 & 2 & 8 \end{pmatrix}$$

is

$$(1436)(257)(8).$$

Let $\left[\begin{matrix} n \\ k \end{matrix} \right]$ be the number of permutations of n symbols which are the product of k disjoint cycles.

A permutation on n symbols can have at most n cycles and if it does then it is the identity. So

$$\left[\begin{matrix} n \\ n \end{matrix} \right] = 1$$

and if $k > n$

$$\left[\begin{matrix} n \\ k \end{matrix} \right] = 0.$$

A permutation with only one cycle is a cycle of length n . There are $n!$ ways to order the symbols: each n -cycle appears n times in the different orderings so

$$\begin{bmatrix} n \\ 1 \end{bmatrix} = (n - 1)!$$

It is not easy to count permutations with a given number of cycles but we can find a recurrence.

Theorem (Recurrence for Stirling I). *The Stirling numbers $\begin{bmatrix} n \\ k \end{bmatrix}$ satisfy the following recurrence*

$$\begin{bmatrix} n \\ k \end{bmatrix} = (n - 1) \begin{bmatrix} n - 1 \\ k \end{bmatrix} + \begin{bmatrix} n - 1 \\ k - 1 \end{bmatrix}$$

for $n \geq 1$ and $k \geq 1$.

Proof We divide the permutations on n symbols with k cycles into two groups: those which fix n and those in which n appears in a longer cycle.

The permutations that fix n are all obtained by choosing a permutation of the first $n - 1$ symbols which contains $k - 1$ cycles and appending the trivial cycle (n) .

The permutations which do not fix n can be obtained in a unique way as follows. Choose a permutation of the first $n - 1$ symbols which contains k cycles and now insert the symbol n into the appropriate place in one of the cycles. In a cycle of length r there are r possible locations to insert n . So for each permutation of length $n - 1$ there are $n - 1$ possible permutations of length n arising from it.

Consequently the permutations of n symbols with k cycles can be paired off with one copy of the permutations of $n - 1$ symbols with $k - 1$ cycles; and $n - 1$ copies of the permutations of $n - 1$ symbols with k cycles.

So

$$\begin{bmatrix} n \\ k \end{bmatrix} = (n - 1) \begin{bmatrix} n - 1 \\ k \end{bmatrix} + \begin{bmatrix} n - 1 \\ k - 1 \end{bmatrix}.$$

□

This makes it easy to calculate values level by level.

$$\begin{bmatrix} n \\ k \end{bmatrix} = (n-1) \begin{bmatrix} n-1 \\ k \end{bmatrix} + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}.$$

	$k = 0$	1	2	3	4
$n = 1$	0	1			
2	0	1	1		
3	0	2	3	1	
4	0	6	11	6	1

From the recurrence we can immediately produce a generating function.

Theorem (The Stirling numbers of the first kind). For each n

$$\sum_{k=1}^n \left[\begin{matrix} n \\ k \end{matrix} \right] x^k = x(x+1)(x+2)\dots(x+n-1).$$

The numbers $\left[\begin{matrix} n \\ k \end{matrix} \right]$ are catchily named the “unsigned Stirling numbers of the first kind”.

Proof It is immediate that the formula holds for $n = 1$. $\left[\begin{matrix} 1 \\ 1 \end{matrix} \right] = 1$ and
if $k \neq 1$ $\left[\begin{matrix} 1 \\ k \end{matrix} \right] = 0$.

To establish the general case we use induction on n .

Assume inductively that

$$\sum_{k=1}^{n-1} \binom{n-1}{k} x^k = x(x+1)\dots(x+n-2)$$

and consider

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

We want to show that it is

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

Now

$$\begin{aligned} & (x + n - 1) \sum_{k=1}^{n-1} \binom{n-1}{k} x^k \\ &= (n-1) \sum_{k=1}^{n-1} \binom{n-1}{k} x^k + \sum_{k=1}^{n-1} \binom{n-1}{k} x^{k+1} \\ &= (n-1) \sum_{k=1}^n \binom{n-1}{k} x^k + \sum_{j=2}^n \binom{n-1}{j-1} x^j \\ &= (n-1) \sum_{k=1}^n \binom{n-1}{k} x^k + \sum_{j=1}^n \binom{n-1}{j-1} x^j \end{aligned}$$

since $\binom{n-1}{n} = \binom{n-1}{0} = 0$.

Hence

$$\begin{aligned} & (x + n - 1) \sum_{k=1}^{n-1} \begin{bmatrix} n - 1 \\ k \end{bmatrix} x^k \\ &= \sum_{k=1}^n \left((n - 1) \begin{bmatrix} n - 1 \\ k \end{bmatrix} + \begin{bmatrix} n - 1 \\ k - 1 \end{bmatrix} \right) x^k \\ &= \sum_{k=1}^n \begin{bmatrix} n \\ k \end{bmatrix} x^k \end{aligned}$$

by the recurrence relation.

This completes the inductive step. □

Summary

The Stirling number of the second kind $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ is the number of partitions of a set of n symbols into k non-empty subsets.

Theorem (Recurrence for Stirling II). *The Stirling numbers $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ satisfy the following recurrence*

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = k \left\{ \begin{matrix} n-1 \\ k \end{matrix} \right\} + \left\{ \begin{matrix} n-1 \\ k-1 \end{matrix} \right\}$$

for $n, k \geq 1$.

Theorem (The Stirling numbers of the second kind). *For each $n \geq 1$ and $k \geq 0$*

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{j=0}^k \binom{k}{j} (-1)^j (k-j)^n.$$

The n^{th} Bell number B_n is the total number of partitions of a set of size n :

$$B_n = \sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\}.$$

Theorem (The Bell numbers). *The Bell numbers are given by*

$$B_n = e^{-1} \sum_{k=0}^{\infty} \frac{k^n}{k!}.$$

Theorem (Exponential generating function for Bell).

$$\sum_{n=0}^{\infty} \frac{B_n}{n!} x^n = \exp(e^x - 1).$$

The unsigned Stirling number of the first kind $\left[\begin{matrix} n \\ k \end{matrix} \right]$ is the number of permutations of n symbols which are the product of k disjoint cycles.

Theorem (Recurrence for Stirling I). *The Stirling numbers $\left[\begin{matrix} n \\ k \end{matrix} \right]$ satisfy the following recurrence*

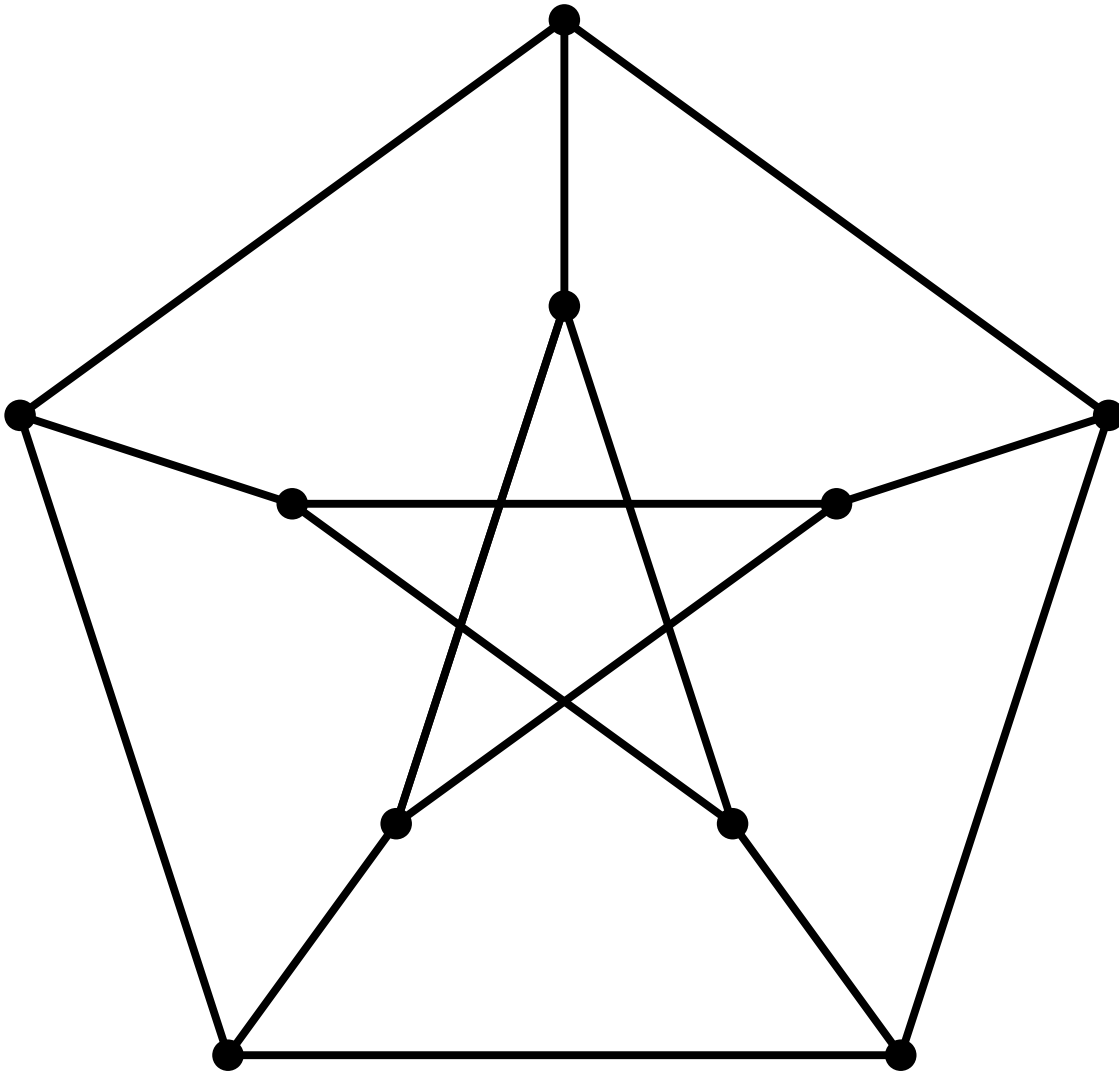
$$\left[\begin{matrix} n \\ k \end{matrix} \right] = (n - 1) \left[\begin{matrix} n - 1 \\ k \end{matrix} \right] + \left[\begin{matrix} n - 1 \\ k - 1 \end{matrix} \right]$$

for $n \geq 1$ and $k \geq 1$.

Theorem (The Stirling numbers of the first kind). *For each n*

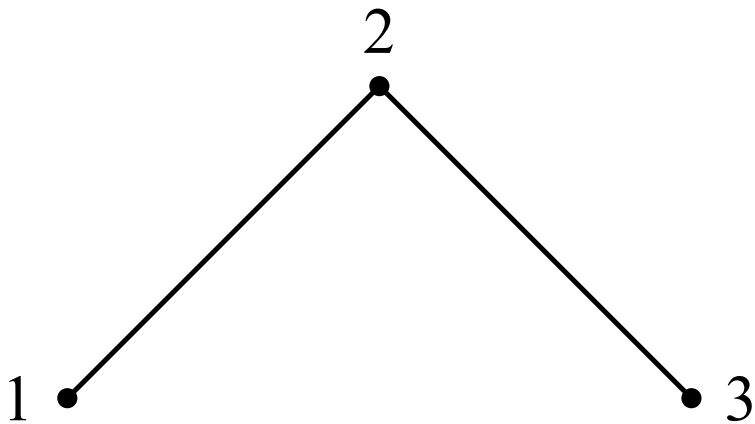
$$\sum_{k=1}^n \left[\begin{matrix} n \\ k \end{matrix} \right] x^k = x(x + 1)(x + 2) \dots (x + n - 1).$$

Volume II. Graph Theory



Chapter 6. Basic theory: Euler trails and circuits

A graph G is a collection of **vertices** $V = \{v_1, v_2, \dots, v_n\}$ together with a set of **edges** E each of which is a pair of vertices. For example if $V = \{1, 2, 3\}$ and $E = \{\{1, 2\}, \{2, 3\}\}$ we get

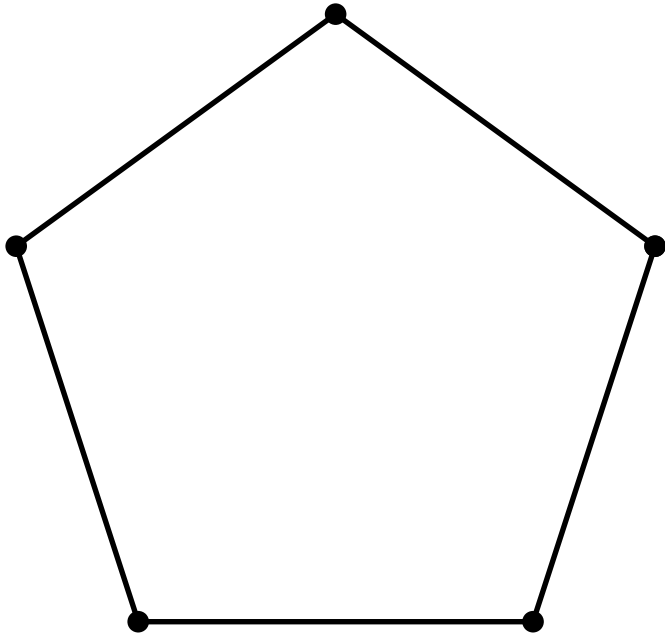


We don't allow loops, multiple edges or directed edges unless we explicitly say so. A number of special graphs turn up a lot:

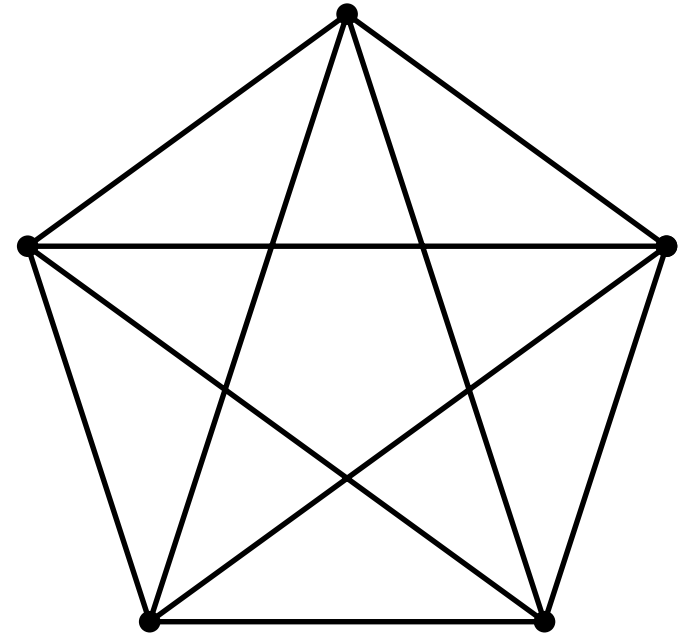
The **path** P_n of length n which has $n + 1$ vertices:



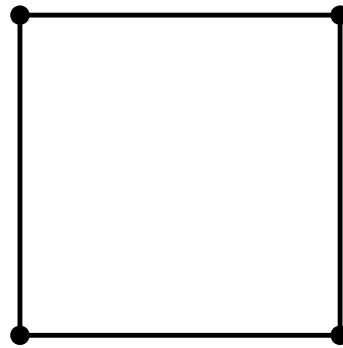
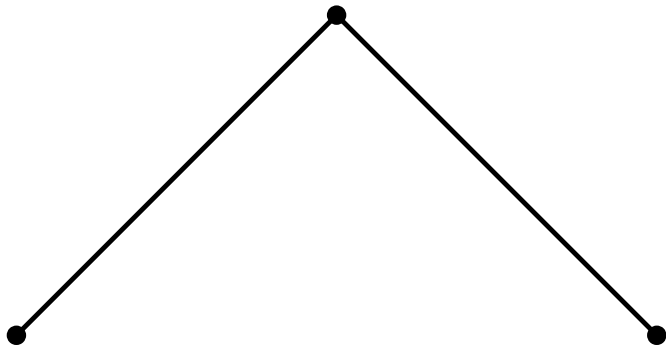
The **cycle** C_n of length n :



The **complete graph** K_n :



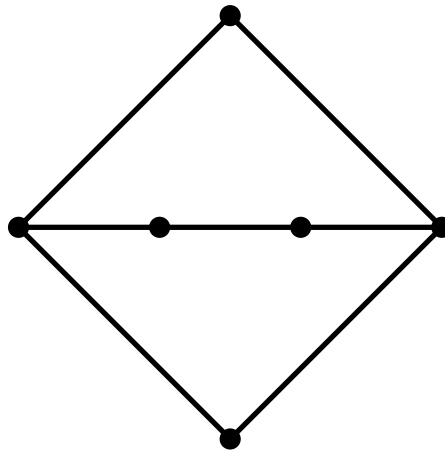
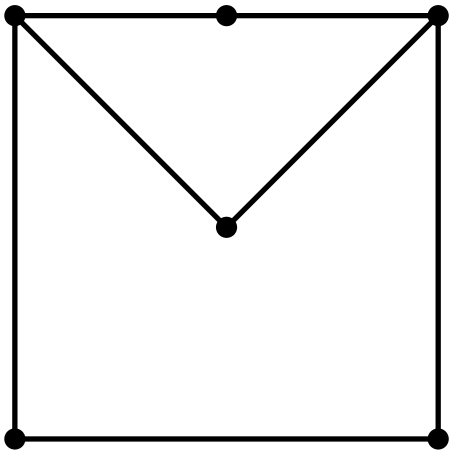
A graph G is said to be **connected** if there is a path in G between any pair of vertices. Each graph can be decomposed into connected components: maximal connected subgraphs. **HW**



This is a graph

It has 4 components.

Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are isomorphic if there is a bijection ϕ from V_1 to V_2 with the property that $\{x, y\} \in E_1$ if and only if $\{\phi(x), \phi(y)\} \in E_2$.

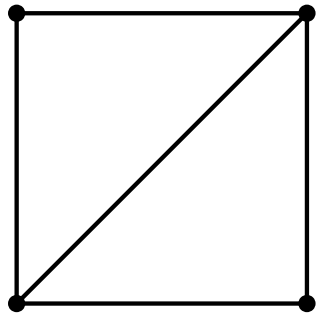


If $G = (V, E)$ is a graph then a **subgraph** of G is a graph (V', E') where $V' \subset V$ and $E' \subset E$ (and the elements of E' are pairs of elements of V').

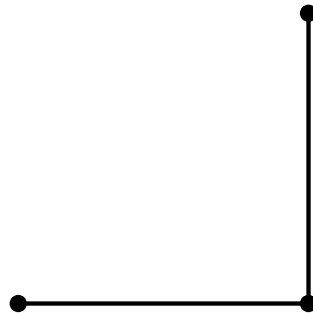
Note that we do not necessarily include all the edges of G that connect vertices in V' .

An **induced** subgraph of (V, E) is a graph (V', E') where $V' \subset V$ and E' consists of all pairs in E that are subsets of V' : we include all the edges of G that we can.

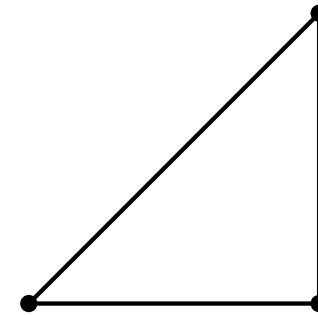
G



Subgraph



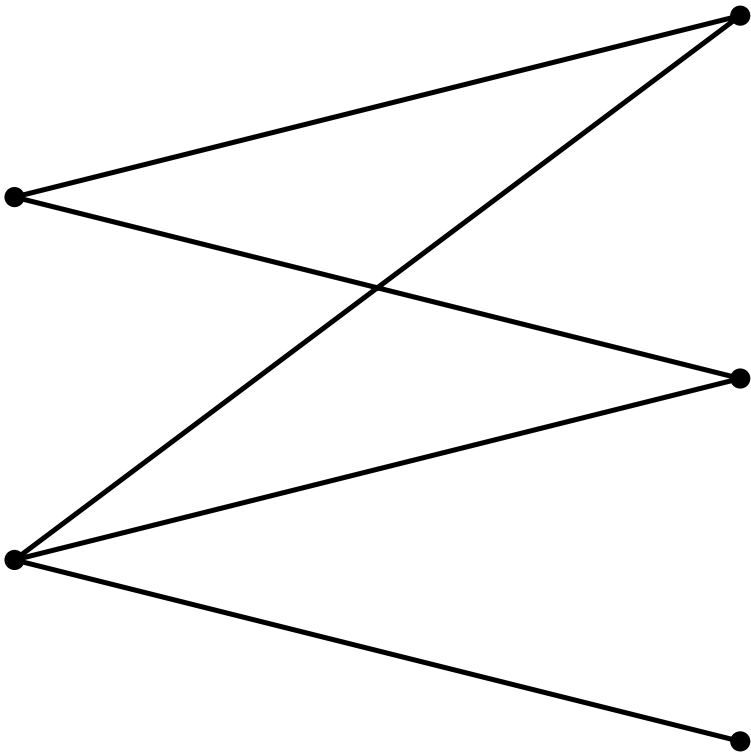
Induced



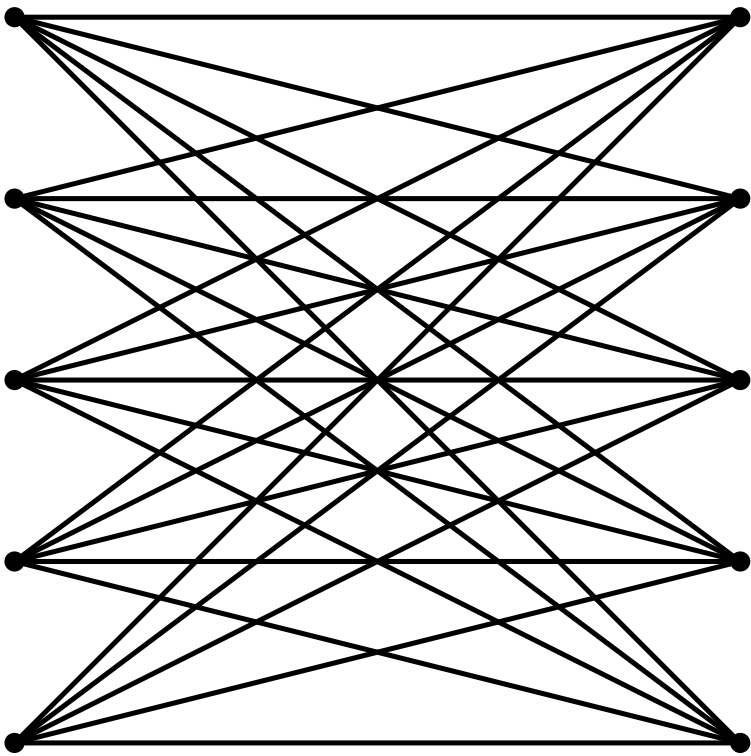
A **walk** in a graph is a sequence of vertices, each one adjacent to the next, possibly with repetition. It is **closed** if its first and last vertices are the same.

A **path** is a walk which uses distinct vertices. A **cycle** is a closed walk which uses distinct vertices except at the ends.

A graph is **bipartite** if its vertex set can be partitioned into two parts A and B in such a way that all edges cross from A to B : (none is inside either part).



A further example that turns up frequently is the **complete bipartite graph** K_{mn} :



There is a simple characterisation of bipartite graphs which is quite easy to prove. It is not a powerful tool but it is instructive to know.

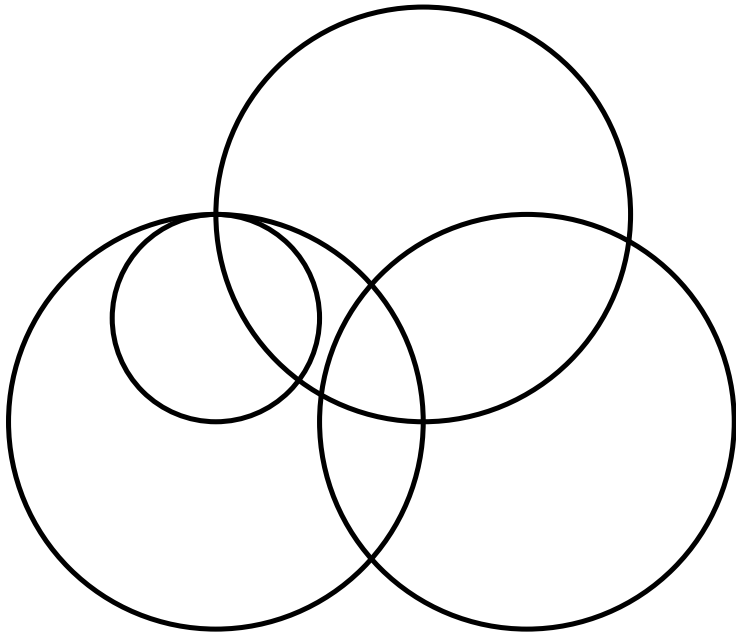
Theorem (Characterisation of bipartite graphs). *A graph is bipartite if and only if it contains no odd cycles.*

It is easy to see that the condition is necessary. A cycle must cross from one vertex class to the other or back an even number of times because it ends where it starts. So a bipartite graph cannot contain odd cycles.

The other direction is a slightly trickier.

To make the proof clearer we start with a lemma.

Lemma (Odd walk lemma). *If a graph contains a closed walk of odd length then it contains a cycle of odd length.*



This is not too hard to prove but it is slightly harder than you might think.

Theorem (Characterisation of bipartite graphs). *A graph is bipartite if and only if it contains no odd cycles.*

Proof If the graph is bipartite then each cycle must cross from one part to the other or back an even number of times, since it ends where it starts. So each cycle has an even number of edges.

On the other hand suppose there are no odd cycles. We may assume that the graph is connected, otherwise we handle each component separately. Pick a vertex s and for each vertex v consider the shortest path from s to v . If this shortest path has even length then put v into part A but if it has odd length then put v into part B . We need to check that all edges cross between the two parts.

We need to check that all edges cross between the two parts. Suppose on the contrary that there are two adjacent vertices in part A , v_1 and v_2 . We can form a closed walk starting at v_1 , walking to s along the path of even length, walking back to v_2 along the path of even length and then stepping from v_2 back to v_1 . This walk has **odd** length so by the lemma the graph contains an odd cycle.

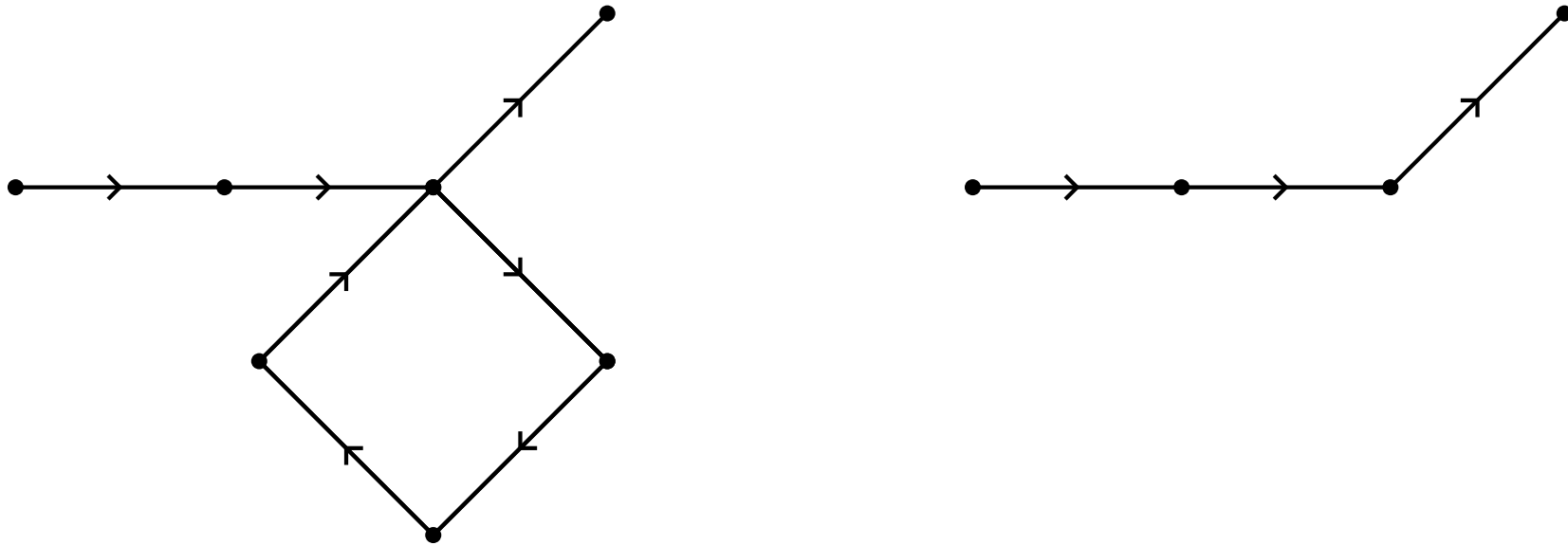
In the same way, there would be an odd cycle if there were adjacent vertices in part B . □

Lemma (Odd walk lemma). *If a graph contains a closed walk of odd length then it contains a cycle of odd length.*

Proof Let us use the word “2-cycle” to mean a walk of the form xyx in which we walk along an edge and immediately back again.

Pick a vertex on the walk and start walking. Eventually you hit a vertex you have already visited. The first time you do so, you have completed a cycle or a “2-cycle”. If this cycle is odd then you have finished. If not you can discard it from the walk and leave a shorter closed walk of odd length.

If this cycle is odd then you have finished. If not you can discard it from the walk and leave a shorter closed walk of odd length.

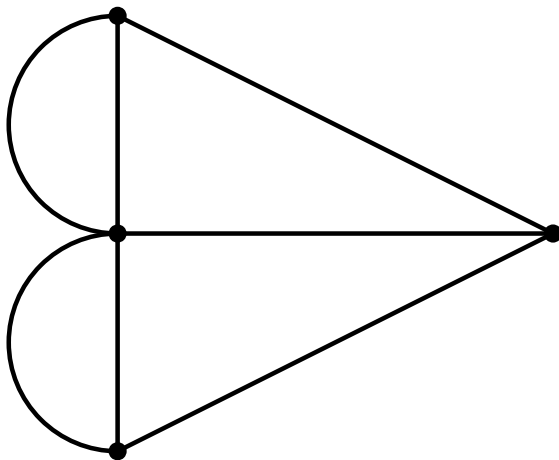


This process cannot continue indefinitely so at some point we form an odd cycle. □

Euler trails and circuits

In 1735 Euler posed a problem which became famous as the beginning of graph theory: the bridges of Königsberg problem. He asked whether it is possible to make a tour of the city, crossing each of its bridges exactly once.

The problem is equivalent to the following: is it possible to walk from vertex to vertex in the following **multigraph**, traversing each edge exactly once? A multigraph is like a graph but with multiple edges.



A walk in which all the edges are distinct is called a **trail**: if it closed it is called a **circuit**.

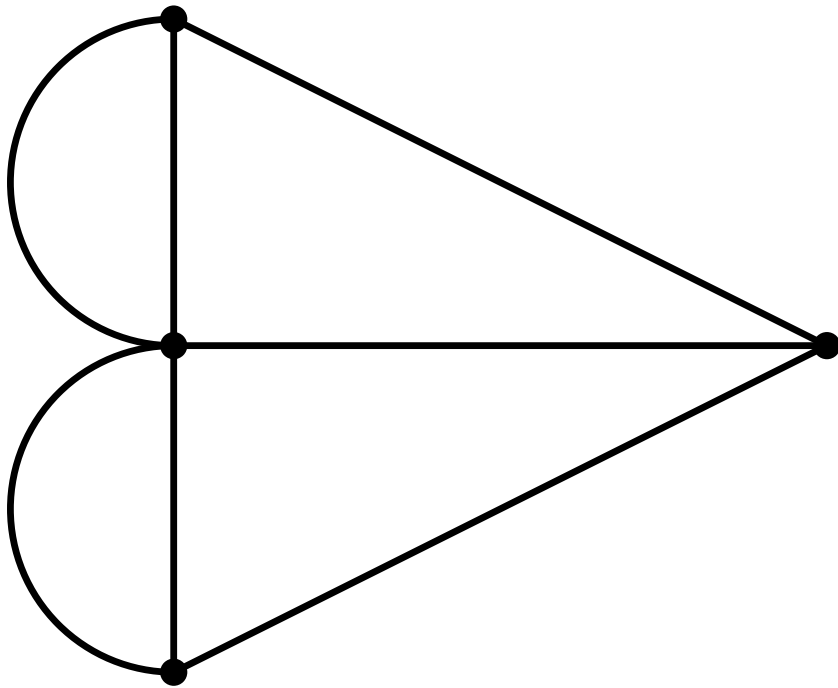
Unlike a path or a cycle a trail and a circuit can revisit vertices: but they do not use the same edge twice.

Euler pointed out that a trail enters and leaves a vertex every time we visit that vertex, unless the vertex is the start or end of the walk. Since the trail he asked for uses all the edges exactly once, we conclude that all vertices other than the start and finish have an even number of edges coming out of them.

Let us say that for a graph (or multigraph) G and a vertex v of G the **degree** of v is the number of edges containing v .

Euler's remark is that if his walking tour exists, then all but two of the vertices must have even degree.

The original Königsberg multigraph is



Since in fact the degrees are 3, 3, 3 and 5 there can be no such tour.

It turns out that for a general graph, Euler's condition is sufficient for the trail to exist as well as being necessary. To begin with let's make a small observation which helps us to fix ideas.

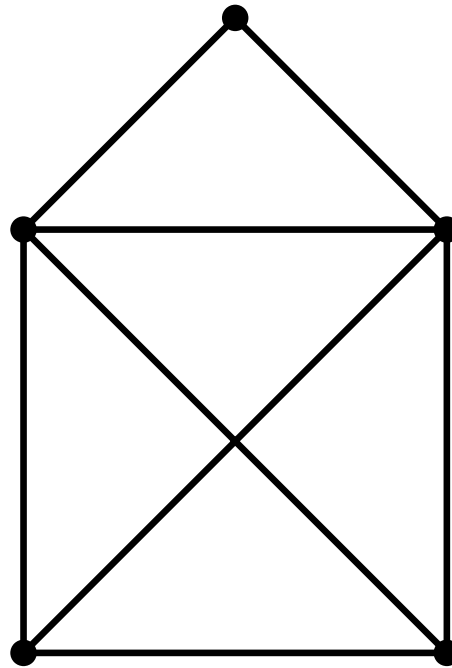
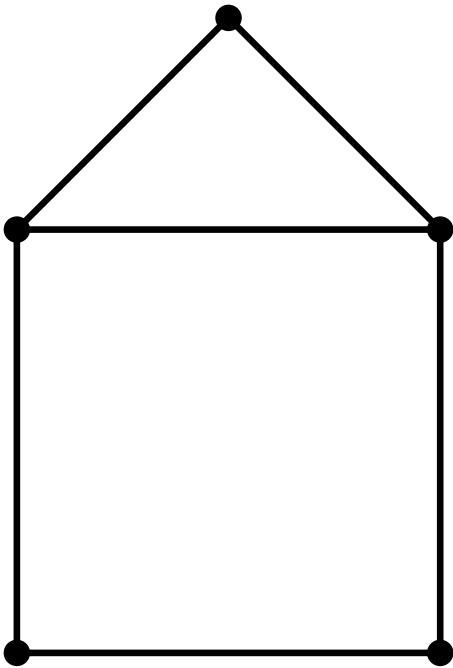
Lemma (Handshaking lemma). *The number of vertices of odd degree in a graph is even.*

Proof HW

□

Theorem (Euler circuits). A connected graph G (or multigraph) has an Euler trail if and only if it has just two vertices of odd degree, and an Euler circuit if and only if it has none.

Example The house and the tudor house.



Theorem (Euler circuits). *A connected graph G (or multigraph) has an Euler trail if and only if it has just two vertices of odd degree, and an Euler circuit if and only if it has none.*

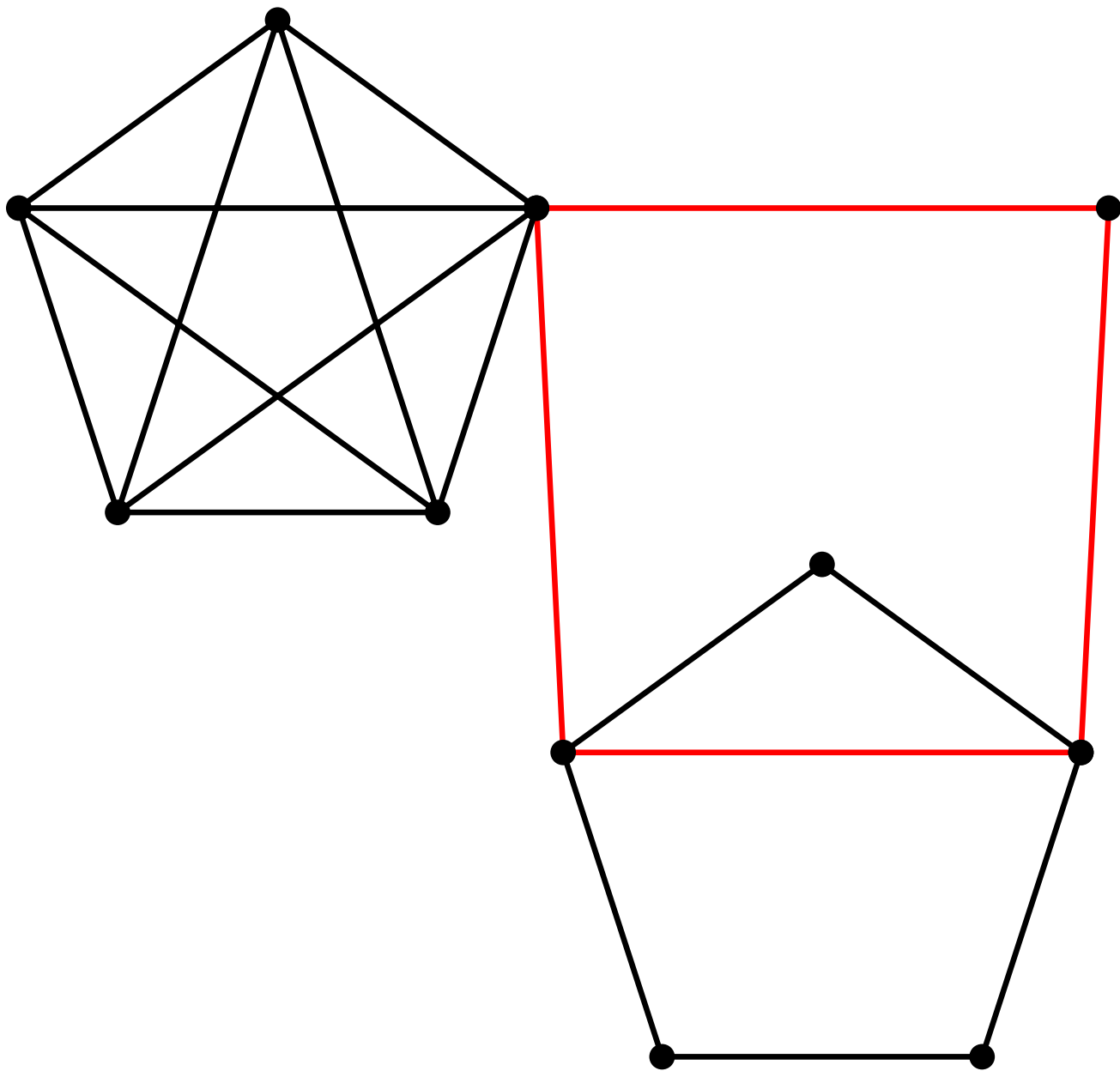
Proof If the graph has an Euler trail then it cannot have more than two vertices of odd degree because the trail leaves each vertex as many times as it enters, except for the vertices at the start and finish. Similarly if the graph has an Euler circuit.

In the other direction we shall begin with the Euler circuit. To prove that the condition is sufficient we use induction on the number of edges. If there are no edges there is nothing to prove.

So suppose that G is connected, that every vertex has even degree and that G is more than just one vertex. Then every vertex of the graph has degree at least 2. I claim that we can find a cycle in G .

Start walking from a vertex. Until you hit a vertex that you have already visited there is always a way to continue. When you **do** hit a vertex for the second time you have completed a cycle: call it C .

Once we remove the edges of C we might disconnect the graph. Our original graph is built from components of the new graph linked together by the cycle C .



Each component of the new graph has fewer edges than G and every vertex in it has even degree. So each one has an Euler circuit or is a single vertex. Moreover each component contains a vertex of C which must be visited by its Euler circuit since the component is connected: so in each component the Euler circuit visits a vertex of C .

Each component of the new graph has an Euler circuit that visits a vertex of C .

So we can string together all the smaller Euler circuits and the cycle C to make an Euler circuit for G .

Finally if we have two vertices x and y with odd degree we can join them through a new vertex u . We then find an Euler circuit in the new graph and delete u to leave an Euler trail from x to y . \square

Summary

A graph G is a collection of **vertices** $V = \{v_1, v_2, \dots, v_n\}$ together with a set of **edges** E each of which is a pair of vertices.

A graph G is said to be **connected** if there is a path in G between any pair of vertices. Each graph can be decomposed into connected components.

Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are isomorphic if there is a bijection ϕ from V_1 to V_2 with the property that $\{x, y\} \in E_1$ if and only if $\{\phi(x), \phi(y)\} \in E_2$.

If $G = (V, E)$ is a graph then a **subgraph** of G is a graph (V', E') where $V' \subset V$ and $E' \subset E$ (and the elements of E' are pairs of elements of V').

An **induced** subgraph of (V, E) is a graph (V', E') where $V' \subset V$ and E' consists of all pairs in E that are subsets of V' : we include all the edges of G that we can.

A **walk** in a graph is a sequence of vertices, each one adjacent to the next, possibly with repetition. It is **closed** if its first and last vertices are the same.

A **path** is a walk which uses distinct vertices. A **cycle** is a closed walk which uses distinct vertices except at the ends.

A graph is **bipartite** if its vertex set can be partitioned into two parts A and B in such a way that all edges cross from A to B : (none is inside either part).

Theorem (Characterisation of bipartite graphs). *A graph is bipartite if and only if it contains no odd cycles.*

Lemma (Odd walk lemma). *If a graph contains a closed walk of odd length then it contains a cycle of odd length.*

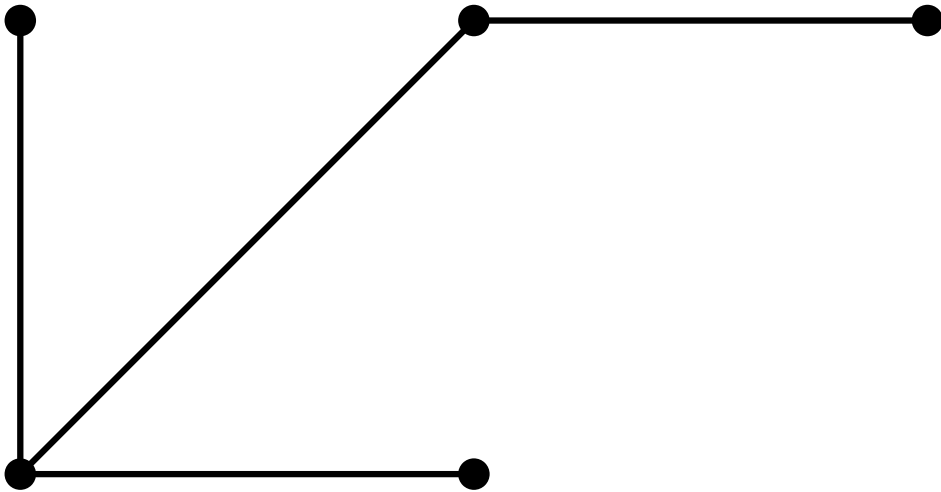
A walk in which all the edges are distinct is called a **trail**: if it closed it is called a **circuit**.

Lemma (Handshaking lemma). *The number of vertices of odd degree in a graph is even.*

Theorem (Euler circuits). *A connected graph G (or multigraph) has an Euler trail if and only if it has just two vertices of odd degree, and an Euler circuit if and only if it has none.*

Chapter 7. Trees, spanning trees and cycles

A **tree** is a graph that is connected but contains no cycle. For example



If a graph is connected and you add another edge you will automatically have a cycle: because there was already a path between the two ends of the new edge.

If a graph is acyclic and you remove an edge then the result is disconnected: if it were connected then adding back the edge you removed would create a cycle.

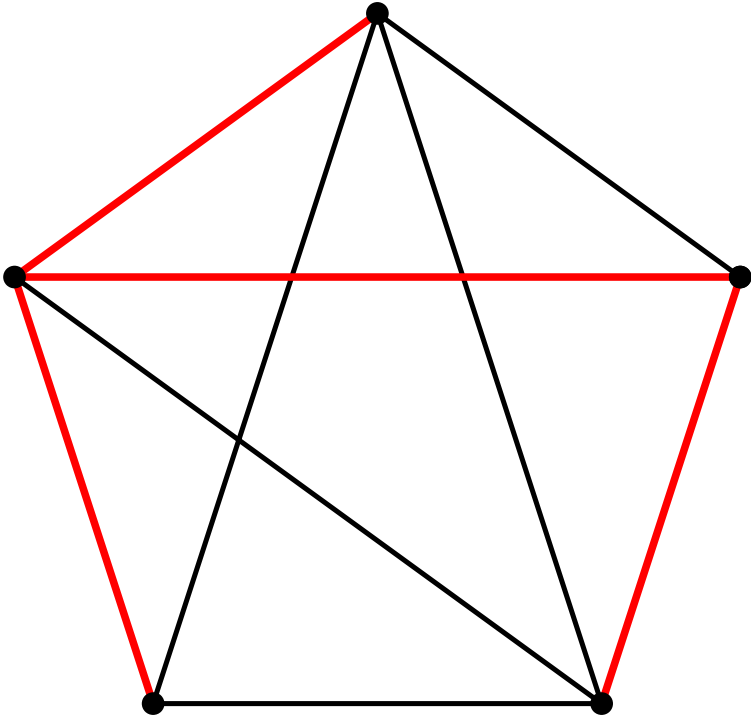
On the other hand if a connected graph does contain a cycle then you **can** remove an edge of this cycle without disconnecting the graph.

Finally, if a graph is not connected then there are edges you can add without producing any new cycle: join two vertices in the graph that are not already connected by a path.

We have shown that:

Lemma. *A graph G is a tree if and only if it is a maximal acyclic graph and if and only if it is a minimal connected graph.*

A **spanning tree** of a graph G is a tree in G that connects all the vertices of G .



Lemma (Spanning Trees). *Every connected graph contains a spanning tree.*

Lemma (Spanning Trees). *Every connected graph contains a spanning tree.*

Proof Choose a minimal connected subgraph of G containing all the vertices. If it contained a cycle we could throw out an edge without disconnecting it. □

Lemma (Tree properties). For a graph G with n vertices, any two of the following implies the third (and hence that G is a tree).

- G has $n - 1$ edges
- G is connected
- G is acyclic

Compare this with what you already know about bases in finite-dimensional vector spaces.

Theorem. *For a set A in a vector space V of dimension n , any two of the following implies the third (and hence that A is a basis).*

- *A has n elements*
- *A spans V*
- *A is linearly independent*

This is more than an analogy.

Lemma (Tree properties). For a graph G with n vertices, any two of the following implies the third (and hence that G is a tree).

- G has $n - 1$ edges
- G is connected
- G is acyclic

Proof To begin with let us check that every tree on n vertices has $n - 1$ edges. Assume inductively that this holds for smaller numbers of vertices and let T be a tree on n vertices.

We saw that a graph in which every vertex has degree at least 2 must contain a cycle, so T contains a vertex of degree 1: a leaf. If we remove this vertex and its attaching edge we produce a tree on the remaining $n - 1$ vertices which by the inductive hypothesis has $n - 2$ edges. So T has $n - 1$ edges.

Now suppose that G is connected and has $n - 1$ edges. G contains a spanning tree which must be the whole graph.

Finally suppose G is acyclic and has $n - 1$ edges. If it has k connected components with r_1, r_2, \dots, r_k vertices respectively then

$$r_1 + r_2 + \dots + r_k = n.$$

These components are trees so they have

$$r_1 - 1, \quad r_2 - 1, \quad \dots$$

edges respectively and so

$$r_1 + r_2 + \dots + r_k - k = n - 1.$$

But this sum is $n - k$ so in fact $k = 1$ and the graph is connected. \square

Remark The proof that each tree on n vertices has $n - 1$ edges used induction: we pulled off a leaf and looked at what was left.

Notice that the inductive step is a process of **reduction**: we reduce the case we are looking at, to a case we have already solved.

The **logic** of an inductive argument appears to go upwards: “we know something at this level so we can prove something at the next level”. But **our job** in proving the inductive step is the opposite: to reduce what we don’t know, to what we do.

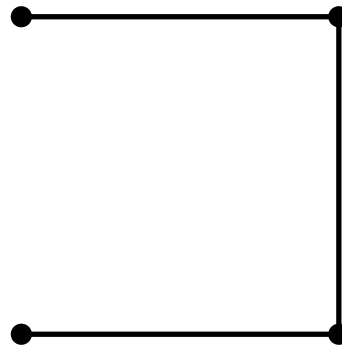
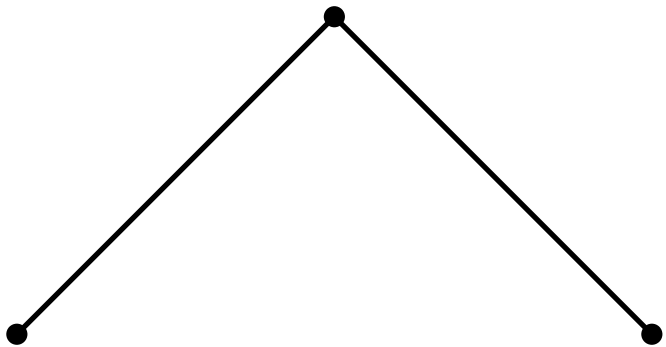
BAD Here's a graph for which I know the result: if I add a bit more I get a bigger one for which I now know the result.

GOOD Here's a graph for which I don't know the result: if I remove a bit I get a smaller one for which I do: and I can use that to get the result for the original.

The first one builds ever bigger graphs for which the statement is true but doesn't prove it for **all** graphs.

Every tree is a bipartite graph because it contains no odd cycle. It isn't usually helpful to think of it this way.

The word **forest** is used to describe a graph which has no cycles but may be disconnected:

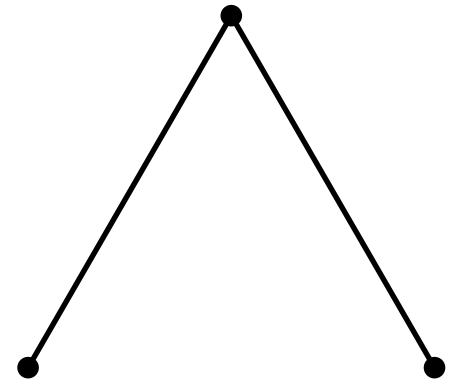
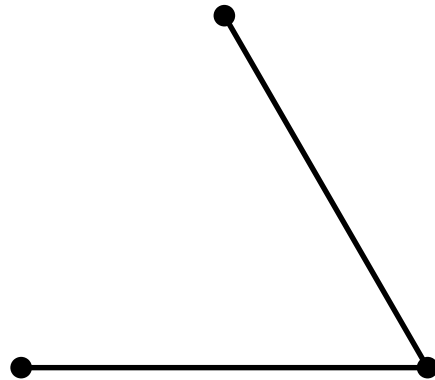
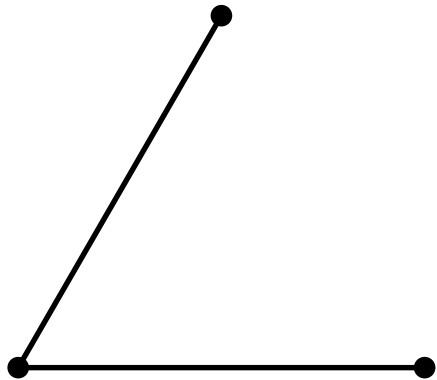


Counting trees

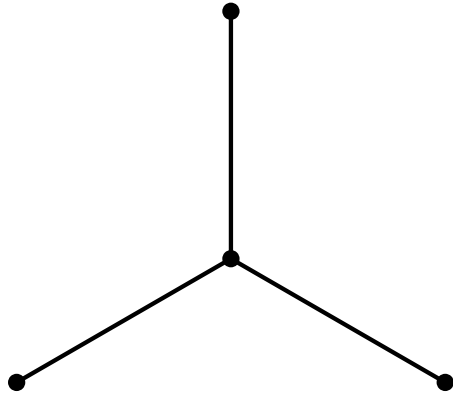
We return for a moment to enumeration.

How many different trees are there on n labelled vertices?

For $n = 3$ there are 3:



For $n = 4$ there are two **types** of tree (two isomorphism classes of tree):



There are 4 examples of the propeller type: you just have to decide which is the central vertex.

There are 12 examples of the path P_3 because there are $4! = 24$ orderings of the vertices but backwards and forwards give the same path.

For $n = 3$ there are 3 trees.

For $n = 4$ there are $16 = 4^2$ trees altogether.

For $n = 5$ it is not too hard to calculate that there are $125 = 5^3$.

On this basis we might make the conjecture that for n vertices there are n^{n-2} trees. This is true and is usually known as Cayley's formula although Cayley referenced Burchardt for it.

Although the formula is simple, it is quite tough to prove.

There is a trick proof by Prüfer who found a bijection between the family of trees on n labelled vertices and the set of sequences of length $n - 2$ with entries from $\{1, 2, \dots, n\}$.

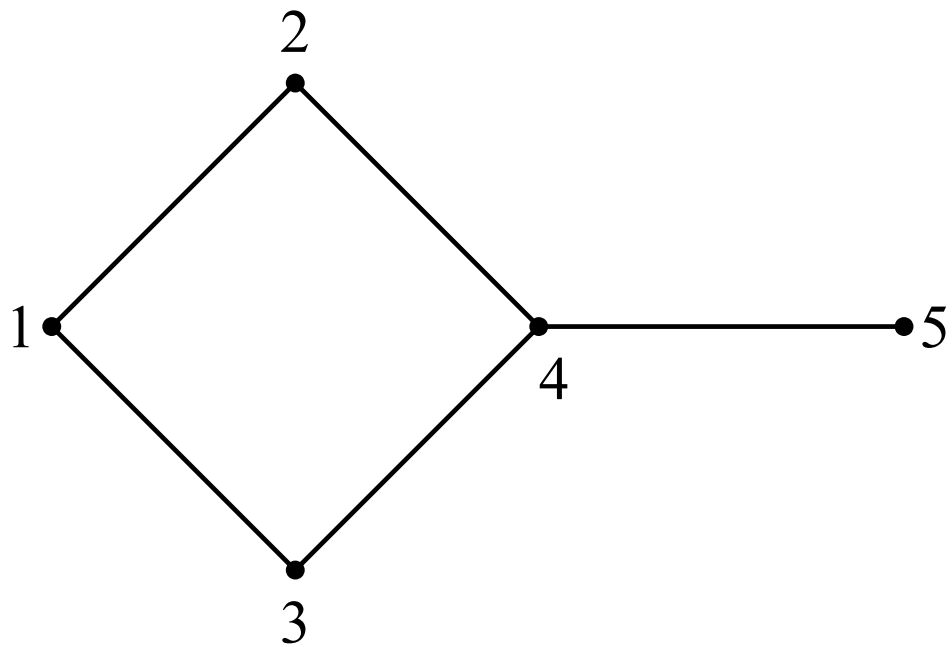
The bijection is called Prüfer encoding.

Instead we shall prove a much more general fact called the Matrix Tree Theorem which is related to electrical networks.

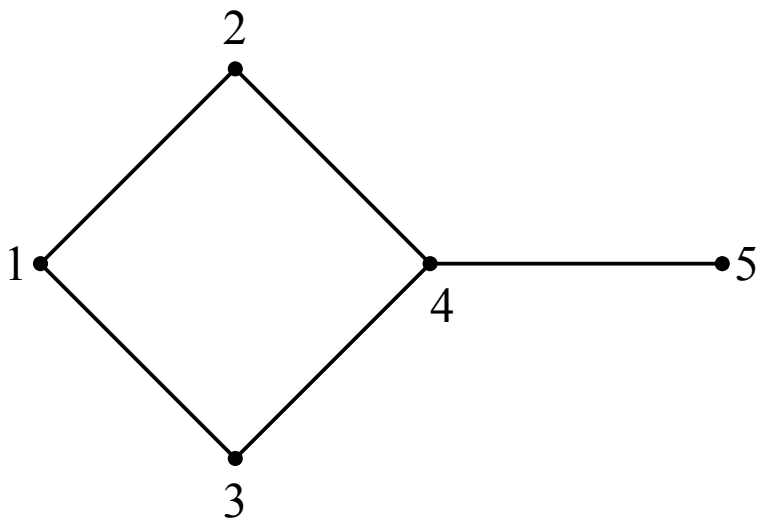
If G is a graph on the vertices $1, 2, \dots, n$ and for each i the vertex i has degree d_i then the **Laplacian** of G is the symmetric matrix (a_{ij}) given by

$$a_{ij} = \begin{cases} d_i & \text{if } i = j \\ -1 & \text{if } ij \text{ is an edge} \\ 0 & \text{if } i \neq j \text{ and } ij \text{ is not an edge} \end{cases}$$

For example if G is the graph



then



$$L = \begin{pmatrix} 2 & -1 & -1 & 0 & 0 \\ -1 & 2 & 0 & -1 & 0 \\ -1 & 0 & 2 & -1 & 0 \\ 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} .$$

Theorem (Kirchhoff's Matrix Tree Theorem). Let L be the Laplacian of a graph G . Then the number of spanning trees of G is any $(n - 1) \times (n - 1)$ principal minor of L .

The determinant of a principal $(n - 1) \times (n - 1)$ submatrix

$$\left(\begin{array}{cccc|c} 2 & -1 & -1 & 0 & 0 \\ -1 & 2 & 0 & -1 & 0 \\ -1 & 0 & 2 & -1 & 0 \\ 0 & -1 & -1 & 3 & -1 \\ \hline 0 & 0 & 0 & -1 & 1 \end{array} \right).$$

Theorem (Cayley's formula). *If $n \geq 2$ there are n^{n-2} trees on n vertices.*

Proof We want to find the number of spanning trees of the complete graph K_n on n vertices. That means we want to calculate the $(n - 1) \times (n - 1)$ minor

$$\det \begin{pmatrix} n-1 & -1 & -1 & \dots & -1 \\ -1 & n-1 & -1 & \dots & -1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -1 & \dots & -1 & -1 & n-1 \end{pmatrix}.$$

This matrix is

$$nI_{n-1} - J_{n-1}$$

where I_k is the $k \times k$ identity and J_k is the $k \times k$ matrix of **1**s. The eigenvalues of J_{n-1} are

$$(n - 1, 0, 0, 0, \dots, 0).$$

So the eigenvalues of $nI_{n-1} - J_{n-1}$ are

$$(1, n, n, n, \dots, n).$$

Hence the determinant is n^{n-2} .

□

Warning: If A and B are symmetric matrices we can't easily compute the eigenvalues of $A + B$ in terms of those of A and B . But if one of them is a multiple of the identity then we have no problem (because every vector is an eigenvector of the identity).

The proof of Kirchhoff's Theorem involves a little multilinear algebra.

Fix n and for each pair $\{i, j\}$ with $i < j$ let e_{ij} be the vector

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

which has a 1 in the i^{th} place and -1 in the j^{th} .

If G is a graph on n vertices with m edges we may form an incidence matrix \tilde{B} : an $n \times m$ matrix in which the columns are the vectors e_{ij} for which ij is an edge.

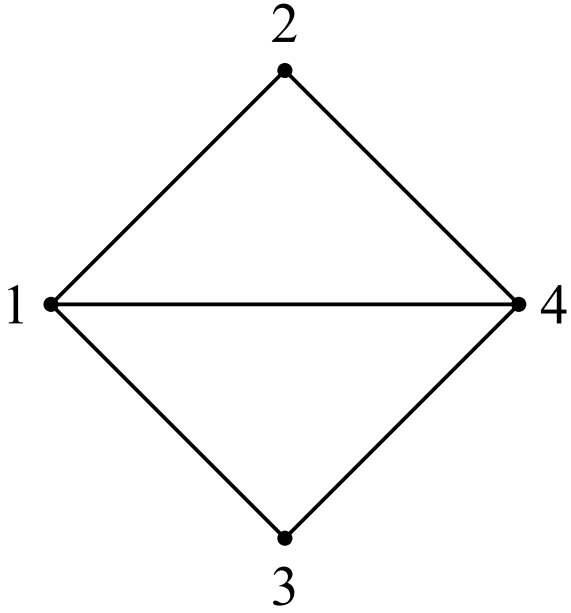
For example the matrix

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & -1 & -1 \end{pmatrix}$$

corresponds to the graph

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & -1 & -1 \end{pmatrix}$$

corresponds to the graph



The proof of Kirchhoff's Theorem uses 3 steps.

Step 3: It is easy to check that $\tilde{B} \cdot \tilde{B}^T = L$.

Step 2: We can relate the minors of L to the minors of \tilde{B} using a standard principle of multilinear algebra: the Cauchy-Binet Theorem.

Step 1: We can express the number of spanning trees of G in terms of the minors of \tilde{B} .

Step 1

Suppose we choose a collection of $n - 1$ of the e_{ij} vectors and put them side by side to form a matrix

$$W = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Note that the sum of the rows is $(0, 0, 0)$. Now look at the determinants of the $(n - 1) \times (n - 1)$ submatrices. I claim that these all have the same size. **Same absolute value.**

You are familiar with this fact if $n = 3$. Consider two vectors

$$\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

These are both perpendicular to the constant vector $(1, 1, 1)^T$.

So their vector product is in the direction of the constant vector.

But the entries of the vector product are the 2×2 minors. So these minors all have the same size.

Back to the general case.

Suppose we choose a collection of $n - 1$ of the e_{ij} vectors and put them side by side to form a matrix

$$W = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Note that the sum of the rows is $(0, 0, 0)$. Now look at the determinants of the $(n - 1) \times (n - 1)$ submatrices. I claim that these all have the same size.

Let's consider the top $(n - 1) \times (n - 1)$ determinant. If we add the second and third rows to the top one the determinant doesn't change.

$$\det \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = \det \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

But the new top row is the sum of the old top 3 rows which is the negative of the old 4th row because in the original 4×3 matrix the rows added up to 0.

If we change the sign of the new top row and move it to the bottom we get a determinant which has the same size and this is the determinant of the bottom 3 rows of the original 4×3 matrix.

$$\det \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

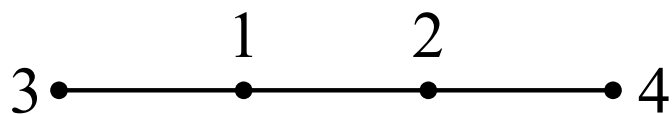
In the same way we can show that any two of the $(n - 1) \times (n - 1)$ minors have the same size.

Now suppose that we have a graph on vertices $(1, 2, \dots, n)$ with $n - 1$ edges. Form the matrix using e_{ij} for each of the edges we have present.

So

$$W(G) = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

corresponds to



I claim that the minors of $W(g)$ are zero if G contains a cycle but are ± 1 if G is a tree.

Firstly, if there is a cycle then the edge vectors from this cycle obey a linear relation. For example the cycle 1354 yields

$$e_{13} + e_{35} - e_{54} - e_{41} = 0.$$

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & -1 & 0 \end{pmatrix}$$

On the other hand suppose the $n - 1$ edges form a tree. Choose a leaf and suppose after reordering the rows that it is the vertex 1.

Then the matrix looks as follows

$$\left(\begin{array}{c|cccc} 1 & 0 & 0 & 0 \\ \hline * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{array} \right).$$

After we delete the first row and column we get another tree matrix and we may assume inductively that its $(n - 2) \times (n - 2)$ minors are ± 1 .

But now if we expand the top minor of the original matrix along the top row we get the top minor of the smaller tree.

Suppose we are given a graph G with m edges and we form the $n \times m$ incidence matrix \tilde{B} using the edge vectors.

Cross out one of the rows: say the bottom one. Now form all the $(n - 1) \times (n - 1)$ determinants of the resulting $(n - 1) \times m$ matrix. The squares of these determinants will be 1 or 0 depending upon whether the $n - 1$ columns correspond to a tree in G : a spanning tree of G .

So the number of spanning trees of G is the sum of the squares of the $(n - 1) \times (n - 1)$ determinants of the incidence matrix with a row deleted.

Step 2

Let B be a $k \times m$ matrix with $k \leq m$. I claim that the sum of the squares of the $k \times k$ determinants is a natural geometric quantity: it doesn't change if we apply a rotation to the columns or to the rows of B .

In the case $k = 2$ and $m = 3$ there are three determinants:

$$\begin{pmatrix} 2 & 3 & 1 \\ 6 & 5 & 1 \end{pmatrix}$$

$2 \times 5 - 6 \times 3$ for example.

These three determinants are the entries of the vector product of the two rows. The sum of the squares of the determinants is the square of the length of the vector product and this depends only upon the lengths of the rows and the angle between them.

In the general case I shall prove the following.

Theorem (Cauchy-Binet). Let B be a $k \times m$ matrix with $k \leq m$. The sum of the squares of the $k \times k$ minors of B is

$$\det(B.B^T).$$

Proof Let B consist of columns $u^{(1)}, u^{(2)}$ and so on. Then for $1 \leq i \leq k$ and $1 \leq r \leq m$

$$(B)_{ir} = u_i^{(r)}.$$

This means that BB^T is the matrix whose ij entry is $\sum_{r=1}^m u_i^{(r)} u_j^{(r)}$.

For each set $\sigma = \{j_1, j_2, \dots, j_k\}$ of k indices let B_σ be the $k \times k$ matrix

$$\begin{pmatrix} | & & | \\ u^{(j_1)} & \dots & u^{(j_k)} \\ | & & | \end{pmatrix}.$$

We want to show that

$$\det \left(\sum_{r=1}^m u_i^{(r)} u_j^{(r)} \right) = \sum_{|\sigma|=k} (\det B_\sigma)^2$$

where the sum is over all subsets of $\{1, 2, \dots, m\}$ of size k .

We can extend the sum over all **multisets** of size k because if we repeat a column the determinant is zero.

Consider the determinant

$$\det \left(\sum_{r=1}^m x_r u_i^{(r)} u_j^{(r)} \right) = \det B \cdot \begin{pmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & 0 & \dots \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & x_m \end{pmatrix} \cdot B^T.$$

This is a homogeneous polynomial of degree k in the variables x_1, x_2, \dots, x_m so it is a combination of the monomials

$$\prod_{r \in \sigma} x_r$$

where σ is a multiset of k indices (of columns).

$$\det \left(\sum_{r=1}^m x_r u_i^{(r)} u_j^{(r)} \right) = \sum_{|\sigma|=k} d_\sigma \prod_{r \in \sigma} x_r.$$

The theorem will follow if we check that the coefficient d_σ is the square of the determinant of B_σ .

Because then we have

$$\det \left(\sum_{r=1}^m x_r u_i^{(r)} u_j^{(r)} \right) = \sum_{|\sigma|=k} (\det B_\sigma)^2 \prod_{r \in \sigma} x_r$$

and we just set all the x_r to be 1.

$$\det \left(\sum_{r=1}^m x_r u_i^{(r)} u_j^{(r)} \right) = \sum_{|\sigma|=k} d_\sigma \prod_{r \in \sigma} x_r.$$

We want the coefficient of $\prod_{r \in \sigma} x_r$ to be zero whenever the multiset contains a repeated index. Suppose that were not the case: that there is a multiset using at most $k - 1$ variables which has non-zero coefficient.

Set all the other variables to zero. On the right hand side we have a polynomial with at least one non-zero coefficient.

On the left hand side is a $k \times k$ matrix which is a sum of $k - 1$ matrices of rank 1: so it has rank at most $k - 1$ and so its determinant is zero for every choice of the variables.

Therefore

$$\det \left(\sum_{r=1}^m x_r u_i^{(r)} u_j^{(r)} \right) = \sum_{|\sigma|=k} d_\sigma \prod_{r \in \sigma} x_r$$

where the sum only uses sets: we only need monomials that are products of k distinct indices.

That means that the coefficient d_σ can be calculated by setting all the variables indexed by σ to 1 and all the other variables to 0.

So

$$d_\sigma = \det \left(\sum_{r \in \sigma} u_i^{(r)} u_j^{(r)} \right).$$

This quantity is the determinant of $B_\sigma B_\sigma^T$.

$$\det \left(\sum_{r=1}^m x_r u_i^{(r)} u_j^{(r)} \right) = \sum_{|\sigma|=k} \det(B_\sigma B_\sigma^T) \prod_{r \in \sigma} x_r.$$

But B_σ is a $k \times k$ matrix so by the multiplicative property of the determinant this is $(\det(B_\sigma))^2$ as required. □

Step 3

We now have the following. If G is a graph on n vertices with m edges form the $n \times m$ matrix \tilde{B} whose columns are the vectors e_{ij} corresponding to edges in G .

Delete one of the rows of \tilde{B} to obtain B .

Then the number of spanning trees in G is $\det B \cdot B^T$.

To finish the proof of Kirchhoff's Theorem I just need to check that the matrices $B.B^T$ are the principal $(n-1) \times (n-1)$ submatrices of the Laplacian L .

In the HW you are asked to prove that

$$\tilde{B}.\tilde{B}^T = L.$$

We want to check that if you cross out a row of \tilde{B} and the corresponding column of \tilde{B}^T you get a principal $(n-1) \times (n-1)$ submatrix of L . Look at the picture.

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & -1 & -1 & -1 & 0 \end{array} \right) \left(\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{array} \right)$$

$$= \left(\begin{array}{ccc|c} 3 & -1 & -1 & -1 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 2 & -1 \\ \hline -1 & -1 & -1 & 3 \end{array} \right)$$

Summary

A **tree** is a graph that is connected but contains no cycle.

Lemma. *A graph G is a tree if and only if it is a maximal acyclic graph and if and only if it is a minimal connected graph.*

A **spanning tree** of a graph G is a tree in G that connects all the vertices of G .

Lemma (Spanning Trees). *Every connected graph contains a spanning tree.*

Lemma (Tree properties). For a graph G with n vertices, any two of the following implies the third (and hence that G is a tree).

- G has $n - 1$ edges
- G is connected
- G is acyclic

If G is a graph on the vertices $1, 2, \dots, n$ and for each i the vertex i has degree d_i then the **Laplacian** of G is the symmetric matrix (a_{ij}) given by

$$a_{ij} = \begin{cases} d_i & \text{if } i = j \\ -1 & \text{if } ij \text{ is an edge} \\ 0 & \text{if } i \neq j \text{ and } ij \text{ is not an edge} \end{cases}$$

Theorem (Kirchhoff's Matrix Tree Theorem). *Let L be the Laplacian of a graph G . Then the number of spanning trees of G is any $(n - 1) \times (n - 1)$ principal minor of L .*

Theorem (Cayley's formula). *If $n \geq 2$ there are n^{n-2} trees on n vertices.*

If we choose a collection of $n - 1$ of the e_{ij} vectors and put them side by side to form a matrix then the determinants of the $(n - 1) \times (n - 1)$ submatrices all have the same size. The $(n - 1) \times (n - 1)$ minors are zero if the edges form a cycle but are ± 1 if they form a tree.

Suppose we are given a graph G with m edges and we form the $n \times m$ incidence matrix \tilde{B} using the edge vectors. Then the number of spanning trees of G is the sum of the squares of the $(n - 1) \times (n - 1)$ determinants of the incidence matrix with a row deleted.

Theorem (Cauchy-Binet). Let B be a $k \times m$ matrix with $k \leq m$. The sum of the squares of the $k \times k$ minors of B is

$$\det(B.B^T).$$

If G is a graph on n vertices with m edges form the $n \times m$ matrix \tilde{B} whose columns are the vectors e_{ij} corresponding to edges in G . If L is the Laplacian of G then

$$\tilde{B}.\tilde{B}^T = L.$$

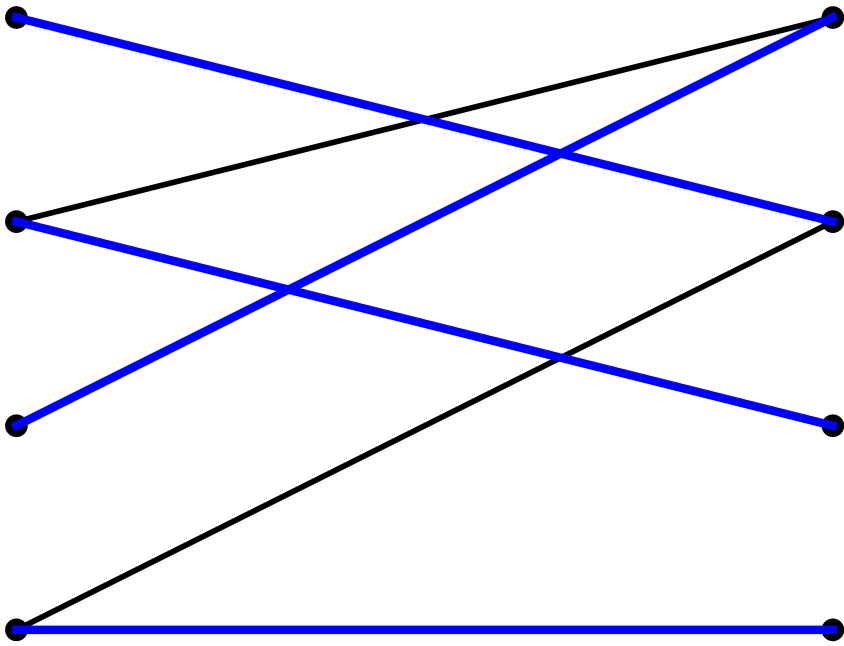
Chapter 8. Hall's Theorem

One of the earliest theorems that was consciously a part of combinatorics is Hall's Marriage Theorem.

We shall discuss it rather briefly but it belongs to a larger circle of ideas of which the most important is the Max-flow Min-cut Theorem.

There are n boys and n girls. They have to pair off for the school dance and we insist that a boy and girl can only make a date if they know one another. Can it be done?

We can draw a bipartite graph of the boys on one side and girls on the other with an edge between any two who are acquainted.



The picture also shows a possible arrangement of dates.

What could go wrong?

If there is a boy who doesn't know any girls then we are stuck.

If there are 2 boys who between them know only 1 girl then again we have a problem.

In general we will certainly need that for each k between 1 and n , every set of k boys between them know at least k girls.

Hall's Theorem says that this condition is sufficient.

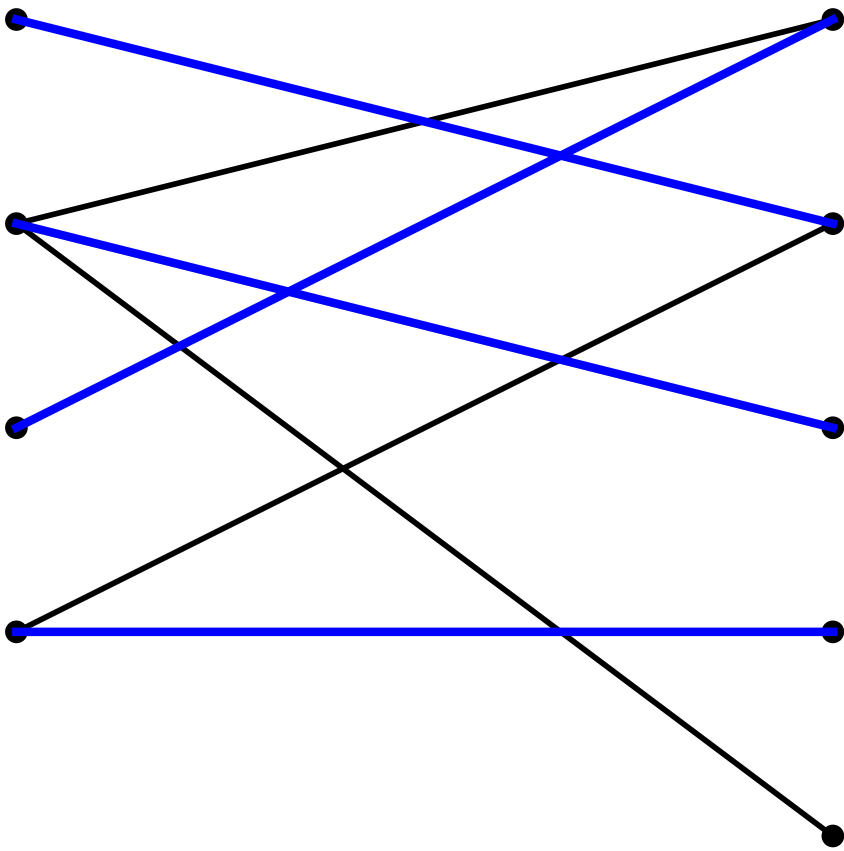
If G is a bipartite graph with vertex sets A and B then a **complete matching** from A into B is a set of disjoint edges which cover the vertices of A . One edge coming out of each vertex of A .

Theorem (Hall's Marriage Theorem). Let G be a bipartite graph with vertex classes A and B . For each subset $U \subset A$ let $\Gamma(U)$ be the set of neighbours of vertices in U :

$$\Gamma(U) = \{b : ab \text{ is an edge for some } a \in U\}.$$

If for every $U \subset A$ the set $\Gamma(U)$ is at least as large as U then G contains a complete matching from A into B .

The hypothesis is not symmetric: we look at neighbours of sets in A .
The conclusion is not symmetric as written. If we know that A and B are of equal size then we get a pairing of the elements of A with those of B . But the theorem applies even if B is larger.

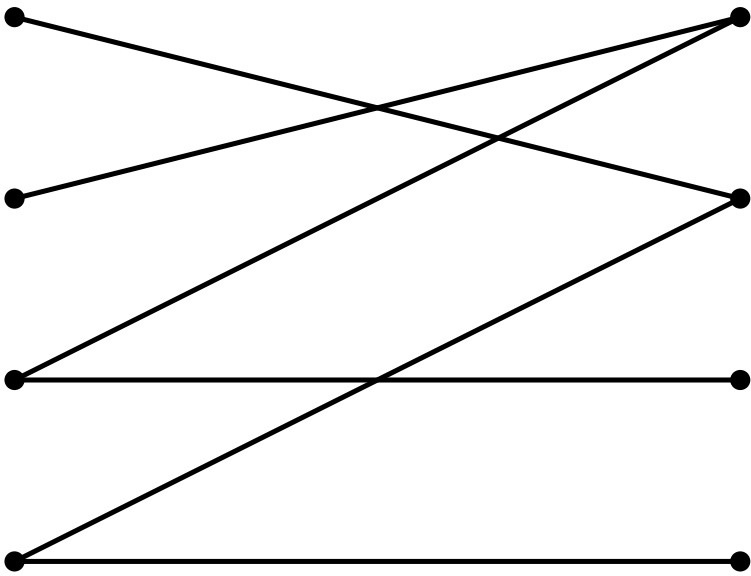


Proof We shall use induction on the size of A . If A contains just one vertex with at least 1 neighbour then we can match it.

Suppose that A contains n elements and assume inductively that the result holds whenever there are fewer than n elements. We shall break into two cases:

1. There is a non-empty proper subset $U \subsetneq A$ for which $\Gamma(U)$ has the **same** number of elements as U .
2. For every non-empty proper subset $U \subsetneq A$ the set $\Gamma(U)$ contains at least $|U| + 1$ elements.

In case 1, we match the critical set U into $\Gamma(U)$ using the inductive hypothesis. **What else can we do?** I claim that once we remove U and $\Gamma(U)$ from the graph what is left still satisfies the hypothesis of Hall's Theorem (and has a smaller initial set).



If $V \subset A - U$ had fewer than $|V|$ neighbours in $B - \Gamma(U)$ then $U \cup V$ would have fewer neighbours in G than $|U| + |V|$ violating the original assumption on G .

So the new graph satisfies the hypothesis of the theorem and we can match the vertices of $A - U$ into $B - \Gamma(U)$.

Together with the matching of U we get a complete matching of A into B .

In case 2 let us pick one vertex x of A arbitrarily and match it with a neighbour, y say. I claim that once I remove x and y from G , the remaining graph still satisfies the assumption of the theorem and has a smaller initial set.

Any subset V of the new initial set is a proper subset of A and so it had at least $|V| + 1$ neighbours in G . Even with y removed it still has at least $|V|$ neighbours.

So by the inductive hypothesis again we can match the vertices of $A - \{x\}$ into $B - \{y\}$ and thus get a complete matching of A .

□

Hall's Theorem is one of a number of statements in combinatorics in which a condition that is obviously necessary turns out to be sufficient (perhaps somewhat surprisingly). Almost all of these examples arise from some sort of **duality** and this is no exception. Hall's Theorem can be deduced from the Max-flow Min-cut Theorem which is a special form of the duality theorem for linear programming.

Hall's Theorem has several extensions including a subtle result called Tutte's 1-factor Theorem.

Bistochastic matrices

An $n \times n$ matrix (a_{ij}) is called bistochastic if its entries are non-negative and all its row and column sums are 1.

For example

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/3 & 1/6 \\ 1/6 & 1/3 & 1/2 \end{pmatrix}.$$

These matrices are important in the study of Markov chains.

The simplest examples are the **permutation** matrices

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

which have a single **1** in each row and column.

As a linear map on \mathbf{R}^3 this matrix permutes the coordinates

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} z \\ x \\ y \end{pmatrix}.$$

Theorem (Extreme points of the bistochastic matrices). *Each $n \times n$ bistochastic matrix is a convex combination of permutation matrices.*

$$\begin{pmatrix} 1/2 & 1/6 & 1/3 \\ 1/3 & 1/2 & 1/6 \\ 1/6 & 1/3 & 1/2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + \frac{1}{6} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

The weights are positive numbers adding up to 1.

Proof For each dimension n we use induction on the number of non-zero entries of the matrix.

An $n \times n$ bistochastic matrix must have at least n entries and if it has exactly n it is a permutation matrix.

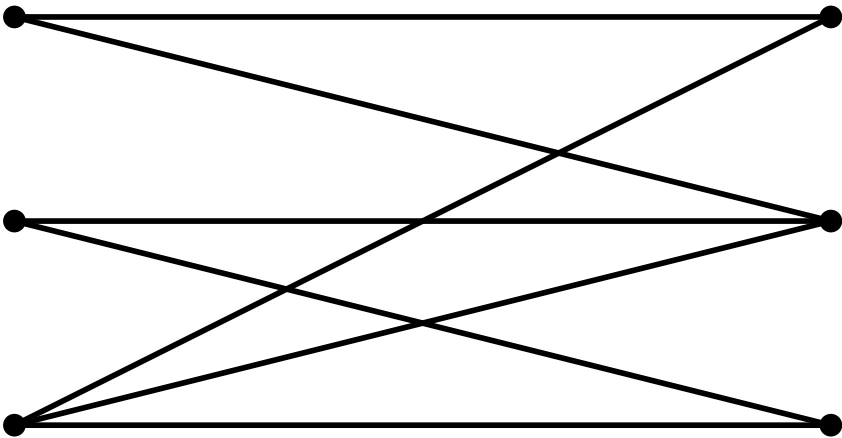
Suppose that $A = (a_{ij})$ is bistochastic. Form a bipartite graph with vertex sets $\{s_1, s_2, \dots, s_n\}$ and $\{t_1, t_2, \dots, t_n\}$ and an edge from s_i to t_j if $a_{ij} > 0$.

$$\begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/3 & 2/3 \\ 1/2 & 1/6 & 1/3 \end{pmatrix}$$

For example

$$\begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/3 & 2/3 \\ 1/2 & 1/6 & 1/3 \end{pmatrix}$$

corresponds to



Form a bipartite graph with vertex sets $\{s_1, s_2, \dots, s_n\}$ and $\{t_1, t_2, \dots, t_n\}$ and an edge from s_i to t_j if $a_{ij} > 0$.

I claim that this graph satisfies the condition of Hall's Theorem.

Let U be a set of k rows. The sum of all the entries in these rows is k . Since each column sum is 1 the non-negative entries in these rows cannot all lie in fewer than k columns.

So by Hall's Theorem there is a matching in the graph. This matching corresponds to a permutation matrix: a matrix with a single **1** in each row and column: call it P .

For example if A is

$$\begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/3 & 2/3 \\ 1/2 & 1/6 & 1/3 \end{pmatrix}$$

then we can find a permutation matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

in which all the ones fall in positions where $a_{ij} > 0$.

Let $p > 0$ be the smallest entry of A in any position where P has a **1**.

If $p = 1$ then $A = P$ and there is nothing to check. If $p < 1$ then $A - pP$ has non-negative entries but fewer non-zero entries than A has.

$$\begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/3 & 2/3 \\ 1/2 & 1/6 & 1/3 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1/6 & 1/2 & 0 \\ 0 & 0 & 2/3 \\ 1/2 & 1/6 & 0 \end{pmatrix}.$$

The row and column sums of $A - pP$ are all equal to $1 - p$ since we have removed p from each row and column.

So the matrix

$$\frac{1}{1-p}(A - pP)$$

is bistochastic and by the inductive hypothesis we may write it as a convex combination of permutations

$$\lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_r P_r.$$

Then we can write A as

$$A = pP + (1-p)\lambda_1 P_1 + (1-p)\lambda_2 P_2 + \cdots + (1-p)\lambda_r P_r$$

completing the inductive step. □

Summary

If G is a bipartite graph with vertex sets A and B then a **complete matching** from A into B is a set of disjoint edges which cover the vertices of A . One edge coming out of each vertex of A .

Theorem (Hall's Marriage Theorem). Let G be a bipartite graph with vertex classes A and B . For each subset $U \subset A$ let $\Gamma(U)$ be the set of neighbours of vertices in U :

$$\Gamma(U) = \{b : ab \text{ is an edge for some } a \in U\}.$$

If for every $U \subset A$ the set $\Gamma(U)$ is at least as large as U then G contains a complete matching from A into B .

Theorem (Extreme points of the bistochastic matrices). *Each $n \times n$ bistochastic matrix is a convex combination of permutation matrices.*

Chapter 9. Ramsey theory

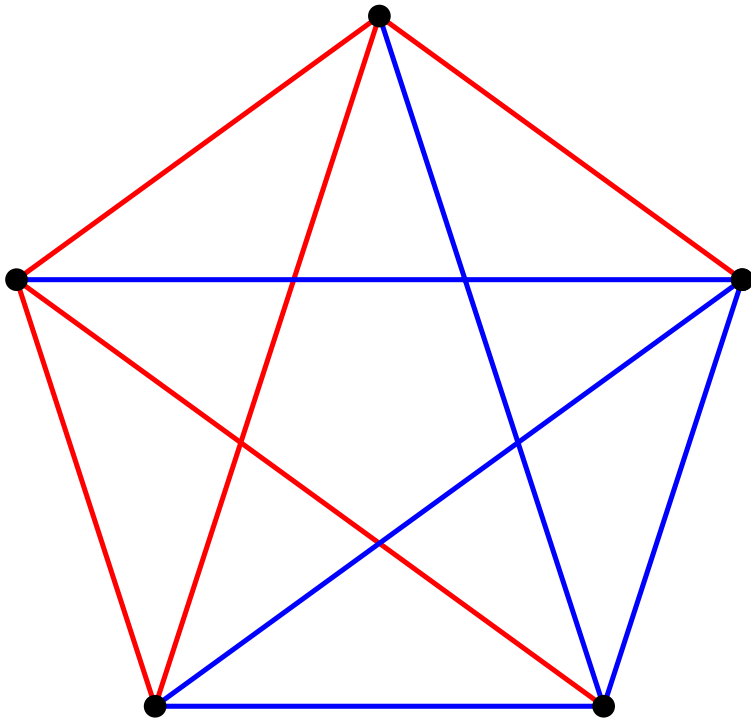
Over the last 50 years Ramsey Theory has become one of the most important parts of combinatorics.

It no longer belongs just to graph theory but this is the natural place to look at it first.

The original work of Ramsey dates from 1928 and was largely ignored at the time. It was really popularised by Erdős in the 1950s (following work he had done with Szekeres).

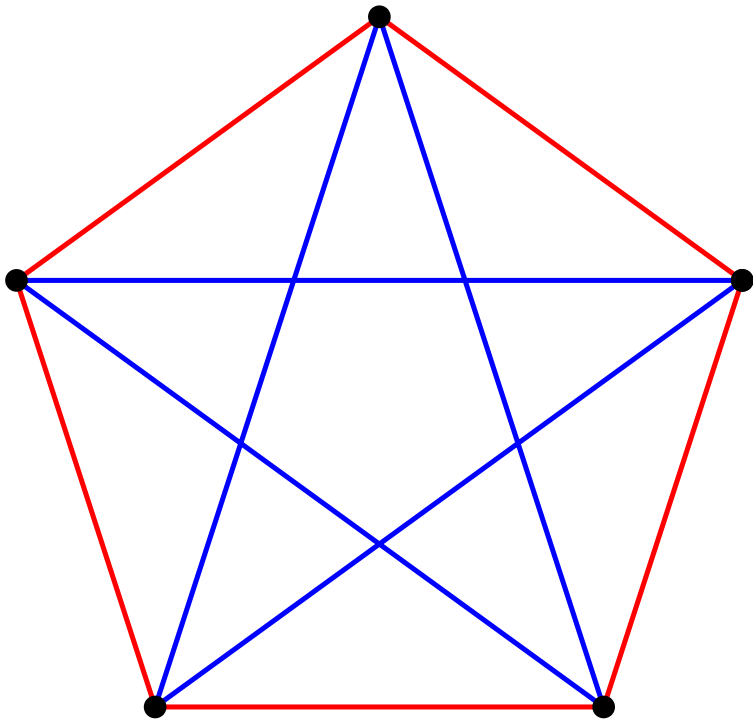
The basic idea is that however chaotic you try to make a structure, there will still be isolated pockets of order.

Start with a complete graph of order 5 say. This is a very regular structure. Now divide the edges into two classes: let's call them **red** and **blue**.



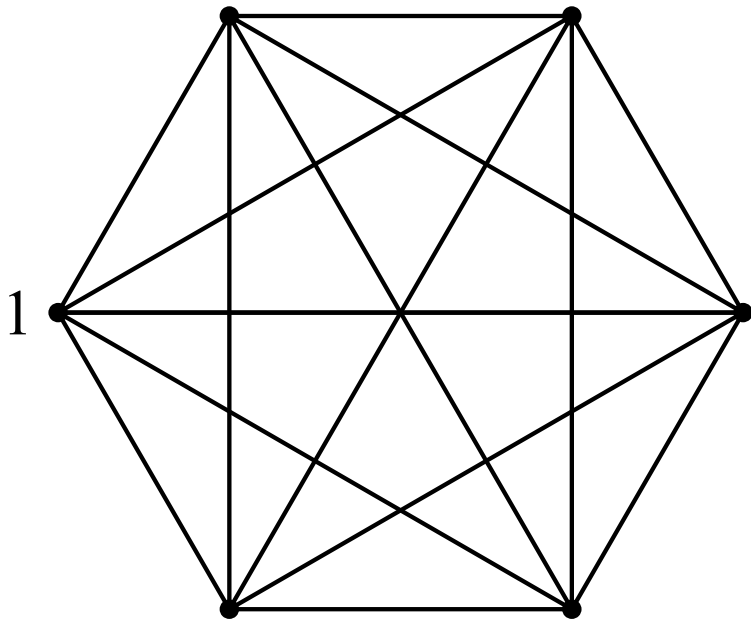
The structure now looks disordered but there are still some pockets of order: a red triangle on the left for example.

If we are careful we can colour the edges in such a way that there are no red triangles and no blue ones. Admittedly the structure now looks quite orderly.



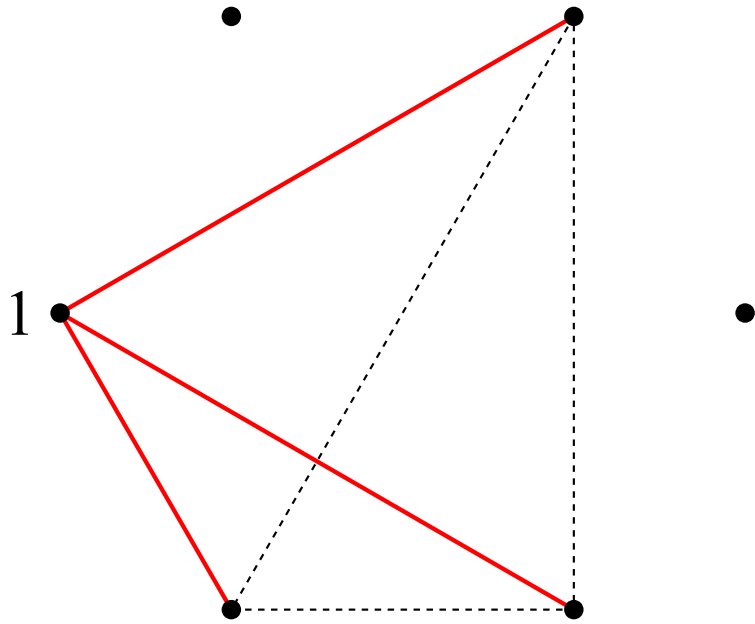
What happens if we try to do this for K_6 ?

Can we colour the edges of K_6 with two colours so that there is no monochromatic triangle?



Consider vertex 1: it has 5 edges incident with it. Either at least 3 are red or at least 3 are blue. Suppose the first.

There are at least 3 red edges incident with vertex 1.



How do we colour the edges joining the 3 vertices at the other ends?

If any **one** of them is red then we have a red triangle. If they are **all** blue then we have a blue triangle.

Theorem ($R(3, 3)$). *If the edges of K_6 are coloured red and blue then either there is a red triangle or there is a blue triangle.*

It is possible to 2-colour the edges of K_5 without forming a monochromatic triangle.

If s and t are integers greater than or equal to 2 we set $R(s, t)$ to be the least number n so that no matter how we 2-colour the edges of the complete graph K_n then we find either a red K_s or a blue K_t .

The little theorem just proved says that $R(3, 3) = 6$.

In the proof of the $R(3, 3)$ Theorem we used the fact that of 5 edges coloured red or blue there must be at least 3 red or at least 3 blue.

This is an example of what is usually called the pigeonhole principle. If you deliver more than mk letters to m pigeonholes, at least 1 pigeonhole will receive more than k letters.

We placed $5 > 2 \times 2$ edges into 2 colour classes: one colour received more than 2 edges.

It is not obvious that $R(s, t)$ is finite for all values of s and t . The fundamental Ramsey Theorem says that it is. So if you 2-colour a large enough complete graph you can find either a red K_{10} or a blue one.

It is clear that $R(s, t)$ is an increasing function of s and t : it is harder to find larger monochromatic subgraphs.

It is also clear that $R(s, 2) = R(2, s) = s$ since if you 2-colour the graph K_s you either find a blue edge K_2 or the whole thing is red K_s .

The proof of Ramsey's Theorem begins with a key lemma which mirrors the argument above.

Theorem (Ramsey recurrence). For s and t at least 3

$$R(s, t) \leq R(s - 1, t) + R(s, t - 1).$$

Proof Let $n = R(s - 1, t) + R(s, t - 1)$ and 2-colour K_n . Pick a vertex. It has degree $n - 1$ so either it has at least $R(s - 1, t)$ red edges incident with it or at least $R(s, t - 1)$ blue ones.

Suppose the first holds and examine the $R(s - 1, t)$ vertices at the other ends.

The complete graph on these vertices either contains a blue K_t or a red K_{s-1} which together with the first vertex yields a red K_s in the whole graph.

Similarly if the second option holds.

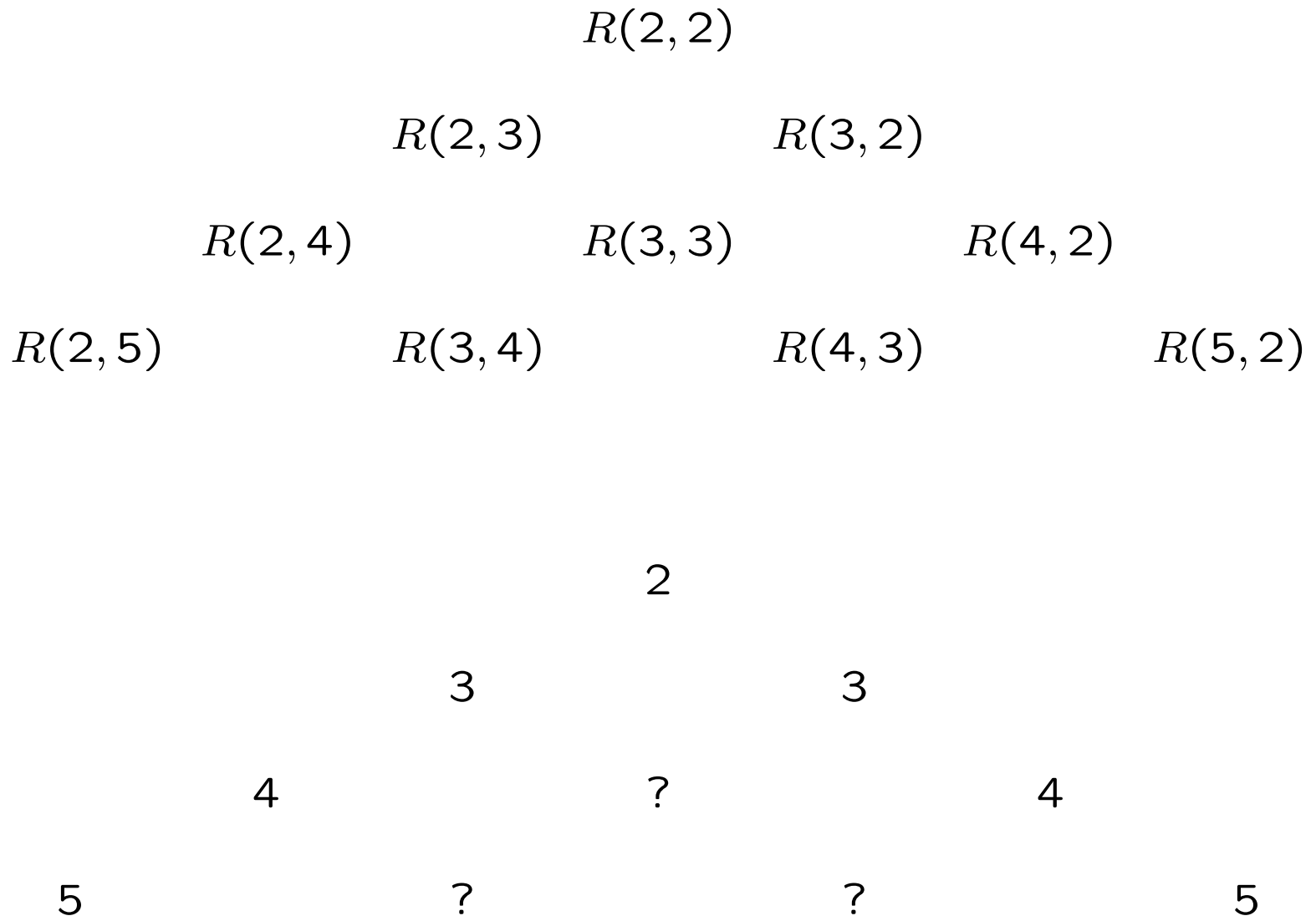
□

From the recurrence inequality it is clear that $R(s, t)$ is finite for all s and t since we can estimate each Ramsey number by others “below” it. The next theorem gives us an actual estimate.

Theorem (Ramsey bound). *For s and t at least 2*

$$R(s, t) \leq \binom{s + t - 2}{s - 1}.$$

Where does the binomial coefficient come from? Look at the Ramsey numbers arranged carefully. I claim that the inequality exactly matches the recurrence relation for binomial coefficients.



Theorem (Ramsey bound). For s and t at least 2

$$R(s, t) \leq \binom{s+t-2}{s-1}.$$

Proof We know that $R(s, 2) = s = \binom{s}{s-1}$ so we have the bound if $t = 2$ and similarly if $s = 2$. We also have

$$R(s, t) \leq R(s-1, t) + R(s, t-1).$$

We shall prove the estimate by induction on $s+t$: **The level**

Assuming inductively that the estimate for $R(p, q)$ holds whenever $p, q \geq 2$ and $p + q < s + t$ we have

$$\begin{aligned} R(s, t) &\leq R(s - 1, t) + R(s, t - 1) \\ &\leq \binom{s + t - 3}{s - 2} + \binom{s + t - 3}{s - 1} \\ &= \binom{s + t - 2}{s - 1}. \end{aligned}$$

□

In particular the theorem shows that $R(s, s) \leq \binom{2s-2}{s-1} \leq 4^{s-1}$.

Observe that in order to prove this we had to introduce the “off-diagonal” Ramsey numbers $R(4, 3)$ and so on.

It is extremely hard to determine exact values of the Ramsey numbers: $R(4, 4) = 18$ but $R(5, 5)$ is unknown.

It is quite easy to extend Ramsey's Theorem to colourings with more colours $R(s, t, u)$ and so on.

We shall prove one exact result due to Chvátal, Erdős and Spencer involving trees.

We need a lemma to start with.

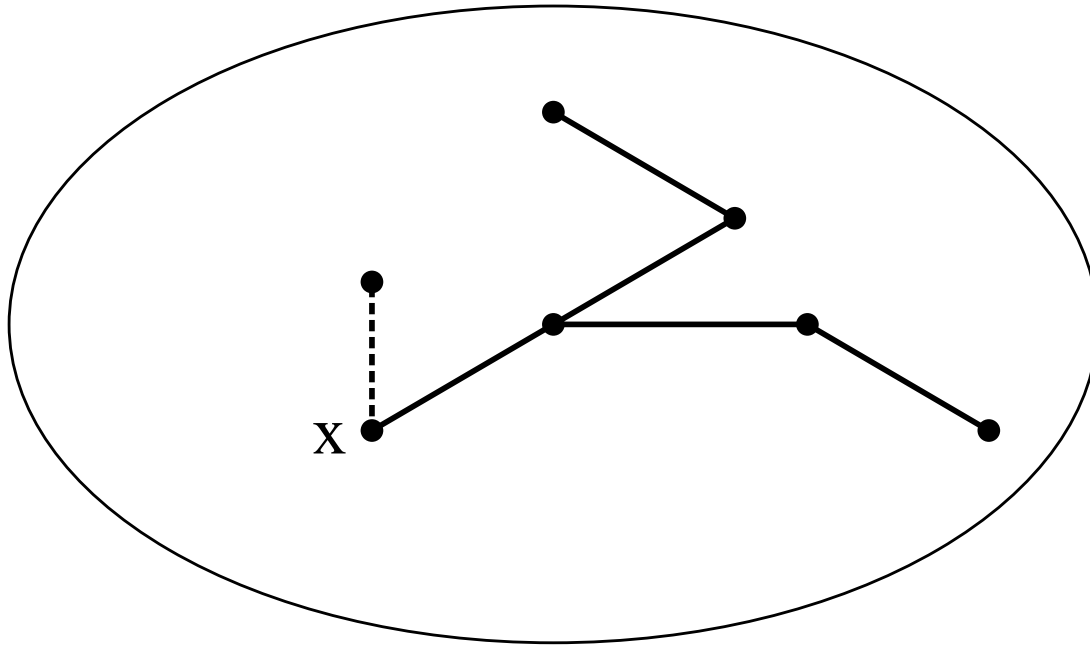
Lemma (Universal graphs for trees). *If G is a graph in which every vertex has degree at least $t - 1$ then it contains a copy of every tree of order t .*

Lemma (Universal graphs for trees). *If G is a graph in which every vertex has degree at least $t - 1$ then it contains a copy of every tree of order t .*

Proof If $t = 2$ the statement is obvious. Assume that t is larger and let T be a tree of order t .

Pull off a leaf. What is left is a smaller tree so we may assume inductively that we can find a copy of it in G .

Pull off a leaf and copy the rest of the tree into G .



Examine the vertex x of G corresponding to the point where our leaf was attached. x has degree at least $t - 1$ in G while our cut tree has only $t - 2$ other vertices. So x is adjacent to at least one more vertex in G and we may extend to get a copy of T . □

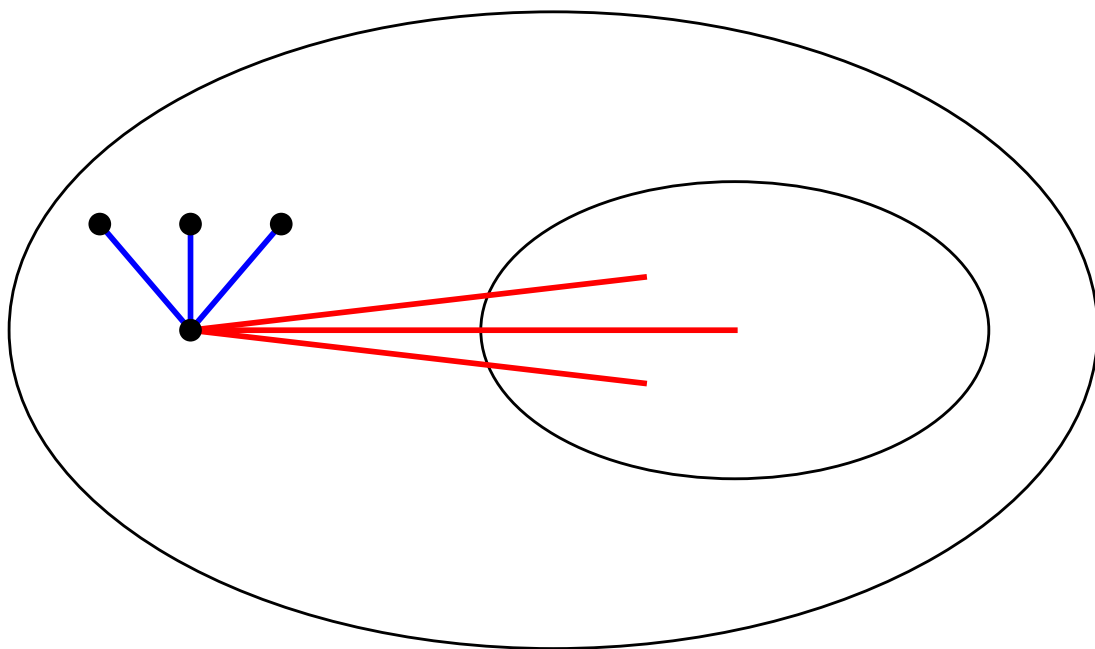
Theorem (Ramsey for a complete graph against trees). Let $n = (s - 1)(t - 1) + 1$. If you 2-colour K_n then you find either a red K_s or a blue copy of every tree of order t .

Note that if $t = 2$ this recovers our earlier remark $R(s, 2) = s$.

Proof By the lemma, if the blue graph has a subgraph in which every vertex has degree at least $t - 1$ then we can find every tree in blue. So suppose that the blue edges contain no subgraph in which every vertex has degree at least $t - 1$. I claim that the graph contains a red K_s .

$n > (s-1)(t-1)$ and we 2-colour K_n . Every blue subgraph has a vertex with degree at most $t-2$.

Let v_1 be a vertex in G of blue degree only $t-2$. Put it on one side and throw out its blue neighbours.



Look at what is left. Pick a vertex v_2 with blue degree at most $t - 2$. Put it to one side and throw out its blue neighbours.

Each time we are reducing the graph by only $t - 1$ vertices so the process continues for more than $s - 1$ steps.

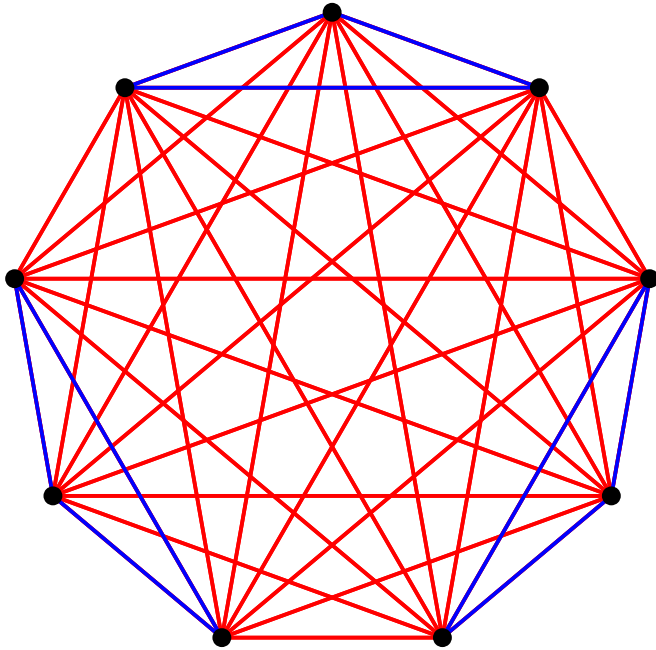
We build a sequence of vertices v_1, v_2, \dots, v_s with the property that each one is connected by red edges to everything that comes after it.

This gives us a red K_s .

□

Theorem (Ramsey for a complete graph against trees). Let $n = (s - 1)(t - 1) + 1$. If you 2-colour K_n then you find either a red K_s or a blue copy of every tree of order t .

The bound in this theorem is tight. Take $s - 1$ copies of a blue K_{t-1} and join them all together with red edges. For example if $t = 4$ and $s = 4$.



Before moving on to lower bounds we look at one example where the bounds are better.

Theorem (Erdős-Szekeres). *A sequence of length $n^2 + 1$ either contains an increasing subsequence of length $n + 1$ or a decreasing one of length $n + 1$.*

From Ramsey's Theorem you get a much worse bound. If (x_i) is the sequence consider the complete graph on $n^2 + 1$ vertices and colour the edge ij red if $x_i x_j$ are in the same order as ij and blue if they are in the opposite order.

An increasing subsequence is a red K_{n+1} : a decreasing subsequence is a blue one.

Theorem (Erdős-Szekeres). *A sequence of length $n^2 + 1$ either contains an increasing subsequence of length $n + 1$ or a decreasing one of length $n + 1$.*

Proof Consider the $n^2 + 1$ points (u_i, d_i) where u_i is the length of the longest increasing subsequence starting at i and d_i is the length of the longest decreasing subsequence starting at i .

If $i < j$ and $x_i \leq x_j$ then $u_i > u_j$ since any increasing subsequence starting at j can be extended by putting i in front. On the other hand, if $x_i \geq x_j$ then $d_i > d_j$.

Therefore the $n^2 + 1$ points are all different and so at least one of them must have a coordinate larger than n . □

To give a flavour of what comes after this I shall quote a famous theorem of van der Waerden.

Theorem (van der Waerden). *Suppose you 2-colour the integers $\{1, 2, 3, \dots\}$. Then you can find arbitrarily long monochromatic arithmetic progressions.*

3, 17, 31, or **11, 21, 31.**

Random colourings

We saw that $R(s, s) \leq 4^{s-1}$. Is it really that large or could it be much smaller?

It is not easy to describe colourings without large monochromatic subgraphs: we don't have a good way to write down something chaotic.

We can prove that chaotic things exist using random methods.

Theorem (Erdős lower bound for $R(s, s)$). *Let $s \geq 3$. Then*

$$R(s, s) > 2^{(s-1)/2}.$$

This argument was the start of a huge field called random graphs which has now merged with the mathematical study of models in statistical physics.

The idea is to colour the edges of K_n randomly. We colour each one red with probability $1/2$ and blue with probability $1/2$ and we colour the edges independently: the choice for one edge is not affected by the choices for the others.

Proof Colour the edges of K_n red or blue independently at random with probability $1/2$ each.

What is the chance that a fixed set of s vertices form a red K_s ? There are $s(s-1)/2$ edges to go red: so the chance is

$$2^{-s(s-1)/2}.$$

The chance that a fixed set of s vertices is monochromatic is just twice as big:

$$2 \times 2^{-s(s-1)/2}.$$

What is the expected number of monochromatic K_s graphs?

There are $\binom{n}{s}$ places where it could occur. So the expected number is

$$\binom{n}{s} 2 \times 2^{-s(s-1)/2} < \frac{2n^s}{s! 2^{s(s-1)/2}}.$$

If we arrange for this number to be less than 1 then there will be colourings without any monochromatic K_s .

If $n \leq 2^{(s-1)/2}$ then we succeed. We can find a bad colouring if $n \leq 2^{(s-1)/2}$ so $R(s, s) > 2^{(s-1)/2}$. □

Compare the proof given above with our calculation of the expected number of matching birthdays with k people.

Recall that expectations are easy.

We just needed the chance of a given K_s being monochromatic and the number of K_s graphs.

Summary

If s and t are integers greater than or equal to 2 we set $R(s, t)$ to be the least number n so that no matter how we 2-colour the edges of the complete graph K_n then we find either a red K_s or a blue K_t .

Theorem ($R(3, 3)$). $R(3, 3) = 6$.

For each s , we have $R(s, 2) = s$.

Theorem (**Ramsey recurrence**). For s and t at least 3

$$R(s, t) \leq R(s - 1, t) + R(s, t - 1).$$

Theorem (Ramsey bound). For s and t at least 3

$$R(s, t) \leq \binom{s + t - 2}{s - 1}.$$

Lemma (Universal graphs for trees). If G is a graph in which every vertex has degree at least $t - 1$ then it contains a copy of every tree of order t .

Theorem (Ramsey for a complete graph against trees). Let $n = (s - 1)(t - 1) + 1$. If you 2-colour K_n then you find either a red K_s or a blue copy of every tree of order t .

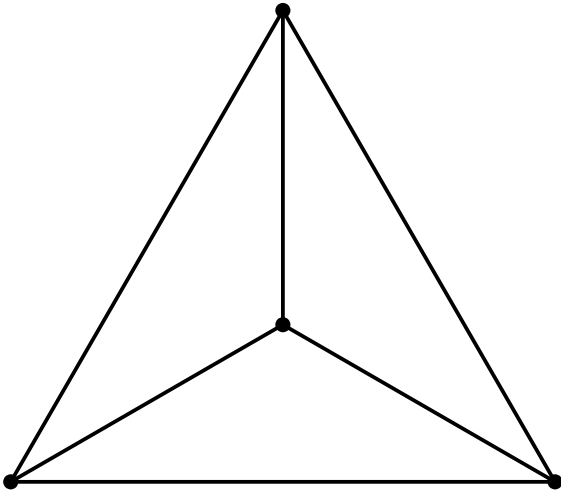
Theorem (Erdős-Szekeres). *A sequence of length $n^2 + 1$ either contains an increasing subsequence of length $n + 1$ or a decreasing one of length $n + 1$.*

Theorem (Erdős lower bound for $R(s, s)$). *Let $s \geq 3$. Then*

$$R(s, s) \geq 2^{\lfloor (s-1)/2 \rfloor}.$$

Chapter 10. Planar graphs

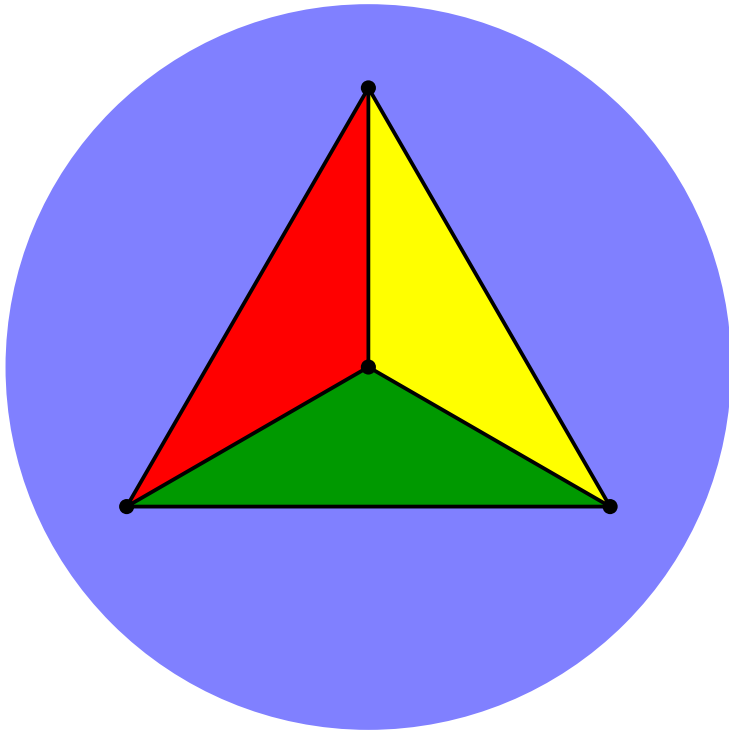
A **planar** graph is a graph that can be drawn in the plane in such a way that edges don't cross one another.



A **plane** graph is such a graph actually drawn (or embedded) in the plane.

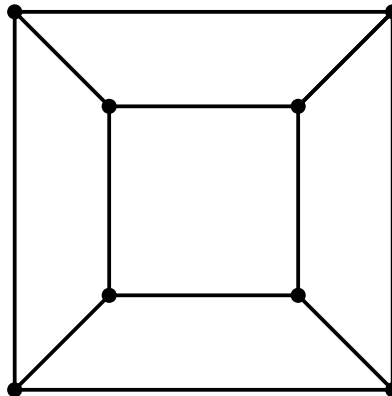
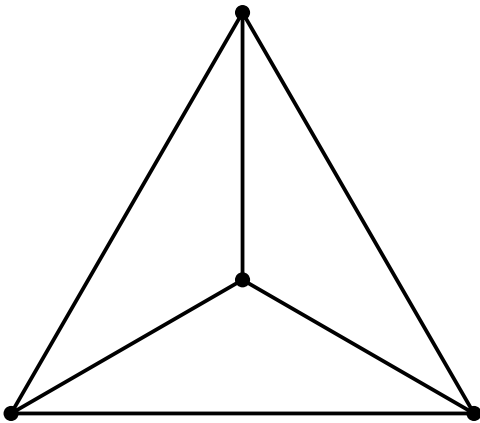
One can show that every planar graph can be drawn using only straight lines for the edges.

If G is a plane graph and we remove the vertices and edges from the plane, what is left falls into connected parts called faces.



Euler produced a famous formula relating the number of vertices, edges and faces of a convex polyhedron.

Plane graphs can represent such polyhedra: squash them. For example the tetrahedron and the cube.



Euler's formula works for any plane graph which is connected.

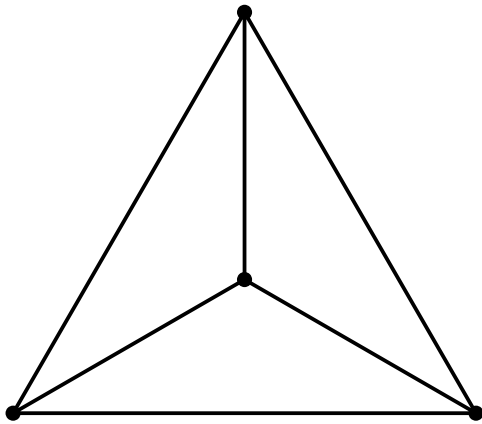
Theorem (Euler's formula). *Let G be a connected plane graph with v vertices, e edges and f faces. Then*

$$f + v = e + 2.$$

Proof We shall apply induction on the number of faces. If $f = 1$ the graph cannot contain a cycle: so it is tree and $v = e + 1$. Therefore the formula holds.

Now suppose our graph has more than one face. Choose a cycle in the graph and an edge in this cycle. The cycle separates the plane into two so the edge lies in two different faces. If we remove that edge we decrease the number of edges by 1 and the number of faces by 1 without changing v . So by induction the formula holds for the larger graph. □

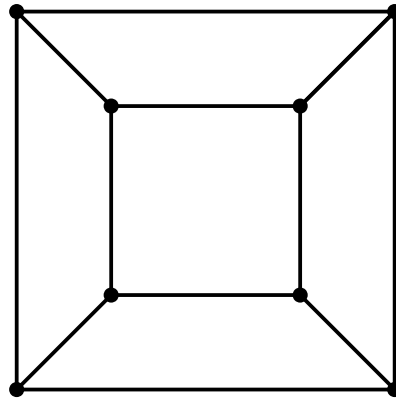
Examples.



$$f = 4$$

$$v = 4$$

$$e = 6$$



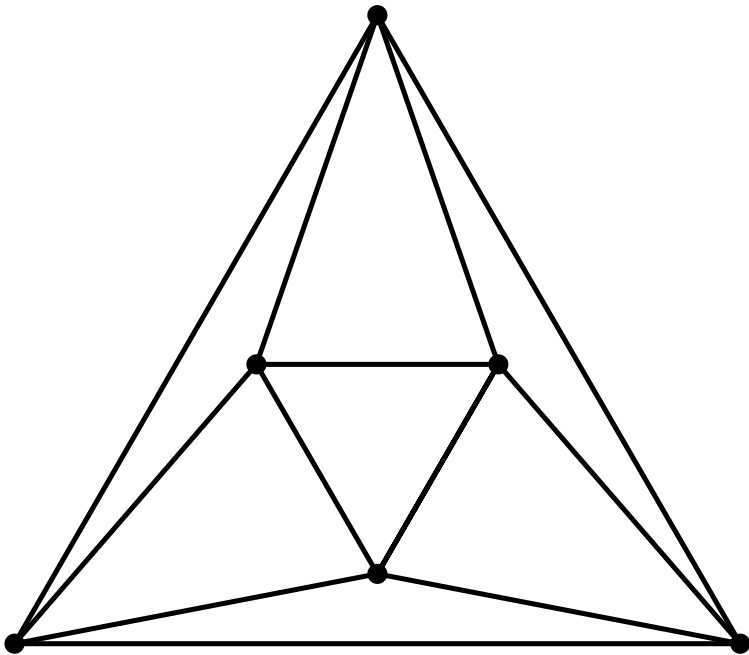
$$f = 6$$

$$v = 8$$

$$e = 12$$

The cube has faces with 4 edges so it is not a **maximal** plane graph.
We can add some edges crossing the faces.

A maximal plane graph with at least 3 vertices has only triangular faces.



Lemma (Maximal plane graphs). A maximal plane graph on $n \geq 3$ vertices has $3n - 6$ edges.

Proof Each edge belongs to 2 faces and each face has 3 edges. So $2e = 3f$.

By Euler's formula

$$\frac{2}{3}e + v = e + 2$$

and this gives

$$v - 2 = \frac{1}{3}e.$$

□

Tetrahedron: 4 faces, 6 edges.

Octahedron: 8 faces, 12 edges.

Icosahedron: 20 faces, $20 \times 3 - 2 = 30$ edges.

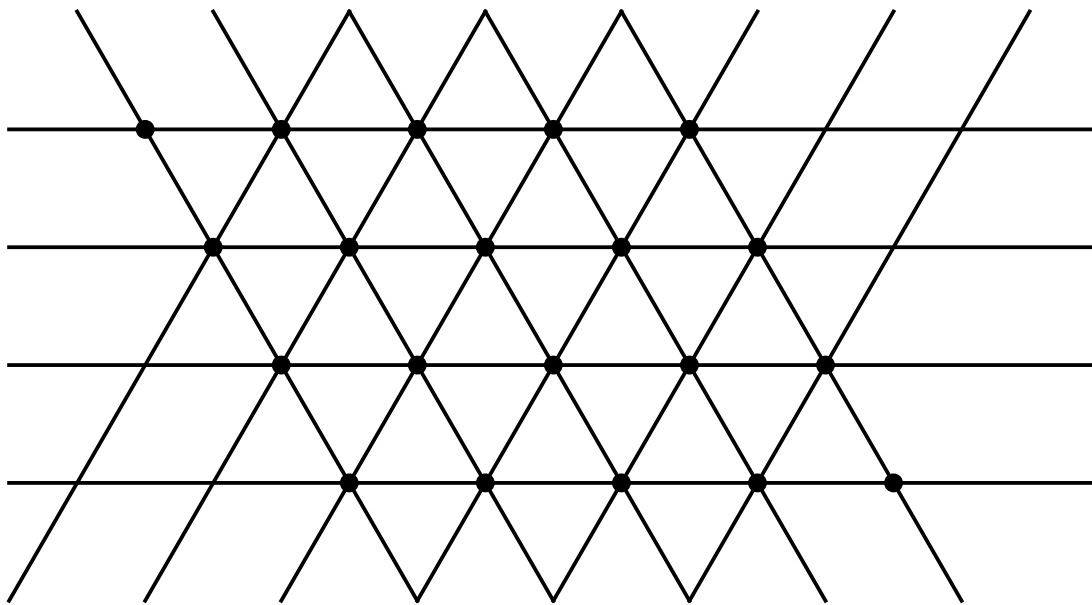
Fishy?

Each edge belongs to 2 faces and each face has 3 edges. So $2e = 3f$.

Consider the ordered pairs (f, e) where f is a face and e is an edge of that face. The total number of ordered pairs is $3f$ and it is also $2e$.

Lemma (Maximal plane graphs). *A maximal plane graph on $n \geq 3$ vertices has $3n - 6$ edges.*

As a consequence we get that every plane graph has a vertex with degree less than 6. You can't improve upon this in the limit.

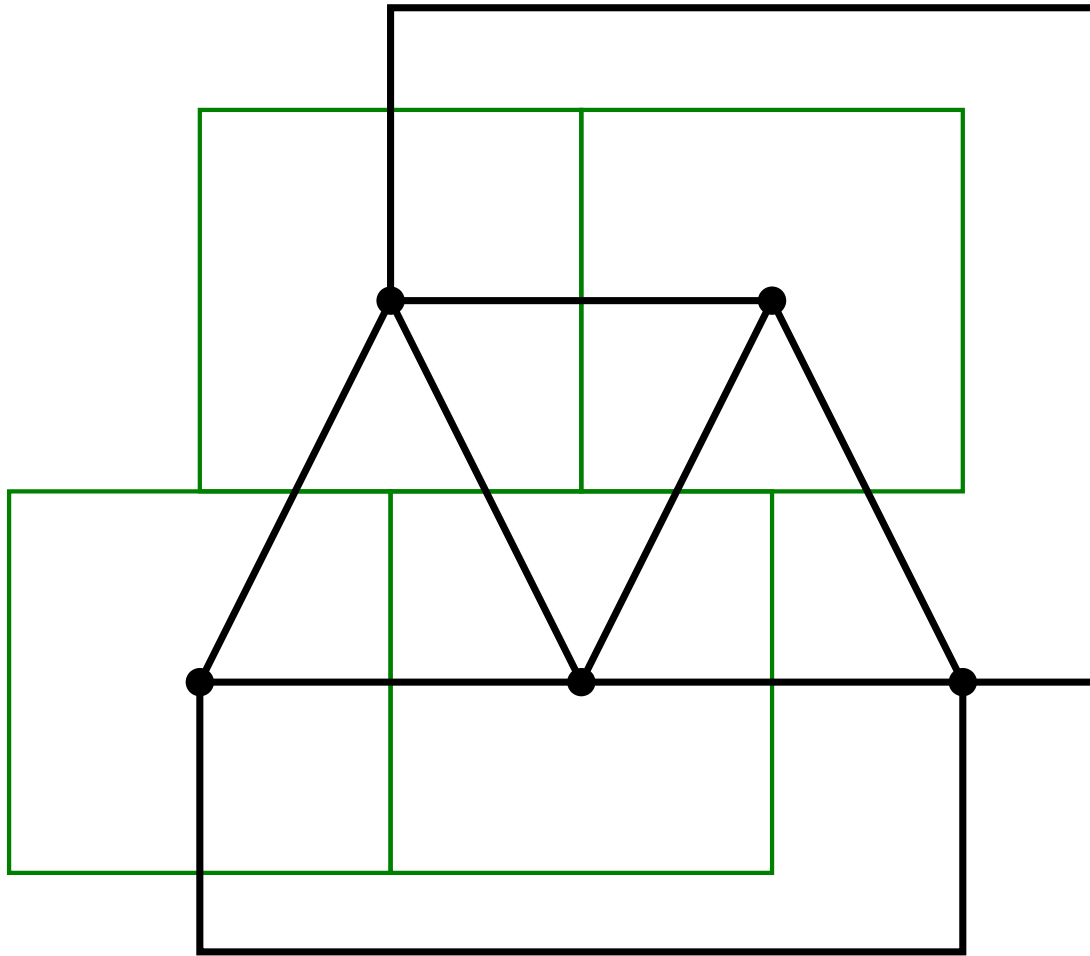


As we remarked at the start of the course, one of the most famous problems in combinatorics was the 4-colour problem.

Is it possible to colour every plane map with 4 colours so that neighbouring countries have different colours?

We can rephrase this in terms of plane graphs. Is it possible to colour the **vertices** of a plane graph with 4 colours so that adjacent vertices have different colours?

We can rephrase this in terms of plane graphs. Is it possible to colour the **vertices** of a plane graph with 4 colours so that adjacent vertices have different colours?



This was answered in the affirmative by Appel and Haken in 1976.
They used a computer aided reduction of hundreds of critical cases.

We shall not prove it in this course.

We shall prove that 5 colours are sufficient.

Theorem (The 6 Colour Theorem). *Every plane graph can be coloured with 6 colours.*

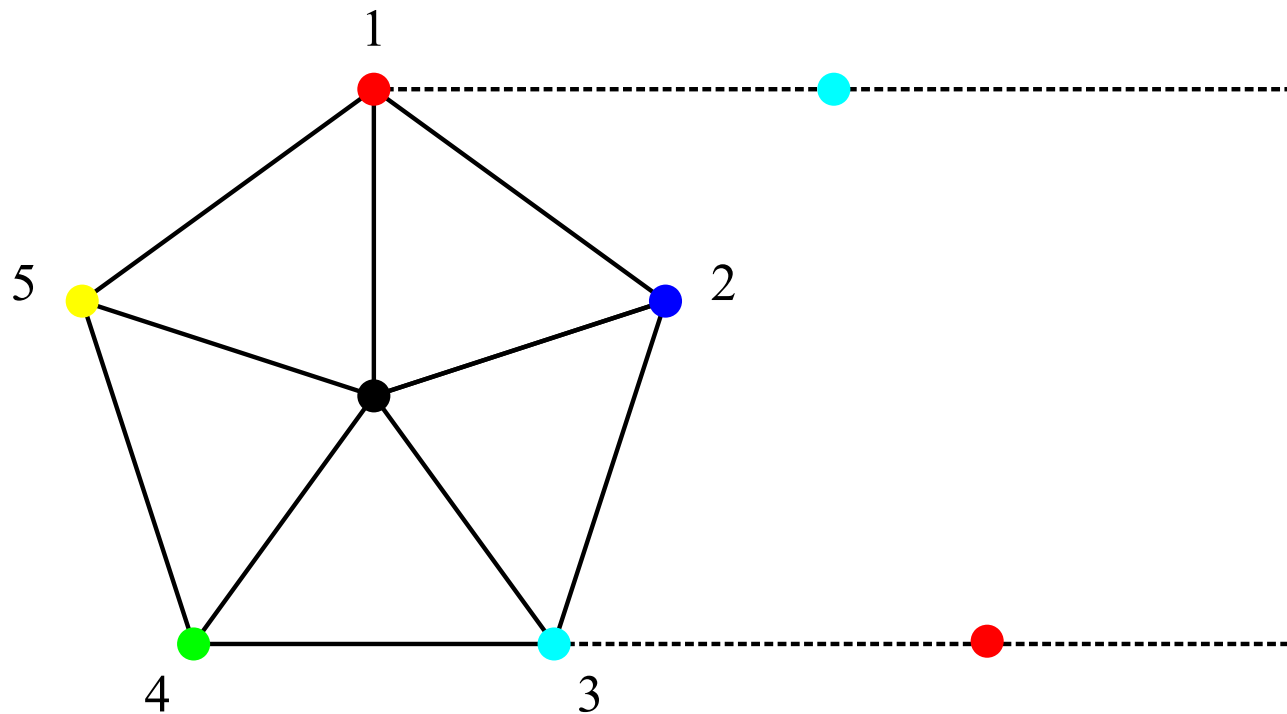
Proof Find a vertex of degree at most 5. Colour the rest by induction. Now colour the one you have left. □

The 5 colour theorem also uses the fact that we have a vertex of degree 5 (or less). We colour the rest of the graph by induction and then try to handle the last vertex. The trick is to build what are called Kempe chains if the last vertex can't be coloured.

Theorem (The 5 Colour Theorem). *Every plane graph can be coloured with 5 colours.*

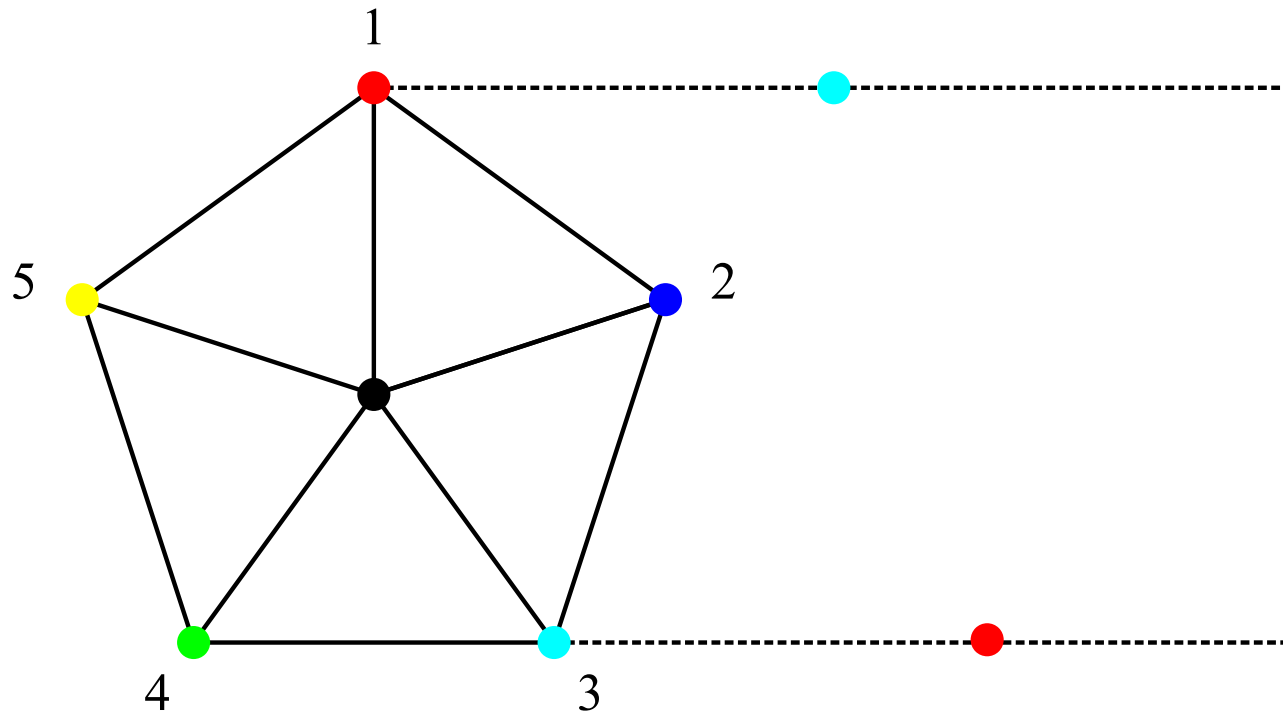
Proof Suppose that there is a plane graph which cannot be coloured with 5 colours and choose the one with fewest vertices. Find a vertex x of degree at most 5. Colour the rest by induction. If x has degree 4 or less we can colour it. If not, look at the 5 neighbours of x and label them 1, 2, 3, 4, 5 cyclicly. If they are not coloured with all 5 different colours then we can again colour x .

So we may assume that they all have different colours (labelled the same as the vertices).



Consider all the vertices of the whole graph coloured 1 and 3. These form an induced subgraph. If vertices 1 and 3 belong to different components of this subgraph then we can switch colours 1 and 3 in one of those components so now vertices 1 and 3 get the same colour. Then we have a spare colour for x .

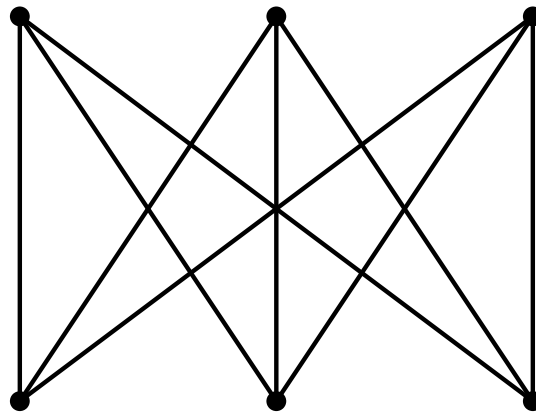
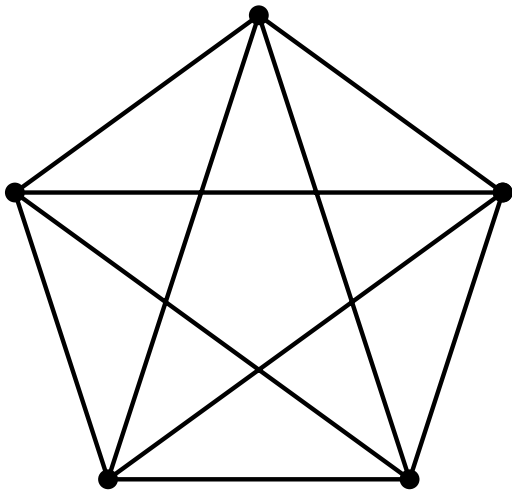
So we can assume that there is a path coloured with 1 and 3 from vertex 1 to vertex 3.



But now if we apply the same argument with colours 2 and 4 we generate a 2-4 coloured path from vertex 2 to vertex 4. But this path must cross the 1-3 coloured path at a vertex which has no colour. \square

Non-planar graphs

There are two famous non-planar graphs that turn up in recreational books. The complete graph K_5 and the complete bipartite graph $K_{3,3}$.

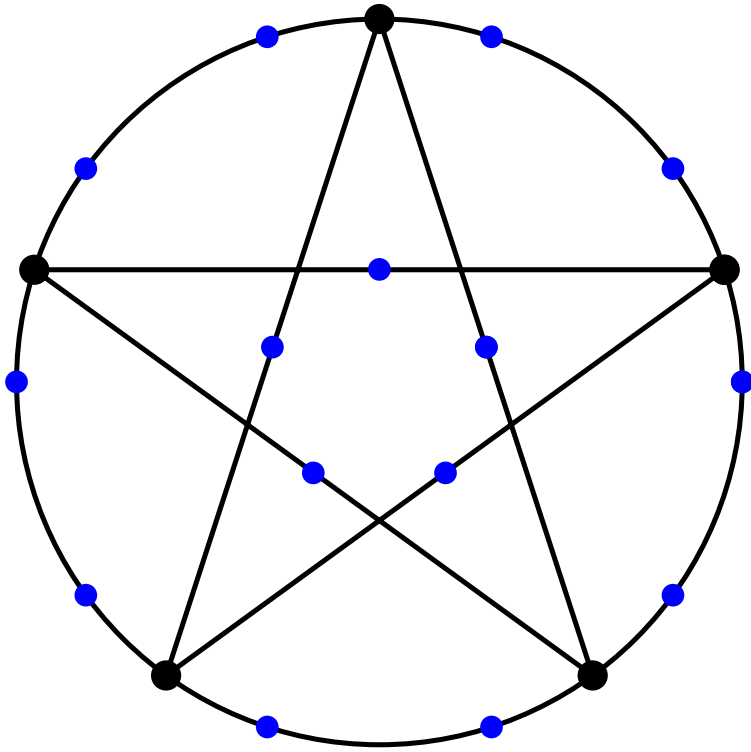


Theorem (Non-planar graphs). *The graphs K_5 and $K_{3,3}$ are not planar.*

Proof A planar graph on n vertices has at most $3n - 6$ edges. If $n = 5$ this maximum is 9. But K_5 has 10 edges.

The graph $K_{3,3}$ has $v = 6$ vertices and $e = 9$ edges. If we draw it in the plane then by Euler's formula it has $f = 5$ faces. But since it is a bipartite graph it contains no triangles: so each face has **at least** 4 edges. Each edge belongs to **at most** 2 faces. This means that the number of edges is at least twice the number of faces giving a contradiction. □

Since K_5 is not planar neither is the following graph in which we replace some edges by paths.



Such a graph is called a subdivision of K_5 . The two graphs K_5 and $K_{3,3}$ turn out to be the only (minimal) obstructions to planarity apart from this subdivision issue.

Theorem (Kuratowski's Theorem). *A graph fails to be planar if and only if it contains a subdivision of K_5 or $K_{3,3}$.*

The proof is not hard but it is a pain.

Summary

A **planar** graph is a graph that can be drawn in the plane in such a way that edges don't cross one another.

Theorem (Euler's formula). *Let G be a connected plane graph with v vertices, e edges and f faces. Then*

$$f + v = e + 2.$$

Lemma (Maximal plane graphs). *A maximal plane graph on $n \geq 3$ vertices has $3n - 6$ edges.*

Theorem (The 5 Colour Theorem). *Every plane graph can be coloured with 5 colours.*

Theorem (Non-planar graphs). *The graphs K_5 and $K_{3,3}$ are not planar.*

Theorem (Kuratowski's Theorem). *A graph fails to be planar if and only if it contains a subdivision of K_5 or $K_{3,3}$.*

Appendix. Boolean functions

A **Boolean** function is a function from $\{0, 1\}^n \rightarrow \{0, 1\}$. To each input sequence (x_1, x_2, \dots, x_n) of bits it assigns an output bit. **Bit means binary digit.**

These functions originally arose in connection with propositional logic. Each coordinate is a statement: “The sun is shining.”: “I feel happy.” They now form the basis for circuit design.

Each statement might be true or false. The function f examines a compound proposition such as “The sun is shining **and** I feel happy.”

Each statement might be true or false. The function f examines a compound proposition such as “The sun is shining **and** I feel happy.”

It outputs true or false depending upon the logical structure of the combination.

The **AND** function only outputs true if each coordinate is true. If one of the coordinate statements is false the compound proposition is false.

We shall replace FALSE by 0 and TRUE by 1. So the **AND** function of two variables looks like

$$(0, 0) \mapsto 0$$

$$(0, 1) \mapsto 0$$

$$(1, 0) \mapsto 0$$

$$(1, 1) \mapsto 1$$

For $n = 2$ each Boolean function assigns an output to each of

$$(0, 0), (0, 1), (1, 0), (1, 1).$$

So there are 2^4 possible functions. In general for sequences of length n there are 2^{2^n} .

For $n = 2$ some functions are

(0, 0)	0	1	0	0
(0, 1)	0	1	0	1
(1, 0)	1	0	0	1
(1, 1)	1	0	1	1
(X, Y)	X	\bar{X}	$X \wedge Y$	$X \vee Y$

\bar{X} means **NOT** X also written $\neg X$.

$X \wedge Y$ means X **AND** Y .

$X \vee Y$ means X **OR** Y .

The **AND** and **OR** functions can be thought of as binary operations and it is easy to check that they are associative so it makes sense to write

$$X \wedge Y \wedge Z$$

and thereby create Boolean functions of more than two variables.

$X \wedge Y \wedge Z$ outputs 1 only at the sequence (1, 1, 1).

We use the word **conjunction** for a string connected by **AND** and **disjunction** for a string connected by **OR**.

An elementary conjunction is an expression using **AND** applied to atoms and their negatives: for example

$$X_1 \wedge X_3 \wedge X_7 \quad \text{or} \quad X_1 \wedge \overline{X_2} \wedge X_5.$$

An elementary disjunction is the same but with \vee .

An elementary conjunction such as

$$X_1 \wedge \overline{X_2} \wedge X_3 \wedge \dots \wedge X_n$$

which involves every variable or its negation exactly once, outputs 1 at just a single sequence: in this case

$$(1, 0, 1, \dots, 1).$$

This means that we can write any Boolean function as a disjunction of elementary conjunctions.

Suppose that f outputs 1 at $(1, 0, 0)$, $(0, 1, 1)$ and $(1, 1, 1)$ but nowhere else.

Then the function is

$$(X_1 \wedge \bar{X}_2 \wedge \bar{X}_3) \vee (\bar{X}_1 \wedge X_2 \wedge X_3) \vee (X_1 \wedge X_2 \wedge X_3).$$

A disjunction of elementary conjunctions is called a **disjunctive normal form** or **DNF**.

Observe that

$$\overline{X \wedge Y} = \overline{X} \vee \overline{Y} \quad \text{and}$$

$$\overline{X \vee Y} = \overline{X} \wedge \overline{Y}.$$

This means that every function can be written as a conjunction of elementary disjunctions: it can be written in **conjunctive normal form** or **CNF**. We just write $\neg f$ in DNF and when we negate, we switch **AND** with **OR**.

We say that the functions \wedge, \vee, \neg together form a complete basis for the Boolean functions meaning that any such function can be built from them. In fact \wedge, \neg form a complete basis since we can make \vee from them.

It was noticed in the early days of computing that there are single functions that form a basis. Probably the best known one is **NAND**. It is usually written \uparrow .

$$X \uparrow Y = \neg(X \wedge Y).$$

In order to prove that this function alone forms a complete basis it suffices to express \neg because then we can get \wedge . But $X \uparrow X = \neg(X \wedge X) = \neg X$.

In a similar way **NOR** forms a complete basis.

We say that a formula in variables X_1, X_2, \dots, X_n is **satisfiable** if there is a choice of values for the variables which makes the formula true. In other words, a Boolean function f is satisfiable if there are sequences to which it assigns 1.

If we are presented with a function in DNF it is easy to check whether it is satisfiable. The formula will yield 1 if one of the elementary conjunctions is true.

An elementary conjunction $X_1 \wedge X_3 \wedge \dots$ will always be satisfiable unless it contains both X_k and $\neg X_k$ for some k . So we can look at each conjunction in turn and see whether it contains a contradiction.

It is very much more difficult to check the satisfiability of a CNF. We need to find an assignment of 0 and 1 to the variables so that every elementary disjunction is true.

For example:

$$(\overline{X_1} \vee X_3) \wedge (\overline{X_2} \vee \overline{X_3}) \wedge (X_1 \vee \overline{X_4}) \wedge (X_2 \vee \overline{X_4}) \wedge (X_4 \vee X_1)$$

is satisfied only by (1, 0, 1, 0) whereas

$$(\overline{X_1} \vee X_3) \wedge (\overline{X_2} \vee \overline{X_3}) \wedge (X_1 \vee \overline{X_4}) \wedge (X_2 \vee \overline{X_4}) \wedge (X_4 \vee X_2)$$

is satisfied only by (0, 1, 0, 0). **Exercise.**

You can't look at the disjunctions one at a time: you need consistency.

However if you tell me that you have found a sequence where my CNF is satisfied I can easily check whether you are correct because I just need to check each disjunction to see whether they are all true.

The satisfiability problem for CNFs is an example of a problem in **NP**.

The acronym stands for non-deterministic polynomial time and the original definition depends upon the idea of a non-deterministic Turing machine.

Roughly speaking, being in NP means that you can check whether someone has found a solution, without using more steps than a polynomial function of the size of the input: in this case the number n of variables.

If we restrict to the case in which each disjunction has k variables or their negations, the problem of satisfiability is called k -SAT.

In this case there are at most $(2n)^k$ possible clauses and each one can be checked for correctness in k steps. So k -SAT can be **checked** in polynomial time.

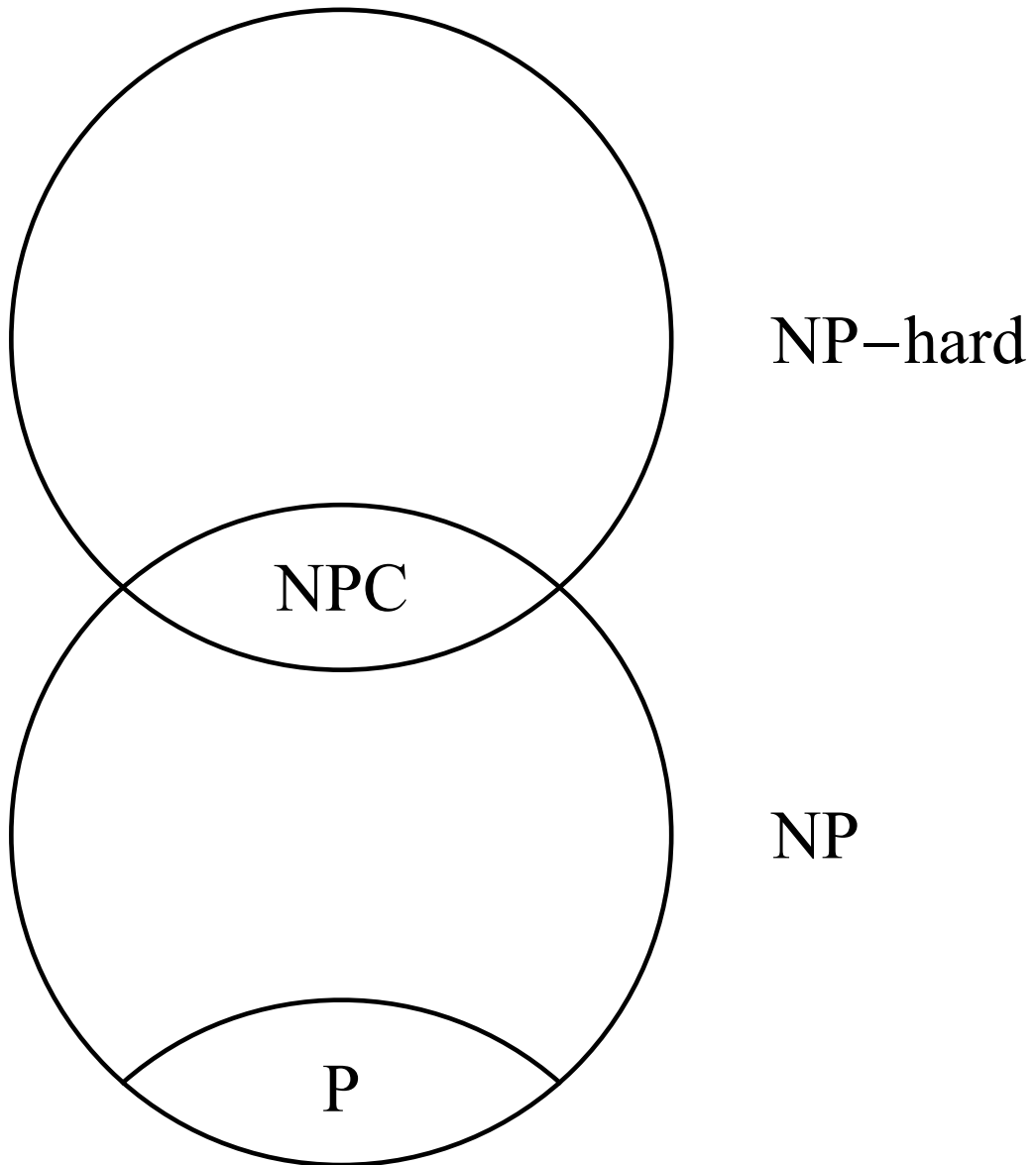
If $k = 2$ the problem can be **solved** with a number of steps that is only polynomial in the number of variables n . So the example above was actually quite easy in computer science terms. The 2-SAT problem is said to be in the space P.

If $k = 3$ then the satisfiability problem is **NP-hard**: meaning that if you could solve it efficiently then you could solve all other NP problems more or less equally efficiently. It is at least as hard as any problem in NP.

More precisely, you can reduce any other problem in NP to the 3-SAT problem, in polynomial time.

3-SAT is **NP-complete** meaning that it is in NP itself and is NP-hard. It has exactly the maximum difficulty of a problem in NP.

These spaces of problems look as follows:



The 2-SAT problem can be solved in polynomial time because we can reduce it to the problem of finding whether there is a path between a given pair of vertices in a graph.

The latter is a simple but quite useful example. If the graph has n vertices there are $(n - 2)!$ possible routes between a given pair of vertices. It would be extremely time consuming to check each route to see whether all the edges are actually present.

Instead we can check in the following way. Suppose the initial vertex is x and the target vertex is y . Put x into box 1.

Look at all n vertices to see which are neighbours of x . Place the neighbours into box 2.

Now check the neighbours of each of these and place any vertex that you haven't already seen into box 3.

Repeat the process.

If you hit y then you know that there is a path from x to y .

After n steps or fewer you will not produce any new vertices. If you haven't seen y by this time then there is **no** path from x to y .

Each step requires at most n^2 checks.

Summary

A **Boolean** function is a function from $\{0, 1\}^n \rightarrow \{0, 1\}$. To each input sequence (x_1, x_2, \dots, x_n) of bits it assigns an output bit.

Examples

$(0, 0)$	0	1	0	0
$(0, 1)$	0	1	0	1
$(1, 0)$	1	0	0	1
$(1, 1)$	1	0	1	1
(X, Y)	X	\overline{X}	$X \wedge Y$	$X \vee Y$

A disjunction of elementary conjunctions is called a **disjunctive normal form** or **DNF**.

A conjunction of elementary disjunctions is called a **conjunctive normal form** or **CNF**.

Each function can be written in DNF or CNF.

It is easy to check the satisfiability of a DNF.

NP is the space of problems for which we can check the validity of a solution in polynomial time.

If a problem is in NP and is NP-hard then it is NP-complete.

Satisfiability for CNF's with 3 or more symbols in each disjunction is NP-complete.