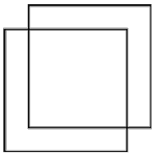
Combinatorics and Graph Theory 1 & 2

Irena Penev

Computer Science Institute
of Charles University (IÚUK)



INFORMATICKÝ ÚSTAV
UNIVERZITY KARLOVY
Matematicko–fyzikální fakulta
Univerzita Karlova

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

Preface

This manuscript is based on a set of lecture notes that I prepared and used for teaching Combinatorics and Graph Theory 1 & 2 at the Faculty of Mathematics and Physics, Charles University, during the academic years 2020/21 and 2021/22. The first eight chapters cover the material from the first semester (Combinatorics and Graph Theory 1), and the remaining eleven chapters cover the material from the second semester (Combinatorics and Graph Theory 2). The main prerequisite for this two-semester sequence is the successful completion of the Discrete Mathematics course (in particular, familiarity with basic graph theory is assumed). Moreover, some chapters assume familiarity with Mathematical Analysis and Linear Algebra. Proofs that appear in this manuscript were taken (often with modification) from a number of texts, listed in the bibliography. However, any errors that remain are mine alone. It is my hope that this manuscript will be of use to the students taking these two courses in the future, as well as to the instructors teaching them.

September 2022 Irena Penev

Contents

Chapter 1

Asymptotic notation. Estimates of factorials and binomial coefficients

1.1 Asymptotic notation

We often need to make statements such as that, for example, the function n^2 is "greater" than the function $1000n$, and "roughly the same" as the function $n^2 + n\sqrt{n}$. Let us try to formalize this.

Given functions $f, g : \mathbb{N} \to \mathbb{R}$ (in practice, we generally assume that f, g are positive-valued), notation

$$f(n) \quad = \quad O(g(n))$$

means that there exist constants $n_0 \in \mathbb{N}$ and $C \in \mathbb{R}$ such that for all $n \in \mathbb{N}$, if $n \geq n_0$, then

$$|f(n)| \quad \leq \quad Cg(n).$$

This is illustrated in Figure 1.1.

Example 1.1.1.

1. $10n^2 + 5 = O(n^2)$;
2. $\ln n + 5 = O(\log n)$;
3. $\ln n + 5 = O(n)$;
4. $n\sqrt{n} = O(n^2)$.

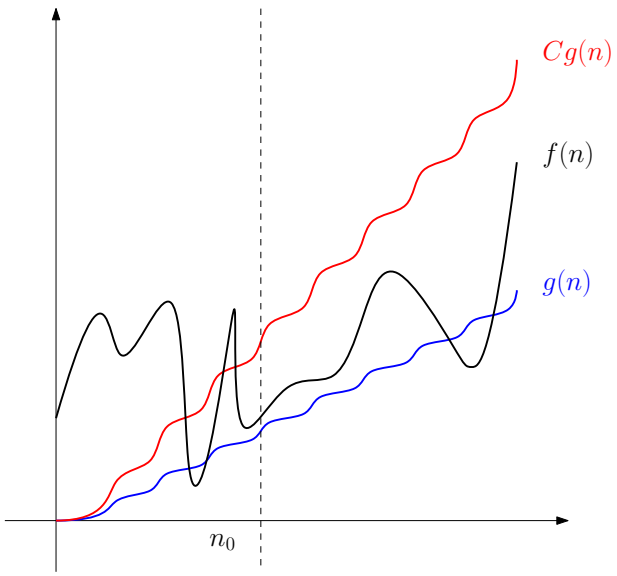


Figure 1.1: $f(n) = O(g(n))$.

There are several other often-used kinds of notation, summarized below.

Notation	Definition
$f(n) = O(g(n))$	$\exists n_0 \in \mathbb{N}$, $C \in \mathbb{R}$ s.t. $\forall n \in \mathbb{N}$, if $n \geq n_0$ then $\lvert f(n)\rvert \leq Cg(n)$
$f(n) = o(g(n))$	$\lim\limits_{n\to\infty} \frac{f(n)}{g(n)} = 0$
$f(n) = \Omega(g(n))$	$g(n) = O(f(n))$
$f(n) = \Theta(g(n))$	$f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$
$f(n) \sim g(n)$	$\lim\limits_{n\to\infty} \frac{f(n)}{g(n)} = 1$

Note that $f(n) = \Theta(g(n))$ is **not** the same as $f(n) \sim g(n)$. For instance, $2n^2 = \Theta(n^2)$, but $2n^2 \not\sim n^2$.

Example 1.1.2.

1. $12n^2 + n = O(n^2)$
2. $n = o(n^2)$
3. $\frac{1}{12}n^3 = \Omega(n^2)$

4. $\frac{1}{12}n^2 = \Theta(n^2)$

5. $5n^2 + n \sim 5n^2 + \log n$

Further $f(n) = g(n) + O(h(n))$ means that $f(n) - g(n) = O(h(n))$. For example, $n^4 + 3n^2 = n^4 + O(n^2)$ because $3n^2 = O(n^2)$. We use similar notation for the symbols o, Ω, and Θ from the table above.

Here is some more commonly used notation.

Notation	Meaning
$O(1)$	constant (or bounded above by a constant)
$O(\log n)$	logarithmic (or sublogarithmic)
$O(n)$	linear (or sublinear)
$O(n^2)$	quadratic (or subquadratic)
$O(n^3)$	cubic (or subcubic)
$n^{O(1)}$	polynomial (or subpolynomial)
$2^{O(n)}$	exponential (or subexponential)

1.2 Estimating factorials

For a positive integer n, we define $n!$ (read "n factorial") to be

$$n! \quad := \quad n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 2 \cdot 1.$$

Furthermore, as a convention, we set $0! = 1$.

$n!$ is the number of ways that n distinct objects can be arranged in a sequence: there are n choices for the first term of the sequence, $n-1$ choices for the second, $n-2$ for the third, etc. For instance, there are $3! = 6$ ways to arrange the elements of the set $\{a, b, c\}$ in a sequence, namely:

(1) a, b, c (3) b, a, c (5) c, a, b

(2) a, c, b (4) b, c, a (6) c, b, a

For small values of n, computing $n!$ is quite straightforward:

- $0! = 1$
- $1! = 1$
- $2! = 2 \cdot 1 = 2$
- $3! = 3 \cdot 2 \cdot 1 = 6$

- $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$
- $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$
- $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$
- $7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5040$
- $8! = 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 40320$
- $9! = 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 362880$

However, as we see from the list above, $n!$ is a very fast increasing function, and computing it for even moderately large n is impractical. Nevertheless, in applications, it is often useful to know roughly how big $n!$ is, that is, how it compares to various other functions of n. Obviously,[1]

$$n! \quad \leq \quad n^n$$

for all non-negative integers n. In this section, we will obtain two better estimates for $n!$, as follows:

(i) $n^{n/2} \leq n! \leq (\frac{n+1}{2})^n$ for all non-negative integers n;

(ii) $e(\frac{n}{e})^n \leq n! \leq en(\frac{n}{e})^n$ for all positive integers n.

For non-negative real numbers x and y, the *arithmetic mean* of x and y is $\frac{x+y}{2}$, and the *geometric mean* of x and y is $\sqrt{xy}$. To prove (i), we will use the inequality of arithmetic and geometric means (below).

Inequality of arithmetic and geometric means. *All non-negative real numbers x and y satisfy*

$$\sqrt{xy} \quad \leq \quad \tfrac{x+y}{2}.$$

Proof. For non-negative real numbers x and y, we have the following sequence of equivalences:

$$\begin{aligned}
& (\sqrt{x} - \sqrt{y})^2 && \geq && 0 \\
\iff \quad & x - 2\sqrt{xy} + y && \geq && 0 \\
\iff \quad & x + y && \geq && 2\sqrt{xy} \\
\iff \quad & \tfrac{x+y}{2} && \geq && \sqrt{xy}.
\end{aligned}$$

Since the first inequality above is obviously true, so is the last one. □

[1] Recall that for all real numbers r, we have that $r^0 = 1$. In particular, $0^0 = 1$.

We are now ready to prove (i).

Theorem 1.2.1. *For all non-negative integers n, the following holds:*

$$n^{n/2} \quad \leq \quad n! \quad \leq \quad (\tfrac{n+1}{2})^n$$

Proof. For $n = 0$ and $n = 1$, the statement is obviously true. So, fix an integer $n \geq 2$.

We first prove the upper bound, as follows:

$$\begin{aligned}
n! &= \sqrt{\Big(n \cdot (n-1) \cdot \cdots \cdot 2 \cdot 1\Big)\Big(1 \cdot 2 \cdot \cdots \cdot (n-1) \cdot n\Big)} \\
&= \sqrt{\Big(n \cdot 1\Big)\Big((n-1) \cdot 2\Big) \ldots \Big(2 \cdot (n-1)\Big)\Big(1 \cdot n\Big)} \\
&= \Big(\sqrt{n \cdot 1}\Big)\Big(\sqrt{(n-1) \cdot 2}\Big) \ldots \Big(\sqrt{2 \cdot (n-1)}\Big)\Big(\sqrt{1 \cdot n}\Big) \\
&\overset{(*)}{\leq} \tfrac{n+1}{2} \cdot \tfrac{(n-1)+2}{2} \cdot \cdots \cdot \tfrac{2+(n-1)}{2} \cdot \tfrac{1+n}{2} \\
&= (\tfrac{n+1}{2})^n,
\end{aligned}$$

where (*) follows from the inequality of arithmetic and geometric means.

It remains to prove the lower bound. First, we claim that for all $i \in \{1, \ldots, n\}$, we have that

$$i(n+1-i) \quad \geq \quad n.$$

Indeed, if $i = 1$ or $i = n$, then $i(n+1-i) = n$. On the other hand, for $i \in \{2, \ldots, n-1\}$, we have that $\min\{i, n+1-i\} \geq 2$ and $\max\{i, n+1-i\} \geq \frac{i+(n+1-i)}{2} \geq \frac{n}{2}$, and consequently,

$$i(n+1-i) \quad = \quad \min\{i, n+1-i\} \cdot \max\{i, n+1-i\} \quad \geq \quad 2 \cdot \tfrac{n}{2} \quad = \quad n,$$

as we had claimed. We now compute:

$$\begin{aligned} n! &= \sqrt{\Big(1\cdot 2\cdot\cdots\cdot(n-1)\cdot n\Big)\Big(n\cdot(n-1)\cdot\cdots\cdot 2\cdot 1\Big)} \\ &= \sqrt{\Big(1\cdot n\Big)\Big(2\cdot(n-1)\Big)\cdot\cdots\cdot\Big((n-1)\cdot 2\Big)\Big(n\cdot 1\Big)} \\ &= \sqrt{\prod_{i=1}^{n}\Big(\underbrace{i\cdot(n+1-i)}_{\geq n}\Big)} \\ &\geq \sqrt{n^n} \\ &= n^{n/2}, \end{aligned}$$

which is what we needed. □

It remains to prove (ii). We begin with the following proposition.

Proposition 1.2.2. *For all real numbers x, the following inequality holds:*

$$1+x \leq e^x.$$

Proof. Let $f:\mathbb{R}\to\mathbb{R}$ be given by $f(x)=e^x-x-1$. Then $f'(x)=e^x-1$, and we have the following table:

x	$-\infty$	$(-\infty,0)$	0	$(0,+\infty)$	$+\infty$
$f'(x)$		$-$	0	$+$	
$f(x)$		$\searrow$	min	$\nearrow$	

So, $f(x)$ reaches a global minimum at $x=0$. Since $f(0)=0$, it follows that $f(x)\geq 0$ for all $x\in\mathbb{R}$, and the result follows. □

We will also need the well-known fact that

$$(1+\tfrac{1}{n})^n \leq e$$

for all positive integers n.[2]

We are now ready to prove (ii).

Theorem 1.2.3. *For all positive integers n, the following holds:*

$$e(\tfrac{n}{e})^n \;\leq\; n! \;\leq\; en(\tfrac{n}{e})^n.$$

Proof. We proceed by induction on n. The claim is clearly true for $n = 1$. Now, fix a positive integer n, and assume inductively that $e(\frac{n}{e})^n \leq n! \leq en(\frac{n}{e})^n$. We must show that $e(\frac{n+1}{e})^{n+1} \leq (n+1)! \leq e(n+1)(\frac{n+1}{e})^{n+1}$.

We first establish the upper bound, that is, we prove that $(n+1)! \leq e(n+1)(\frac{n+1}{e})^{n+1}$. We first compute:

$$\begin{aligned}
(n+1)! &= (n+1)\cdot n! \\
&\leq (n+1)\cdot en(\tfrac{n}{e})^n && \text{by the induction hypothesis} \\
&= \Big(e(n+1)(\tfrac{n+1}{e})^{n+1}\Big)\cdot(\tfrac{n}{n+1})^{n+1}e.
\end{aligned}$$

It now remains to prove that $(\frac{n}{n+1})^{n+1}e \leq 1$, for then we will obtain precisely the inequality that we need. We prove this as follows:

$$\begin{aligned}
(\tfrac{n}{n+1})^{n+1}e &= (1-\tfrac{1}{n+1})^{n+1}e \\
&\leq (e^{-\frac{1}{n+1}})^{n+1}e && \text{by Proposition 1.2.2, for } x = -\tfrac{1}{n+1} \\
&= 1.
\end{aligned}$$

It remains to establish the lower bound, i.e. to prove that $e(\frac{n+1}{e})^{n+1} \leq (n+1)!$. For this, we compute:

$$\begin{aligned}
e(\tfrac{n+1}{e})^{n+1} &= (n+1)(\tfrac{n}{e})^n\cdot(1+\tfrac{1}{n})^n \\
&\leq (n+1)(\tfrac{n}{e})^n\cdot e && \text{because } (1+\tfrac{1}{n})^n \leq e \\
&\leq (n+1)\cdot n! && \text{by the induction hypothesis} \\
&= (n+1)!,
\end{aligned}$$

which is what we needed. □

[2] As you saw in Analysis, the sequence $\{(1+\frac{1}{n})^n\}_{n=1}^{\infty}$ is strictly increasing and bounded above, and so by the Monotone Sequence Theorem, it converges. The constant e is defined as the limit of this sequence, i.e. $e := \lim_{n\to\infty}(1+\frac{1}{n})^n$, and the inequality follows.

We complete this section by giving the following formula (without proof).

Stirling's formula. $\lim\limits_{n\to\infty} \frac{\sqrt{2\pi n}\,(\frac{n}{e})^n}{n!} = 1.$

Using the notation that we introduced in section 1.1, Stirling's formula states that

$$n! \;\sim\; \sqrt{2\pi n}\,\left(\frac{n}{e}\right)^n.$$

So, for very large values of n, the function $f(n) = \sqrt{2\pi n}\,(\frac{n}{e})^n$ is a good approximation for $n!$.

1.3 Estimating binomial coefficients

For integers n and k such that $n \geq k \geq 0$, we define the number $\binom{n}{k}$, read "n choose k," as follows:

$$\binom{n}{k} \;=\; \frac{n(n-1)\ldots(n-k+1)}{k\cdot(k-1)\cdot\ldots\cdot 1} \;=\; \prod_{i=0}^{k-1} \frac{n-i}{k-i}.$$

Note that this implies that

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

and consequently,

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

$\binom{n}{k}$ is the number of k-element subsets of an n-element set.[3] For example, the number of 3-element subsets of the 5-element set $\{a, b, c, d, e\}$ is $\binom{5}{3} = 10$; those subsets are:

(1) $\{a, b, c\}$ (3) $\{a, b, e\}$ (5) $\{a, c, e\}$ (7) $\{b, c, d\}$ (9) $\{b, d, e\}$

(2) $\{a, b, d\}$ (4) $\{a, c, d\}$ (6) $\{a, d, e\}$ (8) $\{b, c, e\}$ (10) $\{c, d, e\}$

We note that for all non-negative integers n, we have that $\binom{n}{0} = 1$. In particular, $\binom{0}{0} = 1$.

Numbers $\binom{n}{k}$ are called *binomial coefficients*. You are already familiar with the Binomial theorem (stated below).

[3] Indeed, there are $n(n-1)\ldots(n-k+1)$ sequences of k different elements of an n-element set: there are n ways to select the first element, $n-1$ ways to select the second element, ..., and $n-k+1$ ways to select the k-th element. Since every k-element set can be ordered in $k!$ ways, there are exactly $\frac{n(n-1)\ldots(n-k+1)}{k!} = \binom{n}{k}$ many k-element subsets of an n-element set.

Binomial theorem. *For all integers $n \geq 0$, and all real numbers x and y, the following holds:*

$$\begin{aligned}(x+y)^n &= \sum_{k=0}^{n} \tbinom{n}{k} x^{n-k} y^k \\ &= \tbinom{n}{0} x^n + \tbinom{n}{1} x^{n-1} y + \cdots + \tbinom{n}{n-1} x y^{n-1} + \tbinom{n}{n} y^n.\end{aligned}$$

Similarly to factorials, binomial coefficients are easy to compute for small values of n and k. However, even for moderately large n and k, computing $\binom{n}{k}$ becomes impractical. So, as in the case of factorials, we would like to obtain some useful estimates (convenient upper and lower bounds) for binomial coefficients.

1.3.1 Estimating the binomial coefficient $\binom{n}{k}$

Our goal is to prove the following theorem.

Theorem 1.3.1. *For all integers n and k such that $n \geq k \geq 1$, the following holds:*

$$\left(\tfrac{n}{k}\right)^k \leq \tbinom{n}{k} \leq \left(\tfrac{en}{k}\right)^k.$$

Theorem 1.3.1 readily follows from Propositions 1.3.2 and 1.3.3 (below). Proposition 1.3.2 establishes the lower bound from Theorem 1.3.1, and Proposition 1.3.3 establishes the upper bound.[4]

Proposition 1.3.2. *For all integers n and k such that $n \geq k \geq 1$, we have that*

$$\left(\tfrac{n}{k}\right)^k \leq \tbinom{n}{k}$$

Proof. Fix integers n, k such that $n \geq k \geq 1$. We observe that for all $i \in \{0, \ldots, k-1\}$, we have that $\frac{n-i}{k-i} \geq \frac{n}{k}$,[5] and so

$$\tbinom{n}{k} = \prod_{i=0}^{k-1} \tfrac{n-i}{k-i} \geq \prod_{i=0}^{k-1} \tfrac{n}{k} = \left(\tfrac{n}{k}\right)^k,$$

which is what we needed. □

[4] In fact, the inequality from Proposition 1.3.3 is stronger than the upper bound from Theorem 1.3.1.

[5] Indeed, this is equivalent to $(n-i)k \geq n(k-i)$, which is in turn equivalent to $ni \geq ki$, which is true since $n \geq k$ and $i \geq 0$.

Proposition 1.3.3. *For all integers n and k such that $n \geq k \geq 1$, we have that:*

$$\sum_{i=0}^{k} \binom{n}{i} \leq (\tfrac{en}{k})^k.$$

Proof. Fix integers n and k such that $n \geq k \geq 1$.

Claim. *For all real numbers x such that $0 < x \leq 1$, we have that*

$$\sum_{i=0}^{k} \binom{n}{i} \leq \tfrac{(1+x)^n}{x^k}.$$

Proof of the Claim. Fix a real number x such that $0 < x \leq 1$. By the Binomial theorem, we have that

$$\begin{aligned} (1+x)^n &= \sum_{i=0}^{n} \binom{n}{i} x^i \\ &\geq \sum_{i=0}^{k} \binom{n}{i} x^i \qquad \text{since } n \geq k \text{ and } x > 0. \end{aligned}$$

Dividing by x^k, we then obtain

$$\begin{aligned} \tfrac{(1+x)^n}{x^k} &\geq \sum_{i=0}^{k} \binom{n}{i} \tfrac{1}{x^{k-i}} \\ &\geq \sum_{i=0}^{k} \binom{n}{i} \qquad \text{because } 0 < x \leq 1. \end{aligned}$$

This proves the Claim. ♦

We now apply the Claim to $x := \frac{k}{n}$, and we obtain

$$\begin{aligned} \sum_{i=0}^{k} \binom{n}{i} &\leq (1+\tfrac{k}{n})^n (\tfrac{n}{k})^k \qquad \text{by the Claim for } x = \tfrac{k}{n} \\ &\leq (e^{k/n})^n (\tfrac{n}{k})^k \qquad \text{by Proposition 1.2.2 for } x = \tfrac{k}{n} \\ &= (\tfrac{en}{k})^k, \end{aligned}$$

which is what we needed. □

1.3.2 Estimating the binomial coefficient $\binom{2n}{n}$

Note that for all integers n and k such that $n \geq k \geq 1$, we have that

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

This implies that[6] for even n, we have that

$$\binom{n}{0} < \binom{n}{1} < \ldots < \binom{n}{n/2} > \ldots > \binom{n}{n-1} > \binom{n}{n},$$

whereas for odd n, we have that

$$\binom{n}{0} < \binom{n}{1} < \ldots < \binom{n}{\lfloor n/2 \rfloor} = \binom{n}{\lceil n/2 \rceil} > \ldots > \binom{n}{n-1} > \binom{n}{n}.$$

In particular, $\binom{n}{\lfloor n/2 \rfloor} = \binom{n}{\lceil n/2 \rceil}$ is maximum among the binomial coefficients $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$. For this reason, it is of particular interest to find good estimates for the behavior of binomial coefficients of the form $\binom{n}{\lfloor n/2 \rfloor}$.

Theorem 1.3.4. *For all integers $m \geq 1$, we have that*

$$\frac{2^{2m}}{2\sqrt{m}} \leq \binom{2m}{m} \leq \frac{2^{2m}}{\sqrt{2m}}$$

Proof. Fix an integer $m \geq 1$, and set

$$P := \frac{1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2m-1)}{2 \cdot 4 \cdot 6 \cdot \cdots \cdot (2m)}.$$

Then

$$\begin{aligned} P &= \frac{1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2m-1)}{2 \cdot 4 \cdot 6 \cdot \cdots \cdot (2m)} \\ &= \frac{1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2m-1)}{2 \cdot 4 \cdot 6 \cdot \cdots \cdot (2m)} \cdot \frac{2 \cdot 4 \cdot \cdots \cdot (2m)}{2 \cdot 4 \cdot \cdots \cdot (2m)} \\ &= \frac{(2m)!}{2^{2m}(m!)^2} \\ &= \frac{1}{2^{2m}}\binom{2m}{m}. \end{aligned}$$

It now suffices to show that

$$\frac{1}{2\sqrt{m}} \leq P \leq \frac{1}{\sqrt{2m}},$$

for the result then follows immediately.

[6] Check this!

We first establish the upper bound for P. For this, we observe that

$$\begin{aligned} 1 &\geq (1-\tfrac{1}{2^2})(1-\tfrac{1}{4^2})\dots(1-\tfrac{1}{(2m)^2}) \\ &= \tfrac{2^2-1}{2^2}\cdot\tfrac{4^2-1}{4^2}\cdot\dots\cdot\tfrac{(2m)^2-1}{(2m)^2} \\ &= \tfrac{1\cdot 3}{2^2}\cdot\tfrac{3\cdot 5}{4^2}\cdot\dots\cdot\tfrac{(2m-1)(2m+1)}{(2m)^2} \\ &= (2m+1)P^2, \end{aligned}$$

and consequently, $P^2 \leq \frac{1}{2m+1}$. This, in turn, implies that

$$P \leq \tfrac{1}{\sqrt{2m+1}} \leq \tfrac{1}{\sqrt{2m}},$$

which is what we needed.

It remains to establish our lower bound for P. The proof is similar as for the upper bound. We observe that

$$\begin{aligned} 1 &\geq (1-\tfrac{1}{3^2})(1-\tfrac{1}{5^2})\dots(1-\tfrac{1}{(2m-1)^2}) \\ &= \tfrac{3^2-1}{3^2}\cdot\tfrac{5^2-1}{5^2}\cdot\dots\cdot\tfrac{(2m-1)^2-1}{(2m-1)^2} \\ &= \tfrac{2\cdot 4}{3^2}\cdot\tfrac{4\cdot 6}{5^2}\cdot\dots\cdot\tfrac{(2m-2)(2m)}{(2m-1)^2} \\ &= \tfrac{1}{2(2m)P^2}. \end{aligned}$$

This implies that

$$P \geq \tfrac{1}{2\sqrt{m}},$$

which is what we needed. This completes the argument. □

Finally, we note that using Stirling's formula (which we stated without proof), we can obtain an even better approximation of $\binom{2m}{m}$, as follows:

$$\lim_{m\to\infty}\left(\left(\tfrac{2^{2m}}{\sqrt{\pi m}}\right)/\binom{2m}{m}\right) = 1.$$

Using the notation from section 1.1, this formula becomes

$$\binom{2m}{m} \sim \frac{2^{2m}}{\sqrt{\pi m}}.$$

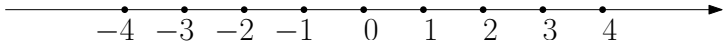
So, for very large values of m, the function $g(m) = \frac{2^{2m}}{\sqrt{\pi m}}$ is a good approximation for $\binom{2m}{m}$.

1.4 An application: random walks

Recall from Analysis that the series $\sum_{m=1}^{\infty} \frac{1}{m}$ is called the *harmonic series*, and that it diverges to infinity, i.e.

$$\sum_{m=1}^{\infty} \frac{1}{m} \quad = \quad \infty.$$

Let us now consider an application of our estimate for binomial coefficients. We consider the integer number line ($\mathbb{Z}$). We begin our walk at the origin (i.e. 0), and at each step we move at random either one step to the left (-1) or one step to the right ($+1$).



One example of such a walk might be

$$0, 1, 2, 3, 2, 3, 2, 1, 0, -1, -2, -1, -2, -1, 0, 1, \ldots$$

We would like to estimate the number of times that we return to the origin in such a walk. Obviously, we can only return to the origin after an even number of steps.[7] There are 2^{2m} random walks of length $2m$, and exactly $\binom{2m}{m}$ of those walks end at the origin.[8] So, the probability of returning to the origin after exactly $2m$ steps is

$$\frac{\binom{2m}{m}}{2^{2m}}.$$

This means that in an infinite random walk, the expected number of returns to the origin is

$$\sum_{m=1}^{\infty} \frac{\binom{2m}{m}}{2^{2m}}.$$

By Theorem 1.3.4, we have that

$$\sum_{m=1}^{\infty} \frac{\binom{2m}{m}}{2^{2m}} \quad \geq \quad \sum_{m=1}^{\infty} \frac{1}{2\sqrt{m}} \quad \overset{(*)}{=} \quad \infty,$$

[7] After an odd number of steps, our position is an odd integer!

[8] Indeed, we must go left exactly m times, and right exactly m times. Out of $2m$ moves, we have $\binom{2m}{m}$ ways of selecting the m leftward moves (the other m moves are rightward).

where for (*) we used the fact that

$$\sum_{m=1}^{\infty} \frac{1}{2\sqrt{m}} \quad = \quad \frac{1}{2} \sum_{m=1}^{\infty} \frac{1}{\sqrt{m}} \quad \geq \quad \frac{1}{2} \sum_{m=1}^{\infty} \frac{1}{m} \quad = \quad \infty.$$

Thus, we can expect that in an infinite one-dimensional random walk starting at the origin, we will return to the origin an infinite number of times.

Chapter 2

Generating functions

2.1 Partial fraction decomposition

We begin with an example, and then we explain the general principle. It is easy to check that

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

Verifying that the equality above is correct is quite easy; but how do we compute the expression on the right, given the expression on the left? We proceed as follows. The numerator is of strictly smaller degree than the denominator,[1] and the denominator is expressed as a product of linear terms. So, we write

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

By multiplying both sides by $x^2(x-1)$, we obtain

$$1 \quad = \quad (A+C)x^2 + (-A+B)x - B.$$

The left-hand-side and the right-hand-side are identical as polynomials, and so they have exactly the same coefficients. So, we get the following system of linear equations:

$$A + C = 0, \qquad -A + B = 0, \qquad -B = 1.$$

By solving the system, we obtain

$$A = -1, \qquad B = -1, \qquad C = 1,$$

[1] This is important! If the degree of the numerator is greater or equal to the degree of the denominator, then this will not work.

and we deduce that

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

Now, let us try to generalize the example above. Suppose $p(x)$ and $q(x)$ are polynomials with complex coefficients[2] such that $\deg p(x) < \deg q(x)$. Next, suppose that $q(x)$ can be factored as

$$q(x) \quad = \quad c(x-\alpha_1)^{\beta_1} \dots (x-\alpha_t)^{\beta_t},$$

where c is a non-zero complex number, $\alpha_1, \dots, \alpha_t$ are pairwise distinct complex numbers, and $\beta_1, \dots, \beta_t$ are positive integers.[3] In this case,[4] there exist complex numbers $A_{1,1}, \dots, A_{1,\beta_1}, \dots, A_{t,1}, \dots, A_{t,\beta_t}$ such that

$$\frac{p(x)}{q(x)} \quad = \quad \frac{A_{1,1}}{x-\alpha_1} + \dots + \frac{A_{1,\beta_1}}{(x-\alpha_1)^{\beta_1}} + \dots + \frac{A_{t,1}}{x-\alpha_t} + \dots + \frac{A_{t,\beta_t}}{(x-\alpha_t)^{\beta_t}}.$$

We find the numbers $A_{1,1}, \dots, A_{1,\beta_1}, \dots, A_{t,1}, \dots, A_{t,\beta_t}$ by multiplying both sides by $q(x)$, then writing the resulting polynomials on both sides in the standard form,[5] and finally, setting corresponding coefficients equal to each other. This yields a system of linear equations, and we obtain the coefficients $A_{1,1}, \dots, A_{1,\beta_1}, \dots, A_{t,1}, \dots, A_{t,\beta_t}$ by solving this system.

For example, for the rational expression $\frac{x^5-7x+1}{7(x-2)^3(x+1)^2(x+2)^4}$, we would get the equation

$$\frac{x^5-7x+1}{7(x-2)^3(x+1)^2(x+2)^4}$$

$$= \quad \frac{A}{x-2} + \frac{B}{(x-2)^2} + \frac{C}{(x-2)^3} + \frac{D}{x+1} + \frac{E}{(x+1)^2} + \frac{F}{x+2} + \frac{G}{(x+2)^2} + \frac{H}{(x+2)^3} + \frac{I}{(x+2)^4},$$

though computing $A, \dots, I$ by hand would take quite some time.

Let us now consider a computationally easier example:

$$\frac{3x^2+4}{x^3(x+1)^2}.$$

[2] In examples that we consider, we will work only with real numbers. However, the method works exactly the same way for complex numbers.

[3] Note that in the example from the beginning of the section, we have $c = 1$, $t = 2$, $\alpha_1 = 0$, $\alpha_2 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$.

[4] We omit the proof, but you can try to convince yourself that this is true.

[5] That is to say, in the form $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$, where $a_n, \dots, a_0$ are complex numbers.

The polynomial in the numerator is of strictly smaller degree than the polynomial in the denominator, and so there exist numbers A, B, C, D, E such that

$$\frac{3x^2+4}{x^3(x+1)^2} = \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x^3} + \frac{D}{x+1} + \frac{E}{(x+1)^2}.$$

After multiplying both sides by $x^3(x+1)^2$, we get

$$3x^2 + 4 = Ax^2(x+1)^2 + Bx(x+1)^2 + C(x+1)^2 + Dx^3(x+1) + Ex^3,$$

and after writing the polynomial on the right-hand-side in standard form, we get

$$\begin{aligned} 3x^2 + 4 &= (A+D)x^4 + (2A+B+D+E)x^3 + \\ &\quad +(A+2B+C)x^2 + (B+2C)x + C. \end{aligned}$$

The polynomial on the left-hand-side and the one on the right-hand-side have the same coefficients, which yields the following system of linear equations:

$$\begin{array}{rcrcrcrcl} A & & & & & + & D & = & 0 \\ 2A & + & B & & & + & D & + & E & = & 0 \\ A & + & 2B & + & C & & & = & 3 \\ & & B & + & 2C & & & = & 0 \\ & & & & C & & & = & 4 \end{array}$$

By solving the system, we obtain

$$A = 15, \quad B = -8, \quad C = 4, \quad D = -15, \quad E = -7.$$

So, we have that

$$\frac{3x^2+4}{x^3(x+1)^2} = \frac{15}{x} - \frac{8}{x^2} + \frac{4}{x^3} - \frac{15}{x+1} - \frac{7}{(x+1)^2}.$$

As pointed out earlier in the section, we can perform the procedure described above only on rational expressions of the form $\frac{p(x)}{q(x)}$, where $p(x)$ has strictly smaller degree than $q(x)$. If $\deg p(x) \geq \deg q(x)$, then we first perform polynomial division, and then we perform our procedure on the remainder. For instance,

$$\frac{3x^4-3x^3+1}{x^2(x-1)} \overset{(*)}{=} 3x + \frac{1}{x^2(x-1)} \overset{(**)}{=} 3x - \frac{1}{x} - \frac{1}{x^2} + \frac{1}{x-1},$$

where (*) is obtained by dividing polynomials, and (**) is from the calculation performed at the beginning of the section.

2.2 The Taylor series: a review

Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$, let $a \in A$, and assume that A contains (as a subset) some open neighborhood of a.[6] Assume furthermore that f is infinitely differentiable at a.[7] Then the *Taylor series* of f centered at a is the series

$$T^{f,a}(x) \quad = \quad \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

The Taylor series $T^{f,0}(x)$ (here, we have $a = 0$) is called the *Maclaurin series*.

For a real number α and a non-negative integer k, we define

$$\binom{\alpha}{k} \quad := \quad \frac{\alpha(\alpha-1)\ldots(\alpha-k+1)}{k!}.$$

In particular, $\binom{\alpha}{0} = 1$.

Here are the Maclaurin series of some familiar functions:

(i) $T^{\exp(x),0}(x) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \ldots$;

(ii) $T^{\sin x,0}(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!} + \ldots$;

(iii) $T^{\cos x,0}(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n\frac{x^{2n}}{(2n)!} + \ldots$;

(iv) $T^{\ln(1+x),0}(x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1}\frac{x^n}{n} + \ldots$;

(v) $T^{(1+x)^\alpha,0}(x) = \binom{\alpha}{0} + \binom{\alpha}{1}x + \binom{\alpha}{2}x^2 + \cdots + \binom{\alpha}{n}x^n + \ldots$, where α is a fixed real number;

(vi) $T^{\frac{1}{1-x},0}(x) = 1 + x + x^2 + \cdots + x^n + \ldots$.

Let us verify (v). Fix a real number α. It is easy to verify by induction[8] that for all positive integers k, we have that

$$\frac{d^k}{dx^k}(1+x)^\alpha \quad = \quad \alpha(\alpha-1)\ldots(\alpha-k+1)(1+x)^{\alpha-k},$$

and consequently,

$$\frac{\frac{d^k}{dx^k}(1+x)^\alpha\Big|_{x=0}}{k!} \quad = \quad \frac{\alpha(\alpha-1)\ldots(\alpha-k+1)}{k!} \quad = \quad \binom{\alpha}{k},$$

[6] So, there exists some $\delta > 0$ such that $(a-\delta, a+\delta) \subseteq A$.

[7] f is *infinitely differentiable* at a if the n-th derivative $f^{(n)}(a)$ exists for all $n \geq 0$. (In particular, f is differentiable, and therefore continuous, at a.) By definition, $f^{(0)} = f$.

[8] Check this!

where as usual, $\frac{d^k}{dx^k}(1+x)^\alpha$ denotes the k-th derivative of the function $(1+x)^\alpha$,[9] and $\frac{d^k}{dx^k}(1+x)^\alpha\Big|_{x=0}$ is the k-th derivative of $(1+x)^\alpha$ evaluated at $x=0$. So, (v) holds.

We remark that these series do not necessarily converge for all values of x. Furthermore, in general, it is possible that $T^{f,a}(x)$ converges, but does **not** converge to $f(x)$. Nonetheless, we do have the following:

(1) $\exp(x) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \ldots$ for all $x \in \mathbb{R}$;

(2) $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!} + \ldots$ for all $x \in \mathbb{R}$;

(3) $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n\frac{x^{2n}}{(2n)!} + \ldots$ for all $x \in \mathbb{R}$;

(4) $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1}\frac{x^n}{n} + \ldots$ for all $x \in (-1,1]$;

(5) $(1+x)^\alpha = \binom{\alpha}{0} + \binom{\alpha}{1}x + \binom{\alpha}{2}x^2 + \cdots + \binom{\alpha}{n}x^n + \ldots$ for $x \in (-1,1)$, where α is a fixed real number;

(6) $\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \ldots$ for $x \in (-1,1)$.

For a non-zero constant a, a positive integer t, and a sufficiently small value of x, we can substitute ax^t for x in the above equations. So, for example, by substituting $2x^3$ for x in (6), we get that

$$\frac{1}{1-2x^3} \quad = \quad 1 + 2x^3 + 4x^6 + \cdots + 2^n x^{3n} + \ldots$$

(as long as x is sufficiently small). When working with generating functions (see the section 2.3 below), we will not worry about exactly how small x needs to be to make our equations work; we simply need that they work for values of x in some (no matter how small) open neighborhood of zero. We also note that (6) follows from (5) for $\alpha = -1$, with $-x$ substituted for x;

[9] The zeroth derivative of a function is simply the function itself.

indeed,

$$\begin{aligned}
\tfrac{1}{1-x} &= (1-(-x))^{-1} \\
&= \sum_{n=0}^{\infty} \tbinom{-1}{n}(-x)^n && \text{by (5)} \\
&= \sum_{n=0}^{\infty} \tfrac{(-1)(-2)\ldots(-1-n+1)}{n!}(-x)^n \\
&= \sum_{n=0}^{\infty} \tfrac{(-1)^n n!}{n!}(-1)^n x^n \\
&= \sum_{n=0}^{\infty} x^n,
\end{aligned}$$

which is precisely (6).

Finally, we remark that the identity from (5) is sometimes called the "Generalized Binomial Theorem." Note that if α is a non-negative integer, then $\binom{\alpha}{k} = 0$ for all integers $k > \alpha$, and we get that

$$(1+x)^\alpha = \tbinom{\alpha}{0} + \tbinom{\alpha}{1}x + \cdots + \tbinom{\alpha}{\alpha}x^\alpha,$$

which is what we also get from the usual (finite) Binomial Theorem. However, if α is negative or not an integer, then the series from (5) is indeed infinite.

2.3 Generating functions

2.3.1 A motivating example

We motivate our study of generating function with the following question: How many ways are there to pay 21 Kč, assuming we have six 1 Kč coins, five 2 Kč coins, and four 5 Kč coins?[10] Here, we are looking for the number of solutions to the equation $i_1 + i_2 + i_5 = 21$, with $i_1 \in \{0,1,2,3,4,5,6\}$, $i_2 \in \{0,2,4,6,8,10\}$, and $i_5 \in \{0,5,10,15,20\}$. Indeed, i_1 is the amount paid with 1 Kč coins, i_2 is the amount paid with 2 Kč coins, and i_5 is the amount paid with 5 Kč coins. Now, we note that the number of solutions is precisely the coefficient in front of x^{21} in the following polynomial:

$$\begin{aligned}
p(x) = {} & (1+x+x^2+x^3+x^4+x^5+x^6)\times(1+x^2+x^4+x^6+x^8+x^{10}) \\
& \times(1+x^5+x^{10}+x^{15}+x^{20})
\end{aligned}$$

[10] Here, we assume that all coins of the same value are the same. So, if we happened to use three 1 Kč coins, we do not care which particular three we chose.

Indeed, we obtain x^{21} by selecting some x^{i_1} from the first term of the product, some x^{i_2} from the second, and some x^{i_5} from the third, in such a way that $i_1 + i_2 + i_5 = 21$. The number of ways of selecting i_1, i_2, i_5 is precisely the coefficient in front of x^{21} in the polynomial $p(x)$. By using computer software,[11] we see that this coefficient is 9. So, there are 9 ways to make our payment. More generally, for each non-negative integer n, let a_n be the number of ways to pay nKč using our coins; then a_n is precisely the coefficient in front of x^n in the polynomial $p(x)$, i.e.

$$p(x) \quad = \quad \sum_{n=0}^{\infty} a_n x^n.$$

We call $p(x)$ the "generating function" of the sequence $\{a_n\}_{n=0}^{\infty}$. In this particular case, $p(x)$ is a polynomial,[12] but in general, it is a (potentially infinite) series. (A formal definition of a generating function is given in section 2.3.2 below).

It might seem that the use of polynomials in the example above does not simplify the problem. Indeed, if you compute by hand, it is easier to simply enumerate all the solutions. However, polynomials are more convenient if we wish to use a computer. More importantly, we can use a similar idea to solve more complicated problems.

2.3.2 Generating functions as power series

Suppose $\{a_n\}_{n=0}^{\infty}$ is some infinite sequence of real numbers.[13] The *generating function* of this sequence is the power series

$$\sum_{n=0}^{\infty} a_n x^n.$$

For example, the generating function of the constant sequence $1, 1, 1, 1, 1, \ldots$ is

$$1 + x + x^2 + x^3 + \ldots \quad = \quad \sum_{n=0}^{\infty} x^n$$

We recognize the above sequence as the Maclaurin series of the function $\frac{1}{1-x}$. So, the generating function of $1, 1, 1, 1, 1, \ldots$ is $\frac{1}{1-x}$.

[11] Or by hand, if you are in the mood to compute.

[12] This is because we only have 36 Kč, and so $a_n = 0$ for all integers $n \geq 37$.

[13] Actually, this also works for complex numbers. However, we shall restrict ourselves to examples with real numbers.

2.3.3 Generating functions and recursively defined sequences

For a positive integer k, a *homogeneous linear difference equation of degree k* is an equation of the form

$$y_{n+k} = a_{k-1}y_{n+k-1} + a_{k-2}y_{n+k-2} + \cdots + a_1y_{n+1} + a_0y_n,$$

where $a_{k-1}, \ldots, a_0$ are fixed constants. Often, sequences are defined recursively, by specifying the values of the first k terms, and by a homogeneous linear difference equation of degree k.

For example, the famous *Fibonacci sequence* $\{F_n\}_{n=0}^{\infty}$ is defined recursively as follows:

- $F_0 = 0$, $F_1 = 1$;
- $F_{n+2} = F_n + F_{n+1}$ for all integers $n \geq 0$.

(Numbers F_n are called the *Fibonacci numbers*.) So, we defined the Fibonacci sequence using a second degree homogeneous linear difference equation.

Often, we are given a recursively defined sequence, and we would like to find a closed formula for the n-th term of the sequence. For example, suppose we are given a sequence $\{a_n\}_{n=0}^{\infty}$, defined recursively as follows:

- $a_0 = 1$
- $a_{n+1} = 2a_n$ for all integers $n \geq 0$.

This sequence is defined via a first degree homogeneous linear difference equation. A closed formula for the general term of the sequence $\{a_n\}_{n=0}^{\infty}$ is

$$a_n = 2^n \qquad \text{for all integers } n \geq 0.$$

This example was easy (we could simply guess the formula, and verify by induction that it works). But often, this is not so easy. For instance, it is not at all obvious what the closed formula for F_n, the the n-th Fibonacci number, should be. As we shall see, such a formula can be found using generating functions.

In theory, generating functions can be used to find the closed formula of the general term of a sequence defined via any homogeneous linear difference equation. However, in practice, if our difference equation is of high degree, this may be difficult or impossible to do due to problems with factoring polynomials of high degree.[14] Here, we show how this can be done for

[14]The quadratic equation allows us to easily factor second degree polynomials. There are also formulas for factoring third and fourth degree polynomials. However, there is no general formula for factoring fifth (and higher) degree polynomials.

sequences defined via second degree homogeneous linear difference equations. We begin with the Fibonacci sequence.

Example 2.3.1. *Find a closed formula for F_n ($n \geq 0$), where F_n is the n-th Fibonacci number.*

Solution. We consider the generating function $f(x) = \sum_{n=0}^{\infty} F_n x^n$ for the sequence $\{F_n\}_{n=0}^{\infty}$. We now manipulate this function as follows:

$$\begin{aligned}
f(x) &= \sum_{n=0}^{\infty} F_n x^n \\
&= F_0 + F_1 x + x^2 \sum_{n=0}^{\infty} F_{n+2} x^n \\
&= x + x^2 \sum_{n=0}^{\infty} (F_n + F_{n+1}) x^n \\
&= x + (x^2 \sum_{n=0}^{\infty} F_n x^n) + (x^2 \sum_{n=0}^{\infty} F_{n+1} x^n) \\
&= x + (x^2 \sum_{n=0}^{\infty} F_n x^n) + (x \sum_{n=0}^{\infty} F_{n+1} x^{n+1}) \\
&= x + (x^2 \sum_{n=0}^{\infty} F_n x^n) + (x \sum_{n=0}^{\infty} F_n x^n) \qquad \text{because } F_0 = 0 \\
&= x + x^2 f(x) + x f(x).
\end{aligned}$$

So, we have obtained the equation

$$f(x) = x + x^2 f(x) + x f(x),$$

which, in turn, yields

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

We now compute:

$$\begin{aligned}
f(x) &= -\tfrac{x}{x^2+x-1} \\
&= -\tfrac{x}{(x-\frac{-1-\sqrt{5}}{2})(x-\frac{-1+\sqrt{5}}{2})} && \text{via the quadratic equation} \\
&= -\tfrac{\frac{1+\sqrt{5}}{2\sqrt{5}}}{x-\frac{-1-\sqrt{5}}{2}} - \tfrac{\frac{-1+\sqrt{5}}{2\sqrt{5}}}{x-\frac{-1+\sqrt{5}}{2}} && \text{via partial fractions} \\
&= -\tfrac{1}{\sqrt{5}}\Big(\tfrac{\frac{1+\sqrt{5}}{2}}{x-\frac{-1-\sqrt{5}}{2}} + \tfrac{\frac{-1+\sqrt{5}}{2}}{x-\frac{-1+\sqrt{5}}{2}}\Big) \\
&= -\tfrac{1}{\sqrt{5}}\Big(\tfrac{1}{1+x\frac{2}{1+\sqrt{5}}} - \tfrac{1}{1+x\frac{2}{1-\sqrt{5}}}\Big) \\
&= -\tfrac{1}{\sqrt{5}}\Big(\tfrac{1}{1-x\frac{1-\sqrt{5}}{2}} - \tfrac{1}{1-x\frac{1+\sqrt{5}}{2}}\Big) \\
&= \tfrac{1}{\sqrt{5}}\Big((-\sum_{n=0}^{\infty}(\tfrac{1-\sqrt{5}}{2})^n x^n) + (\sum_{n=0}^{\infty}(\tfrac{1+\sqrt{5}}{2})^n x^n)\Big) && \text{via Maclaurin expansion} \\
&= \sum_{n=0}^{\infty} \tfrac{(1+\sqrt{5})^n-(1-\sqrt{5})^n}{2^n\sqrt{5}} x^n.
\end{aligned}$$

Recall that $f(x) = \sum_{n=0}^{\infty} F_n x^n$. So, for all non-negative integers n, we have that

$$F_n = \tfrac{(1+\sqrt{5})^n-(1-\sqrt{5})^n}{2^n\sqrt{5}}.$$

We can easily check that the answer is correct by induction. Indeed,

$$\tfrac{(1+\sqrt{5})^0-(1-\sqrt{5})^0}{2^0\sqrt{5}} = 0 = F_0$$

$$\tfrac{(1+\sqrt{5})^1-(1-\sqrt{5})^1}{2^1\sqrt{5}} = 1 = F_1$$

and so the formula is correct for $n = 0$ and $n = 1$. For the induction step, we fix an integer $n \geq 0$, and we assume that

$$F_n = \tfrac{(1+\sqrt{5})^n-(1-\sqrt{5})^n}{2^n\sqrt{5}} \quad \text{and} \quad F_{n+1} = \tfrac{(1+\sqrt{5})^{n+1}-(1-\sqrt{5})^{n+1}}{2^{n+1}\sqrt{5}}.$$

Then

$$\begin{aligned}
F_{n+2} &= F_n + F_{n+1} \\
&= \tfrac{(1+\sqrt{5})^n-(1-\sqrt{5})^n}{2^n\sqrt{5}} + \tfrac{(1+\sqrt{5})^{n+1}-(1-\sqrt{5})^{n+1}}{2^{n+1}\sqrt{5}} \\
&= \tfrac{1}{\sqrt{5}}\Big((\tfrac{1+\sqrt{5}}{2})^n(1+\tfrac{1+\sqrt{5}}{2}) - (\tfrac{1-\sqrt{5}}{2})^n(1+\tfrac{1-\sqrt{5}}{2})\Big) \\
&= \tfrac{1}{\sqrt{5}}\Big((\tfrac{1+\sqrt{5}}{2})^n\tfrac{3+\sqrt{5}}{2} - (\tfrac{1-\sqrt{5}}{2})^n\tfrac{3-\sqrt{5}}{2}\Big) \\
&= \tfrac{1}{\sqrt{5}}\Big((\tfrac{1+\sqrt{5}}{2})^n(\tfrac{1+\sqrt{5}}{2})^2 - (\tfrac{1-\sqrt{5}}{2})^n(\tfrac{1-\sqrt{5}}{2})^2\Big) \\
&= \tfrac{1}{\sqrt{5}}\Big((\tfrac{1+\sqrt{5}}{2})^{n+2} - (\tfrac{1-\sqrt{5}}{2})^{n+2}\Big) \\
&= \tfrac{(1+\sqrt{5})^{n+2}-(1-\sqrt{5})^{n+2}}{2^{n+2}\sqrt{5}},
\end{aligned}$$

and so the formula is correct for $n+2$. □

The *golden ratio* is the number

$$\phi = \tfrac{1+\sqrt{5}}{2}.$$

Our solution to Example 2.3.1 implies that the n-th Fibonacci number ($n \geq 0$) satisfies[15]

$$F_n = \tfrac{\phi^n-(1-\phi)^n}{\sqrt{5}} = \tfrac{\phi^n-(-\phi)^{-n}}{\sqrt{5}}.$$

Example 2.3.2. *Let $\{a_0\}_{n=0}^{\infty}$ be a sequence defined recursively as follows:*

- $a_0 = 0$ *and* $a_1 = 1$*;*
- $a_{n+2} = -a_n + 2a_{n+1}$ *for all integers* $n \geq 0$.

Find a closed formula for a_n ($n \geq 0$).

Solution. We consider the generating function $a(x) = \sum_{n=0}^{\infty} a_n x^n$ for the

[15] Check this!

sequence $\{a_n\}_{n=0}^{\infty}$, and we compute:

$$\begin{aligned}
a(x) &= \sum_{n=0}^{\infty} a_n x^n \\
&= a_0 + a_1 x + x^2 \sum_{n=0}^{\infty} a_{n+2} x^n \\
&= x + x^2 \sum_{n=0}^{\infty} (-a_n + 2a_{n+1}) x^n \\
&= x - \left(x^2 \sum_{n=0}^{\infty} a_n x^n \right) + \left(2x \sum_{n=0}^{\infty} a_{n+1} x^{n+1} \right) \\
&= x - \left(x^2 \sum_{n=0}^{\infty} a_n x^n \right) + \left(2x \sum_{n=0}^{\infty} a_n x^n \right) \qquad \text{because } a_0 = 0 \\
&= x - x^2 a(x) + 2x a(x)
\end{aligned}$$

Thus, we have obtained the equation

$$a(x) = x - x^2 a(x) + 2x a(x),$$

which yields

$$a(x) = \tfrac{x}{(x-1)^2}.$$

We now compute:

$$
\begin{aligned}
a(x) &= \tfrac{x}{(x-1)^2} \\
&= -\tfrac{1}{1-x} + \tfrac{1}{(1-x)^2} && \text{via partial fractions} \\
&= -\Big(\sum_{n=0}^{\infty} x^n\Big) + \Big(\sum_{n=0}^{\infty} \tbinom{-2}{n}(-x)^n\Big) && \text{via Maclaurin expansion} \\
&= -\Big(\sum_{n=0}^{\infty} x^n\Big) + \Big(\sum_{n=0}^{\infty} \tfrac{(-2)(-3)\ldots(-2-n+1)}{n!}(-x)^n\Big) \\
&= -\Big(\sum_{n=0}^{\infty} x^n\Big) + \Big(\sum_{n=0}^{\infty} \tfrac{(-1)^n(n+1)!}{n!}(-1)^n x^n\Big) \\
&= -\Big(\sum_{n=0}^{\infty} x^n\Big) + \Big(\sum_{n=0}^{\infty} (n+1)x^n\Big) \\
&= \sum_{n=0}^{\infty} nx^n
\end{aligned}
$$

Since $a(x) = \sum_{n=0}^{\infty} a_n$, we deduce that $a_n = n$ for all integers $n \geq 0$.[16]

We can easily check that our formula is correct by induction. Indeed, $a_0 = 0$ and $a_1 = 1$ by construction, and so the formula is correct for $n = 0$ and $n = 1$. For the induction step, we fix an integer $n \geq 0$, we assume inductively that $a_n = n$ and $a_{n+1} = n + 1$, and we observe that

$$\begin{aligned} a_{n+2} &= -a_n + 2a_{n+1} \\ &= -n + 2(n+1) \\ &= n + 2, \end{aligned}$$

and so the formula is correct for $n + 2$. This completes the induction. □

Sometimes, generating functions can be used to find a closed formula for the general term of a recursively defined sequence, even if the recurrence is not given by a homogeneous linear difference equation. We now look at one such example.

Example 2.3.3. *Let $\{a_n\}_{n=0}^{\infty}$ be a sequence defined recursively as follows:*

[16] Alternatively, we could have proceeded as follows:

$$\begin{aligned} a(x) &= \tfrac{x}{(x-1)^2} \\ &= x\tfrac{1}{(1-x)^2} \\ &= x\sum_{n=0}^{\infty}\tbinom{-2}{n}(-x)^n && \text{via Maclaurin expansion} \\ &= x\sum_{n=0}^{\infty}\tfrac{(-2)(-3)\ldots(-2-n+1)}{n!}(-x)^n \\ &= x\sum_{n=0}^{\infty}\tfrac{(-1)^n(n+1)!}{n!}(-1)^n x^n \\ &= x\sum_{n=0}^{\infty}(n+1)x^n \\ &= \sum_{n=0}^{\infty}(n+1)x^{n+1} \\ &= \sum_{n=0}^{\infty} nx^n, \end{aligned}$$

and so $a_n = n$ for all integers $n \geq 0$.

- $a_0 = 1$*;*
- $a_{n+1} = 7a_n + 6^{n+1}$ *for all integers* $n \geq 0$.

Find a closed formula for a_n*.*

Solution. We consider the generating function $a(x) = \sum_{n=0}^{\infty} a_n x^n$ for the sequence $\{a_n\}_{n=0}^{\infty}$. We manipulate this function as follows:

$$\begin{aligned}
a(x) &= \sum_{n=0}^{\infty} a_n x^n \\
&= a_0 + \sum_{n=0}^{\infty} a_{n+1} x^{n+1} \\
&= 1 + \sum_{n=0}^{\infty} (7a_n + 6^{n+1}) x^{n+1} \\
&= 1 + 7x\Big(\sum_{n=0}^{\infty} a_n x^n\Big) + \Big(\sum_{n=1}^{\infty} 6^n x^n\Big) \\
&= 7x\Big(\sum_{n=0}^{\infty} a_n x^n\Big) + \Big(\sum_{n=0}^{\infty} 6^n x^n\Big) \\
&= 7xa(x) + \tfrac{1}{1-6x}.
\end{aligned}$$

So, we have obtained the equation

$$a(x) = 7xa(x) + \tfrac{1}{1-6x},$$

which implies that

$$a(x) = \tfrac{1}{(7x-1)(6x-1)}.$$

We now compute

$$\begin{aligned}
a(x) &= \tfrac{1}{(7x-1)(6x-1)} \\
&= \tfrac{7}{1-7x} - \tfrac{6}{1-6x} && \text{via partial fractions} \\
&= (7\sum_{n=0}^{\infty} 7^n x^n) - (6\sum_{n=0}^{\infty} 6^n x^n) \\
&= \sum_{n=0}^{\infty} (7^{n+1} - 6^{n+1}) x^n.
\end{aligned}$$

Recall that $a(x) = \sum_{n=0}^{\infty} a_n x^n$. So, we get that

$$a_n = 7^{n+1} - 6^{n+1}$$

for all integers $n \geq 0$.

We can check that this formula is correct by induction. Clearly,

$$7^{0+1} - 6^{0+1} = 1 = a_0,$$

and so the formula is correct for $n = 0$. Now, fix a non-negative integer n, and assume that $a_n = 7^{n+1} - 6^{n+1}$. Then

$$\begin{aligned} a_{n+1} &= 7a_n + 6^{n+1} \\ &= 7(7^{n+1} - 6^{n+1}) + 6^{n+1} \\ &= 7^{n+2} - 7 \cdot 6^{n+1} + 6^{n+1} \\ &= 7^{n+2} - 6^{n+2}. \end{aligned}$$

This completes the induction. □

2.4 Basic operations with generating functions

We now consider some ways of combining generating functions. Suppose $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ are sequences of real (or complex) numbers, and suppose $a(x) = \sum_{n=0}^{\infty} a_n x^n$ and $b(x) = \sum_{n=0}^{\infty} b_n x^n$ are the corresponding generating functions. Further, suppose α is a real (or complex) constant. Then we have the following.

1. The generating function of the sequence $\{a_n + b_n\}_{n=0}^{\infty}$ is $a(x) + b(x)$.
2. The generating function of the sequence $\{a_n - b_n\}_{n=0}^{\infty}$ is $a(x) - b(x)$.
3. The generating function of the sequence $\{\alpha a_n\}_{n=0}^{\infty}$ is $\alpha a(x)$.
4. For an integer $k \geq 1$, the generating function of the sequence

$$\underbrace{0, \ldots, 0}_{k}, a_0, a_1, a_2, \ldots$$

is $x^k a(x)$.

5. For an integer $k \geq 1$, the generating function of the sequence $\{a_{n+k}\}_{n=0}^{\infty}$, i.e. the sequence $a_k, a_{k+1}, a_{k+2}, \ldots$, is $\frac{1}{x^k}\Big(a(x) - \sum_{i=0}^{k-1} a_i x^i\Big)$.[17]

6. The generating function of the sequence $\{\alpha^n a_n\}_{n=0}^{\infty}$ is $c(x) = a(\alpha x)$.[18]

7. For an integer $k \geq 1$, the generating function of the sequence

$$a_0, \underbrace{0, \ldots, 0}_{k}, a_1, \underbrace{0, \ldots, 0}_{k}, a_2, \underbrace{0, \ldots, 0}_{k}, a_3, \ldots$$

   is $a(x^{k+1})$.[19]

8. The generating function of the sequence $\{(n+1)a_{n+1}\}_{n=0}^{\infty}$, i.e. the sequence $a_1, 2a_2, 3a_3, 4a_4, \ldots$, is $a'(x)$. The generating function for the sequence $0, a_0, \frac{1}{2}a_1, \frac{1}{3}a_2, \frac{1}{4}a_3, \ldots$ is $\int_0^x a(t)dt$. (We differentiate and integrate power series term-by-term.)

9. The function $c(x) = a(x)b(x)$ is the generating function of the sequence $\{c_n\}_{n=0}^{\infty}$, where $c_n = \sum_{i=0}^{n} a_i b_{n-i}$ for each integer $n \geq 0$.[20]

Example 2.4.1. *Let $\{a_n\}_{n=0}^{\infty}$ be a sequence, and let $a(x)$ be its generating function. Find the generating function of the sequence $a_0, 0, a_2, 0, a_4, \ldots$ in terms of the function $a(x)$.*

Solution. We observe that $a_0, 0, a_2, 0, a_4, \ldots$ is the sum of the following two sequences: $\{\frac{a_n}{2}\}_{n=0}^{\infty}$ and $\{\frac{(-1)^n a_n}{2}\}_{n=0}^{\infty}$. The generating function of $\{\frac{a_n}{2}\}_{n=0}^{\infty}$ is $\frac{1}{2}a(x)$, and the generating function of $\{\frac{(-1)^n a_n}{2}\}_{n=0}^{\infty}$ is $\frac{1}{2}a(-x)$. So, the generating function of $a_0, 0, a_2, 0, a_4, \ldots$ is $\frac{a(x)+a(-x)}{2}$. □

Example 2.4.2. *Find (the closed form of) the generating function of the sequence $1, 1, 2, 2, 4, 4, 8, 8, 16, 16, \ldots$, i.e. the sequence $\{2^{\lfloor n/2 \rfloor}\}_{n=0}^{\infty}$.*

[17] For example, the generating function of the sequence $a_3, a_4, a_5, \ldots$ is

$$\frac{1}{x^3}\Big(a(x) - (a_0 + a_1 x + a_2 x^2)\Big).$$

[18] For instance, since $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$ is the generating function of $1, 1, 1, 1, 1, \ldots$, we see that $\frac{1}{1-2x}$ is the generating function of $1, 2, 4, 8, 16, \ldots$.

[19] For instance, the generating function of the sequence $a_0, 0, 0, a_1, 0, 0, a_2, 0, 0, a_3, \ldots$ is $a(x^3)$.

[20] So, $c_0 = a_0 b_0$, $c_1 = a_0 b_1 + a_1 b_0$, $c_2 = a_0 b_2 + a_1 b_1 + a_2 b_0$, etc.

Solution. Recall that the generating function of the sequence $1, 2, 4, 8, 16, \ldots$ is $\frac{1}{1-2x}$. The generating function of $1, 0, 2, 0, 4, 0, 8, 0, \ldots$ is $\frac{1}{1-2x^2}$, and the generating function of $0, 1, 0, 2, 0, 4, 0, 8, 0, \ldots$ is $\frac{x}{1-2x^2}$. So, the generating function of $1, 1, 2, 2, 4, 4, 8, 8, 16, 16, \ldots$ is the sum of these two functions, i.e. $\frac{1+x}{1-2x^2}$. □

Example 2.4.3. *Find (the closed form of) the generating function of the sequence* $1^2, 2^2, 3^2, 4^2, \ldots$, *i.e. the sequence* $\{(n+1)^2\}_{n=0}^{\infty}$.

Solution. The generating function of the sequence $1, 1, 1, 1, \ldots$ is $\frac{1}{1-x}$. By differentiating, we see that $\frac{d}{dx}(\frac{1}{1-x}) = \frac{1}{(1-x)^2}$ is the generating function of the sequence $1, 2, 3, 4, \ldots$, i.e. the sequence $\{n+1\}_{n=0}^{\infty}$. By differentiating again, we see that $\frac{d}{dx}(\frac{1}{(1-x)^2}) = \frac{2}{(1-x)^3}$ is the generating sequence of the sequence $1 \cdot 2, 2 \cdot 3, 3 \cdot 4, 4 \cdot 5, \ldots$, i.e. the sequence $\{(n+1)(n+2)\}_{n=0}^{\infty}$. Now, $(n+1)^2 = (n+1)(n+2) - (n+1)$ for all integers $n \geq 0$, and we have computed the generating functions for the sequences $\{(n+1)(n+2)\}_{n=0}^{\infty}$ and $\{n+1\}_{n=0}^{\infty}$. So, the generating function of $\{(n+1)\}_{n=0}^{\infty}$ is

$$a(x) \quad = \quad \frac{2}{(1-x)^3} - \frac{1}{(1-x)^2}.$$

□

2.5 An application of generating functions: counting binary trees

In this section, we consider binary trees of the sort that are often used in data structures. For our purposes, we can define binary trees recursively as follows: a *binary tree* is either empty (i.e. contains no nodes), or consists of a designated node r (called the *root*), plus an ordered pair (T_L, T_R) of binary trees, where T_L and T_R (called the *left subtree* and the *right subtree*, respectively) have disjoint sets of nodes and do not contain the node r (see Figure 2.1). The empty binary tree has zero nodes, and if a binary tree T consists of a root r and an ordered pair (T_L, T_R) of binary trees, then the number of nodes of T is $1 + n_L + n_R$, where n_L is the number of nodes of T_L, and n_R is the number of nodes of T_R.

For each integer $n \geq 0$, let b_n be the number of binary trees on n nodes, and let $b(x) = \sum_{n=0}^{\infty} b_n x^n$ be the generating function of the sequence $\{b_n\}_{n=0}^{\infty}$. It is easy to check that $b_0 = 1$, $b_1 = 1$, $b_2 = 2$, and $b_3 = 5$ (see Figure 2.2). Now, let us find a recursive formula for b_n ($n \geq 1$). The number of binary

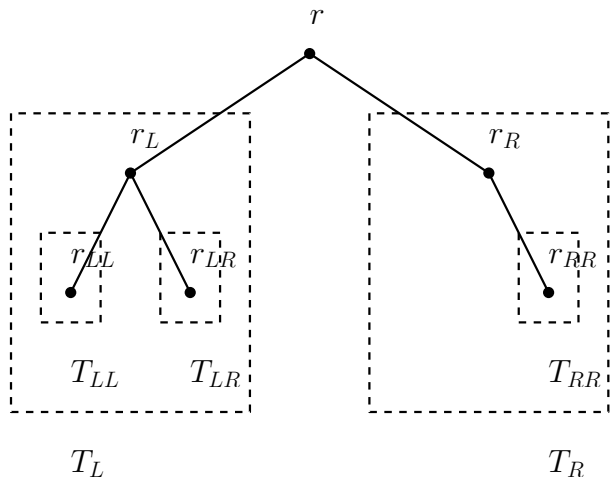


Figure 2.1: A binary tree on six nodes. Note that T_{RL} (the left subtree of the right subtree) is empty.

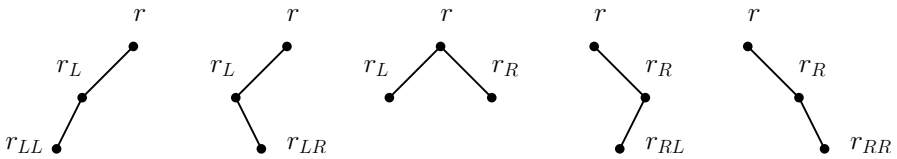


Figure 2.2: All the binary trees on three nodes.

trees on $n \geq 1$ nodes is equal to the number of ordered pairs (T_L, T_R) of binary trees such that T_L and T_R together have $n-1$ nodes. Thus, for all integers $n \geq 1$, we have that

$$b_n \quad = \quad b_0 b_{n-1} + b_1 b_{n-2} + \cdots + b_{n-1} b_0 \quad = \quad \sum_{k=0}^{n-1} b_k b_{n-k-1}.$$

Since $b_0 = 1$, this implies that

$$b(x) \quad = \quad 1 + x b(x)^2.$$

Using the quadratic formula,[21] we get that either

$$b(x) = \tfrac{1-\sqrt{1-4x}}{2x} \qquad \text{or} \qquad b(x) = \tfrac{1+\sqrt{1-4x}}{2x}.$$

We must now determine which of these two formulas is the correct one.

Since $b_0 = 1$, we have that $\lim_{x \to 0^+} b(x) = 1$. Since

$$\begin{aligned} \lim_{x \to 0^+} \tfrac{1-\sqrt{1-4x}}{2x} &= \lim_{x \to 0^+} \left(\tfrac{1-\sqrt{1-4x}}{2x} \cdot \tfrac{1+\sqrt{1-4x}}{1+\sqrt{1-4x}} \right) \\ &= \lim_{x \to 0^+} \tfrac{1-(1-4x)}{2x(1+\sqrt{1-4x})} \\ &= \lim_{x \to 0^+} \tfrac{2}{1+\sqrt{1-4x}} \\ &= 1, \end{aligned}$$

whereas

$$\lim_{x \to 0^+} \tfrac{1+\sqrt{1-4x}}{2x} \quad = \quad \infty,$$

we deduce that[22]

$$b(x) \quad = \quad \tfrac{1-\sqrt{1-4x}}{2x}.$$

[21] Here, we treat $b(x)$ as the variable and x as a constant.

[22] Since we have x in the denominator, $b(0)$ is not defined, and in particular, $b(x)$ does not have a Maclaurin series. However, as we shall see, the constant term in the Maclaurin expansion of $1 - \sqrt{1-4x}$ is zero, and so we can simply divide the resulting series by $2x$ and thus obtain another series, which is precisely the generating function (expressed as a power series) of the sequence $\{b_n\}_{n=0}^{\infty}$.

By the Generalized Binomial Theorem, we have that

$$\begin{aligned}
\sqrt{1-4x} &= \sum_{n=0}^{\infty} \tbinom{1/2}{n}(-4x)^n \\
&= \sum_{n=0}^{\infty} (-4)^n \tbinom{1/2}{n} x^n \\
&= 1 + \sum_{n=1}^{\infty} (-4)^n \tbinom{1/2}{n} x^n \qquad \text{because } (-4)^0 \tbinom{-1/2}{0} x^0 = 1 \\
&= 1 + x \sum_{n=0}^{\infty} (-4)^{n+1} \tbinom{1/2}{n+1} x^n
\end{aligned}$$

and consequently,

$$1 - \sqrt{1-4x} = -x \sum_{n=0}^{\infty} (-4)^{n+1} \tbinom{1/2}{n+1} x^n.$$

It follows that

$$b(x) = \tfrac{1-\sqrt{1-4x}}{2x} = \sum_{n=0}^{\infty} (-\tfrac{1}{2})(-4)^{n+1} \tbinom{1/2}{n+1} x^n.$$

Thus, for all non-negative integers n, we have that

$$b_n = (-\tfrac{1}{2})(-4)^{n+1} \tbinom{1/2}{n+1}.$$

Let us now try to obtain a nicer formula for b_n. For an integer $n \geq 0$, we

compute:

$$\begin{aligned}
b_n &= (-\tfrac{1}{2})(-4)^{n+1}\tbinom{1/2}{n+1} \\
&= (-\tfrac{1}{2})(-4)^{n+1}\tfrac{(\frac{1}{2})(\frac{1}{2}-1)(\frac{1}{2}-2)\ldots(\frac{1}{2}-n)}{(n+1)!} \\
&= (-\tfrac{1}{2})(-4)^{n+1}\tfrac{(\frac{1}{2})(-\frac{1}{2})(-\frac{3}{2})\ldots(-\frac{2n-1}{2})}{(n+1)!} \\
&= (-\tfrac{1}{2})(-4)^{n+1}(-1)^n(\tfrac{1}{2})^{n+1}\tfrac{1\cdot 3\cdot\cdots\cdot(2n-1)}{(n+1)!} \\
&= 2^n\cdot\tfrac{1\cdot 3\cdot\cdots\cdot(2n-1)}{(n+1)!} \\
&= \tfrac{2\cdot 4\cdot\cdots\cdot(2n)}{n!}\cdot\tfrac{1\cdot 3\cdot\cdots\cdot(2n-1)}{(n+1)!} \\
&= \tfrac{(2n)!}{n!(n+1)!} \\
&= \tfrac{1}{n+1}\tbinom{2n}{n}.
\end{aligned}$$

So, we have obtained that, for all integers $n \geq 0$, the number of binary trees on n nodes is

$$b_n = \tfrac{1}{n+1}\tbinom{2n}{n}.$$

We remark that the numbers $\frac{1}{n+1}\binom{2n}{n}$ above have a special name: they are called *Catalan numbers*.

2.6 An application of generating functions: random walks

We consider the following infinite random walk on the integer line $\mathbb{Z}$: we begin our walk at 1, and at each step, we move at random either two units to the right ($+2$) or one unit to the left (-1). We would like to determine the probability that we reach the origin at some point in our walk.

We proceed as follows. For each integer $n \geq 0$, let P_n be the probability that we reach the origin after at most n steps. Obviously, $\{P_n\}_{n=0}^{\infty}$ is a non-decreasing sequence, and it is bounded above by 1. So, by the Monotone Sequence Theorem, it converges. Let

$$P := \lim_{n\to\infty} P_n.$$

Then P is the probability that we need to compute.

Now, for each integer $n \geq 0$, let a_n be the number of n-step walks in which we reach the origin for the first time after precisely n steps.[23] Since the total number of n-step walks is 2^n, we see that

$$P_n = \sum_{i=0}^{n} \frac{a_i}{2^i}$$

for all non-negative integers n, and consequently,

$$P = \sum_{n=0}^{\infty} \frac{a_n}{2^n}.$$

Let $a(x) = \sum_{n=0}^{\infty} a_n x^n$ be the generating function for the sequence $\{a_n\}_{n=0}^{\infty}$. Note that

$$P = a\left(\tfrac{1}{2}\right).$$

For our solution, it will be useful to consider random walks that start at points other than 1 (but proceed according to the same rules: at each step, we move at random either two units to the right or one unit to the left). For an integer $n \geq 0$, let b_n be the number of n-step random walks (following our rules) starting at 2 and ending at the origin, without reaching the origin at any point during the walk (except at the very end). In such a walk, we cannot reach the origin without first reaching 1, and then reaching the origin from there. So, if we are to reach the origin for the first time after precisely n steps, starting at 2, then there must be some $k \in \{1, \ldots, n-1\}$ such that we reach 1 for the first time after precisely k steps,[24] and then starting at 1, we reach the origin for the first time after $n-k$ steps.[25] For each choice of $k \in \{1, \ldots, n-1\}$, there are $a_k a_{n-k}$ such walks, and so

$$\begin{aligned} b_n &= \sum_{k=1}^{n-1} a_k a_{n-k} \\ &= \sum_{k=0}^{n} a_k a_{n-k} \qquad \text{because } a_0 = 0. \end{aligned}$$

[23] So, at the end of our n-step walk, we are at the origin, and furthermore, we were not at the origin after k steps for any non-negative integer $k < n$.

[24] There are precisely a_k many k-step walks that have this property: the number of ways to reach 1 from 2 for the first time after k steps is the same as the number of ways to reach the origin from 1 for the first time after k steps.

[25] There are precisely a_{n-k} many $(n-k)$-step walks that have this property.

Now, if $b(x) = \sum_{n=0}^{\infty} b_n x^n$ is the generating function for the sequence $\{b_n\}_{n=0}^{\infty}$, then we get that

$$b(x) \quad = \quad a(x)^2.$$

Next, we consider random walks starting at 3, and moving according to our rules. For each integer $n \geq 0$, let c_n be the number of n-step random walks (following our rules) starting at 3 and ending at the origin, without reaching the origin at any point during the walk (except at the very end). We now argue similarly to the above. In such a walk, we cannot reach the origin before first reaching 2, and then reaching the origin from there. So, if we are to reach the origin for the first time after precisely n steps, starting at 3, then there must be some $k \in \{1, \ldots, n-1\}$ such that we reach 2 for the first time after precisely k steps,[26] and then starting at 2, we reach the origin for the first time after $n-k$ steps.[27] For each choice of $k \in \{1, \ldots, n-1\}$, there are $a_k b_{n-k}$ such walks, and so

$$\begin{aligned} b_n &= \sum_{k=1}^{n-1} a_k b_{n-k} \\ &= \sum_{k=0}^{n} a_k b_{n-k} \qquad \text{because } a_0 = 0 \text{ and } b_0 = 0. \end{aligned}$$

Now, if $c(x) = \sum_{n=0}^{\infty} c_n x^n$ is the generating function for the sequence $\{c_n\}_{n=0}^{\infty}$, then we get that $c(x) = a(x)b(x)$. We already saw that $b(x) = a(x)^2$, and so it follows that

$$c(x) \quad = \quad a(x)^3.$$

We now observe the following. Obviously, $a_0 = 0$ and $a_1 = 1$. Next, if we start at 1, then for an integer $n \geq 2$, there are precisely c_{n-1} ways to reach the origin for the first time after precisely n steps: we must first move two units to the right,[28] and then reach the origin from 3 for the first time after precisely $n-1$ steps. Thus, $a_n = c_{n-1}$ for all integers $n \geq 2$. We now

[26] There are precisely a_k many k-step walks that have this property.

[27] There are precisely b_{n-k} many $(n-k)$-step walks that have this property.

[28] Indeed, if we moved one unit to the left instead, then we would reach the origin after precisely one step, and so (since $n \geq 2$) we would not reach the origin for the first time after n steps.

compute:

$$\begin{aligned}
a(x) &= a_0 + a_1 x + \sum_{n=2}^{\infty} a_n x^n \\
&= x + x \sum_{n=2}^{\infty} a_n x^{n-1} && \text{because } a_0 = 0 \text{ and } a_1 = 1 \\
&= x + x \sum_{n=2}^{\infty} c_{n-1} x^{n-1} && \text{because } a_n = c_{n-1} \text{ for } n \geq 2 \\
&= x + x \sum_{n=1}^{\infty} c_n x^n \\
&= x + x \sum_{n=0}^{\infty} c_n x^n && \text{because } c_0 = 0 \text{ (obvious)} \\
&= x + x c(x).
\end{aligned}$$

We have now obtained the equation $a(x) = x + xc(x)$, and we know from before that $c(x) = a(x)^3$. So, we have that

$$a(x) = x + x a(x)^3.$$

At this point, we could in principle use the cubic equation to compute $a(x)$,[29] and then compute $P = a(\frac{1}{2})$ by plugging in $x = \frac{1}{2}$ into the function a. However, there is a quicker and easier way. Since $P = a(\frac{1}{2})$, we have that

$$P = \tfrac{1}{2} + \tfrac{1}{2} P^3.$$

The equation above has three solutions: 1, $\frac{-1+\sqrt{5}}{2}$, and $\frac{-1-\sqrt{5}}{2}$.[30] Obviously, $P \geq 0$, and so $P \neq \frac{-1-\sqrt{5}}{2}$. To simplify notation, we set

$$\Phi := \tfrac{-1+\sqrt{5}}{2}.$$

(Note that $\Phi = \phi^{-1}$, where $\phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio.) We now have that either $P = 1$ or $P = \Phi$.

[29] But note that we would get three solutions, and we would have to figure out which one is the correct one.

[30] The equation $P = \frac{1}{2} + \frac{1}{2}P^3$ is equivalent to the equation $P^3 - 2P + 1 = 0$. Obviously, 1 is a root of the latter equation. We find the other two roots by first factoring $P^3 - 2P + 1 = (P-1)(P^2 + P - 1)$, and then using the quadratic equation to find the other two roots.

Let us show that $P \neq 1$, i.e. that $a(\frac{1}{2}) \neq 1$. First, $a(x) = \sum_{n=0}^{\infty} a_n x^n$ has non-negative coefficients and converges for $x = \frac{1}{2}$. So, the function a is continuous and increasing on the interval $[0, \frac{1}{2}]$.[31] Obviously, $a(0) = a_0 = 0$. Suppose that $a(\frac{1}{2}) = 1$. Since $0 < \Phi < 1$, and since a is continuous on $[0, \frac{1}{2}]$, the Intermediate Value Theorem guarantees that there exists some $x_0 \in (0, \frac{1}{2})$ such that $a(x_0) = \Phi$. Since Φ is a root of the equation $P = \frac{1}{2} + \frac{1}{2}P^3$, we have that

$$\Phi \quad = \quad \tfrac{1}{2} + \tfrac{1}{2}\Phi^3.$$

On the other hand, we know that $a(x_0) = x_0 + x_0 a(x_0)^3$, and so

$$\Phi \quad = \quad x_0 + x_0\Phi^3.$$

It follows that

$$\tfrac{1}{2} + \tfrac{1}{2}\Phi^3 \quad = \quad x_0 + x_0\Phi^3,$$

which implies that

$$(x_0 - \tfrac{1}{2})(\Phi^3 + 1) \quad = \quad 0,$$

which is false since $x_0 \neq \frac{1}{2}$ and $\Phi^3 \neq -1$. This proves that $P \neq 1$, and it follows that $P = \Phi$, i.e. that

$$P \quad = \quad \tfrac{-1+\sqrt{5}}{2}.$$

[31] This is a little bit informal (and we omit the details), but it should be intuitively obvious.

Chapter 3

Finite projective planes

3.1 Finite projective planes: definition and basic properties

For a set X, the *power set* of X, denoted by $\mathscr{P}(X)$, is the set of all subsets of X. For example, if $X = \{1, 2, 3\}$, then

$$\mathscr{P}(X) \;=\; \Big\{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\Big\}.$$

Obviously, for any set X, we have that $\emptyset \in \mathscr{P}(X)$ and $X \in \mathscr{P}(X)$. Furthermore, if X is finite, then $|\mathscr{P}(X)| = 2^{|X|}$.

A *set system* is an ordered pair $(X, \mathcal{S})$ such that X is a set (called the *ground set*) and $\mathcal{S} \subseteq \mathscr{P}(X)$.

A *finite projective plane* is a set system $(X, \mathcal{P})$ such that X is a finite, and the following three properties are satisfied:

(P0) there exists a 4-element subset $Q \subseteq X$ such that every $P \in \mathcal{P}$ satisfies $|P \cap Q| \leq 2$;

(P1) all distinct $P_1, P_2 \in \mathcal{P}$ satisfy $|P_1 \cap P_2| = 1$;

(P2) for all distinct $x_1, x_2 \in X$, there exists a unique $P \in \mathcal{P}$ such that $x_1, x_2 \in P$.

If $(X, \mathcal{P})$ is a finite projective plane, then members of X are called *points*, and members of $\mathcal{P}$ are called *lines*. For a point $x \in X$ and a line $P \in \mathcal{P}$ such that $x \in P$, we say that the line P is *incident* with the point x, or that P *contains* x, or that P *passes through* x. For distinct points $a, b \in X$, we

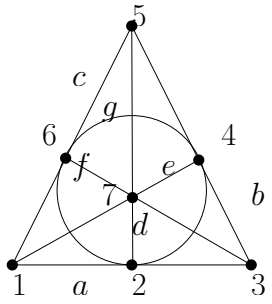


Figure 3.1: The Fano plane.

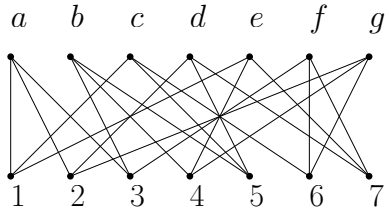


Figure 3.2: The incidence graph of the Fano plane.

denote by $\overline{ab}$ the unique line in $\mathcal{P}$ that contains a and b (the existence and uniqueness of such a line follow from (P2)).

Finite projective planes (defined above) and the usual Euclidean planes (i.e. planes that you studied in high school) have some obvious similarities, but also some obvious differences. In a Euclidean plane, two distinct lines may intersect in at most one point, but distinct, parallel lines have an empty intersection. In finite projective planes, there are no "parallel lines": by (P1), two distinct lines always intersect in exactly one point, called their *point of intersection* or *intersection point*. Property (P2) from the definition of a finite projective plane is the same as for the Euclidean plane.

Example 3.1.1. *Let* $X = \{1, 2, 3, 4, 5, 6, 7\}$ *and* $\mathcal{P} = \{a, b, c, d, e, f, g\}$*, where*

- $a = \{1, 2, 3\}$,
- $b = \{3, 4, 5\}$,
- $c = \{5, 6, 1\}$,
- $d = \{5, 7, 2\}$,
- $e = \{1, 7, 4\}$,
- $f = \{3, 7, 6\}$,
- $g = \{2, 4, 6\}$.

Then $(X, \mathcal{P})$ *is a finite projective plane,*[1] *called the* Fano plane *(see Figure 3.1).*

Note that in Figure 3.1, the seven lines of the Fano plane are represented by six line segments and one circle. However, formally, each line of the Fano plane is simply a set of three points. Drawings such as the one in Figure 3.1 can sometimes be useful for guiding our intuition. However, formal proofs should never rely on such pictures; instead, they should rely solely on the definition of a finite projective plane or on results (propositions, lemmas, theorems) proven about them.[2]

To each finite projective plane $(X, \mathcal{P})$, we associate an "incidence graph" defined as follows. The *incidence graph* of a finite projective plane $(X, \mathcal{P})$ is a bipartite graph with bipartition $(X, \mathcal{P})$,[3] in which $x \in X$ and $P \in \mathcal{P}$ are adjacent if and only if $x \in P$. The incidence graph of the Fano plane is represented in Figure 3.2.

Note that each line of the Fano plane contains the same number of points. As our next proposition shows, this is not an accident.

Proposition 3.1.2. *Let* $(X, \mathcal{P})$ *be a finite projective plane. Then all lines in* $\mathcal{P}$ *have the same number of points.*

Proof. Fix $P_1, P_2 \in \mathcal{P}$. We must show that $|P_1| = |P_2|$.

Claim. There exists a point $z \in X$ such that $z \notin P_1 \cup P_2$.

Proof of the Claim. First, using (P0) from the definition of a finite projective plane, we fix a 4-element subset $Q \subseteq X$ such that for all $P \in \mathcal{P}$, we have that $|Q \cap P| \leq 2$. If $Q \not\subseteq P_1 \cup P_2$, then we take any $z \in Q \setminus (P_1 \cup P_2)$, and we are done. So, assume that $Q \subseteq P_1 \cup P_2$.

Since $|Q| = 4$ and $|Q \cap P_1|, |Q \cap P_2| \leq 2$, we now deduce that $Q \cap P_1$ and $Q \cap P_2$ are disjoint and contain exactly two points each. Set $Q \cap P_1 = \{a, b\}$ and $Q \cap P_2 = \{c, d\}$. We now consider the lines $P_{ac} := \overline{ac}$ and $P_{bd} := \overline{bd}$.[4]

[1] It is easy to check that (P1) and (P2) are satisfied. For (P0), we can take, for instance, $Q = \{1, 3, 5, 7\}$.

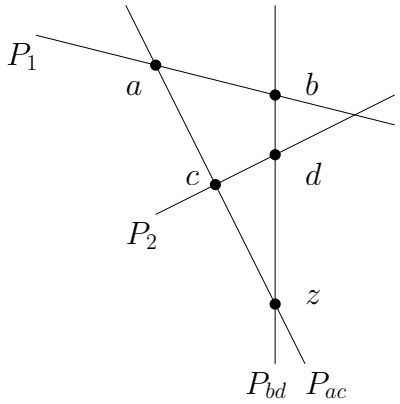
[2] The proofs of those results must, ultimately, rely only on the definition of a finite projective plane.

[3] So, in our incidence graph, X and $\mathcal{P}$ are stable (i.e. independent) sets. (A *stable set*, also called an *independent set*, in a graph G is any set of pairwise non-adjacent vertices of G.)

[4] Recall that, by (P2), there exists a unique line in $\mathcal{P}$ that contains both a and c, and according to our notation, this line is denoted $\overline{ac}$. For convenience, we set $P_{ac} = \overline{ac}$. Similar remarks hold for b, d.

Since no line in $\mathcal{P}$ contains more than two points of Q, and since $a, c \in Q \cap P_{ac}$, we see that $Q \cap P_{ac} = \{a, c\}$. Similarly, $Q \cap P_{bd} = \{b, d\}$. Since P_1, P_2, P_{ac}, P_{bd} have pairwise distinct intersections with the set Q, we see that the lines P_1, P_2, P_{ac}, P_{bd} are pairwise distinct.

Now, by (P1), we have that $|P_{ac} \cap P_{bd}| = 1$; set $P_{ac} \cap P_{bd} = \{z\}$ (see the picture below).

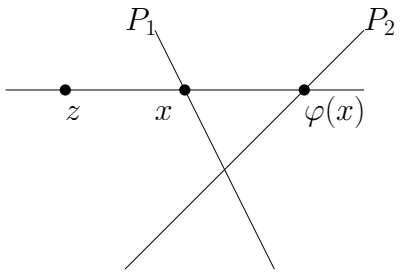


Since $P_{ac} \cap Q$ and $P_{bd} \cap Q$ are disjoint, we see that $z \notin Q$. If $z \in P_1$, then $a, z \in P_1 \cap P_{ac}$, which is impossible because a, z are distinct points,[5] P_1, P_{ac} are distinct lines, and by (P1), any two distinct lines intersect in exactly one point. Thus, $z \notin P_1$, and similarly, $z \notin P_2$. This proves the Claim. ♦

Let z be as in the Claim. We now define a function $\varphi : P_1 \to P_2$, as follows. For all $x \in P_1$, let $\varphi(x)$ be the unique point in the intersection of the lines $\overline{xz}$ and P_2 (see the picture below); by (P1) and (P2), our function φ is well-defined.[6]

[5] We know that $a \neq z$ because $a \in Q$ and $z \notin Q$.

[6] Let us check that this in detail. First, since $z \notin P_1$, we know that for all $x \in P_1$, we have that $x \neq z$, and so by (P2), there is exactly one line (which we denoted by $\overline{xz}$) that passes through x and z; furthermore, since $z \notin P_2$, we have that $\overline{xz}$ and P_2 are distinct lines, and so (P1) guarantees that $\overline{xz}$ and P_2 intersect in exactly one point, and we call this point $\varphi(x)$. Thus, φ is well-defined. (We remark that it is possible that $P_1 = P_2$; in this case, the function φ is simply the identity function on $P_1 = P_2$, that is, $\varphi(x) = x$ for all $x \in P_1$.)



Let us check that $\varphi : P_1 \to P_2$ is surjective (i.e. onto). Fix $y \in P_2$, and let x be the point of intersection of the lines P_1 and $\overline{yz}$.[7] Then y is the point of intersection of lines $\overline{xz}$ and P_2, and it follows that $y = \varphi(x)$. So, $\varphi : P_1 \to P_2$ is surjective. This implies that $|P_1| \geq |P_2|$. By symmetry, we also have that $|P_2| \geq |P_1|$, and we deduce that $|P_1| = |P_2|$. □

The *order* of a finite projective plane $(X, \mathcal{P})$ is the number $|P| - 1$, where P is any line in $\mathcal{P}$.[8] By Proposition 3.1.2, this is well-defined. Note that the Fano plane has order two. Furthermore, the following proposition states that the order of any finite projective plane is at least two.

Proposition 3.1.3. *The order of any finite projective plane is at least two.*

Proof. Let $(X, \mathcal{P})$ be a finite projective plane. It suffices to show that some line in $\mathcal{P}$ passes through at least three points. Using (P0) from the definition of a finite projective plane, we fix a 4-element subset $Q \subseteq X$ such that for all $P \in \mathcal{P}$, we have that $|Q \cap P| \leq 2$. Set $Q = \{a, b, c, d\}$. Consider the lines $P_{ab} := \overline{ab}$ and $P_{cd} := \overline{cd}$. Since Q intersects each line in $\mathcal{P}$ in at most two points, we see that $Q \cap P_{ab} = \{a, b\}$ and $Q \cap P_{cd} = \{c, d\}$; in particular, $P_{ab} \neq P_{cd}$. By (P1), P_{ab} and P_{cd} intersect in exactly one point, call it z. Since $Q \cap P_{ab}$ and $Q \cap P_{cd}$ are disjoint, we see that $z \notin Q$. But now P_{ab} contains at least three points, namely a, b, z. □

Theorem 3.1.4. *Let $(X, \mathcal{P})$ be a finite projective plane of order n.*[9] *Then all the following hold:*

(a) for each point $x \in X$, exactly $n + 1$ lines in $\mathcal{P}$ pass through x;

(b) $|X| = n^2 + n + 1$;

(c) $|\mathcal{P}| = n^2 + n + 1$.

[7] Check that x exists and is unique!

[8] So, if $(X, \mathcal{P})$ is a finite projective plane of order n, then each line in $\mathcal{P}$ contains exactly $n + 1$ points.

[9] By Proposition 3.1.3, we have that $n \geq 2$.

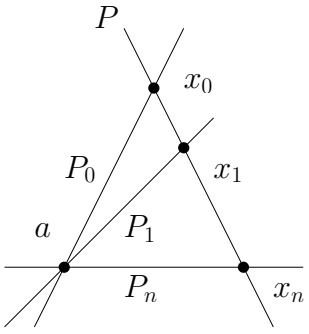
Proof of (a) and (b). We begin by proving an auxiliary claim.

> **Claim.** For every point $x \in X$, there exists a line $P \in \mathcal{P}$ such that $x \notin P$.

Proof of the Claim. Fix a point $x \in X$. Using (P0) from the definition of a finite projective plane, we fix a 4-element subset $Q \subseteq X$ such that for all $P \in \mathcal{P}$, we have that $|Q \cap P| \leq 2$. Clearly, $|Q \setminus \{x\}| \geq 3$. Let $a, b, c \in Q \setminus \{x\}$ be pairwise distinct. It now suffices to show that x belongs to at most one of $\overline{ab}$ and $\overline{ac}$. Suppose otherwise, i.e. suppose that x belongs both to $\overline{ab}$ and to $\overline{ac}$. Then the lines $\overline{ab}$ and $\overline{ac}$ have at least two points (namely, a and x) in common, and so by (P2), we have that $\overline{ab} = \overline{ac}$. But now the line $\overline{ab} = \overline{ac}$ contains at least three points (namely, a, b, c) of Q, a contradiction. This proves the Claim. ♦

We now prove (a). Fix a point $x \in X$. By the Claim, there exists a line $P \in \mathcal{P}$ such that $x \notin P$. Since $(X, \mathcal{P})$ is of order n, we know that $|P| = n+1$; set $P = \{x_0, x_1, \ldots, x_n\}$. By (P2), the lines $\overline{xx_0}, \overline{xx_1}, \ldots, \overline{xx_n}$ are pairwise distinct,[10] and they all contain x. So, there are at least $n+1$ lines passing through x. On the other hand, by (P1), any line passing though x intersects the line P in one of the points $x_0, x_1, \ldots, x_n$, and is therefore (by (P2)) equal to one of $\overline{xx_0}, \ldots, \overline{xx_1}, \overline{xx_n}$. Thus, exactly $n+1$ lines pass through x. This proves (a).

We now prove (b). Fix any line $P \in \mathcal{P}$. Since $(X, \mathcal{P})$ is of order n, we know that $|P| = n+1$; set $P = \{x_0, x_1, \ldots, x_n\}$. Since every line in $\mathcal{P}$ has $n+1$ points, the Claim guarantees that $|X| \geq n+2$; consequently, $P \subsetneqq X$. Fix any $a \in X \setminus P$. For each $i \in \{0, 1, \ldots, n\}$, we set $P_i := \overline{ax_i}$ (see the picture below).



[10] Indeed, suppose that for some distinct $i, j \in \{0, 1, \ldots, n\}$, we had $\overline{xx_i} = \overline{xx_j}$. Now the line $\overline{xx_i} = \overline{xx_j}$ contains both x_i and x_j. On the other hand, the line P contains both x_i and x_j. By (P2), there is exactly one line that contains both x_i and x_j, and we deduce that $P = \overline{xx_i} = \overline{xx_j}$. But this implies that $x \in P$, contrary to the choice of P.

By (P2), lines $P_0, P_1, \ldots, P_n$ are pairwise distinct,[11] and so by (P1), any two of them have exactly one point in common. Since a lies on each of $P_0, P_1, \ldots, P_n$, we see that $P_i \cap P_j = \{a\}$ for all distinct $i, j \in \{0, 1, \ldots, n\}$; consequently, $P_0 \setminus \{a\}, P_1 \setminus \{a\}, \ldots, P_n \setminus \{a\}$ are pairwise disjoint. Now, since $(X, \mathcal{P})$ is of order n, we know that $P_0, P_1, \ldots, P_n$ each have $n+1$ points, and we deduce that

$$\begin{aligned} |P_0 \cup P_1 \cup \cdots \cup P_n| &= |\{a\}| + |P_0 \setminus \{a\}| + |P_1 \setminus \{a\}| + \cdots + |P_n \setminus \{a\}| \\ &= 1 + (n+1)n \\ &= n^2 + n + 1. \end{aligned}$$

It now remains to show that $X = P_0 \cup P_1 \cup \cdots \cup P_n$; in fact, we only need to show that $X \subseteq P_0 \cup P_1 \cup \cdots \cup P_n$, for the reverse inclusion is immediate. Fix a point $x \in X$; we must show that x belongs to at least one of $P_0, P_1, \ldots, P_n$. We may assume that $x \neq a$, for otherwise we are done. The line $R := \overline{xa}$ is distinct from P (because $a \in R$, but $a \notin P$), and so by (P1), $|P \cap R| = 1$. Since $P = \{x_0, x_1, \ldots, x_n\}$, it follows that there exists some $i \in \{0, 1, \ldots, n\}$ such that $P \cap R = \{x_i\}$. Now lines P_i and R have at least two points (namely, a and x_i) in common, and so by (P2), we have that $R = P_i$. Since $x \in R$, we deduce that $x \in P_i$. This completes the argument. □

We postpone the proof of Theorem 3.1.4(c) to the end of section 3.2.

3.2 Duality

In this section, we show (roughly speaking) that by swapping the roles of points and lines of a finite projective plane, we obtain another finite projective plane (called the "dual" of the original finite projective plane).

Let us be more precise. For a set system $(X, \mathcal{S})$, we define the *dual* of $(X, \mathcal{S})$ to be the ordered pair $(Y, \mathcal{T})$, where $Y = \mathcal{S}$ and

$$\mathcal{T} = \Big\{ \{S \in \mathcal{S} \mid x \in S\} \;\mid\; x \in X \Big\}.$$

Example 3.2.1. *Let $X = \{1, 2, 3\}$ and $\mathcal{S} = \{A, B\}$, where $A = \{1, 2\}$ and $B = \{1, 3\}$. Then the dual of $(X, \mathcal{S})$ is $(Y, \mathcal{T})$, where $Y = \{A, B\}$ and $\mathcal{T} = \Big\{\{A, B\}, \{A\}, \{B\}\Big\}$.*[12]

[11] This is analogous to the argument from footnote 10.

[12] Indeed $\{S \in \mathcal{S} \mid 1 \in S\} = \{A, B\}$, $\{S \in \mathcal{S} \mid 2 \in S\} = \{A\}$, and $\{S \in \mathcal{S} \mid 3 \in S\} = \{B\}$.

Theorem 3.2.2 (below) states that the dual of a finite projective plane is again a finite projective plane. Before giving a formal proof, let us try to give some intuition behind this. If $(X, \mathcal{P})$ is a finite projective plane, and $(Y, \mathcal{R})$ is its dual, then the lines of $(X, \mathcal{P})$ become points of $(Y, \mathcal{R})$ (indeed, by definition, $Y = \mathcal{P}$). Furthermore, points of $(X, \mathcal{P})$ correspond to the lines of $(Y, \mathcal{R})$ in a natural way: a point $x \in X$ corresponds to the line $R_x := \{P \in \mathcal{P} \mid x \in P\} \in \mathcal{R}$. The incidence graphs of $(X, \mathcal{P})$ and $(Y, \mathcal{R})$ are isomoprhic (i.e. identical up to a relabeling of the vertices), except that points turn into lines and vice versa.

Theorem 3.2.2. *The dual of a finite projective plane is again a finite projective plane.*

Proof. Let $(X, \mathcal{P})$ be a finite projective plane, and let $(Y, \mathcal{R})$ be its dual. To simplify notation, for all $x \in X$, we set $R_x := \{P \in \mathcal{P} \mid x \in P\}$. We now have that $Y = \mathcal{P}$ and $\mathcal{R} = \{R_x \mid x \in X\}$. Obviously, for all $x \in X$, we have that $R_x \subseteq \mathcal{P} = Y$, and consequently $R_x \in \mathscr{P}(Y)$; thus, $\mathcal{R} \subseteq \mathscr{P}(Y)$, i.e. $(Y, \mathcal{R})$ is a set system. Furthermore, since X is finite, and since $Y = \mathcal{P} \subseteq \mathcal{P}(X)$, we have that Y is finite. It now remains to show that $(Y, \mathcal{R})$ satisfies (P0), (P1), and (P2).

We first prove that $(Y, \mathcal{R})$ satisfies (P0). Since $(X, \mathcal{P})$ is a finite projective plane, (P0) guarantees that there exists a 4-element set $Q \subseteq X$ such that for all $P \in \mathcal{P}$, we have that $|Q \cap P| \leq 2$. Set $Q = \{a, b, c, d\}$. Further, set $P_1 = \overline{ab}$, $P_2 = \overline{bc}$, $P_3 = \overline{cd}$, and $P_4 = \overline{da}$. Since $|Q \cap P| \leq 2$ for all $P \in \mathcal{P}$, we now deduce that $Q \cap P_1 = \{a, b\}$, $Q \cap P_2 = \{b, c\}$, $Q \cap P_3 = \{c, d\}$, and $Q \cap P_4 = \{d, a\}$; in particular, every point of Q belongs to exactly two of P_1, P_2, P_3, P_4. Now, set $Q^* = \{P_1, P_2, P_3, P_4\}$; we must show that no element of $\mathcal{R}$ contains more than two elements of Q^*. Suppose otherwise. Then there exist some $x \in X$ and pairwise distinct $i, j, k \in \{1, 2, 3\}$ such that $P_i, P_j, P_k \in R_x$; consequently, $x \in P_i \cap P_j \cap P_k$. Since each point in Q belongs to exactly two of P_1, P_2, P_3, P_4, whereas x belongs to at least three of them, we see that $x \notin Q$. On the other hand, for any three of P_1, P_2, P_3, P_4, some two of them have a point of Q in common. So, some two of P_i, P_j, P_k, have at least two points in common (namely, one point of Q, plus the point x) and are therefore (by (P2) applied to $(X, \mathcal{P})$) identical, a contradiction. This proves that $(Y, \mathcal{R})$ satisfies (P0).

We next show that $(Y, \mathcal{R})$ satisfies (P1). Fix distinct $R_1, R_2 \in \mathcal{R}$; we must show that $|R_1 \cap R_2| = 1$. By the construction of $\mathcal{R}$, there exist some $x_1, x_2 \in X$ such that $R_1 = R_{x_1}$ and $R_2 = R_{x_2}$; since $R_1 \neq R_2$, we have that $x_1 \neq x_2$. Now, $R_1 \cap R_2 = \{P \in \mathcal{P} \mid x_1, x_2 \in P\}$. By (P2) for $(X, \mathcal{P})$,

there is exactly one $P \in \mathcal{P}$ such that $x_1, x_2 \in P$; so, $R_1 \cap R_2 = \{P\}$, and in particular, $|R_1 \cap R_2| = 1$. Thus, $(Y, \mathcal{R})$ satisfies (P1).

It remains to show that $(Y, \mathcal{R})$ satisfies (P2). Fix distinct $P_1, P_2 \in Y (= \mathcal{P})$; we must show that there is exactly one member of $\mathcal{R}$ that contains both P_1 and P_2. By (P1) for $(X, \mathcal{P})$, we know that $|P_1 \cap P_2| = 1$; set $P_1 \cap P_2 = \{x_0\}$. So, R_{x_0} is the only member of $\mathcal{R}$ that contains both P_1 and P_2. Thus, $(Y, \mathcal{R})$ satisfies (P2). □

Notation: The dual of a finite projective plane $(X, \mathcal{P})$ is sometimes denoted by $(X, \mathcal{P})^*$.

We complete this section by proving Theorem 3.1.4(c), as follows. Let $(X, \mathcal{P})$ be a finite projective plane of order n; we must show that $|\mathcal{P}| = n^2 + n + 1$. By Theorem 3.2.2, $(Y, \mathcal{R}) := (X, \mathcal{P})^*$ is also a finite projective plane. By Theorem 3.1.4(a), we have that for all $x \in X$, there are exactly $n + 1$ lines $P \in \mathcal{P}$ that contain x. It then follows from the construction that all $R \in \mathcal{R}$ satisfy $|R| = n + 1$.[13] So, the finite projective plane $(Y, \mathcal{R})$ is of order n. By Theorem 3.1.4(b), we now have that $|Y| = n^2 + n + 1$. But $Y = \mathcal{P}$, and so $|\mathcal{P}| = n^2 + n + 1$, which is what we needed to show.

3.3 Finite projective planes and Latin squares

For a positive integer n, an $n \times n$ *Latin square* is an $n \times n$ array (or matrix) whose entries are numbers $1, \ldots, n$, and in which each number $1, \ldots, n$ occurs exactly once in each row and in each column. Two 3×3 Latin squares are represented in Figure 3.3. When we write that $[a_{i,j}]_{n \times n}$ is a Latin square, we mean that this Latin square is of size $n \times n$, and that for all $i, j \in \{1, \ldots, n\}$, the (i, j)-th entry (i.e. the entry in the i-th row and j-th column) of the Latin square is $a_{i,j}$. Now, two $n \times n$ Latin squares, say $[a_{i,j}]_{n \times n}$ and $[b_{i,j}]_{n \times n}$, are *orthogonal* if each entry of the matrix matrix obtained by superimposing A on B, i.e. of the matrix $[(a_{i,j}, b_{i,j})]_{n \times n}$, is different. Since an $n \times n$ matrix has n^2 entries, and the Cartesian product $\{1, \ldots, n\} \times \{1, \ldots, n\}$ has exactly n^2 elements, we see that two $n \times n$ Latin squares are orthogonal if and only if each element of $\{1, \ldots, n\} \times \{1, \ldots, n\}$ appears exactly once in the matrix obtained by superimposing the two $n \times n$ Latin squares. For instance, the Latin squares from Figure 3.3 are orthogonal, as we can see from Figure 3.4.

[13] Let us check this. First, for all $x \in X$, we set $R_x = \{P \in \mathcal{P} \mid x \in P\}$, as in the proof of Theorem 3.2.2. Since every point in X belongs to precisely $n + 1$ lines in $\mathcal{P}$, we see that for all $x \in X$, we have that $|R_x| = n + 1$. Since $\mathcal{R} = \{R_x \mid x \in X\}$, we deduce that all members of $\mathcal{R}$ have precisely $n + 1$ elements.

1	2	3
2	3	1
3	1	2

1	2	3
3	1	2
2	3	1

Figure 3.3: Two 3×3 Latin squares.

$(1,1)$	$(2,2)$	$(3,3)$
$(2,3)$	$(3,1)$	$(1,2)$
$(3,2)$	$(1,3)$	$(2,1)$

Figure 3.4: The matrix obtained by superimposing the left (red) 3×3 Latin square from Figure 3.3 onto the right (blue) one.

For a positive integer n, a Latin square $A = [a_{i,j}]_{n\times n}$ and a permutation π of the set $\{1, \ldots, n\}$, we set $\pi(A) = [\pi(a_{i,j})]_{n\times n}$; obviously, $\pi(A)$ is a Latin square. For example, if

$$A = \begin{array}{|c|c|c|}\hline 1 & 3 & 2 \\ \hline 3 & 2 & 1 \\ \hline 2 & 1 & 3 \\ \hline \end{array}$$

and if $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$, then

$$\pi(A) = \begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline 2 & 3 & 1 \\ \hline 3 & 1 & 2 \\ \hline \end{array} \quad .$$

Proposition 3.3.1. *Let $A = [a_{i,j}]_{n\times n}$ and $B = [b_{i,j}]_{n\times n}$ be orthogonal $n \times n$ Latin squares, and let π_A, π_B be permutations of the set $\{1, \ldots, n\}$. Then $\pi_A(A)$ and $\pi_B(B)$ are orthogonal Latin squares.*

Proof. Obvious.[14] □

Theorem 3.3.2. *Let $n \geq 2$ be an integer, and let M be a set of pairwise orthogonal $n \times n$ Latin squares. Then $|M| \leq n - 1$.*

Proof. We may assume that $M \neq \emptyset$, for otherwise, the result is immediate. Set $t = |M|$ and $M = \{A_1, \ldots, A_t\}$; we must show that $t \leq n - 1$. First, for each $i \in \{1, \ldots, t\}$, we let π_i be the permutation of $\{1, \ldots, n\}$ that transforms the first row of A_i into $1, \ldots, n$, and let $A'_i = \pi_i(A_i)$. By Proposition 3.3.1, Latin squares $A'_1, \ldots, A'_t$ are pairwise orthogonal. Now, since 1 is the $(1,1)$-th entry (i.e. the entry in the first row and first column) of all the matrices $A'_1, \ldots, A'_t$, we see that 1 is not the $(2,1)$-th entry (i.e. the entry in the second row and first column) of any of the Latin squares $A'_1, \ldots, A'_t$. Further, no two of $A'_1, \ldots, A'_t$ can have the same number in the $(2,1)$-th entry; indeed, if for some distinct $i, j \in \{1, \ldots, t\}$, we had that the $(2,1)$-th entry of A'_i and A'_j was the same, say k, then (k,k) would be both the $(1,k)$-th and the $(2,1)$-th entry of the matrix obtained by superimposing A'_i and A'_j, contrary to the fact that A'_i and A'_j are orthogonal. So, each of $A'_1, \ldots, A'_t$ has a number from $2, \ldots, n$ in the $(2,1)$-th entry, and no two of $A'_1, \ldots, A'_t$ have the same $(2,1)$-th entry; thus, $t \leq n - 1$. □

Theorem 3.3.3. *Let $n \geq 2$ be an integer. Then the following are equivalent:*

(a) there exists a finite projective plane of order n;

(b) there exists a collection of $n-1$ pairwise orthogonal $n \times n$ Latin squares.

Proof of "(b) $\Longrightarrow$ (a)" (outline). Assume that (b) is true, and let $L_1, \ldots, L_{n-1}$ be pairwise orthogonal $n \times n$ Latin squares. We will give a construction of the corresponding finite projective plane of order n.[15]

Our finite projective plane has $n^2 + n + 1$ points, and we call them $r, s, \ell_1, \ldots, \ell_{n-1}, x_{1,1}, \ldots, x_{1,n}, x_{2,1}, \ldots, x_{2,n}, \ldots, x_{n,1}, \ldots, x_{n,n}$.[16]

Our finite projective plane has $n^2 + n + 1$ lines, and we construct them as follows. One line is $B = \{r, s, \ell_1, \ldots, \ell_{n-1}\}$. Further, for each $i \in \{1, \ldots, n\}$, we have the line $R_i = \{r, x_{i,1}, \ldots, x_{i,n}\}$; and for each $j \in \{1, \ldots, n\}$, we have the line $S_j = \{s, x_{1,j}, \ldots, x_{n,j}\}$.[17] The points and lines constructed

[14] Can you see why?

[15] As an exercise, prove that this construction is correct.

[16] So, we have the points r and s; we have $n - 1$ points ℓ_i; and we have n^2 points $x_{i,j}$. In total, we have $2 + (n-1) + n^2 = n^2 + n + 1$ points.

[17] We remark that for all $i, j \in \{1, \ldots, n\}$, we have that $R_i \cap S_j = \{x_{i,j}\}$. We also remark that, so far, we have constructed $2n+1$ lines, and we need to construct $(n^2+n+1)-(2n+1) = n^2 - n = (n-1)n$ more.

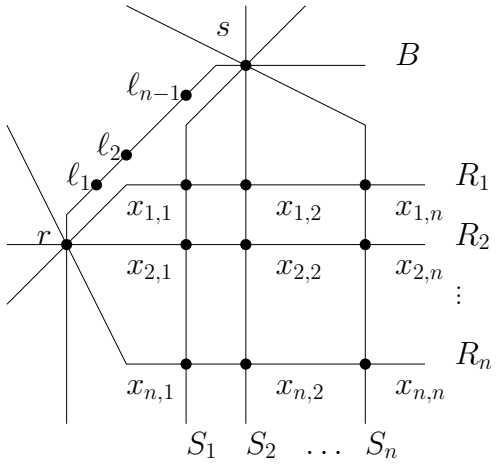


Figure 3.5: Points and lines (except the L_i^j's) of the projective plane from the proof of Theorem 3.3.3.

thus far are represented in Figure 3.5. Now, for each $i \in \{1, \ldots, n-1\}$, the point ℓ_i belongs to the (already constructed) line B, and to n other lines, call them $L_i^1, \ldots, L_i^n$, which we construct as follows. For all $i \in \{1, \ldots, n-1\}$ and $j \in \{1, \ldots, n\}$, we set $L_i^j = \{\ell_i\} \cup \{x_{p,q} \mid 1 \leq p, q \leq n, \text{and the } (p,q)\text{-th entry of } L_i \text{ is } j\}$.

The proof of correctness (i.e. of the fact that we have indeed constructed a finite projective plane) is left as an exercise.[18] □

We remark that the proof of the "(a) $\Longrightarrow$ (b)" part of Theorem 3.3.3 is similar to the "(b) $\Longrightarrow$ (a)" direction, only it goes the other way (from a finite projective plne to a collection of pairwise orthogonal Latin squares). To check your understanding, you can try to give the construction by yourself.

Example 3.3.4. *Let L_1 and L_2 be, respectively, the left (red) and right (blue) Latin Square from Figure 3.3. The finite projective plane of order* 3 *that corresponds to $\{L_1, L_2\}$ is as follows. Its points are*

$$r, s, \ell_1, \ell_2, x_{1,1}, x_{1,2}, x_{1,3}, x_{2,1}, x_{2,2}, x_{2,3}, x_{3,1}, x_{3,2}, x_{3,3}.$$

Its lines are as follows:

[18] We remark, however, that once we have shown that we have indeed constructed a finite projective plane, Theorem 3.1.4 immediately implies that the order of our finite projective plane is n (e.g. because we have $n^2 + n + 1$ points).

- $B = \{r, s, \ell_1, \ell_2\}$;
- $R_1 = \{r, x_{1,1}, x_{1,2}, x_{1,3}\}$;
- $R_2 = \{r, x_{2,1}, x_{2,2}, x_{2,3}\}$;
- $R_3 = \{r, x_{3,1}, x_{3,2}, x_{3,3}\}$;
- $S_1 = \{s, x_{1,1}, x_{2,1}, x_{3,1}\}$;
- $S_2 = \{s, x_{1,2}, x_{2,2}, x_{3,2}\}$;
- $S_3 = \{s, x_{1,3}, x_{2,3}, x_{3,3}\}$;
- $L_1^1 = \{\ell_1, x_{1,1}, x_{2,3}, x_{3,2}\}$;
- $L_1^2 = \{\ell_1, x_{1,2}, x_{2,1}, x_{3,3}\}$;
- $L_1^3 = \{\ell_1, x_{1,3}, x_{2,2}, x_{3,1}\}$;
- $L_2^1 = \{\ell_2, x_{1,1}, x_{2,2}, x_{3,3}\}$;
- $L_2^2 = \{\ell_2, x_{1,2}, x_{2,3}, x_{3,1}\}$;
- $L_2^3 = \{\ell_2, x_{1,3}, x_{2,1}, x_{3,2}\}$.

3.4 An algebraic construction of projective planes

Let $\mathbb{F}$ be any field. As usual, $+$ and $\cdot$ are, respectively, addition and multiplication in $\mathbb{F}$, and 0 and 1 are, respectively, the additive and multiplicative identity in $\mathbb{F}$. We construct the projective plane $\mathbb{F}P^2$ as follows. We begin with the set $T := \mathbb{F}^3 \setminus \{(0,0,0)\}$, i.e. the set of all ordered triples of elements of $\mathbb{F}$, except for the triple whose entries are all zero. We then form a binary relation $\sim$ on T as follows: for $(x_1, y_1, z_1), (x_2, y_2, z_2) \in T$, we have $(x_1, y_1, z_1) \sim (x_2, y_2, z_2)$ if and only if there exists a scalar $\lambda \in \mathbb{F} \setminus \{0\}$ such that $(x_2, y_2, z_2) = \lambda(x_1, y_1, z_1)$.[19] It is easy to see that $\sim$ is an equivalence relation on T.[20] The set of points of $\mathbb{F}P^2$ is $T/_\sim$; in other words, points of $\mathbb{F}P^2$ are the equivalence classes of the equivalence relation $\sim$ on T. We will denote the equivalence class of $(x, y, z) \in T$ by $\overline{(x, y, z)}$, so that $\overline{(x, y, z)} = \{(\lambda x, \lambda y, \lambda z) \mid \lambda \in \mathbb{F} \setminus \{0\}\}$. Thus, the set of points of $\mathbb{F}P^2$ is precisely the set $\{\overline{(x, y, z)} \mid (x, y, z) \in T\}$. Next, for each $(a, b, c) \in T$, we define $P(a, b, c)$ to be the set of all points $\overline{(x, y, z)}$ such that $ax + by + cz = 0$;[21] the lines of $\mathbb{F}P^2$ are precisely the sets $P(a, b, c)$ with $(a, b, c) \in T$. We remark that for all $(a_1, b_1, c_1), (a_2, b_2, c_2) \in T$, we have that $P(a_1, b_1, c_1) = P(a_2, b_2, c_2)$ if and only if $(a_1, b_1, c_1) \sim (a_2, b_2, c_2)$.[22]

Theorem 3.4.1. *For each field $\mathbb{F}$, $\mathbb{F}P^2$ is a projective plane.*[23]

[19] This means that $x_2 = \lambda x_1$, $y_2 = \lambda y_1$, and $z_2 = \lambda z_1$.

[20] Check this!

[21] Note that for all $\lambda \in \mathbb{F} \setminus \{0\}$, we have that $ax + by + cz = 0$ if and only if $a(\lambda x) + b(\lambda y) + c(\lambda z) = 0$, and so this is well-defined.

[22] Check this!

[23] This means that $\mathbb{F}P^2$ satisfies all the conditions from the definition of a finite projective plane, except that the set of points may possibly be infinite.

Proof. We use notation from the construction of $\mathbb{F}P^2$. We must verify that the points and lines of $\mathbb{F}P^2$ satisfy (P0), (P1), and (P2) from the definition of a projective plane.

First, we check that (P0) is satisfied for

$$Q \;:=\; \Big\{\overline{(1,0,0)}, \overline{(0,1,0)}, \overline{(0,0,1)}, \overline{(1,1,1)}\Big\}.$$

We note that each of the following four matrices

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

has rank three.[24] So, if A is any one of the four matrices above, then $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, and consequently, no line of $\mathbb{F}P^2$ contains three (or more) points of Q. So, (P0) is satisfied.

Next, we check that (P1) is satisfied. We fix distinct lines P_1, P_2 of $\mathbb{F}P^2$, and we show that $|P_1 \cap P_2| = 1$. By construction, there exist $(a_1, b_1, c_1), (a_2, b_2, c_2) \in T$ such that $P_1 = P(a_1, b_1, c_1)$ and $P_2 = P(a_2, b_2, c_2)$. Since $P_1 \neq P_2$, we have that $(a_1, b_1, c_1) \not\sim (a_2, b_2, c_2)$, that is, neither one of $(a_1, b_1, c_1), (a_2, b_2, c_2)$ is a scalar multiple of the other. We now use Linear Algebra. We consider the 2×3 matrix

$$A \;=\; \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix}.$$

Since neither row of A is a scalar multiple of the other, we see that $\operatorname{rank}(A) = 2$. On the other hand, by the Rank-Nullity Theorem, we have that $\operatorname{rank}(A) + \dim \ker(A) = 3$. So, $\dim \ker(A) = 1$. Let $\left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \right\}$ be a basis for $\ker(A)$;[25] then $P_1 \cap P_2 = \Big\{\overline{(x, y, z)}\Big\}$, and we deduce that $|P_1 \cap P_2| = 1$. Thus, (P1) is satisfied.

The proof of the fact that (P2) is satisfied is analogous to the proof that (P1) is satisfied.[26] □

[24] Note that each of these matrices was obtained by taking three of the four elements of Q and (essentially) turning them into rows of the matrix. Each selection of three elements of Q corresponds to one of our four matrices.

[25] So, $(x, y, z) \neq (0, 0, 0)$, and we see that $(x, y, z) \in T$. Furthermore, we have that $\ker(A) = \Big\{ \begin{bmatrix} \lambda x \\ \lambda y \\ \lambda z \end{bmatrix} \mid \lambda \in \mathbb{F} \Big\}$.

[26] Check this!

Theorem 3.4.2. *If $\mathbb{F}$ is a finite field, with $|\mathbb{F}| = n$, then $\mathbb{F}P^2$ is a finite projective plane of order n.*

Proof. By Theorem 3.4.1, $\mathbb{F}P^2$ is a projective plane. Furthermore, since $\mathbb{F}$ is finite, it is obvious that the projective plane $\mathbb{F}P^2$ is finite. We must show that the order of $\mathbb{F}P^2$ is n. In view of Theorem 3.1.4, it suffices to show that $\mathbb{F}P^2$ has precisely n^2+n+1 points. Now, note that for all $(x, y, z) \in T$, there exists a unique triple $(x', y', z') \in T$ such that the last non-zero coordinate of (x', y', z') is 1 and $(x, y, z) \sim (x', y', z')$.[27] Now, there are n^2 triples of the form $(x, y, 1)$ in T; there are n triples of the form $(x, 1, 0)$ in T; and there is one triple $(1, 0, 0)$ in T. So, there are $n^2 + n + 1$ equivalence classes of $\sim$, that is, $\mathbb{F}P^2$ has n^2+n+1 points. As we already pointed out, Theorem 3.1.4 now implies that the finite projective plane $\mathbb{F}P^2$ is of order n. □

It is well-known that for all integers $n \geq 2$, there exists a field of size n if and only if n is a power of a prime (that is, if and only if there exist a prime number p and a positive integer k such that $n = p^k$). This, together with Theorem 3.4.2, implies that if $n \geq 2$ is a power of a prime, then there exists a finite projective plane of order n. However, it is not known whether there exists a finite projective plane whose order is not a power of a prime.

[27] For existence, we observe that for all $(x, y, z) \in T$, we have the following:

- if $z \neq 0$, then $(x, y, z) \sim (z^{-1}x, z^{-1}y, 1)$;
- if $z = 0$ and $y \neq 0$, then $(x, y, z) \sim (y^{-1}x, 1, 0)$;
- if $y = z = 0$, then $x \neq 0$ (since x, y, z cannot all be zero) and $(x, y, z) \sim (1, 0, 0)$.

Can you check uniqueness?

Chapter 4

Flows and cuts in networks. Matchings in bipartite graphs

4.1 Network flows and cuts

A *network* is an ordered four-tuple (G, s, t, c), where G is an oriented graph, s and t are two distinct vertices of this graph (called the *source* and *sink*, respectively), and $c : E(G) \to [0, +\infty)$ is a function, called the *capacity function* (see Figure 4.1 for an example). The *capacity* of an edge $e \in E(G)$ is the number $c(e)$.

Networks can be used to model, for example, a system of pipes used to transport some resource, such as water or oil; capacities would be the number of units of volume that a given pipe can transport per unit time.

A *feasible flow* (or simply *flow*) in a network (G, s, t, c) is a function $f : E(G) \to [0, +\infty)$ that satisfies the following two properties (see Figure 4.2 for an example):

- $f(e) \leq c(e)$ for all $e \in E(G)$;[1]
- for all $v \in V(G) \setminus \{s, t\}$, we have $\sum_{(x,v)\in E(G)} f(x, v) = \sum_{(v,y)\in E(G)} f(v, y)$.[2]

The *value* of a flow f is

$$val(f) \quad = \quad \Big(\sum_{(s,x)\in E(G)} f(s, x) \Big) - \Big(\sum_{(x,s)\in E(G)} f(x, s) \Big).$$

[1] This means that flow cannot be higher than capacity.

[2] This means that, for each vertex other than the source and the sink, the in-flow is equal to the out-flow. This condition is called the *conservation of flow* condition.

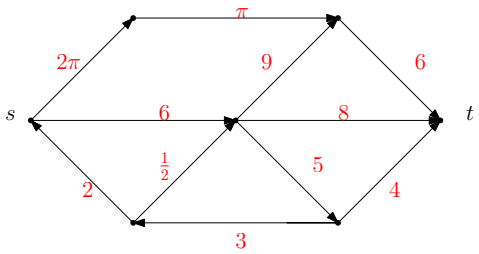


Figure 4.1: A network with capacities in red.

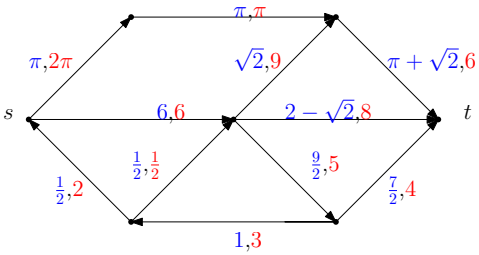


Figure 4.2: A network flow. Flows are in blue and capacities are in red.

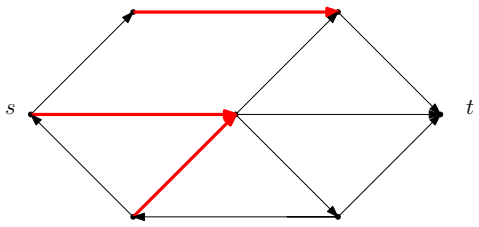


Figure 4.3: A cut in a network. (The edges of the cut are in red.)

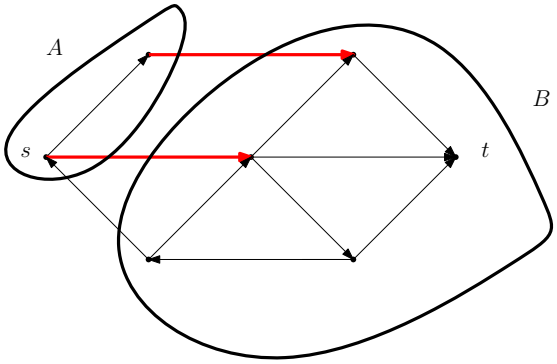


Figure 4.4: A cut $S(A, B)$ in a network. (The edges of the cut are in red.)

For instance, the value of the flow in Figure 4.2 is $(\pi + 6) - \frac{1}{2} = \pi + \frac{11}{2}$. A *maximum flow* in (G, s, t, c) is a flow f^* that has maximum value, i.e. one that satisfies $val(f) \leq val(f^*)$ for all flows f.

Theorem 4.1.1. *Every network (G, s, t, c) has a maximum flow.*

Proof. Omitted. □

Theorem 4.1.1 should certainly seem plausible, and yet it is not entirely obvious how one might prove it (since the number of flows is, typically, infinite). The proof relies on certain results from Analysis, which we omit.

As usual, for a (directed or undirected) graph G and a set $R \subseteq E(G)$, we denote by $G - R$ the graph obtained from G by deleting all edges in R.

An *s, t-cut*, or simply *cut*, in a network (G, s, t, c) is a set $R \subseteq E(G)$ such that $G - R$ contains no directed path from s to t (see Figure 4.3 for an example).[3] The *capacity* of the cut R is $c(R) = \sum_{e \in R} c(e)$.

Our main theorem (proven in the next section) is the following.

Max-flow min-cut theorem. *The maximum value of a flow in a network is equal to the minimum capacity of a cut in that network.*

4.2 Proof of the Max-flow min-cut theorem

We now need some terminology and notation. First, for a network (G, s, t, c), a flow f in that network, and a set of edges $R \subseteq E(G)$, we write

[3] Equivalently, every directed path from s to t in G uses at least one edge of R.

- $c(R) = \sum_{e \in R} c(e)$;
- $f(R) = \sum_{e \in R} f(e)$.

Next, for a directed graph G and disjoint sets $A, B \subseteq V(G)$, we set

$$S(A,B) \quad := \quad \{(a,b) \in E \mid a \in A, b \in B\}.$$

Thus, $S(A,B)$ is the set of all edges from A to B (see Figure 4.4 for an example).[4]

For a network (G,s,t,c), disjoint sets $A, B \subseteq V(G)$, and a flow f, we write

- $c(A,B) = c(S(A,B))$;[5]
- $f(A,B) = f(S(A,B))$.

Proposition 4.2.1. *Let (G,s,t,c) be a network, and let (A,B) be a partition of $V(G)$ such that $s \in A$ and $t \in B$. Then $S(A,B)$ is a cut in (G,s,t,c).*

Proof. Let $P = p_0, p_1, \ldots, p_\ell$, with $p_0 = s$ and $p_\ell = t$, be a directed path in G. We must show that P uses at least one edge of $S(A,B)$. By hypothesis, $p_0 = s \in A$ and $p_\ell = t \in B$; let $i \in \{0, \ldots, \ell - 1\}$ be maximum with the property that $p_i \in A$. Then $p_{i+1} \in B$, and see that $(p_i, p_{i+1}) \in S(A,B)$, i.e. the directed path P uses an edge of $S(A,B)$, which is what we needed to show. □

Proposition 4.2.2. *Let (G,s,t,c) be a network, and let R be a cut in this network. Then there exists a partition (A,B) of $V(G)$ such that $s \in A$, $t \in B$, and $S(A,B) \subseteq R$.*[6]

Proof. Let A be the set of all vertices $v \in V(G)$ such that $G - R$ contains a directed path from s to v, and set $B = V(G) \setminus A$. Clearly, $s \in A$ and $t \in B$.[7] We now claim that $S(A,B) \subseteq R$. Suppose otherwise, and fix an edge

[4] $S(A,B)$ does **not** contain edges from B to A!

[5] According to our notation, $c(S(A,B)) = \sum_{e \in S(A,B)} c(e)$, i.e. $c(A,B)$ is the sum of capacities of all the edges from A to B.

[6] Note that this implies that $c(A,B) \leq c(R)$. Thus, in our proof of the Max-flow min-cut theorem, it will be enough to consider cuts of the form $S(A,B)$, where (A,B) is a partition of $V(G)$, with $s \in A$ and $t \in B$; cuts of this form are sometimes called *elementary cuts*.

[7] The fact that $t \notin A$ follows from the fact that R is a cut in (G,s,t,c), and so there are no directed paths from s to t in $G - R$; so, $t \in B$.

$(x, y) \in S(A, B) \setminus R$. (In particular, $x \in A$ and $y \in B$.) Let $P = p_0, \ldots, p_\ell$, with $p_0 = s$ and $p_\ell = x$, be a directed path in $G - R$. Since $(x, y) \notin R$, we then have that $p_0, \ldots, p_\ell, y$ is a directed path from s to y in $G - R$, and so by construction, we have that $y \in A$, contrary to the fact that $y \in B$. □

Lemma 4.2.3. *Let f be a flow in a network (G, s, t, c), and let (A, B) be a partition of $V(G)$ such that $s \in A$ and $t \in B$. Then*

$$val(f) \quad = \quad f(A, B) - f(B, A).$$

In particular,[8] *we have that*

$$val(f) \quad = \quad \Big(\sum_{(x,t) \in E(G)} f(x, t) \Big) - \Big(\sum_{(t,x) \in E(G)} f(t, x) \Big).$$

Proof. By the definition of a flow, for all vertices $v \in A \setminus \{s\}$, we have that

$$\Big(\sum_{(v,x) \in E(G)} f(v, x) \Big) - \Big(\sum_{(x,v) \in E(G)} f(x, v) \Big) \quad = \quad 0,$$

and consequently,

$$\sum_{v \in A \setminus \{s\}} \left(\Big(\sum_{(v,x) \in E(G)} f(v, x) \Big) - \Big(\sum_{(x,v) \in E(G)} f(x, v) \Big) \right) \quad = \quad 0,$$

On the other hand, for the source s, we have that

$$\Big(\sum_{(s,x) \in E(G)} f(s, x) \Big) - \Big(\sum_{(x,s) \in E(G)} f(x, s) \Big) \quad = \quad val(f).$$

By adding the last two equalities, we get

$$\sum_{v \in A} \left(\Big(\sum_{(v,x) \in E(G)} f(v, x) \Big) - \Big(\sum_{(x,v) \in E(G)} f(x, v) \Big) \right) \quad = \quad val(f).$$

Note that for each edge $(u_1, u_2) \in E(G)$ such that $u_1, u_2 \in A$, the term $f(u_1, u_2)$ appears exactly twice in the sum above: once with the + sign,[9] and one with the − sign.[10] After we cancel out such terms, what remains is precisely $f(A, B) - f(B, A) = val(f)$, which is what we needed to show. □

[8] This happens if we take $A = V(G) \setminus \{t\}$ and $B = \{t\}$.

[9] For this, we take $v = u_1$, $x = u_2$, and $(v, x) \in E(G)$ to add $f(u_1, u_2)$ (via the first sum).

[10] For this, we take $v = u_2$, $x = u_1$, and $(x, v) \in E(G)$ to subtract $f(u_1, u_2)$ (via the second sum).

Corollary 4.2.4. *Let f be a flow in a network (G, s, t, c), and let R be a cut in (G, s, t, c). Then $val(f) \leq c(R)$.*

Proof. By Proposition 4.2.2, there exists a partition (A, B) of $V(G)$ such that $s \in A$, $t \in B$, and $S(A, B) \subseteq R$. Then

$$\begin{aligned} val(f) &= f(A,B) - f(B,A) && \text{by Lemma 4.2.3} \\ &\leq f(A,B) && \text{because } f(e) \geq 0 \text{ for all } e \in E(G) \\ &\leq c(A,B) && \text{because } f(e) \leq c(e) \text{ for all } e \in E(G) \\ &\leq c(R) && \text{because } S(A,B) \subseteq R \text{ and} \\ & && \text{and } c(e) \geq 0 \text{ for all } e \in E(G) \end{aligned}$$

which is what we needed to show. □

We now introduce a key new concept: that of an "augmenting path." First, an (s, t)-*path* in a network (G, s, t, c) is a sequence $v_0, v_1, \ldots, v_\ell$ of pairwise distinct vertices of G such that $v_0 = s$, $v_\ell = t$, and for all $i \in \{0, \ldots, \ell - 1\}$, we have that one of (v_i, v_{i+1}) and (v_{i+1}, v_i) belongs to $E(G)$. Note that an (s, t)-path may, but need not be, a directed (s, t)-path (see the figure below for an example).

Now, given a flow f in the network (G, s, t, c), an (s, t)-path $v_0, v_1, \ldots, v_\ell$ in (G, s, t, c) is said to be an f-*augmenting path* if the following two conditions are satisfied (see Figure 4.5 for an example):

- for all $i \in \{1, \ldots, \ell - 1\}$ such that $(v_i, v_{i+1}) \in E(G)$, we have that $f(v_i, v_{i+1}) < c(v_i, v_{i+1})$;[11]
- for all $i \in \{1, \ldots, \ell - 1\}$ such that $(v_{i+1}, v_i) \in E(G)$, we have that $f(v_{i+1}, v_i) > 0$.[12]

Lemma 4.2.5. *Let f be a flow in a network (G, s, t, c). Then f is a maximum flow if and only if there does not exist an f-augmenting path in (G, s, t, c). Furthermore, if f is a maximum flow, then there exists a cut R in (G, s, t, c) such that $val(f) = c(R)$.*

[11] So, the flow through each edge directed "with the flow" is strictly smaller than the capacity of that edge.

[12] So, the flow through each edge directed "against the flow" is strictly positive.

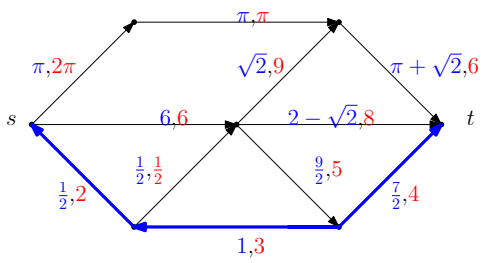


Figure 4.5: An f-augmenting path (edges in blue) in a network (G, s, t, c). (Flows are in blue and capacities are in red.)

Proof. It suffices to prove the following two statements:

(a) if there exists an f-augmenting path in (G, s, t, c), then f is not a maximum flow in (G, s, t, c);

(b) if there does not exist an f-augmenting path in (G, s, t, c), then f is a maximum flow in (G, s, t, c), and furthermore, there exists a cut R in (G, s, t, c) such that $val(f) = c(R)$.

We first prove (a). Suppose that $v_0, \ldots, v_\ell$ (with $v_0 = s$ and $v_\ell = t$) is an f-augmenting path in (G, s, t, c). Now, set

- $\varepsilon_1 := \min\Big(\{c(v_i, v_{i+1}) - f(v_i, v_{i+1}) \mid 0 \leq i \leq \ell - 1, (v_i, v_{i+1}) \in E(G)\} \cup \{\infty\}\Big)$;
- $\varepsilon_2 := \min\Big(\{f(v_{i+1}, v_i) \mid 0 \leq i \leq \ell - 1, (v_{i+1}, v_i) \in E(G)\} \cup \{\infty\}\Big)$;
- $\varepsilon := \min\{\varepsilon_1, \varepsilon_2\}$.[13]

Since $v_0, \ldots, v_\ell$ is an f-augmenting path, we have that $\varepsilon_1, \varepsilon_2 > 0$, and consequently, $\varepsilon > 0$. We now define a new flow f' as follows:

- $f'(v_i, v_{i+1}) = f(v_i, v_{i+1}) + \varepsilon$ for all $i \in \{0, \ldots, \ell - 1\}$ such that $(v_i, v_{i+1}) \in E(G)$;[14]

[13] The reason we have ∞ in the definition of ε_1 and ε_2 is because our f-augmenting path may have only "with-the-flow" or only "against-the-flow" edges, and we cannot take the minimum of an empty set. Note, however, that at least one of ε_1 and ε_2 is a real number (and not ∞), and consequently, ε is a real number.

[14] So, for edges on our augmenting path directed with the flow, we increase the flow by ε.

- $f'(v_{i+1}, v_i) = f(v_{i+1}, v_i) - \varepsilon$ for all $i \in \{0, \ldots, \ell-1\}$ such that $(v_{i+1}, v_i) \in E(G)$;[15]
- $f'(e) = f(e)$ for all other edges e.

It is easy to verify that f' is indeed a feasible flow.[16] Furthermore, by construction, $val(f') = val(f) + \varepsilon$, and so (since $\varepsilon > 0$) we have that $val(f') > val(f)$. Thus, f is not a maximum flow in (G, s, t, c).

It remains to prove (b). For this, we suppose that (G, s, t, c) does not admit an f-augmenting path, and we show that f is a maximum flow. Let A be the set of all vertices $v \in V(G)$ such that there exists a path $v_0, \ldots, v_\ell$ with $v_0 = s$ and $v_\ell = v$, and satisfying the following two properties:[17]

- for all $i \in \{1, \ldots, \ell - 1\}$ such that $(v_i, v_{i+1}) \in E(G)$, we have that $f(v_i, v_{i+1}) < c(v_i, v_{i+1})$;
- for all $i \in \{1, \ldots, \ell - 1\}$ such that $(v_{i+1}, v_i) \in E(G)$, we have that $f(v_{i+1}, v_i) > 0$.

Set $B = V(G) \setminus A$. Clearly, $s \in A$ and $t \in B$.[18] Further, for all $x \in A$ and $y \in B$,

- if $(x, y) \in E(G)$, then $f(x, y) = c(x, y)$, and
- if $(y, x) \in E(G)$, then $f(y, x) = 0$.[19]

Note that this implies that $f(A, B) = c(A, B)$ and $f(B, A) = 0$. But now we have that

$$\begin{aligned} val(f) &= f(A, B) - f(B, A) && \text{by Lemma 4.2.3} \\ &= c(A, B) && \text{because } f(A, B) = c(A, B) \\ & && \text{and } f(B, A) = 0. \end{aligned}$$

[15] So, for edges on our augmenting path directed against the flow, we decrease the flow by ε.

[16] Check this!

[17] Essentially, but somewhat informally, we are choosing A to be the set of all vertices $v \in V(G)$ such that there exists an f-augmenting path from s to v.

[18] If we had $t \in A$, then by the construction of A, there would be an f-augmenting path in (G, s, t, c), a contradiction.

[19] Otherwise, there would be an f-augmenting path from s to y, contrary to the fact that $y \notin A$.

By Proposition 4.2.1, we know that $R := S(A, B)$ is a cut, and by what we just showed, $val(f) = c(A, B) = c(R)$. It now follows from Corollary 4.2.4 that f is a maximum flow in (G, s, t, c).[20] □

We are now ready to prove the Max-flow min-cut theorem, restated below.

Max-flow min-cut theorem. *The maximum value of a flow in a network is equal to the minimum capacity of a cut in that network.*

Proof. Let (G, s, t, c) be a network, and let f be a maximum flow in it (the existence of such a flow is guaranteed by Theorem 4.1.1). By Lemma 4.2.5, there exists a cut R in (G, s, t, c) such that $val(f) = c(R)$. Furthermore, for any cut R' in (G, s, t, c), Corollary 4.2.4 guarantees that $val(f) \leq c(R')$, and consequently, $c(R) \leq c(R')$; thus, R is a cut of minimum capacity in (G, s, t, c). □

4.3 The Ford-Fulkerson algorithm

The proof of Lemma 4.2.5 can easily be converted into an algorithm[21] that finds a maximum flow and a minimum capacity of a cut in an input network. The idea is to repeatedly find augmenting paths and update the flow (increasing its value). When no augmenting path exists, we instead find a cut whose capacity is equal to the value of our flow, which (by Corollary 4.2.4) guarantees that this cut is of minimum capacity.

Before we describe the algorithm, a couple of remarks are in order. First of all, the term "algorithm" is not entirely appropriate here because for some networks, the procedure might not terminate. This, however, can only happen if the capacities are irrational (a concrete example is given at the end of this section).[22] If all capacities are rational, then the algorithm will indeed terminate (see Theorems 4.3.4 and 4.3.5). We also emphasize that, if the algorithm does terminate, then its output is correct.

The procedure that we now describe will be used repeatedly as a subroutine in the Ford-Fulkerson algorithm. Suppose that f is a flow in a network (G, s, t, c). We now either find an f-augmenting path in (G, s, t, c), or we find a cut whose capacity is $val(f)$, as follows:

[20] Indeed, suppose f' is any flow in (G, s, t, c). Then by Corollary 4.2.4, we have that $val(f') \leq R$; since $val(f) = c(R)$, it follows that $val(f') \leq val(f)$.

[21] As we shall see, technically, this is not quite an algorithm.

[22] Note, however, that it is possible that the algorithm terminates even if some (or all) capacities are irrational.

1. Set $A := \{s\}$.
2. While $t \notin A$:
   (a) Either find vertices $x \in A$ and $y \in V(G) \setminus A$ such that
   - $(x, y) \in E(G)$ and $f(x, y) < c(x, y)$, or
   - $(y, x) \in E(G)$ and $f(y, x) > 0$,

   or determine that such x and y do not exist.
   (b) If we found x and y, then we set backpoint$(y) := x$, and we update $A := A \cup \{y\}$.
   (c) Otherwise, we stop and return the cut $S(A, V(G) \setminus A)$.[23]
3. Construct an f-augmenting path by following backpoints starting from t, and return this path.

Example 4.3.1. *Consider the flow f in the network (G, s, t, c) in Figure 4.6. Either find an f-augmenting path, or find a cut whose capacity is $val(f)$.*

Solution. We begin with $A := \{s\}$. We now iterate several times.

1. We select $s \in A$ and $u \in V(G) \setminus A$, and we set $A := \{s, u\}$ and backpoint$(u) := s$.
2. We select $s \in A$ and $w \in V(G) \setminus A$, and we set $A := \{s, u, w\}$ and backpoint$(w) := s$.
3. We select $u \in A$ and $v \in V(G) \setminus A$, and we set $A := \{s, u, w, v\}$ and backpoint$(v) := u$.
4. We select $v \in A$ and $t \in V(G) \setminus A$, and we set $A := \{s, u, w, v, t\}$ and backpoint$(t) := v$.

We now reconstruct our f-augmenting path: s, u, v, t. □

Note that the choices made in our solution to Example 4.3.1 were not unique. For instance, in step 3, we could have made the following choice instead:

3. We select $w \in A$ and $t \in V(G) \setminus A$, and we set $A := \{s, u, w, t\}$ and backpoint$(t) := w$.

This would have yielded the augmenting path s, w, t.

[23] In this case, an argument analogous to the proof of Lemma 4.2.5 guarantees that $c(A, V(G) \setminus A) = val(f)$.

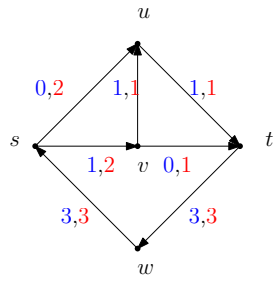


Figure 4.6: The network and flow from Example 4.3.1. Flows are in blue and capacities in red.

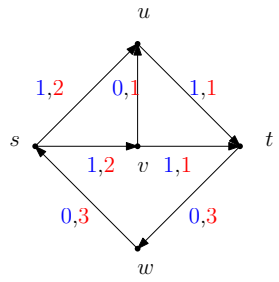


Figure 4.7: The network and flow from Example 4.3.2. Flows are in blue and capacities in red.

Example 4.3.2. *Consider the flow f in the network (G, s, t, c) in Figure 4.7. Either find an f-augmenting path, or find a cut whose capacity is $val(f)$.*

Solution. We begin with $A := \{s\}$. We now iterate several times.

1. We select $s \in A$ and $u \in V(G) \setminus A$, and we set $A := \{s, u\}$ and backpoint$(u) := s$.
2. We select $s \in A$ and $v \in V(G) \setminus A$, and we set $A := \{s, u, v\}$ and backpoint$(v) := s$.

There are now no further vertices that we can select, and $t \notin A$. We now see that $S(A, V(G) \setminus A) = \{(u, t), (v, t)\}$ is a cut whose capacity is 2, which is precisely equal to $val(f)$. □

We now describe the Ford-Fulkerson algorithm, which finds a maximum flow in a network (G, s, t, c). Its steps are as follows:

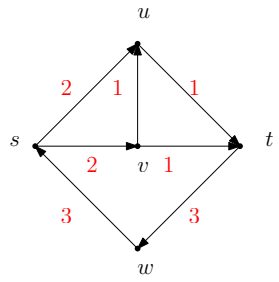


Figure 4.8: The network from Example 4.3.3.

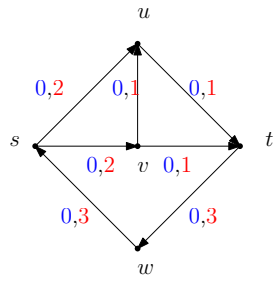
1. Set $f(e) := 0$ for all $e \in E(G)$.
2. While there exists an f-augmenting path in the network:
   (a) Find an f-augmenting path $v_0, \ldots, v_\ell$ (with $v_0 = s$ and $v_\ell = t$).
   (b) Set
   - $\varepsilon_1 := \min\Big(\{c(v_i, v_{i+1}) - f(v_i, v_{i+1}) \mid 0 \leq i \leq \ell - 1, (v_i, v_{i+1}) \in E(G)\} \cup \{\infty\}\Big)$;
   - $\varepsilon_2 := \min\Big(\{f(v_{i+1}, v_i) \mid 0 \leq i \leq \ell - 1, (v_{i+1}, v_i) \in E(G)\} \cup \{\infty\}\Big)$;
   - $\varepsilon := \min\{\varepsilon_1, \varepsilon_2\}$.

   (c) Update f as follows:
   - $f(v_i, v_{i+1}) := f(v_i, v_{i+1}) + \varepsilon$ for all $i \in \{0, \ldots, \ell - 1\}$ such that $(v_i, v_{i+1}) \in E(G)$;[24]
   - $f(v_{i+1}, v_i) := f(v_{i+1}, v_i) - \varepsilon$ for all $i \in \{0, \ldots, \ell - 1\}$ such that $(v_{i+1}, v_i) \in E(G)$.[25]
3. Return f.

Example 4.3.3. *Find a maximum flow and an a cut of minimum capacity in the network represented in Figure 4.8.*

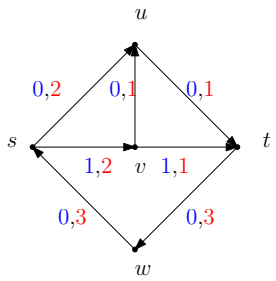
Solution. We first set $f(e) = 0$ for all $e \in E(G)$ (see the figure below, with flows in blue and capacities in red).

[24] So, for edges on our augmenting path directed with the flow, we increase the flow by ε.
[25] So, for edges on our augmenting path directed against the flow, we decrease the flow by ε.

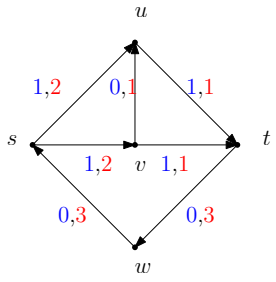


We now iterate several times.

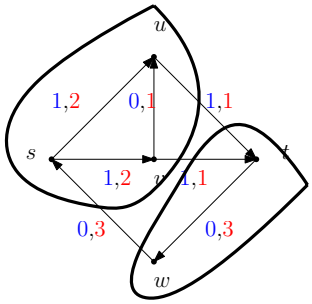
1. We find an augmenting path s, v, t, we get $\varepsilon = 1$, and we update f as in the picture below (flows are in blue and capacities are in red).



2. We find an augmenting path s, u, t, we get $\varepsilon = 1$, and we update f as in the picture below (flows are in blue and capacities are in red).



3. We find a cut $S(\{s, u, v\}, \{w, t\}) = \{(u, t), (v, t)\}$ of capacity is 2, which is precisely equal to $val(f)$.



The flow f is a maximum flow, and the cut $S(\{s,u,v\},\{w,t\}) = \{(u,t),(v,t)\}$ is a minimum capacity cut. □

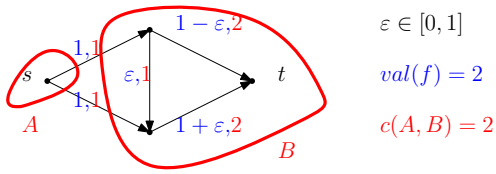
As we already mentioned, if all capacities of the input network are rational, then the Ford-Fulkerson algorithm terminates. Moreover, the output flow through each edge is rational. We first prove this for integer capacities (see Theorem 4.3.4), and then more generally for rational capacities (see Theorem 4.3.5).

Theorem 4.3.4. *Let (G,s,t,c) be a network in which all capacities are non-negative integers. Then, for input (G,s,t,c), the Ford-Fulkerson algorithm terminates and outputs a maximum flow, and furthermore, the output flow through each edge is a non-negative integer. In particular, some maximum flow in (G,s,t,c) has the property that flows through all edges are non-negative integers.*

Proof. If we begin with an integer flow (i.e. a flow f such that $f(e)$ is an integer for each edge e in our network) in the network (G,s,t,c), and we find an augmenting path, then since all capacities are integers, the number ε (defined as in the description of the Ford-Fulkerson algorithm) will be a positive integer; so, the updated flow will still be an integer flow, since the flow through an edge can either remain unchanged, or increase by ε, or decrease by ε. Now, the initial flow created by the Ford-Fulkerson algorithm for the network (G,s,t,c) is the zero-flow (and so in particular, an integer flow), and by what we just proved, after each iteration, the new flow is still an integer flow. In each iteration, the value of the flow increases by a positive integer (namely, by the ε that we compute for that iteration), and possible values of feasible flows are bounded above (e.g. by the sum of capacities). So, there can be only finitely many iterations, and in particular, the algorithm terminates. The fact that the algorithm returns a correct answer follows from its stopping criterion: the algorithm terminates and returns a flow f once

there are no f-augmenting paths, and in this case, Lemma 4.2.5 guarantees that f is a maximum flow. □

Note that Theorem 4.3.4 does **not** state that every maximum flow in a network with integer capacities is an integer flow. It merely guarantees that at least one maximum flow in such a network is an integer flow.[26] For instance, the flow in the picture below is maximum for any value of $\varepsilon \in [0, 1]$, but only two values of ε (namely, $\varepsilon = 0$ and $\varepsilon = 1$) yield an integer flow.



Theorem 4.3.4 is important for certain theoretical applications (we will see this in section 4.4), as well for certain practical applications.[27]

If we replace the word "integer" by the word "rational" in the statemnent of Theorem 4.3.4, we still get a correct statement.

Theorem 4.3.5. *Let (G, s, t, c) be a network in which all capacities are non-negative rational numbers. Then, for input (G, s, t, c), the Ford-Fulkerson algorithm terminates and outputs a maximum flow, and furthermore, the output flow through each edge is an non-negative rational number. In particular, some maximum flow in (G, s, t, c) has the property that flows through all edges are non-negative rational numbers.*

Proof. Let d be a positive integer such that all capacities in (G, s, t, c) are integer multiples of $\frac{1}{d}$.[28] Now the proof is completely analogous to that of Theorem 4.3.4, except that instead of integers, we have integer multiples of $\frac{1}{d}$ (for flows and capacities) throughout.[29] □

The key point of the proof of Theorem 4.3.5 is that there exists some positive integer d such that in each iteration, the value of the flow increases

[26] While the maximum value of a flow in a network is unique, there may be many (possibly infinitely many) flows in the network that have that value, and by definition, all such flows are maximum.

[27] Consider, for example, a network that models a transportation network of trucks, where the capacity of a truck is the number of containers that it can carry. Certainly, we would want a maximum flow that is an integer flow. (A truck should not transport $\frac{7}{3}$ or $\sqrt[3]{\pi}$ containers!)

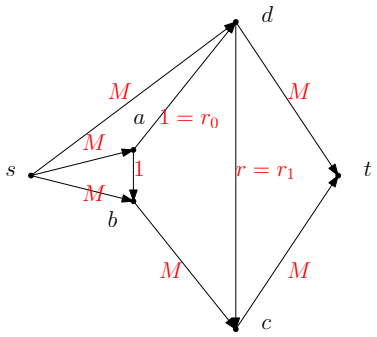
[28] To see that d exists, we can first write all capacities in (G, s, t, c) as fractions, and then we take d to be the least common multiple of the denominators of the capacities.

[29] Check this!

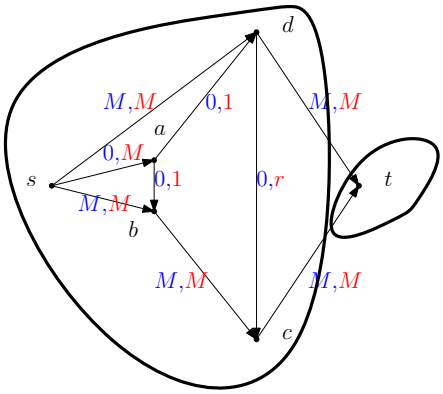
by at least $\frac{1}{d}$, and so there cannot be infinitely many iterations. If (some of) our capacities are irrational, such a d need not exist. Let us give an example of this.[30] First, let $r = \frac{-1+\sqrt{5}}{2}$, and let the sequence $\{r_n\}_{n=0}^{\infty}$ be defined recursively as follows:

- $r_0 = 1$ and $r_1 = r$;
- $r_{n+2} = r_n - r_{n+1}$ for all integers $n \geq 0$.

It is easy to check that $r_n = r^n$ for all integers $n \geq 0$.[31] Let M be some large number (say, $M = 100$). We now consider the network flow below.



The maximum value of a flow in this network is $2M$, as certified by the flow represented below, and the cut $(\{s, a, b, c, d\}, \{t\})$ of capacity $2M$.

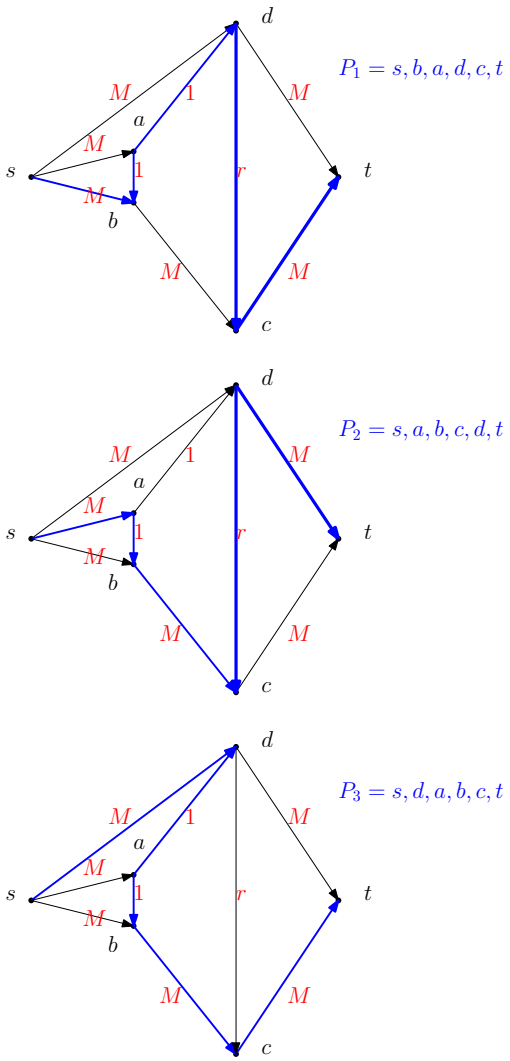


[30] We give only describe the construction. If you'd like a challenge, prove that it actually works. (It's a slightly messy induction.)

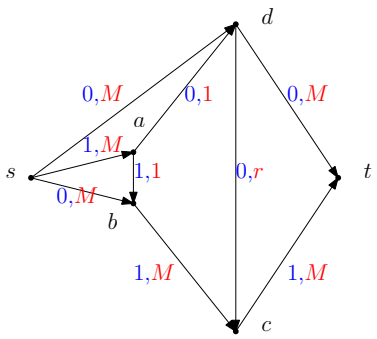
[31] This formula can be obtained using, for example, generating functions. Correctness is easily verified by induction.

We note that the flow above can easily be obtained in two iterations of the Ford-Fulkerson algorithm: we start with the zero flow, then we choose the augmenting path s, d, t (with $\varepsilon = M$), and then we choose the augmenting path s, b, c, t (again with $\varepsilon = M$). However, if we keep choosing "bad" augmenting paths, the algorithm may continue forever, as we describe below.

Let P_1 be the s,t-path s, b, a, d, c, t; let P_2 be the s,t-path s, a, b, c, d, t; and let P_3 be the s,t-path s, d, a, b, c, t.



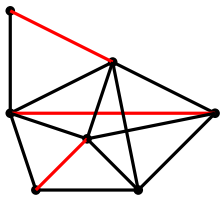
We start with the zero flow f_0, and then we use the augmenting path s, a, b, c, t (with $\varepsilon = 1$), thus obtaining the flow f_1, represented below.



From now on, we cyclically select augmenting paths P_1, P_2, P_3. It can be then shown by induction that the algorithm never terminates,[32] and furthermore, the value of the flows that the algorithm produces converges to $1+2\sum_{n=2}^{\infty} r_n = 3$, whereas the maximum flow in our network has value $2M > 3$.[33]

4.4 Matchings and transversals

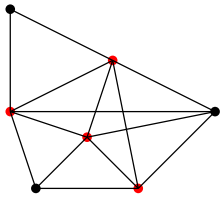
A *matching* in a graph G is a set of edges $M \subseteq E(G)$ such that every vertex of G is incident with at most one edge in M. An example of a matching in a graph is given below (edges of the matching are in red).



A *vertex cover* of a graph G is any set C of vertices of G such that every edge of G has at least one endpoint in C. An example of a vertex cover in a graph is given below (vertices of the vertex cover are in red).

[32] This is, essentially, because ε tends to zero as we keep iterating. Recall that in the case of rational capacities (see Theorem 4.3.5), we could always find an integer $d \geq 1$ such that in each iteration, we had $\varepsilon \geq \frac{1}{d}$. This need not be the case if (some of) our capacities are irrational.

[33] If you want a bit of a challenge, try to prove (by induction) that this is indeed correct.



The Kőnig-Egerváry theorem. *The maximum size of a matching in a bipartite graph is equal to the minimum size of a vertex cover in that graph.*

Proof. Let G be a bipartite graph with bipartition (A, B). Clearly, it suffices to prove the following two statements:

(a) for every matching M and every vertex cover C of G, we have that $|M| \leq |C|$;[34]

(b) there exist a matching M and a vertex cover C of G such that $|M| = |C|$.

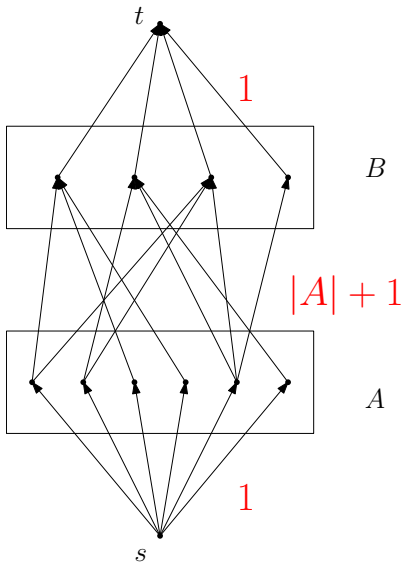
We begin by proving (a). Fix a matching M and a vertex cover C in G. Clearly, every edge of M has at least one endpoint in C. Since no two edges of M share an endpoint, we deduce that $|M| \leq |C|$. This proves (a).

It remains to prove (b). Let s and t be two new vertices, i.e. $s \neq t$ and $s, t \notin V(G)$. We now form a network (G', s, t, c) as follows:

- $V(G') = V(G) \cup \{s, t\}$;
- $E(G') = \{(s, a) \mid a \in A\} \cup \{(a, b) \mid a \in A, b \in B, ab \in E(G)\} \cup \{(b, t) \mid b \in B\}$;
- $c(a, b) = |A| + 1$ for all $(a, b) \in E(G')$, with $a \in A$ and $b \in B$;
- $c(s, a) = 1$ for all $a \in A$;
- $c(b, t) = 1$ for all $b \in B$.

Let f be a maximum flow in (G', s, t, c), and let R be a cut of minimum capacity. By Theorem 4.3.4, we may assume that $f(e)$ is an integer for all $e \in E(G')$. By the Max-flow min-cut theorem, we know that $val(f) = c(R)$. It now suffices to produce a matching of size $val(f)$ and vertex cover of size $c(R)$.

[34] In fact, (a) holds for all graphs, not just bipartite ones. However, there are (non-bipartite) graphs for which (b) fails.



First, we claim that $f(e) \in \{0,1\}$ for all $e \in E(G')$. Clearly, it suffices to show that $f(e) \leq 1$ for all $e \in E(G')$.[35] For all $a \in A$, we have that $f(s,a) \leq c(s,a) = 1$; and for all $b \in B$, we have that $f(b,t) \leq c(b,t) = 1$. Now, fix $a \in A$ and $b \in B$ such that $ab \in E(G)$. The inflow into a is at most 1,[36] and so the outflow is at most 1. So, $f(a,b) \leq 1$. This proves that $f(e) \in \{0,1\}$ for all $e \in E(G')$, as we had claimed.

Now, let $M = \{ab \in E(G) \mid a \in A, b \in B, f(a,b) = 1\}$. Then[37]

$$\begin{aligned} |M| &= |\{(a,b) \in E(G') \mid a \in A, b \in B, f(a,b) = 1\}| \\ &= |\{e \in S_{G'}(A \cup \{s\}, B \cup \{t\}) \mid f(e) = 1\}| \\ &\overset{(*)}{=} f(A \cup \{s\}, B \cup \{t\}) \\ &\overset{(**)}{=} val(f), \end{aligned}$$

where (*) follows from the fact that $f(e) \in \{0,1\}$ for all $e \in E(G)$, and (**) follows from Lemma 4.2.3. Let us check that M is a matching in G. Suppose otherwise. Then at least one of the following holds:

[35] This is because, for all $e \in E(G')$, $f(e)$ is a non-negative integer, and so if $f(e) \leq 1$, then $f(e) \in \{0,1\}$.

[36] This is because (s,a) is the only edge in G' with head a, and $f(s,a) \leq c(s,a) = 1$.

[37] $S_{G'}(A \cup \{s\}, B \cup \{t\})$ is the set of all edges from $A \cup \{s\}$ to $B \cup \{t\}$ in the oriented graph G'; note that all edges in $S_{G'}(A \cup \{s\}, B \cup \{t\})$ are in fact from A to B.

(i) there exist $a \in A$ and $b_1, b_2 \in B$ (with $b_1 \neq b_2$) such that $ab_1, ab_2 \in M$;

(ii) there exist $a_1, a_2 \in A$ (with $a_1 \neq a_2$) and $b \in B$ such that $a_1b, a_2b \in M$.

Suppose first that (i) holds. Then $f(a, b_1) = f(a, b_2) = 1$, and so the outflow from a is at least 2. On the other hand, the inflow into a is at most 1,[38] a contradiction. Suppose now that (ii) holds. then $f(a_1, b) = f(a_2, b) = 1$, and so the inflow into b is at least 2. On the other hand, the outflow from b is at most 1,[39] a contradiction. This proves that M is indeed a matching.

It remains to produce a vertex cover of size $c(R)$. Let C be the set of all vertices in $V(G) = A \cup B$ that are incident with at least one edge of R. Our goal is to show that C is a vertex cover of size at most $c(R)$. First, note that $\{(s, a) \mid a \in A\}$ is a cut in (G', s, t, c) of capacity $|A|$, and so $c(R) \leq |A|$. Since every edge from A to B has capacity $|A| + 1 > c(R)$, we deduce that R does not contain any edges from A to B; then $R = \{(s, a) \mid a \in A \cap C\} \cup \{(b, t) \mid b \in B \cap C\}$. It follows that

$$\begin{aligned} c(R) &= \Big(\sum_{a \in A \cap C} \underbrace{c(s,a)}_{=1}\Big) + \Big(\sum_{b \in B \cap C} \underbrace{c(b,t)}_{=1}\Big) \\ &= |A \cap C| + |B \cap C| \\ &= |C|. \end{aligned}$$

It remains to show that C is a vertex cover of G. Fix adjacent vertices $a \in A$ and $b \in B$; we must show that at least one of a, b belongs to C. Suppose otherwise. It then follows from the construction of C that R contains neither (s, a) nor (b, t); moreover, since R contains no edges from A to B, we see that R does not contain (a, b), either. So, s, a, b, t is a directed path from s to t in $G' - R$, contrary to the fact that R is a cut in (G', s, t, c). This proves that C is indeed a vertex cover of G. This completes the proof of (b). □

Given a bipartite graph G with bipartition (A, B),

- an *A-saturating matching* in G is a matching M in G such that every vertex of A is incident with some edge in M;
- a *B-saturating matching* in G is a matching M in G such that every vertex of B is incident with some edge in M.

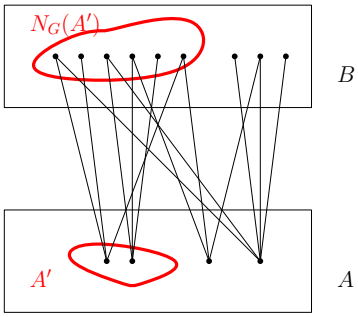
[38] This is because (s, a) is the only edge in G' with head a, and $f(s, a) \leq c(s, a) = 1$.
[39] This is because (b, t) is the only edge in G' with tail b, and $f(b, t) \leq c(b, t) = 1$.

For a graph G and a set $A \subseteq V(G)$, we denote by $N_G(A)$ the set of all vertices in $V(G) \setminus A$ that have a neighbor in A. As a corollary of the Kőnig-Egerváry theorem, we obtain the following.

Hall's theorem (graph theoretic formulation). *Let G be a bipartite graph with bipartition (A, B). Then the following are equivalent:*

(a) all sets $A' \subseteq A$ satisfy $|A'| \leq |N_G(A')|$;

(b) G has an A-saturating matching.



Proof. Suppose first that (b) holds; we must prove that (a) holds. Fix an A-saturating matching M in G, and fix $A' \subseteq A$. Since M is an A-saturating matching, and since A' is a stable set,[40] we know that precisely $|A'|$ edges in M are incident with a vertex in A', and each of those edges has another endpoint in B. No two edges in M share an endpoint, and it follows that exactly $|A'|$ vertices in B are incident with an edge of M that has an endpoint in A'; let B' be the set of all such vertices of B. But clearly, $B' \subseteq N_G(A')$, and so $|N_G(A')| \geq |B'| = |A'|$. This proves (a).

Suppose, conversely, that (a) holds; we must prove that (b) holds. Since all edges of G are between A and B, it suffices to show that G has a matching of size at least $|A|$.[41] By the Kőnig-Egerváry theorem, it is enough to show that any vertex cover of G is of size at least $|A|$. Let C be a vertex cover of G. Then there can be no edges between $A \setminus C$ and $B \setminus C$, and we deduce

[40] A *stable set* (or *independent set*) is a set of pairwise non-adjacent vertices.

[41] Note that any matching in G of size at least $|A|$ is in fact of size precisely $|A|$.

that $N_G(A \setminus C) \subseteq B \cap C$. Now we have the following:

$$\begin{aligned} |A| &= |A \cap C| + |A \setminus C| \\ &\leq |A \cap C| + |N_G(A \setminus C)| \qquad \text{by (a)} \\ &\leq |A \cap C| + |B \cap C| \qquad \text{because } N_G(A \setminus C) \subseteq B \cap C \\ &= |C|. \end{aligned}$$

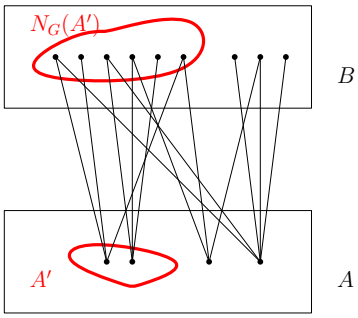
This completes the proof of (b). □

The *degree* of a vertex v in a graph G, denoted by $d_G(v)$, is the number of edges of G that v is incident with.

Corollary 4.4.1. *Let G be a bipartite graph with bipartition (A, B). Assume that G has at least one edge and that for all $a \in A$ and $b \in B$, we have that $d_G(a) \geq d_G(b)$. Then G has an A-saturating matching.*

Proof. We first check that $d_G(a) \geq 1$ for all $a \in A$. Since G has at least one edge, and since every edge of G has one endpoint in A and the other one in B, we see that some vertex $b_0 \in B$ is incident with at least one edge, and therefore satisfies $d_G(b_0) \geq 1$. But then by hypothesis, every vertex $a \in A$ satisfies $d_G(a) \geq d_G(b_0) \geq 1$.

Now, suppose that G does not have an A-saturating matching. Then by Hall's theorem, there exists some $A' \subseteq A$ such that $|A'| > |N_G(A')|$.



Note that every edge in G has at least one endpoint in $(A \setminus A') \cup N_G(A')$,[42]

[42]Indeed, if some edge of G had neither endpoint in $(A \setminus A') \cup N_G(A')$, then one of its endpoints would be in A' and the other one would be in $B \setminus N_G(A')$, a contradiction.

and so

$$\begin{aligned} |E(G)| &\leq \sum_{v \in (A \setminus A') \cup N_G(A')} d_G(v) \\ &\leq \Big(\sum_{a \in A \setminus A'} d_G(a) \Big) + \Big(\sum_{b \in N_G(A')} d_G(b) \Big). \end{aligned}$$

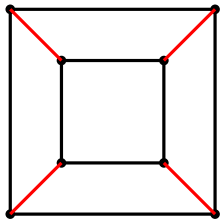
Now, since $A' \subseteq A$ and $N_G(A') \subseteq B$, we know that for all $a \in A'$ and $b \in N_G(A')$, we have that $d_G(a) \geq d_G(b)$. Furthermore, by our choice of A', we have that $|A'| > |N_G(A')|$. Since $d_G(a) \geq 1$ for all $a \in A$, we now deduce that $\sum_{a \in A'} d_G(a) > \sum_{b \in N_G(A')} d_G(b)$, and it follows that

$$\begin{aligned} |E(G)| &\leq \Big(\sum_{a \in A \setminus A'} d_G(a) \Big) + \Big(\sum_{b \in N_G(A')} d_G(b) \Big). \\ &< \Big(\sum_{a \in A \setminus A'} d_G(a) \Big) + \Big(\sum_{a \in A'} d_G(a) \Big) \\ &= \sum_{a \in A} d_G(a) \\ &= |E(G)|, \end{aligned}$$

which is impossible. □

For a non-negative integer k, a graph G is *k-regular* if it all its vertices are of degree k. A graph G is *regular* if there exists some non-negative integer k such that G is k-regular. (In other words, a graph is *regular* if all its vertices are of the same degree.)

A *perfect matching* in a graph G is a matching M such that every vertex of G is incident with an edge in M. An example of a perfect matching is shown below (edges of the perfect matching are in red).



Obviously, not all graphs have perfect matchings. For instance, no graph with an odd number of vertices has a perfect matching. (There are also many

graphs that have an even number of vertices, and yet do not have a perfect matching.)

Corollary 4.4.2. *Every regular bipartite graph that has at least one edge has a perfect matching.*

Proof. Let G be a k-regular ($k \geq 0$) bipartite graph with bipartition (A, B), and assume that G has at least one edge. By Corollary 4.4.1, G has an A-saturating matching. To show that this matching is a perfect matching, it suffices to show that $|A| = |B|$. Since G is k-regular, we see that $|E(G)| = k|A|$ and $|E(G)| = k|B|$; consequently, $k|A| = k|B|$. Since G has at least one edge, we see that $k \neq 0$, and we deduce that $|A| = |B|$. This completes the argument. □

In this section, we have studied matchings in bipartite graphs. There is also a theory of matchings in general graphs. Notably, Tutte's theorem (see section 9.3) gives a necessary and sufficient condition for a graph to have a perfect matching.

We complete this section by giving another formulation of Hall's theorem. We first need a definition. Suppose X and I are sets, and $\{A_i\}_{i\in I}$ is a family of (not necessarily distinct) subsets of X.[43] A *transversal* (or a *system of distinct representatives*) for $(X, \{A_i\}_{i\in I})$ is an injective (i.e. one-to-one) function $f : I \to X$ such that for all $i \in I$, we have that $f(i) \in A_i$. The following readily follows from the graph theoretic formulation of Hall's theorem (the details are left as an exercise).

Hall's theorem (combinatorial formulation). *Let X and I be finite sets, and let $\{A_i\}_{i\in I}$ be a family of (not necessarily distinct) subsets of X. Then the following are equivalent:*

(a) all sets $J \subseteq I$ satisfy $|J| \leq |\bigcup_{j\in J} A_j|$;

(b) $(X, \{A_i\}_{i\in I})$ has a transversal.

4.5 Extending Latin rectangles

For positive integers r and n, with $r \leq n$, an $r \times n$ *Latin rectangle* is an $r \times n$ array (or matrix) whose entries are numbers $1, \ldots, n$, and in which each number $1, \ldots, n$ occurs at most once in each row and each column. One 2×4 Latin rectangle is represented below.

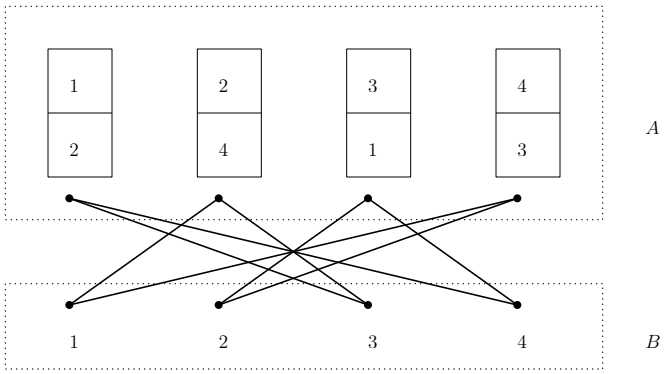
[43] Technically, we have that $A : I \to \mathscr{P}(X)$; for $i \in I$, we write A_i instead of $A(i)$.

1	2	3	4
2	4	1	3

Theorem 4.5.1. *Let r and n be positive integers such that $r < n$. Then every $r \times n$ Latin rectangle can be extended to an $n \times n$ Latin square.*[44]

Proof. Let $L = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$ be an $r \times n$ Latin rectangle.[45] Obviously, it suffices to show that we can extend L to an $(r+1) \times n$ Latin rectangle by adding a row of length n to the bottom of L, for then the result will follow immediately by an easy induction.

Let $A = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ and $B = \{1, \dots, n\}$, and let G be the bipartite graph with bipartition (A, B) in which $\mathbf{a}_i \in A$ and $j \in B$ are adjacent if and only if j is not an entry of the column $\mathbf{a}_i$. For instance, for the Latin rectangle from the beginning of the section, we would get the following bipartite graph:



Each column of L has r entries, and consequently, there are $n - r$ values in B that do not appear in it. So, for all $\mathbf{a}_i \in A$, we have that $d_G(\mathbf{a}_i) = n - r$. Now, fix $j \in B$. We know that j appears exactly once in each row of L, and L has r rows. Consequently, j appears exactly r times in L, and since it cannot appear more than once in any column, we see that it appears in precisely r columns of L. Thus, j fails to appear in precisely $n - r$ columns of L, and consequently, $d_G(j) = n - r$. We have now shown that is $(n-r)$-regular. So, G is a regular bipartite graph, and (since $r < n$) it has at least one edge. Corollary 4.4.2 now implies that G has a perfect matching, call it M. Now,

[44] This means that, for any $r \times n$ Latin rectangle, it is possible to add $n - r$ rows of length n to the bottom of the Latin rectangle that we started with and thus obtain an $n \times n$ Latin square.

[45] This means that $\mathbf{a}_1, \dots, \mathbf{a}_n$ are the columns of our Latin rectangle, in that order.

for each $i \in \{1, \ldots, n\}$, let j_i be the (unique) element of $\{1, \ldots, n\}$ such that $\mathbf{a}_i j_i \in M$. We now add the row $\begin{bmatrix} j_1 & \ldots & j_n \end{bmatrix}$ to the bottom of L, and we thus obtain an $(r+1) \times n$ Latin rectangle, which is what we needed. $\square$

Chapter 5

Graph connectivity

5.1 Vertex and edge connectivity

Notation: For a graph G and a set $X \subsetneq V(G)$, we denote by $G \setminus X$ the graph obtained from G by deleting all vertices of X (as well as all edges that have at least one endpoint in X); if $|X| = 1$, say $X = \{x\}$, we often write $G \setminus x$ instead of $G \setminus X$. For a graph G and a set $F \subseteq E(G)$, we denote by $G - F$ the graph obtained from G by deleting all edges in F; if $|F| = 1$, say $F = \{e\}$, we often write $F - e$ instead of $G - F$.

For a graph G and (not necessarily disjoint) sets $A, B \subseteq V(G)$, an *A-B path* in G, or a *path from A to B* in G, is either a one-vertex path whose sole vertex is in $A \cap B$, or a path on at least two vertices whose one endpoint is in A and whose other endpoint is in B.

Given a graph G and (not necessarily disjoint) sets $A, B \subseteq V(G)$, we say that a set $X \subseteq V(G)$ *separates* A from B in G if every path from A to B in G contains at least one vertex of X. Note that this implies that $A \cap B \subseteq X$.[1]

A graph G is *connected* if for all $x, y \in V(G)$, the graph G contains a path between x and y.

Given a graph G and a non-negative integer k, we say that G is *k-vertex-connected*, or simply *k-connected*, if $|V(G)| \geq k + 1$ and for all $X \subseteq V(G)$ such that $|X| \leq k-1$, we have that $G \setminus X$ is connected. Note that this means that every (non-null) graph is 0-connected, and that every connected graph on at least two vertices is 1-connected.[2] The *vertex-connectivity*, or simply *connectivity*, of a graph G, denoted $\kappa(G)$, is the largest integer k such that

[1] Indeed, if $x \in A \cap B$, then x counts as a one-vertex path from A to B. So, any set of vertices that separates A from B must include $A \cap B$ as a subset.

[2] However, K_1 is **not** 1-connected.

G is k-connected. Clearly, if there exists a set of at most k vertices whose deletion from G yields a disconnected graph, then $\kappa(G) \leq k$. Note also that if $k = \kappa(G)$, then either $G \cong K_{k+1}$ or there exists a set of k vertices whose deletion from G yields a disconnected graph.

Given a graph G and disjoint sets $A, B \subseteq V(G)$, we say that a set $F \subseteq E(G)$ *separates* A from B in G if every path from A to B contains at least one edge of F.

Given a graph G and a non-negative integer ℓ, we say that G is *ℓ-edge-connected* if $|V(G)| \geq 2$ and for all $F \subseteq E(G)$ such that $|F| \leq \ell - 1$, we have that $G - F$ is connected. The *edge-connectivity* of a graph G on at least two vertices, denoted by $\lambda(G)$, is the largest integer ℓ such that G is ℓ-edge-connected. Clearly, if there exists a set of at most ℓ edges whose deletion from G yields a disconnected graph, then $\lambda(G) \leq \ell$. Note that if $\ell = \lambda(G)$, then there exists a set of ℓ edges whose deletion from G yields a disconnected graph.

Proposition 5.1.1. *Let G be a graph on at least two vertices. Then*

(a) for all edges $e \in E(G)$, we have that $\lambda(G) - 1 \leq \lambda(G - e) \leq \lambda(G)$;

(b) for all sets $F \subseteq E(G)$, we have that $\lambda(G - F) \leq \lambda(G)$.

Proof. Clearly, (b) follows from (a) by an easy induction. It remains to prove (a). Fix $e \in E(G)$.

We first show that $\lambda(G - e) \geq \lambda(G) - 1$. Fix $F \subseteq E(G - e)$ such that $|F| \leq \lambda(G) - 2$. Set $F' := F \cup \{e\}$; then $|F'| \leq \lambda(G) - 1$, and we deduce that $G - F'$ is connected. But $(G - e) - F = G - F'$, and we deduce that $(G - e) - F$ is connected. This proves that $\lambda(G - e) \geq \lambda(G) - 1$.

It remains to show that $\lambda(G-e) \leq \lambda(G)$. Fix $F \subseteq E(G)$ with $|F| = \lambda(G)$, such that $G - F$ is disconnected. Set $F' := F \setminus \{e\}$; then $|F'| \leq \lambda(G)$. Note that $(G - e) - F'$ is obtained from $G - F$ by possibly deleting one edge.[3] Since $G - F$ is disconnected, it follows that $(G - e) - F'$ is disconnected. Since $|F'| \leq \lambda(G)$, we deduce that $\lambda(G - e) \leq \lambda(G)$. □

Proposition 5.1.2. *Let G be a graph on at least two vertices. Then*

(a) for all edges $e \in E(G)$, we have that $\kappa(G) - 1 \leq \kappa(G - e) \leq \kappa(G)$;

(b) for all sets $F \subseteq E(G)$, we have that $\kappa(G - F) \leq \kappa(G)$.

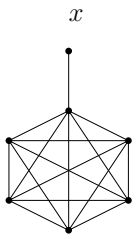
[3]Indeed, if $e \in F$, then $(G - e) - F' = G - F$. On the other hand, if $e \notin F$, then $(G - e) - F' = (G - F) - e$.

Proof. Clearly, (b) follows from (a) by an easy induction. It remains to prove (a). Fix $e \in E(G)$.

We first show that $\kappa(G-e) \geq \kappa(G)-1$. Since G is $\kappa(G)$-connected, we know that G (and consequently, $G-e$ as well) has at least $\kappa(G)+1$ vertices. Now, fix $X \subseteq V(G)$ such that $|X| \leq \kappa(G)-2$; we must show that $(G-e)-X$ is connected. Suppose first that e is incident with some vertex in X. Then $(G-e) \setminus X = G \setminus X$. Since $|X| \leq \kappa(G)-2$, we see that $G \setminus X$ is connected, and it follows that $(G-e) \setminus X$ is connected. It remains to consider the case when e is not incident with any vertex in X. Set $e = x_1x_2$ (i.e. let x_1 and x_2 be the endpoints of e). Set $X_1 := X \cup \{x_1\}$ and $X_2 := X \cup \{x_2\}$. Then $|X_1| = |X_2| = \kappa(G)-1$, and we deduce that $G \setminus X_1$ and $G \setminus X_2$ are connected. Now, since $x_2 \in V(G) \setminus X_1$, and since $G \setminus X_1$ is a connected graph on at least two vertices, we see that x_2 is adjacent to some vertex in $u \in V(G) \setminus X_1$; since $x_1 \in X_1$, we see that $u \neq x_1$. Now, $(G-e) \setminus X$ can be obtained from the connected graph $G \setminus X_2$ by adding to it the vertex x_2 and making it adjacent to all vertices in $N_G(x_2) \setminus \{x_1\}$. Since $u \in N_G(x_2) \setminus \{x_1\}$, we see that x_2 is not an isolated vertex of $(G-e) \setminus X$,[4] and we deduce that $(G-e) \setminus X$ is connected. This proves that $\kappa(G-e) \geq \kappa(G)-1$.

It remains to show that $\kappa(G-e) \leq \kappa(G)$. By definition, $|V(G)| \geq \kappa(G)+1$. If G has precisely $\kappa(G)+1$ vertices, then so does $G-e$, and it follows from the definition that $\kappa(G-e) \leq \kappa(G)$. It remains to consider the case when $|V(G)| \geq \kappa(G)+2$. In this case, there exists a set $X \subseteq V(G)$ of size $\kappa(G)$ such that $G \setminus X$ is disconnected. But then $(G-e) \setminus X$ is disconnected as well, and it follows that $\kappa(G-e) \leq \kappa(G)$. □

We note that, unlike edge deletion, vertex deletion sometimes increases connectivity. For instance, for the graph G represented below, we have that $\kappa(G) = \lambda(G) = 1$, but $\kappa(G \setminus x) = \lambda(G \setminus x) = 5$.



G

Recall that for a graph G, $\delta(G)$ is the minimum and $\Delta(G)$ the maximum

[4] An *isolated vertex* is a vertex that has no neighbors.

degree in G, i.e. $\delta(G) = \min\{d_G(v) \mid v \in V(G)\}$ and $\Delta(G) = \max\{d_G(v) \mid v \in V(G)\}$.

Theorem 5.1.3. *Let G be a graph on at least two vertices. Then $\kappa(G) \leq \lambda(G) \leq \delta(G)$.*

Proof. We first prove that $\lambda(G) \leq \delta(G)$. Fix a vertex $v \in V(G)$ such that $d_G(v) = \delta(G)$, and let F be the set of all edges of G that are incident with v; then $|F| = \delta(G)$. Clearly, $G - F$ is disconnected,[5] and it follows that $\lambda(G) \leq \delta(G)$.

It remains to show that $\kappa(G) \leq \lambda(G)$. Fix a set $F \subseteq E(G)$ such that $|F| = \lambda(G)$ and $G - F$ is disconnected.

> **Claim.** If C is the vertex set of a component of $G - F$, then no edge of F has both its endpoints in C.

Proof of the Claim. Suppose otherwise. Let C be the vertex set of a component of $G - F$,[6] and let $e \in F$ be an edge that has both its endpoints in C. Then $G - (F \setminus \{e\})$ is still disconnected,[7] contrary to the fact that $|F \setminus \{e\}| = |F| - 1 = \lambda(G) - 1$. This proves the Claim. ♦

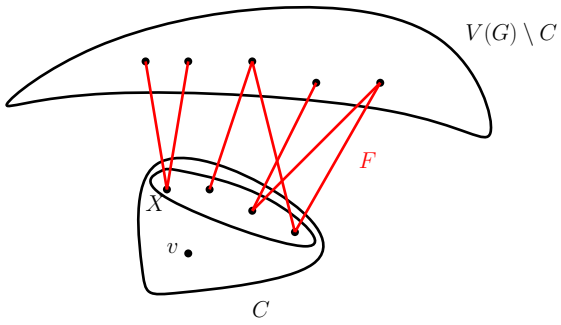
Suppose first that there exists a vertex $v \in V(G)$ that is not incident with any edge in F. Let C be the vertex set of the component of $G - F$ that contains v (see the picture below). Now, let X be the set of all vertices in C that are incident with an edge in F; then $v \in C \setminus X$. By the Claim, no edge in F has both endpoints in C, and so $|X| \leq |F| = \lambda(G)$. Moreover, X separates $C \setminus X$ from $V(G) \setminus C$ in G; since both $C \setminus X$ and $V(G) \setminus C$ are non-empty,[8] it follows that $G \setminus X$ is disconnected. So, $\kappa(G) \leq \lambda(G)$.

[5] We are using the fact that G has at least two vertices.

[6] Since $G - F$ is disconnected, this implies that C and $V(G) \setminus C$ are both non-empty, and there are no edges between them.

[7] This is because there are still no edges between C and $V(G) \setminus C$, and both C and $V(G) \setminus C$ are non-empty.

[8] The fact that $G \setminus X$ is non-empty follows from the fact that $v \in C \setminus X$. The fact that $V(G) \setminus C$ is nonempty follows from the fact that C is the vertex set of a component of the disconnected graph $G - F$.



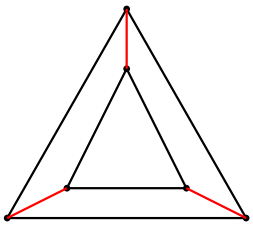
It remains to consider the case when every vertex of G is incident with an edge of F.[9] Fix any $v \in V(G)$; we will show that $d_G(v) \leq \lambda(G)$. Let C be the component of $G - F$ that contains v, and let F_v be the set of edges of F incident with v. Let $u_1, \ldots, u_t$ be the neighbors of v in the component C, and for all $i \in \{1, \ldots, t\}$, let F_i be the set of all edges of F incident with u_i. By supposition, sets $F_v, F_1, \ldots, F_t$ are all non-empty, and by the Claim, they are pairwise disjoint. So,

$$d_G(v) \quad = \quad |F_v| + t \quad \leq \quad |F_v| + |F_1| + \cdots + |F_t| \quad \leq \quad |F| \quad = \quad \lambda(G),$$

as we had claimed. Since we chose v arbitrarily, it now follows that $\Delta(G) \leq \lambda(G)$; we already saw that $\lambda(G) \leq \delta(G) \leq \Delta(G)$, and we now deduce that $\lambda(G) = \Delta(G)$. Now, if G is a complete graph, then $|V(G)| = \Delta(G) + 1$, and we see that $\kappa(G) = \Delta(G) = \lambda(G)$.[10] So assume that G is not complete, and fix some $x \in V(G)$ that has a non-neighbor in G. Then $G \setminus N_G(x)$ is disconnected, and we have that $|N_G(x)| = d_G(x) \leq \Delta(G) = \lambda(G)$. So, $\kappa(G) \leq \lambda(G)$. □

Terminology: A *vertex-cutset* (or simply *cutset*) of a graph G is any set $X \subsetneqq V(G)$ such that $G \setminus X$ has more components than G.[11] Similarly, an

[9] For an example, see the graph below, with the edges of F in red.



[10] Check this!

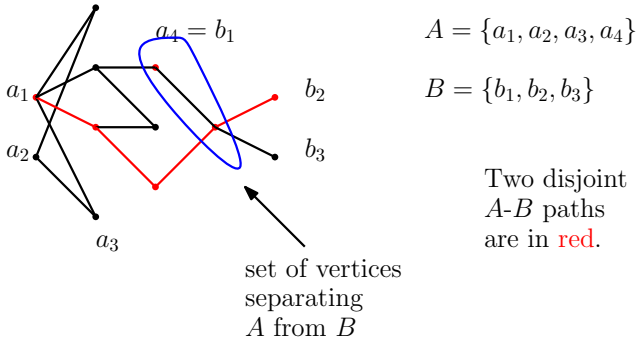
[11] So, if G is connected, then a vertex-cutset of G is any set $X \subsetneqq V(G)$ such that $G \setminus X$ is disconnected.

edge-cutset of G is any set $F \subseteq E(G)$ such that $G - F$ has more components than G.

By definition, no graph G has a vertex-cutset of size strictly smaller than $\kappa(G)$. Similarly, no graph G has an edge-cutset of size strictly smaller than $\lambda(G)$.

5.2 Menger's theorems

Menger's theorem (vertex version). *Let G be a graph, and let $A, B \subseteq V(G)$.*[12] *Then the minimum number of vertices separating A from B in G is equal to the maximum number of pairwise disjoint A-B paths in G.*[13]



Proof. We assume inductively that the theorem holds for graphs that have fewer than $|E(G)|$ edges. More precisely, we assume that for all graphs G' such that $|E(G')| < |E(G)|$, and all sets $A', B' \subseteq V(G')$, the minimum number of vertices separating A' from B' in G' is equal to the maximum number of pairwise disjoint A'-B' paths in G'. We must prove that this holds for G as well. From now on, we let k be the minimum number of vertices separating A from B in G.

First, we claim that there can be no more than k pairwise disjoint paths from A to B in G. Indeed, let $X \subseteq V(G)$ be a set of size k separating A from B in G, and let $\mathcal{P}$ be any collection of pairwise disjoint paths from A to B. Then every path in $\mathcal{P}$ contains at least one vertex of X, and since paths in $\mathcal{P}$ are pairwise disjoint, no two paths in $\mathcal{P}$ contain the same vertex of X. So, $|\mathcal{P}| \leq |X| = k$, as we had claimed.

[12] A and B need not be disjoint.

[13] "Pairwise disjoint" here means that no two of the paths have a vertex in common (and consequently, no two of the paths have an edge in common).

It remains to show that there are at least k pairwise disjoint paths from A to B. Clearly, for any set $X \subseteq V(G)$ separating A from B in G, we have that $A \cap B \subseteq X$; consequently, $|A \cap B| \leq k$. Now, if $E(G) = \emptyset$, then $A \cap B$ separates A from B in G, and so $|A \cap B| = k$; in this case, the vertices of $A \cap B$ form k pairwise disjoint one-vertex paths from A to B, and we are done. From now on, we assume that G has at least one edge, say xy. Let $G_{xy} := G/xy$, i.e. let G_{xy} be the graph obtained from G by contracting the edge xy, and let v_{xy} be the vertex obtained by contracting xy.[14]

G $\qquad$ $G_{xy} = G/xy$

Now, if x or y belongs to A, then let $A' = (A \setminus \{x, y\}) \cup \{v_{xy}\}$, and otherwise, let $A' = A$. Similarly, if x or y belongs to B, then let $B' = (B \setminus \{x, y\}) \cup \{v_{xy}\}$, and otherwise, let $B' = B$.

Let $Y \subseteq V(G_{xy})$ be a minimum-sized set of vertices separating A' from B' in G_{xy}.[15] By the induction hypothesis, there are $|Y|$ many pairwise disjoint paths in G_{xy} from A' to B', and it readily follows[16] that there are at least $|Y|$ many pairwise disjoint paths in G from A to B. So, if $|Y| \geq k$,[17] then we are done. From now on, we assume that $|Y| \leq k - 1$. Then $v_{xy} \in Y$, for otherwise, Y would separate A from B in G,[18] contrary to the fact that $|Y| \leq k - 1$. Now $X := (Y \setminus \{v_{xy}\}) \cup \{x, y\}$ separates A from B in G,[19] and we have that $|X| = |Y| + 1$. Note that this implies that $|X| = k$.[20] Set $X = \{x_1, \ldots, x_k\}$.

[14] Formally, v_{xy} is some ("new") vertex that does not belong to $V(G)$, and G_{xy} is the graph with vertex set $V(G_{xy}) = \Big(V(G) \setminus \{x, y\}\Big) \cup \{v_{xy}\}$ and edge set $E(G_{xy}) = \{e \in E(G) \mid e \text{ is incident neither with } x \text{ nor with } y \text{ in } G\} \cup \{vv_{xy} \mid v \in V(G), v \text{ is adjacent to } x \text{ or } y \text{ in } G\}$.

[15] This means that for all sets $Y' \subseteq V(G_{xy})$ separating A from B in G_{xy}, we have that $|Y| \leq |Y'|$.

[16] Details?

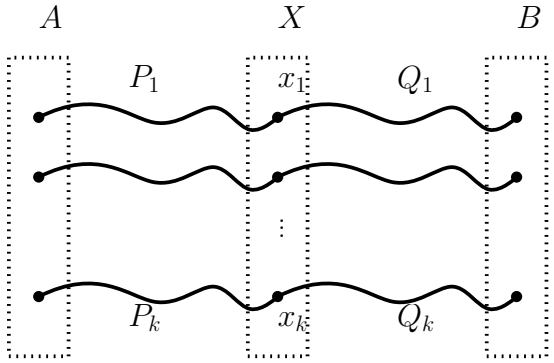
[17] In fact, it is not possible that $|Y| > k$ (details?), but we do not need this stronger fact.

[18] Proof?

[19] Proof?

[20] Indeed, since $|Y| \leq k - 1$, we have that $|X| \leq k$. On the other hand, since X separates A from B in G, we know that $|X| \geq k$. So, $|X| = k$.

We now consider the graph $G - xy$, i.e. the graph obtained from G by deleting the edge xy. Since $x, y \in X$, we know that any set of vertices separating A from X in $G - xy$ also separates A from B in G;[21] consequently, any such set has at least k vertices, and so by the induction hypothesis, there are k pairwise disjoint paths from A to X in $G - xy$, call them $P_1, \ldots, P_k$. Similarly, there are k pairwise disjoint paths from B to X in $G - xy$, call them $Q_1, \ldots, Q_k$. We may assume that for all $i \in \{1, \ldots, k\}$, x_i is an endpoint both of P_i and of Q_i. But now $P_1 - x_1 - Q_1, \ldots, P_k - x_k - Q_k$ are pairwise disjoint paths from A to B,[22] and we are done.



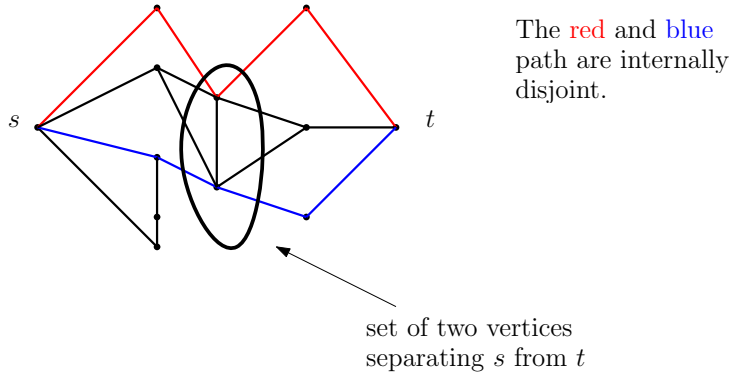
□

Given a graph G and distinct vertices $s, t \in V(G)$, two paths from s to t in G are *internally disjoint* if they have no vertices in common except the endpoints s and t. The following corollary is also often referred to as the vertex version of Menger's theorem.

Corollary 5.2.1. *Let G be a graph, and let $s, t \in V(G)$ be distinct, non-adjacent vertices of G. Then the minimum number of vertices of $V(G) \setminus \{s, t\}$ separating s from t in G is equal to the maximum number of pairwise internally disjoint s-t paths in G.*

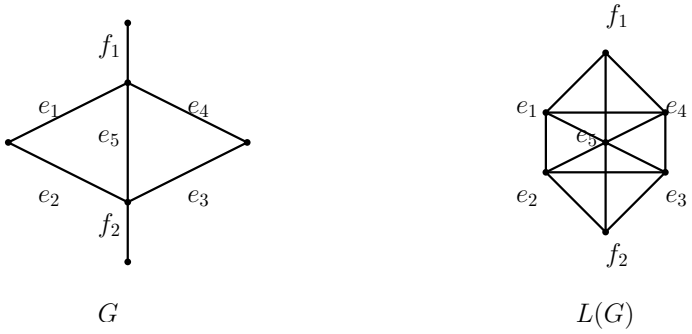
[21] Let us check this. Let Z be any set of vertices separating A from X in $G - xy$, and let $p_1, \ldots, p_t$, with $p_1 \in A$ and $p_t \in B$, be a path from A to B in G. Then some vertex of $p_1, \ldots, p_t$ belongs to X; let $i \in \{1, \ldots, t\}$ be the smallest index such that $p_i \in X$. Then $p_1, \ldots, p_i$ is a path from A to X in $G - xy$. Furthermore, since $p_1, \ldots, p_i$ contains exactly one vertex of X, and since $x, y \in X$, we see that the path $p_1, \ldots, p_i$ does not use the edge xy; consequently, $p_1, \ldots, p_i$ is a path from A to X in $G - xy$, and we deduce that this path (and consequently, the path $p_1, \ldots, p_t$ as well) contains a vertex of Z.

[22] It is clear that the $P_i - x_i - Q_i$'s are walks from A to B in G. But in fact, they are pairwise disjoint paths, for otherwise, there would be a path from A to B in G that uses no vertices of X (details?), contrary to the fact that X separates A from B in G.



Proof. Let $S := N_G(s)$ and $T := N_G(t)$. Obviously, the minimum number of vertices of $V(G) \setminus \{s,t\}$ separating s from t in G is equal to the minimum number of vertices of $V(G) \setminus \{s,t\}$ separating S from T in $G \setminus \{s,t\}$.[23] Similarly, the maximum number of pairwise internally disjoint s-t paths in G is equal to the maximum number of pairwise disjoint S-T paths in G. By Menger's theorem (vertex version), the minimum number of vertices separating S from T in $G \setminus \{s,t\}$ is equal to the maximum number of pairwise disjoint S-T paths in $G \setminus \{s,t\}$. So, the minimum number of vertices of $V(G) \setminus \{s,t\}$ separating s from t in G is equal to the maximum number of pairwise internally disjoint s-t paths in G. □

Our next goal is to prove the edge version of Menger's theorem. The *line graph* of a graph G, denoted by $L(G)$, is the graph whose vertex set is $E(G)$, and in which $e, f \in L(V(G)) = E(G)$ are adjacent if and only if e and f share an endpoint in G. An example is shown below.



[23] Indeed, for any set $X \subseteq V(G) \setminus \{s,t\}$, we have that X separates s from t in G if and only if X separates S from T in $G \setminus \{s,t\}$.

Our goal is to use line graphs to derive the edge version of Menger's theorem from the vertex version of the theorem.

An *induced path* (or *chordless path*) in a graph G is a path $P = p_1, \ldots, p_r$ such that for all distinct $i, j \in \{1, \ldots, r\}$, we have that $p_i p_j \in E(G)$ if and only if $|i - j| = 1$. Note that if G contains a path P between vertices x and y, then G contains an induced path Q between x and y such that $V(Q) \subseteq V(P)$.[24]

Proposition 5.2.2. *Let G be a graph, let $s, t \in V(G)$ be distinct vertices of G, let S be the set of all edges of G incident with s, and let T be the set of all edges of G incident with t. Let $X \subseteq E(G)$. Then X separates s from t in G if and only if X separates S from T in $L(G)$.*[25]

Proof. We will prove that X fails to separate s from t in G if and only if X fails to separate S from T in $L(G)$.

Suppose first that X does not separate s from t in G; we must show that X does not separate S from T in $L(G)$. Fix a path $v_1, \ldots, v_r$ in G, with $v_1 = s$ and $v_r = t$, that does not use any edge of X.[26] But now $v_1v_2, v_2v_3, \ldots, v_{r-1}v_r$ is a path from S to T in $L(G)$ that does not use any vertex (in $L(G)$) in X. So, X does not separate S from T in $L(G)$.

Suppose now that X does not separate S from T in $L(G)$; we must show that X does not separate s from t in G. Fix an induced path $e_1, \ldots, e_r$ in $L(G)$, with $e_1 \in S$ and $e_r \in T$, that does not contain any vertex (in $L(G)$) from X.[27] For each $i \in \{1, \ldots, r-1\}$, let v_i be a common endpoint of e_i and e_{i+1}.[28] Since the path $e_1, \ldots, e_r$ in $L(G)$ is induced, we see that $s, v_1, \ldots, v_{r-1}, t$ is a path in G that uses only edges $e_1, \ldots, e_r$, and therefore, fails to use any edge from X. So, X does not separate s from t in G. □

Two paths in a graph G are *edge-disjoint* if they have no edges in common. An example is shown below.

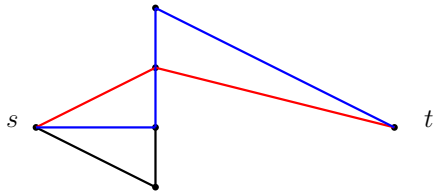
[24] Let us check this. Suppose that $P = p_1, \ldots, p_r$, with $p_1 = x$ and $p_r = y$, is a path in G. Of all paths in G between x and y using only vertices in $V(P) = \{p_1, \ldots, p_r\}$, let $Q = q_1, \ldots, q_s$ (with $q_1 = x$ and $q_s = y$) be one of minimum length. But then it is clear that the path Q is induced (details?).

[25] Note that the elements of X are edges of G, but vertices of $L(G)$.

[26] Such a path exists because X does not separate s from t in G.

[27] Since X does not separate S from T in $L(G)$, we know that $L(G)$ contains a path from S to T that uses no vertices of X. If we choose $e_1, \ldots, e_r$ to be a minimum-length path with these properties, then we see that the path $e_1, \ldots, e_r$ is induced in $L(G)$.

[28] Such a vertex exists because e_i and e_{i+1} are adjacent vertices of $L(G)$, and consequently, they are edges of G that share an endpoint.



The red and blue path are edge-disjoint.

Proposition 5.2.3. *Let G be a graph, let $s,t \in V(G)$ be distinct vertices of G, let S be the set of all edges in G incident with s, and let T be the set of all edges in G incident with t. Let ℓ be a non-negative integer. Then the following are equivalent:*

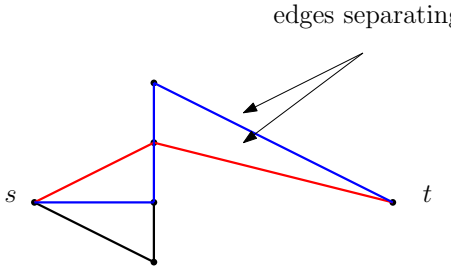
(i) there exist ℓ pairwise edge-disjoint s-t paths in G;

(ii) there exist ℓ pairwise disjoint S-T paths in $L(G)$.

Proof. Suppose first that (i) holds, and fix ℓ pairwise edge-disjoint s-t paths in G, say $P_1, \dots, P_\ell$. For all $i \in \{1, \dots, \ell\}$, set $P_i = v_1^i, \dots, v_{r_i}^i$ (with $v_1^i = s$ and $v_{r_i}^i = t$). Now, for all $i \in \{1, \dots, \ell\}$, set $P_i^L = v_1^i v_2^i, v_2^i v_3^i, \dots, v_{r_i-1}^i v_{r_i}^i$. Clearly, $P_1^L, \dots, P_\ell^L$ are pairwise disjoint S-T paths in $L(G)$.

Suppose now that (ii) holds, and fix ℓ pairwise disjoint S-T paths in $L(G)$, say $Q_1^L, \dots, Q_\ell^L$; we may assume that the paths Q_i^L are induced in $L(G)$.[29] For all $i \in \{1, \dots, \ell\}$, set $Q_i^L = e_1^i, \dots, e_{r_i}^i$. Now, for all $i \in \{1, \dots, \ell\}$ and $j \in \{1, \dots, r_i\}$, let v_j^i be a common vertex of the edges e_j^i and e_{j+1}^i in G, and set $Q_i = s, v_1^i, \dots, v_{r_i-1}^i, t$. Then $Q_1, \dots, Q_\ell$ are pairwise edge-disjoint s-t paths in G,[30] and we are done. □

Menger's theorem (edge version). *Let G be a graph, and let $s,t \in V(G)$ be distinct vertices of G. Then the minimum number of edges separating s from t in G is equal to the maximum number of pairwise edge-disjoint s-t paths in G.*



[29] Why may we assume this?

[30] We are using the fact that the paths $Q_1^L, \dots, Q_\ell^L$ are induced in $L(G)$.

Proof. Let S be the set of all edges in G incident with s, and let T be the set of all edges in G incident with t. By Proposition 5.2.2, the minimum number of edges separating s from t in G is equal to the minimum number of vertices separating S from T in $L(G)$. By Proposition 5.2.3, the maximum number of pairwise edge-disjoint s-t paths in G is equal to the maximum number of pairwise disjoint S-T paths in G. By Menger's theorem (vertex version), the minimum number of vertices separating S from T in $L(G)$ is equal to the maximum number of pairwise disjoint S-T paths in G. We now deduce that the minimum number of edges separating s from t in G is equal to the maximum number of pairwise edge-disjoint s-t paths in G. □

The global version of Menger's theorem. *Let G be a graph on at least two vertices, and let $k, \ell \geq 0$ be integers.*

(a) G is k-connected if and only if for all distinct $s, t \in V(G)$, there exist k pairwise internally disjoint s-t paths in G.

(b) G is ℓ-edge-connected if and only if for all distinct $s, t \in V(G)$, there exist ℓ pairwise edge-disjoint s-t paths in G.

Proof. We first prove (a). Suppose that G is k-connected, and let s and t be distinct vertices of G.

Suppose first that s and t are non-adjacent. Since G is k-connected, s and t cannot be separated by fewer than k vertices of $V(G) \setminus \{s, t\}$; so, by Corollary 5.2.1, there are k internally disjoint paths between s and t.

Suppose now that s and t are adjacent. Set $G' := G - st$.[31] By Proposition 5.1.2, G' is $(k-1)$-connected. Now s and t are distinct and non-adjacent in G', and they cannot be separated (in G') by fewer than $k-1$ vertices of $V(G') \setminus \{s, t\}$; so, Corollary 5.2.1 guarantees that there are $k-1$ internally disjoint paths between s and t in G'. These $k-1$ paths, plus the one-edge path s, t, form k internally disjoint paths in G.

Suppose now that there are k internally disjoint paths between any two distinct vertices of G; we must show that G is k-connected.

Let us first show that $|V(G)| \geq k+1$. By hypothesis, G has at least two vertices; fix any distinct vertices $s, t \in V(G)$. Then there are k internally disjoint paths between them, and all but possibly one of those paths have an internal vertex;[32] so, these k paths together have at least $k-1$ internal

[31] So, G' is the graph obtained from G by deleting the edge st.

[32] If s and t are adjacent, then s, t is a path between s and t with no internal vertices. However, any other path between s and t has at least one internal vertex.

vertices, and it follows that $|V(G)| \geq (k-1)+2 = k+1$,[33] which is what we needed.

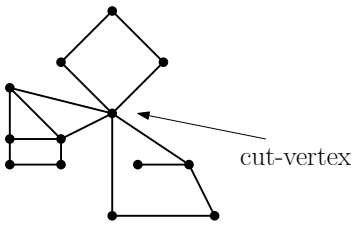
It remains to show that for all sets $X \subseteq V(G)$ such that $|X| \leq k-1$, we have that $G \setminus X$ is connected. Suppose otherwise, and fix some $X \subseteq V(G)$ such that $|X| \leq k-1$ and $G \setminus X$ is disconnected. Then $G \setminus X$ has at least two components, and we choose vertices s and t from distinct components of $G \setminus X$. Now $X \subseteq V(G) \setminus \{s,t\}$ separates s from t, and so by Corollary 5.2.1, there can be at most $|X| \leq k-1$ internally disjoint paths between s and t in G. But this contradicts the fact that there are k internally disjoint paths between any two distinct vertices of G. This completes the proof of (a).

We now prove (b). Suppose first that G is ℓ-edge-connected. Fix distinct vertices $s, t \in V(G)$. Since G is ℓ-edge-connected, s cannot be separated from t with fewer than ℓ edges of G, and so by Menger's theorem (edge version), there are at least ℓ pairwise edge-disjoint paths between s and t in G.

Suppose now that G is not ℓ-edge connected. Then there exists a set $F \subseteq E(G)$ such that $|F| \leq \ell - 1$ and $G - F$ is disconnected. Since $G - F$ is disconnected, it has at least two components; let s and t be vertices from distinct components of $G - F$. Now F separates s from t, and in particular, s can be separated from t by at most $|F| \leq \ell - 1$ edges of G. So, by Menger's theorem (edge version), there are at most $\ell - 1$ pairwise edge-disjoint paths between s and t in G. This completes the proof of (b). □

5.3 2-connected graphs and ear decomposition

A *cut-vertex* of a graph G is any vertex $v \in V(G)$ such that $G \setminus v$ has more components than G.

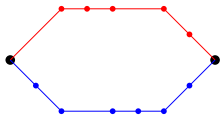


Recall that, for a non-negative integer k, a graph G is *k-connected* if $|V(G)| \geq k+1$ and for all $S \subseteq V(G)$ such that $|S| \leq k-1$, we have that $G \setminus S$ is connected. So, a graph is 2-connected if it has at least three vertices, is connected, and has no cut-vertices.

[33] We are counting the $k-1$ internal vertices of our paths, plus the endpoints s and t.

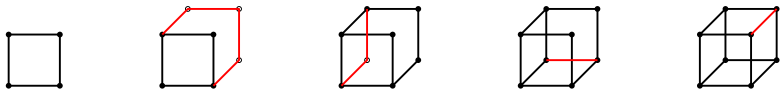
Lemma 5.3.1. *Let G be a graph on at least two vertices. Then G is 2-connected if and only if any two distinct vertices lie on a common cycle.*[34]

Proof. By Menger's theorem (global version), a graph on at least two vertices is 2-connected if and only if for any pair of distinct vertices, there are two internally disjoint paths between them. But obviously, two distinct vertices lie on a common cycle if and only if there are two internally-disjoint paths between them. The result now follows.



□

In this section, we give a full structural description of 2-connected graphs. A *path addition* (sometimes called *open ear addition*) to a graph H is the addition to H of a path between two distinct vertices of H in such a way that no internal vertex and no edge of the path belongs to H. In the picture below, we show how the cube graph can be constructed by starting with a cycle of length four and then repeatedly adding paths (the path/open ear added at each step is in red).



The Ear lemma. *A graph is 2-connected if and only if it is a cycle or can be obtained from a cycle by repeated path addition.*

Proof. We first prove the "if" (i.e. "$\Longleftarrow$") part of the lemma. Clearly, cycles are 2-connected (indeed, every cycle has at least three vertices, is connected, and has no cut-vertices).[35] Further, if a graph G can be obtained from a 2-connected graph H by adding a path, then G has at least three vertices (because H does), and it is easy to see that G is connected and has no cut-vertices;[36] so, G is 2-connected. It now follows by an easy induction (e.g. on the number of paths added) that any graph obtained from a cycle by repeated path addition is 2-connected. This proves the "if" part of the lemma.

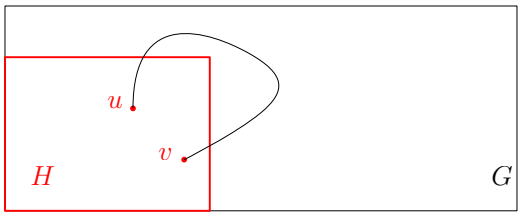
[34] Note that if G has at least two vertices, and any two distinct vertices lie on a common cycle, then in particular, G contains a cycle, and therefore, G has at least three vertices.

[35] Alternatively, this follows from Lemma 5.3.1.

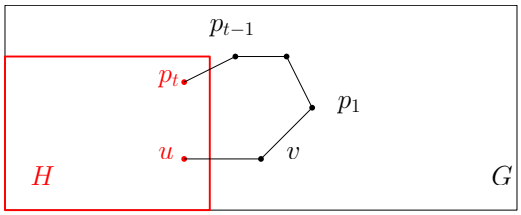
[36] Check this!

It remains to prove the "only if" (i.e. "$\Longrightarrow$") part of the lemma. Fix a 2-connected graph G. By Lemma 5.3.1, G contains a cycle. Now, let H be a maximal subgraph of G that either is a cycle or can be obtained from a cycle by repeated path addition.[37] We must show that $H = G$.

First, we claim that H is an induced subgraph of G.[38] If not, then there exist distinct vertices $u, v \in V(H)$ that are adjacent in G, but not in H; but then the graph obtained from H by adding the one-edge path u, v contradicts the maximality of H. So, H is indeed an induced subgraph of G.



It remains to show that $V(H) = V(G)$. Suppose otherwise. Then since G is connected, there is at least one edge between $V(H)$ and $V(G) \setminus V(H)$; fix adjacent vertices $u \in V(H)$ and $v \in V(G) \setminus V(H)$. Since G is 2-connected, we know that $G \setminus u$ is connected; consequently, there is a path in $G \setminus u$ from v to some vertex in $V(H) \setminus \{u\}$; let $P = v, p_1, \ldots, p_t$ ($t \geq 1$) be a path in $G \setminus u$ with $p_t \in V(H) \setminus \{u\}$; we may assume that $p_1, \ldots, p_{t-1} \in V(G) \setminus V(H)$.[39] But now the graph obtained from H by adding the path $u, v, p_1, \ldots, p_t$ contradicts the maximality of H.



This proves that $V(H) = V(G)$. Since we already know that H is an induced subgraph of G, it follows that $H = G$. This proves the "only if" part of the lemma. □

[37] This means that no subgraph H^* of G that either is a cycle or can be obtained from a cycle by repeated path addition contains H as a proper subgraph.

[38] A graph H is an *induced subgraph* of a graph G if $V(H) \subseteq V(G)$, and for all distinct $u, v \in V(H)$, we have that $uv \in E(H)$ if and only if $uv \in E(G)$.

[39] Otherwise, we fix a minimal index $i \in \{1, \ldots, t-1\}$ such that $p_i \in V(H)$, and we consider the path $v, p_1, \ldots, p_i$ instead of $v, p_1, \ldots, p_t$.

Chapter 6

Triangle-free graphs and graphs without a C_4 subgraph. Cayley's formula. Sperner's theorem

6.1 Graphs without K_3 as a subgraph

A graph is said to be *triangle-free* if it does not contain K_3 as a subgraph.

The following theorem is a special case of "Turán's theorem," gives a formula for the maximum number of edges in any K_n-free graph (see section 19.1).

Mantel's theorem. *Let n be a positive integer. Then*

(a) any triangle-free graph on n vertices has at most $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil = \lfloor \frac{n^2}{4} \rfloor$ edges;

(b) there exists a triangle-free graph on n vertices that has precisely $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil = \lfloor \frac{n^2}{4} \rfloor$ edges.

Proof. First, let us check that $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil = \lfloor \frac{n^2}{4} \rfloor$. If n is even, then this is obvious. If n is odd, then there exists a non-negative integer k such that $n = 2k + 1$, we compute

$$\lfloor \tfrac{n}{2} \rfloor \lceil \tfrac{n}{2} \rceil \;=\; \lfloor \tfrac{2k+1}{2} \rfloor \lceil \tfrac{2k+1}{2} \rceil \;=\; k(k+1) \;=\; k^2 + k$$

and

$$\lfloor \tfrac{n^2}{4} \rfloor \;=\; \lfloor \tfrac{(2k+1)^2}{4} \rfloor \;=\; \lfloor \tfrac{4k^2+4k+1}{4} \rfloor \;=\; k^2 + k,$$

and we deduce that $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil = \lfloor \frac{n^2}{4} \rfloor$.

For (b), we observe that the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ is triangle-free[1] and has precisely n vertices and $\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil$ edges.

It remains to prove (a). We assume inductively that the claim holds for graphs on fewer than n vertices, i.e. that for all positive integers $\widetilde{n} < n$, any triangle-free graph on $\widetilde{n}$ vertices has at most $\lfloor \frac{\widetilde{n}^2}{4} \rfloor$ edges. It is clear that (a) holds for $n = 1$ and $n = 2$. So, we assume that $n \geq 3$, we fix a triangle-free graph G on n vertices, and we show that G has at most $\lfloor \frac{n^2}{4} \rfloor$ edges. If G has no edges, then we are done. So assume that G has at least one edge, say uv. Then $G \setminus \{u, v\}$ is triangle-free and has $n-2$ vertices, and so by the induction hypothesis, it has at most $\lfloor \frac{(n-2)^2}{4} \rfloor$ edges. Further, since G is triangle-free and uv is an edge of G, a vertex in $V(G) \setminus \{u, v\}$ can be adjacent to at most one of u, v; so, the number of edges between $\{u, v\}$ and $V(G) \setminus \{u, v\}$ is at most $|V(G) \setminus \{u, v\}| = n - 2$. Since the edges of G are precisely the edges of $G \setminus \{u, v\}$, plus the edges between $\{u, v\}$ and $V(G) \setminus \{u, v\}$, plus the edge uv, we see that

$$\begin{aligned} |E(G)| &\leq \lfloor \tfrac{(n-2)^2}{4} \rfloor + (n-2) + 1 \\ &= \lfloor \tfrac{n^2-4n+4}{4} \rfloor + n - 1 \\ &= \lfloor \tfrac{n^2}{4} \rfloor, \end{aligned}$$

which is what we needed to show. □

6.2 Graphs without C_4 as a subgraph

In what follows, we will use the Cauchy-Schwarz inequality (below).

The Cauchy-Schwarz inequality. *All real numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$ satisfy*

$$\Big(\sum_{i=1}^{n} a_i b_i \Big)^2 \leq \Big(\sum_{i=1}^{n} a_i^2 \Big) \Big(\sum_{i=1}^{n} b_i^2 \Big).$$

Proof. Omitted. □

An *isolated vertex* is a vertex that has no neighbors.

Theorem 6.2.1. *Let n be a positive integer. Any graph on n vertices that does not contain C_4 as a subgraph has at most $\frac{1}{2}(n + n^{3/2})$ edges.*

[1] Indeed, all bipartite graphs are triangle free.

Proof. Let G be a graph on n vertices, and assume that G does not contain C_4 as a subgraph. Clearly, we may assume that G has no isolated vertices.[2] Let $d_1, \ldots, d_n$ be the degrees of the vertices of G;[3] since G is has no isolated veetices, we see that $d_1, \ldots, d_n \geq 1$.

Let $M := \left\{(v, A) \mid v \in V(G), A \in \binom{N_G(v)}{2}\right\}$.[4] We will count the number of elements of M in two ways.

First, for each $v \in V(G)$, there are precisely $\binom{d_G(v)}{2}$ choices of A such that $(v, A) \in M$. So, $|M| = \sum\limits_{v \in V(G)} \binom{d_G(v)}{2} = \sum\limits_{i=1}^{n} \binom{d_i}{2}$.

We now bound $|M|$ above, as follows. Note that the second coordinate of any element of M is simply an element of $\binom{V(G)}{2}$; since $|V(G)| = n$, there are at most $\binom{n}{2}$ choices for the second coordinate of an element of M. On the other hand, since G contains no C_4 as a subgraph, we see that no two distinct elements of M have the same second coordinate. Indeed, suppose that (v_1, A) and (v_2, A) are distinct elements of M; we then set $A = \{u_1, u_2\}$, we and observe that v_1, u_1, v_2, u_2, v_1 is a (not necessarily induced) C_4 in G, a contradiction. So, $|M| \leq \binom{n}{2}$.

We now have that

$$\sum_{i=1}^{n} \binom{d_i}{2} \quad \leq \quad \binom{n}{2}.$$

Obviously, $\binom{n}{2} \leq \frac{n^2}{2}$, and since $d_1, \ldots, d_n \geq 1$, we see that $\binom{d_i}{2} \geq \frac{(d_i-1)^2}{2}$ for all $i \in \{1, \ldots, n\}$; consequently,

$$\sum_{i=1}^{n} \frac{(d_i-1)^2}{2} \quad \leq \quad \sum_{i=1}^{n} \binom{d_i}{2} \quad \leq \quad \binom{n}{2} \quad \leq \quad \frac{n^2}{2},$$

and it follows that

$$\sum_{i=1}^{n} (d_i - 1)^2 \quad \leq \quad n^2.$$

[2] Why?

[3] The d_i's are not necessarily distinct; d_i is the degree of the i-th vertex of G.

[4] In other words, M is the set of all ordered pairs $(v, \{u_1, u_2\})$ such that $v \in V(G)$, and $u_1, u_2 \in V(G)$ are two distinct neighbors of v. Note also that $(v, \{u_1, u_2\}) \in M$ if and only if u_1, v, u_2 is a (not necessarily induced) two-edge path of G. So, $|M|$ is in fact the number of (not necessarily induced) two-edge paths in G.

We now compute:

$$\begin{aligned}
\sum_{i=1}^{n}(d_i-1) &= \sum_{i=1}^{n}(d_i-1)\cdot 1 \\
&\leq \sqrt{\sum_{i=1}^{n}(d_i-1)^2}\sqrt{\sum_{i=1}^{n}1^2} && \text{by the Cauchy-Schwarz inequality} \\
&= \sqrt{\sum_{i=1}^{n}(d_i-1)^2}\sqrt{n} \\
&\leq \sqrt{n^2}\sqrt{n} && \text{because } \sum_{i=1}^{n}(d_i-1)^2 \leq n^2 \\
&= n^{3/2}.
\end{aligned}$$

It now follows that

$$|E(G)| = \tfrac{1}{2}\sum_{i=1}^{n} d_i = \tfrac{1}{2}\Big(n+\sum_{i=1}^{n}(d_i-1)\Big) \leq \tfrac{1}{2}(n+n^{3/2}),$$

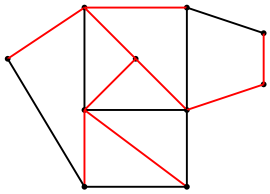
which is what we needed to show. □

6.3 Cayley's formula for the number of spanning trees of a complete graph

Recall that a *forest* is an acyclic graph (i.e. a graph that has no cycles), and a *tree* is a connected forest. A *leaf* is a vertex of degree one, i.e. a vertex that has exactly one neighbor. In what follows, we will use the well-known fact that every tree on at least two vertices has a leaf.[5] It is clear that if v is a leaf of a tree T, then $T \setminus v$ is still a tree.

[5]In fact, every tree on at least two vertices has at least two leaves. Let us prove this. Suppose that T is a tree on at least two vertices, and let $P = p_1, \ldots, p_t$ be a path of maximum length in T. Since T has at least one edge (because it is connected and has at least two vertices), we know that $t \geq 2$. We claim that p_1 and p_t are leaves of T; by symmetry, it suffices to show that p_1 is a leaf. Obviously, p_1 is adjacent to p_2 in T. Further, if p_1 were adjacent to some p_i with $i \in \{3, \ldots, t\}$, then $p_1, p_2, \ldots, p_i, p_1$ would be a cycle in T, contrary to the fact that T is a tree. Finally, if p_1 were adjacent to some vertex $v \in V(T) \setminus \{p_1, \ldots, p_t\}$, then the path $v, p_1, \ldots, p_t$ would contradict the maximality of P. So, p_2 is the only neighbor of p_1 in T, and it follows that p_1 is a leaf of T.

A *spanning tree* of a connected graph G is a tree T that is a subgraph of G, and satisfies $V(T) = V(G)$. An example is given below (the edges of the spanning tree are in red).



Now, suppose we are given a labeled complete graph on n ($n \geq 2$) vertices (say, with vertices labeled $1, \ldots, n$). We would like to count the number of spanning trees in this graph; equivalently, we would like to count the number of trees on the vertex set $\{1, \ldots, n\}$. There is only one spanning tree for K_2, and it is easy to see that there are three spanning trees for K_3. For K_4, there are 16 spanning trees, represented in Figure 6.1 (only the edges of the trees are represented; the remaining edges of the K_4 are not shown).

Our goal in this section is to prove the following.

Cayley's formula. *For all $n \geq 2$, the number of spanning trees of K_n is n^{n-2}.*

There are a number of known proofs of Cayley's formula; here, we give the one that uses the so called "Prüfer codes."

We will show that for all finite sets $S \subseteq \mathbb{N}$ with $|S| \geq 2$, the number of trees on the vertex set S is $|S|^{|S|-2}$ (see Lemma 6.3.4). Obviously, this will immediately imply Cayley's formula, since the number of spanning trees of K_n is equal to the number of trees on the vertex set $\{1, \ldots, n\}$.

To simplify terminology, we will say that a tree is an *integer tree* if all its vertices are positive integers.[6] We now define the *Prüfer code* of integer trees on at least two vertices recursively, as follows:

- for any integer tree T on exactly two vertices, the Prüfer code of T, denoted by $P(T)$, is the empty sequence;
- for any integer tree T on at least three vertices, we define the Prüfer code of T to be $P(T) := a_i, P(T \setminus i)$, where i is the smallest leaf of T, and a_i is the unique neighbor of i in T.[7]

[6] Note, however, that this is **not** standard terminology. (There is no standard terminology for such trees.) We use the term "integer tree" as a convenient shorthand in this section.

[7] So, $P(T)$ is obtained by adding a_i to the front of $P(T \setminus i)$.

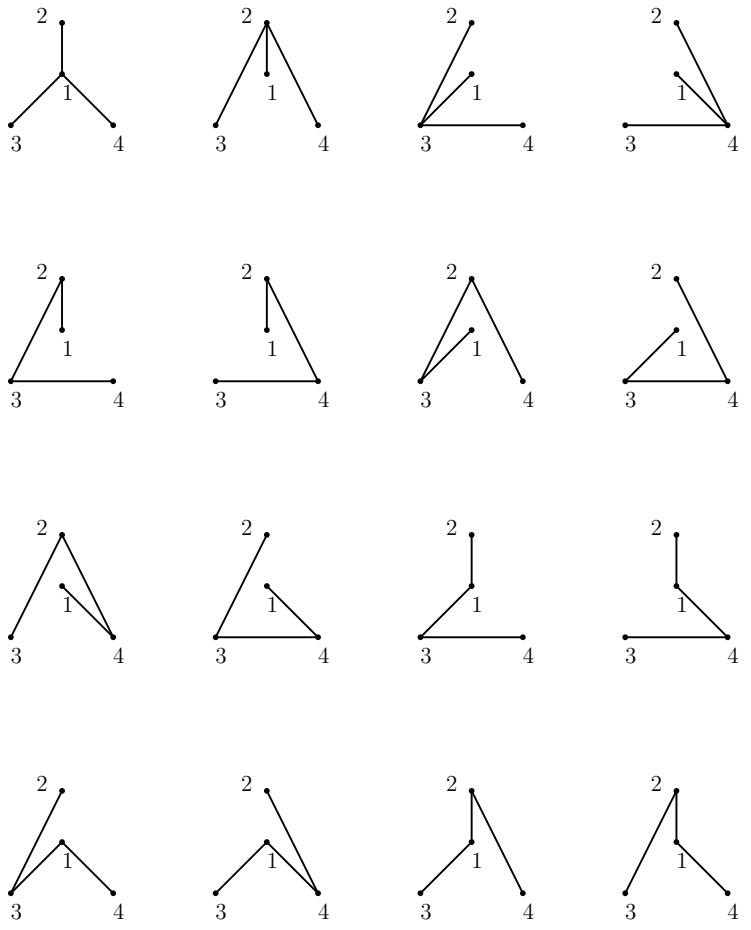
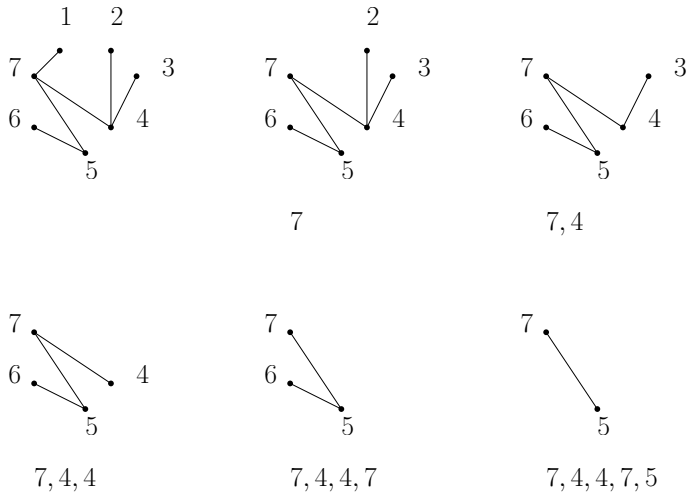


Figure 6.1: Spanning trees of K_4, or equivalently, trees on the vertex set $\{1, 2, 3, 4\}$.

An example is given below (the Prüfer code of the tree in the top left corner is $7, 4, 4, 7, 5$, and the procedure for finding it is shown below).



Now, our goal is to show that given a set $S \subseteq \mathbb{N}$ (with $|S| = n \geq 2$), the function $T \mapsto P(T)$ is a bijection from the set of all trees with vertex set S, to the set of all sequences of length $n-2$ with terms in S.[8]

Lemma 6.3.1. *If T is an integer tree on at least two vertices, then every non-leaf of T appears in $P(T)$, and none of the leaves do.*

Proof. We prove the lemma by induction on the number of vertices of the integer tree T. If T is a 2-vertex integer tree, then both its vertices are leaves, and by definition, $P(T)$ is the empty sequence; so the lemma is true for 2-vertex integer trees. Now, fix an integer $n \geq 2$, and assume inductively that the lemma holds for integer trees on n vertices. Let T be an integer tree on $n+1$ vertices. Let i be the smallest leaf of T, and let a_i be the unique neighbor of i. Since T is connected and has at least three vertices, adjacent vertices cannot both be leaves of T, and so a_i is a non-leaf of T. By construction, $P(T) = a_i, P(T \setminus i)$, and so the non-leaf a_i of T appears in $P(T)$, whereas the leaf i of T does not. Note that for $v \in V(T) \setminus \{i, a_i\}$, we have that $d_T(v) = d_{T \setminus i}(v)$, and so each vertex of T other than i and a_i is a leaf in T if and only if it is a leaf in $T \setminus i$. The result now follows from the induction hypothesis. □

Lemma 6.3.2. *If two integer trees have the same vertex set and the same Prüfer code, then they are identical.*

[8] Obviously, there are precisely n^{n-2} such sequences.

Proof. We proceed by induction on the number of vertices. There is only one tree on a fixed two-element vertex set, and so the lemma clearly holds for 2-vertex integer trees. Now, fix an integer $n \geq 2$, and suppose inductively that the lemma is true for integer trees on n vertices. Fix $S \subseteq \mathbb{N}$ with $|S| = n+1$, and let T_1 and T_2 be trees with vertex-set S and identical Prüfer code P. P is of length $n-1$, and so at least two members of S do not appear in P; let i be the smallest integer in S that does not appear in P. Let a_i be the first term of P, and let P_i be obtained from P by deleting its first term. By Lemma 6.3.1, i is the smallest leaf of both T_1 and T_2, and by the definition of the Prüfer code, a_i is the unique neighbor of i in both T_1 and T_2. Further, we have that $P(T_1 \setminus i) = P(T_2 \setminus i) = P_i$, and so by the induction hypothesis, $T_1 \setminus i = T_2 \setminus i$. Since i has the same neighborhood in T_1 and in T_2, it follows that $T_1 = T_2$. □

Lemma 6.3.3. *If $n \geq 2$ is an integer, and if $S \subseteq \mathbb{N}$ with $|S| = n$, then every sequence of length $n-2$, all of whose terms are in S, is the Prüfer code of some tree with vertex-set S.*

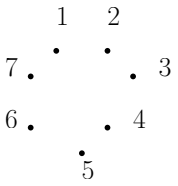
Proof. We proceed by induction on n.

Suppose first that $S \subseteq \mathbb{N}$ satisfies $|S| = 2$, and let P be a sequence of length $2-2=0$, all of whose terms are in S. Then P is the empty sequence. Let T be the unique tree on the vertex-set S. Then $P(T) = P$.

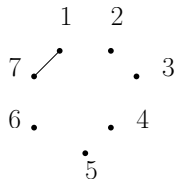
Now, fix an integer $n \geq 2$, and suppose inductively that the lemma is true for n. We need to show that it holds for $n+1$. Let $S \subseteq \mathbb{N}$ be such that $|S| = n+1$, and let P be a sequence of length $n-1$, all of whose terms are in S. Let i be the smallest member of S that does not appear in P, and let a_i be the first term of P. Let P_i be the sequence obtained by deleting the first term from P. By the induction hypothesis, there is a tree T_i with vertex-set $S \setminus \{i\}$ and Prüfer code P_i. Let T be the tree obtained by adding the vertex i to T_i, and making i adjacent to a_i and to no other vertex of T_i. Now $P(T) = P$. This completes the induction. □

Note that the proof of Lemma 6.3.3 in fact gives us a recipe for "decoding" a given Prüfer code, i.e. for finding the tree to which the code corresponds. For an integer $n \geq 2$, an n-element set $S \subseteq \mathbb{N}$, and an $(n-2)$-term sequence P, with terms in S, we proceed as follows. If $n \geq 3$, then we let i be the smallest element of S that is not in P, and we let a_i be the first term of P. We make i and a_i adjacent, we delete i from S, and we delete the first term of P. We repeat the process until S has only two elements left, and P is the empty sequence. At this point, we make the last two remaining elements of S adjacent. An example is given below: the tree on the vertex

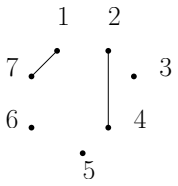
set $S = \{1, 2, 3, 4, 5, 6, 7\}$ whose Prüfer code is $7, 4, 4, 7, 5$ is the tree on the bottom of the picture (e is the empty sequence).



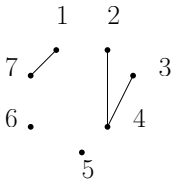
$P = 7, 4, 4, 7, 5$
$S = \{1, 2, 3, 4, 5, 6, 7\}$



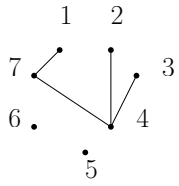
$P = 4, 4, 7, 5$
$S = \{2, 3, 4, 5, 6, 7\}$



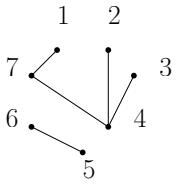
$P = 4, 7, 5$
$S = \{3, 4, 5, 6, 7\}$



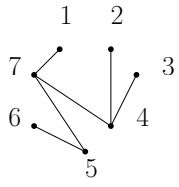
$P = 7, 5$
$S = \{4, 5, 6, 7\}$



$P = 5$
$S = \{5, 6, 7\}$



$P = e$
$S = \{5, 7\}$



Putting Lemmas 6.3.2 and 6.3.3 together, we obtain the following.

Lemma 6.3.4. *Let $n \geq 2$ be an integer, and let $S \subseteq \mathbb{N}$ be such that $|S| = n$. Then the number of trees on the vertex set S is n^{n-2}.*

Proof. By Lemmas 6.3.2 and 6.3.3, the mapping $T \mapsto P(T)$ is a bijection from the set of all integer trees on the vertex set S to the set of all $(n-2)$-term sequences, all of whose terms are elements of S. There are precisely n^{n-2} sequences of length $n-2$, with terms in S, and it follows that there are precisely n^{n-2} trees on the vertex set S. □

Cayley's formula follows immediately from Lemma 6.3.4, since the number of spanning trees of K_n is precisely the number of trees on the vertex set $\{1, \ldots, n\}$.

6.3.1 Cayley's formula via determinants

In this subsection, we give (without proof) a formula for computing the number of spanning trees of **any** graph on the vertex set $\{1, \ldots, n\}$.

Suppose that $n \geq 2$ is an integer, and that G is a graph on the vertex set $\{1, \ldots, n\}$. Then the *Laplacian* of G is the matrix $Q = [q_{i,j}]_{n \times n}$ given by

$$q_{i,j} = \begin{cases} d_G(i) & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } ij \in E(G) \\ 0 & \text{if } i \neq j \text{ and } ij \notin E(G) \end{cases}$$

We now need some notation. Suppose $A = [a_{i,j}]_{n \times m}$ is a matrix, and suppose $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$; then $A_{i,j}$ is the matrix obtained from A by deleting the i-th row and j-th column. In particular, $A_{1,1}$ is the matrix obtained from A by deleting the first row and first column.

Theorem 6.3.5. *Let $n \geq 2$ be an integer, let G be any graph on the vertex set $\{1, \ldots, n\}$, and let Q be the Laplacian of G. Then the number of spanning trees of G is precisely* $\det(Q_{1,1})$.

Proof. Omitted. □

Example 6.3.6. *Using Theorem 6.3.5, prove Cayley's formula.*

Solution. Fix an integer $n \geq 2$, and consider the complete graph on the vertex set $\{1, \ldots, n\}$. Then the Laplacian of this graph is the $n \times n$ matrix

$$Q = \begin{bmatrix} n-1 & -1 & -1 & \ldots & -1 \\ -1 & n-1 & -1 & \ldots & -1 \\ -1 & -1 & n-1 & \ldots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \ldots & n-1 \end{bmatrix}_{n \times n}.$$

The matrix $Q_{1,1}$ has exactly the same form, only it is of size $(n-1) \times (n-1)$:

$$Q_{1,1} = \begin{bmatrix} n-1 & -1 & -1 & \ldots & -1 \\ -1 & n-1 & -1 & \ldots & -1 \\ -1 & -1 & n-1 & \ldots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \ldots & n-1 \end{bmatrix}_{(n-1) \times (n-1)}.$$

We now compute the determinant of $Q_{1,1}$:

$$\begin{aligned}
\det(Q_{1,1}) &= \begin{vmatrix} n-1 & -1 & -1 & \dots & -1 \\ -1 & n-1 & -1 & \dots & -1 \\ -1 & -1 & n-1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \dots & n-1 \end{vmatrix}_{(n-1)\times(n-1)} \\
&\overset{(*)}{=} \begin{vmatrix} n-1 & -1 & -1 & \dots & -1 \\ -n & n & 0 & \dots & 0 \\ -n & 0 & n & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -n & 0 & 0 & \dots & n \end{vmatrix}_{(n-1)\times(n-1)} \\
&\overset{(**)}{=} \begin{vmatrix} 1 & -1 & -1 & \dots & -1 \\ 0 & n & 0 & \dots & 0 \\ 0 & 0 & n & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & n \end{vmatrix}_{(n-1)\times(n-1)} \\
&\overset{(***)}{=} n^{n-2},
\end{aligned}$$

where (*) is obtained by subtracting the first row from all the subsequent ones, (**) is obtained by adding to the first column the sum of all subsequent ones, and (***) is obtained by multiplying the diagonal entries of the upper triangular matrix that we obtained. By Theorem 6.3.5, we now have that the number of spanning trees of K_n is precisely n^{n-2}, which proves Cayley's formula. □

6.4 Sperner's theorem

For a partially ordered set $(X, \leq)$,

- a *chain* in $(X, \leq)$ is any set $\mathcal{C} \subseteq X$ such that for all $x_1, x_2 \in \mathcal{C}$, we have that either $x_1 \leq x_2$ or $x_2 \leq x_1$.[9]

[9]This definition works both for finite and for infinite X. Note also that $\emptyset$ is a chain in $(X, \leq)$. However, if X is finite and $\mathcal{C}$ is a non-empty chain in $(X, \leq)$, then $\mathcal{C}$ can be ordered as $\mathcal{C} = \{x_1, \dots, x_t\}$ so that $x_1 \leq \dots \leq x_t$.

- a *maximal chain* in $(X, \leq)$ is a chain $\mathcal{C}$ in $(X, \leq)$ such that there is no chain $\mathcal{C}'$ in $(X, \leq)$ with the property that $\mathcal{C} \subsetneq \mathcal{C}'$;
- an *antichain* in $(X, \leq)$ is any set $\mathcal{A} \subseteq X$ such that for all distinct $x_1, x_2 \in \mathcal{A}$, we have that $x_1 \not\leq x_2$ and $x_2 \not\leq x_1$.

Note that a chain and an antichain in $(X, \leq)$ can have at most one element in common.[10]

Here, we are interested in a special case of the above. As usual, for a set X, we denote by $\mathscr{P}(X)$ the power set (i.e. the set of all subsets) of X. Clearly, for any set X, we have that $\subseteq_{\mathscr{P}(X)} := \{(A, B) \mid A, B \in \mathscr{P}(X), A \subseteq B\}$ is a partial order on X. To simplify notation, in what follows, we write $(\mathscr{P}(X), \subseteq)$ instead of $(\mathscr{P}(X), \subseteq_{\mathscr{P}(X)})$. We apply the above definitions to $(\mathscr{P}(X), \subseteq)$, as follows. For a set X,

- a *chain* in $(\mathscr{P}(X), \subseteq)$ is any set $\mathcal{C}$ of subsets of X such that for all $C_1, C_2 \in \mathcal{C}$, we have that either $C_1 \subseteq C_2$ or $C_2 \subseteq C_1$.[11]
- a *maximal chain* in $(\mathscr{P}(X), \subseteq)$ is a chain in $(\mathscr{P}(X), \subseteq)$ such that there is no chain $\mathcal{C}'$ in $(\mathscr{P}(X), \subseteq)$ with the property that $\mathcal{C} \subsetneq \mathcal{C}'$;
- an *antichain* in $(\mathscr{P}(X), \subseteq)$ is any set $\mathcal{A}$ of subsets of X such that for all distinct $A_1, A_2 \in \mathcal{A}$, we have that $A_1 \not\subseteq A_2$ and $A_2 \not\subseteq A_1$.[12]

As before, note that a chain and an antichain in $(\mathscr{P}(X), \subseteq)$ can have at most one element in common.

Example 6.4.1. *Let $X = \{1, 2, 3, 4\}$. All the following are chains in $(\mathscr{P}(X), \subseteq)$:*[13]

- $\{\{2, 4\}, \{1, 2, 4\}\}$;[14]
- $\{\emptyset, \{1\}, \{1, 2\}, \{1, 2, 3\}, X\}$.[15]

[10] Indeed, if distinct elements x_1, x_2 belong to a chain of $(X, \leq)$, then $x_1 \leq x_2$ or $x_2 \leq x_1$. On the other hand, if they belong to an antichain of $(X, \leq)$, then $x_1 \not\leq x_2$ and $x_2 \not\leq x_1$. So, distinct elements x_1 and x_2 cannot simultaneously belong to a chain and an antichain of $(X, \leq)$.

[11] This definition works both for finite and for infinite X. Note also that $\emptyset$ is a chain in $(\mathscr{P}(X), \subseteq)$. However, if X is finite and $\mathcal{C}$ is a non-empty chain in $(\mathscr{P}(X), \subseteq)$, then $\mathcal{C}$ can be ordered as $\mathcal{C} = \{C_1, \ldots, C_t\}$ so that $C_1 \subseteq \cdots \subseteq C_t$.

[12] Equivalently: $A_1 \setminus A_2$ and $A_2 \setminus A_1$ are both non-empty.

[13] Note, however, that there are many other chains in $(\mathscr{P}(X), \subseteq)$ as well.

[14] Note that this chain is **not** maximal, since we can add (for example) the set $\{2\}$ to it and obtain a larger chain.

[15] This chain is maximal.

- $\{\emptyset, \{4\}, \{2,4\}, \{1,2,4\}, X\}$;[16]

Further, the following are all antichains in $(\mathscr{P}(X), \subseteq)$*:*[17]

- $\{\emptyset\}$;
- $\{X\}$;
- $\{\{1,2\}, \{2,3\}, \{1,3,4\}\}$;
- $\{\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}\}$.

Sperner's theorem. *Let n be a non-negative integer, and let X be an n-element set. Then any antichain in $(\mathscr{P}(X), \subseteq)$ has at most $\binom{n}{\lfloor n/2 \rfloor}$ elements. Furthermore, this bound is tight, that is, there exists an antichain in $(\mathscr{P}(X), \subseteq)$ that has precisely $\binom{n}{\lfloor n/2 \rfloor}$ elements.*

Proof. First, we note that the set of all $\lfloor n/2 \rfloor$-element subsets of X is an antichain in $(\mathscr{P}(X), \subseteq)$, and this antichain has precisely $\binom{n}{\lfloor n/2 \rfloor}$ elements. It remains to show that any antichain in $(\mathscr{P}(X), \subseteq)$ has at most $\binom{n}{\lfloor n/2 \rfloor}$ elements.

> **Claim 1.** There are precisely $n!$ maximal chains in $(\mathscr{P}(X), \subseteq)$.

Proof of Claim 1. Clearly, any maximal chain in $(\mathscr{P}(X), \subseteq)$ is of the form $\{\emptyset, \{x_1\}, \{x_1, x_2\}, \ldots, \{x_1, x_2, \ldots, x_n\}\}$, where $x_1, \ldots, x_n$ is some ordering of the elements of X. There are precisely $n!$ such orderings, and so the number of maximal chains in $(\mathscr{P}(X), \subseteq)$ is $n!$. ♦

> **Claim 2.** For every set $A \subseteq X$, the number of maximal chains of $(\mathscr{P}(X), \subseteq)$ containing A is precisely $|A|!(n - |A|)!$.

Proof of Claim 2. Set $k = |A|$. As in the proof of Claim 1, we observe that any maximal chain in $(\mathscr{P}(X), \subseteq)$ is of the form $\{\emptyset, \{x_1\}, \{x_1, x_2\}, \ldots, \{x_1, x_2, \ldots, x_n\}\}$, where $x_1, \ldots, x_n$ is some ordering of the elements of X; this chain contains A if and only if $A = \{x_1, \ldots, x_k\}$ (and therefore, $X \setminus A = \{x_{k+1}, \ldots, x_n\}$). The number of ways of ordering A is $k!$, and the number of ways of ordering $X \setminus A$ is $(n-k)!$. So, the total number of chains of $(\mathscr{P}(X), \subseteq)$ containing A is precisely $k!(n-k)!$. ♦

Now, fix an antichain $\mathcal{A}$ in $(\mathscr{P}(X), \subseteq)$. We form the matrix M whose rows are indexed by the elements of $\mathcal{A}$, and whose columns are indexed by

[16] This chain is maximal.

[17] Note, however, that there are many other antichains in $(\mathscr{P}(X), \subseteq)$ as well.

the maximal chains of $(\mathscr{P}(X), \subseteq)$, and in which the $(A, \mathcal{C})$-th entry is 1 if $A \in \mathcal{C}$ and is 0 otherwise.[18] Our goal is to count the number of 1's in the matrix M in two ways.

First, by Claim 2, for any $A \in \mathcal{A}$, the number of maximal chains of $(\mathscr{P}(X), \subseteq)$ containing A is precisely $|A|!(n-|A|)!$; so, the number of 1's in the row of M indexed by A is precisely $|A|!(n-|A|)!$. Thus, the number of 1's in the matrix M is precisely

$$\sum_{A \in \mathcal{A}} |A|!(n-|A|)!.$$

On the other hand, by Claim 1, the number of columns of M is precisely $n!$. Furthermore, no chain of $(\mathscr{P}(X), \subseteq)$ contains more than one element of the antichain $\mathcal{A}$, and so no column of M contains more than one 1. So, the total number of 1's in the matrix M is at most $n!$. We now have that

$$\sum_{A \in \mathcal{A}} |A|!(n-|A|)! \quad \leq \quad n!,$$

and consequently,

$$\sum_{A \in \mathcal{A}} \frac{|A|!(n-|A|)!}{n!} \quad \leq \quad 1.$$

On the other hand, for all $A \subseteq X$ (and in particular, for all $A \in \mathcal{A}$), we have that

$$\frac{|A|!(n-|A|)!}{n!} \quad = \quad \frac{1}{\frac{n!}{|A|!(n-|A|)!}} \quad = \quad \frac{1}{\binom{n}{|A|}} \quad \overset{(*)}{\geq} \quad \frac{1}{\binom{n}{\lfloor n/2 \rfloor}},$$

where (*) follows from the fact that $\binom{n}{k} \leq \binom{n}{\lfloor n/2 \rfloor}$ for all $k \in \{0, \ldots, n\}$.[19] We now have that

$$1 \quad \geq \quad \sum_{A \in \mathcal{A}} \frac{|A|!(n-|A|)!}{n!} \quad \geq \quad \sum_{A \in \mathcal{A}} \frac{1}{\binom{n}{\lfloor n/2 \rfloor}} \quad \geq \quad |\mathcal{A}| \frac{1}{\binom{n}{\lfloor n/2 \rfloor}},$$

which yields $|\mathcal{A}| \leq \binom{n}{\lfloor n/2 \rfloor}$. This completes the argument. □

[18] Here, $A \in \mathcal{A}$, $\mathcal{C}$ is a maximal chain in $(\mathscr{P}(X), \subseteq)$, and the $(A, \mathcal{C})$-th entry of M is the entry in the row indexed by A and column indexed by $\mathcal{C}$.

[19] This was discussed in section 1.3.

Chapter 7

Ramsey theory

7.1 The Pigeonhole principle

The Pigeonhole Principle. *Let $n_1, \ldots, n_t$ ($t \geq 1$) be non-negative integers, and let X be a set of size at least $1 + n_1 + \cdots + n_t$. If $(X_1, \ldots, X_t)$ is any partition of X,*[1] *then there exists some $i \in \{1, \ldots, t\}$ such that $|X_i| > n_i$.*[2]

Proof. Suppose otherwise, and fix a partition $(X_1, \ldots, X_t)$ such that $|X_i| \leq n_i$ for all $i \in \{1, \ldots, t\}$. But then

$$1 + n_1 + \cdots + n_t \quad \leq \quad |X| \quad = \quad |X_1| + \cdots + |X_t| \quad \leq \quad n_1 + \cdots + n_t,$$

a contradiction. □

As an immediate corollary, we obtain the following.

Corollary 7.1.1. *Let n and t be positive integers. Let X be an n-element set, and let $(X_1, \ldots, X_t)$ be any partition of X.*[3] *Then there exists some $i \in \{1, \ldots, t\}$ such that $|X_i| \geq \lceil \frac{n}{t} \rceil$.*

Proof. By the Pigeonhole Principle, we need only show that $n \geq 1 + t(\lceil \frac{n}{t} \rceil - 1)$. If $t \,|\, n$,[4] then $\lceil \frac{n}{t} \rceil = \frac{n}{t}$, and we have that

$$1 + t(\lceil \tfrac{n}{t} \rceil - 1) \quad \leq \quad 1 + t(\tfrac{n}{t} - 1) \quad = \quad n - t + 1 \quad \leq \quad n,$$

[1] Here, we allow sets $X_1, \ldots, X_t$ to possibly be empty.

[2] If one thinks of elements of X as "pigeons" and sets $X_1, \ldots, X_t$ as "pigeonholes," then the Pigeonhole Principle states that some pigeonhole X_i receives more than n_i pigeons.

[3] Here, we allow sets $X_1, \ldots, X_t$ to possibly be empty.

[4] "$t \,|\, n$" means that n is divisible by t.

which is what we needed. Suppose now that $t \not| n$, so that $\lceil \frac{n}{t} \rceil - 1 = \lfloor \frac{n}{t} \rfloor$. Then let $m = \lfloor \frac{n}{t} \rfloor$ and $\ell = n - mt$; since $t \not| n$, we have that $\ell \geq 1$. But now

$$1 + t(\lceil \tfrac{n}{t} \rceil - 1) \;=\; 1 + t(\lfloor \tfrac{n}{t} \rfloor) \;=\; 1 + tm \;\leq\; \ell + tm \;=\; n,$$

and we are done. □

We remark that Corollary 7.1.1 is also often referred to as the Pigeonhole Principle.

7.2 Ramsey numbers

A *clique* in a graph G is any set of pairwise adjacent vertices of G. The *clique number* of G, denoted by $\omega(G)$, is the maximum size of a clique in G.

A *stable set* (or *independent set*) in a graph G is any set of pairwise non-adjacent vertices of G. The *stability number* (or *independence number*) of G, denoted by $\alpha(G)$, is the maximum size of a stable set in G.

We begin with a simple proposition, which we will then generalize.

Proposition 7.2.1. *Let G be a graph on at least six vertices. Then either $\omega(G) \geq 3$ or $\alpha(G) \geq 3$.*

Proof. Let u be any vertex of G. Then $|V(G) \setminus \{u\}| \geq 5$, and so (by the Pigeonhole Principle) either u has at least three neighbors or it has at least three non-neighbors.

Suppose first that u has at least three neighbors. If at least two of those neighbors, say u_1 and u_2, are adjacent, then $\{u, u_1, u_2\}$ is a clique of G of size three, and we deduce that $\omega(G) \geq 3$. On the other hand, if no two neighbors of u are adjacent, then they together form a stable set of size at least three, and we deduce that $\alpha(G) \geq 3$.

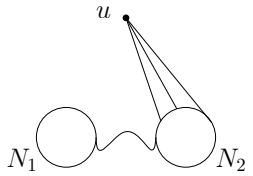
Suppose now that u has at least three non-neighbors. If at least two of those non-neighbors, say u_1 and u_2, are non-adjacent, then $\{u, u_1, u_2\}$ is a stable set of G of size three, and we deduce that $\alpha(G) \geq 3$. On the other hand, if the non-neighbors of u are pairwise adjacent, then they together form a clique of size at least three, and we deduce that $\omega(G) \geq 3$. □

As usual, for a graph G and a vertex u, $N_G(u)$ is the set of all neighbors of u in G, and we set $N_G[u] := \{u\} \cup N_G(u)$; $N_G(u)$ is called the *open neighborhood* (or simply *neighborhood*) of u in G, and $N_G[u]$ is called the *closed neighborhood* of u in G. Our next theorem generalizes Proposition 7.2.1.

Theorem 7.2.2. *Let k and ℓ be positive integers, and let G be a graph on at least $\binom{k+\ell-2}{k-1}$ vertices.*[5] *Then either $\omega(G) \geq k$ or $\alpha(G) \geq \ell$.*

Proof. We may assume inductively that for all positive integers k', ℓ' such that $k' + \ell' < k + \ell$, all graphs G' on at least $\binom{k'+\ell'-2}{k'-1}$ vertices satisfy either $\omega(G') \geq k'$ or $\alpha(G') \geq \ell'$.

If $k = 1$ or $\ell = 1$, then the result is immediate.[6] So, we may assume that $k, \ell \geq 2$. Now, set $n = \binom{k+\ell-2}{k-1}$, $n_1 = \binom{k+\ell-3}{k-1}$, and $n_2 = \binom{k+\ell-3}{k-2}$; then $n = n_1 + n_2$, and consequently, $n - 1 = 1 + (n_1 - 1) + (n_2 - 1)$. Fix any vertex $u \in V(G)$, and set $N_1 = V(G) \setminus N_G[u]$ and $N_2 = N_G(u)$.



Since (N_1, N_2) is a partition of $V(G) \setminus \{u\}$, and since $|V(G) \setminus \{u\}| \geq n - 1 = 1 + (n_1 - 1) + (n_2 - 1)$, the Pigeonhole Principle guarantees that either $|N_1| \geq n_1$ or $|N_2| \geq n_2$.

Suppose first that $|N_1| \geq n_1$, i.e. $|N_1| \geq \binom{k+(\ell-1)-2}{k-1}$. Then by the induction hypothesis, either $\omega(G[N_1]) \geq k$ or $\alpha(G[N_1]) \geq \ell - 1$. In the former case, we have that $\omega(G) \geq \omega(G[N_1]) \geq k$, and we are done. So, suppose that $\alpha(G[N_1]) \geq \ell - 1$, and let S be a stable set of $G[N_1]$ of size $\ell - 1$. Then $\{u\} \cup S$ is a stable set of size ℓ in G, we deduce that $\alpha(G) \geq \ell$, and again we are done.

Suppose now that $|N_2| \geq n_2$, i.e. $|N_2| \geq \binom{(k-1)+\ell-2}{k-2}$. Then by the induction hypothesis, either $\omega(G[N_2]) \geq k - 1$ or $\alpha(G[N_2]) \geq \ell$. In the latter case, we have that $\alpha(G) \geq \alpha(G[N_2]) \geq \ell$, and we are done. So, suppose that $\omega(G[N_2]) \geq k - 1$, and let C be a clique of $G[N_2]$ of size $k - 1$. But then $\{u\} \cup C$ is a clique of size k in G, we deduce that $\omega(G) \geq k$, and again we are done. □

For positive integers k and ℓ, we denote by $R(k, \ell)$ the smallest integer n such that every graph G on at least n vertices satisfies either $\omega(G) \geq k$ or $\alpha(G) \geq \ell$. The existence of $R(k, \ell)$ follows immediately from Theorem 7.2.2. Numbers $R(k, \ell)$ (with $k, \ell \geq 1$) are called *Ramsey numbers*.

[5] Note that $\binom{k+\ell-2}{k-1} = \binom{k+\ell-2}{\ell-1}$.

[6] Indeed, it is clear that $\omega(G) \geq 1$ and $\alpha(G) \geq 1$. So, if $k = 1$, then $\omega(G) \geq k$; and if $\ell = 1$, then $\alpha(G) \geq \ell$.

It is easy to see that for all $k, \ell \geq 1$, we have that[7]

$$R(1,\ell) = 1 \qquad R(k,1) = 1$$

$$R(2,\ell) = \ell \qquad R(k,2) = k$$

Furthermore, we have $R(3,3) = 6$. Indeed, by Proposition 7.2.1, $R(3,3) \leq 6$. On the other hand, $\omega(C_5) = 2$ and $\alpha(C_5) = 2$, and so $R(3,3) > 5$. Thus, $R(3,3) = 6$. The exact values of a few other Ramsey numbers are known, but no general formula for $R(k,\ell)$ is known. Note, however, that Theorem 7.2.2 gives an upper bound for Ramsey numbers, namely,

$$R(k,\ell) \quad \leq \quad \tbinom{k+\ell-2}{k-1}$$

for all positive integers k, ℓ.

We complete this section by giving a lower bound for the Ramsey number $R(k,k)$. Interestingly, this lower bound is obtained using probabilistic methods.

Theorem 7.2.3. *For all integers $k \geq 3$, we have that $R(k,k) > 2^{k/2}$.*

Proof. Since $\omega(C_5) = 2$ and $\alpha(C_5) = 2$, we see that $R(3,3) > 5 > 2^{3/2}$ and $R(4,4) > 5 > 2^{4/2}$. Thus, the claim holds for $k = 3$ and $k = 4$. From now on, we assume that $k \geq 5$. We will show that there exists a graph G on $\lfloor 2^{k/2} \rfloor$ vertices such that $\omega(G), \alpha(G) < k$. This is enough, because it implies that $R(k,k) > \lfloor 2^{k/2} \rfloor$, and consequently, that $R(k,k) > 2^{k/2}$ (because Ramsey numbers are integers).

Let G be a graph on $n := \lfloor 2^{k/2} \rfloor$ vertices, with adjacency as follows: between any two distinct vertices, we (independently) put an edge with probability $\frac{1}{2}$ (and a non-edge with probability $\frac{1}{2}$).

For any set of k vertices of G, the probability that this set is a clique is $(\frac{1}{2})^{\binom{k}{2}}$; there are $\binom{n}{k}$ subsets of $V(G)$ of size k, and the probability that at least one of them is a clique is at most $\binom{n}{k}(\frac{1}{2})^{\binom{k}{2}}$. So, the probability that $\omega(G) \geq k$ is at most $\binom{n}{k}(\frac{1}{2})^{\binom{k}{2}}$. Similarly, the probability that $\alpha(G) \geq k$ is at most $\binom{n}{k}(\frac{1}{2})^{\binom{k}{2}}$. Thus, the probability that G satisfies at least one of

[7] Check this!

$\omega(G) \geq k$ and $\alpha(G) \geq k$ is at most

$$\begin{aligned}
2\binom{n}{k}(\tfrac{1}{2})^{\binom{k}{2}} &\leq 2(\tfrac{en}{k})^k(\tfrac{1}{2})^{\binom{k}{2}} && \text{by Theorem 1.3.1} \\
&\leq \frac{2(\frac{e2^{k/2}}{k})^k}{2^{k(k-1)/2}} && \text{because } n = \lfloor 2^{k/2} \rfloor \\
&= 2(\frac{e2^{k/2}}{k2^{(k-1)/2}})^k \\
&= 2(\frac{e\sqrt{2}}{k})^k \\
&< 1 && \text{because } k \geq 5.
\end{aligned}$$

Thus, the probability that G satisfies neither $\omega(G) \geq k$ nor $\alpha(G) \geq k$ is strictly positive. So, there must be at least one graph on $n = \lfloor 2^{k/2} \rfloor$ vertices whose clique number and stability number are both strictly smaller than k. This completes the argument. □

7.3 Ramsey's theorem (hypergraph version)

First, we need some notation. We denote by $\mathbb{N}$ the set of all positive integers.[8] For a positive integer n, we set $[n] = \{1, \ldots, n\}$. For a set X and a non-negative integer k, we denote by $\binom{X}{k}$ the set of all subsets of X of size k. In particular, $\binom{X}{2}$ is the set of all subsets of X of size two. Note that this means that if G is a (simple) graph, then $E(G) \subseteq \binom{V(G)}{2}$.

Recall that for positive integers k and ℓ, the Ramsey number $R(k, \ell)$ is the smallest $N \in \mathbb{N}$ such that every graph G on at least N vertices satisfies either $\omega(G) \geq k$ or $\alpha(G) \geq \ell$. Here is a slightly different way to think about Ramsey numbers. Clearly, any graph G corresponds to a complete graph on the same vertex set, whose edges are colored black or white, with an edge of the complete graph colored black if it was an edge of the graph G, and colored white otherwise. With this set-up, it is easy to see that $R(k, \ell)$ (with $k, \ell \in \mathbb{N}$) is the smallest $N \in \mathbb{N}$ such that any complete graph on at least N vertices, whose edges are colored black or white, has either a monochromatic[9] black complete subgraph of size k, or a monochromatic white complete subgraph of size ℓ. Now, let us suppose that instead of colors

[8] In some texts, $\mathbb{N}$ is used to denote the set of all non-negative integers. Here, it is the set of all positive integers.

[9] Here, "monochromatic" simply means that all edges are colored with the same color.

black and white, we use colors 1 and 2. Then a coloring of the complete graph on the vertex set X is simply a function $c : \binom{X}{2} \to [2]$.[10] We now see that $R(k, \ell)$ (with $k, \ell \in \mathbb{N}$) is the smallest $N \in \mathbb{N}$ such that for all finite sets X with $|X| \geq N$, and all colorings $c : \binom{X}{2} \to [2]$, either there exists a set $A_1 \in \binom{X}{k}$ such that c assigns color 1 to each set in $\binom{A_1}{2}$, or there exists a set $A_2 \in \binom{X}{\ell}$ such that c assigns color 2 to each set in $\binom{A_2}{2}$.[11]

This can be generalized!

Recall that, for a set X, we denote by $\mathscr{P}(X)$ the *power set* of X, i.e. the set of all subsets of X. A *hypergraph* is an ordered pair $H = (V(H), E(H))$, where $V(H)$ is some non-empty finite set,[12] and $E(H) \subseteq \mathscr{P}(V(H)) \setminus \{\emptyset\}$. As in the graph case, members of $V(H)$ are called *vertices* and members of $E(H)$ are called *edges* of the hypergraph H.[13] For a positive integer p, a hypergraph is *p-uniform* if all its edges have precisely p vertices. A hypergraph is *uniform* if it is p-uniform for some p. So, if H is a p-uniform hypergraph, then $E(H) \subseteq \binom{V(H)}{p}$. Note that this means that a graph is simply a 2-uniform hypergraph.

Given $p, t, k_1, \ldots, k_t \in \mathbb{N}$, the *Ramsey number* $R^p(k_1, \ldots, k_t)$ is the smallest $N \in \mathbb{N}$ (if it exists) such that for all finite sets X with $|X| \geq N$, and all colorings (i.e. functions) $c : \binom{X}{p} \to [t]$,[14] there exist an index $i \in [t]$ and a set $A_i \in \binom{X}{k_i}$ such that c assigns color i to each element of $\binom{A_i}{p}$.[15] As our next theorem shows, the Ramsey numbers $R^p(k_1, \ldots, k_t)$ are always defined. We will give two proofs of this theorem. The first is more elementary (it proceeds by induction on p), but also somewhat messy. The second one (given in section 7.5) relies on the "infinite version" of Ramsey's theorem (see section 7.4); this second proof is more "advanced" (i.e. it requires more sophisticated mathematical results), but it is also more elegant.

Ramsey's theorem (hypergraph version). *For all $p, t, k_1, \ldots, k_t \in \mathbb{N}$, the number $R^p(k_1, \ldots, k_t)$ exists.*

Proof. We fix $t \in \mathbb{N}$, and we proceed by induction on p.

First, for $p = 1$, we fix $k_1, \ldots, k_t \in \mathbb{N}$, and we set $N = (k_1 - 1) + \cdots + (k_t - 1) + 1$. Fix any finite set X with $|X| \geq N$, and any coloring $c : \binom{X}{p} \to [t]$.

[10] Note that the edge set of the complete graph on vertex set X is precisely the set $\binom{X}{2}$.

[11] Note that "$A_1 \in \binom{X}{k}$" simply means that A_1 is a k-element subset of X. Similarly, "$A_2 \in \binom{X}{\ell}$" simply means that A_2 is an ℓ-element subset of X.

[12] Occasionally, $V(H)$ is allowed to be empty.

[13] So, an edge of a hypergraph can be any non-empty subset of vertices of the hypergraph.

[14] So, c is an assignment of colors to the edges of the "complete" p-uniform hypergraph on vertex set X.

[15] With this set-up, we have that $R(k, \ell) = R^2(k, \ell)$.

Now, for all $i \in [t]$, set $C_i = \{x \in X \mid c(\{x\}) = i\}$. Then $(C_1, \ldots, C_t)$ is a partition of X, and $|X| \geq N = (k_1 - 1) + \cdots + (k_t - 1) + 1$. So, by the Pigeonhole Principle, there is some $i \in [t]$ such that $|C_i| \geq k_i$. Now, let A_i be any subset of C_i such that $|A_i| = k_i$; so, $A_i \in \binom{X}{k_i}$. By construction, c assigns color i to each element of $\binom{A_i}{p}$. So, $R^1(k_1, \ldots, k_t)$ exists, and we see that the theorem holds for $p = 1$.

Now, fix $p \in \mathbb{N}$, and assume inductively that the Ramsey number $R^p(k_1, \ldots, k_t)$ is defined for all $k_1, \ldots, k_t \in \mathbb{N}$. We must show that the number $R^{p+1}(k_1, \ldots, k_t)$ is defined for all $k_1, \ldots, k_t \in \mathbb{N}$.

Fix $k_1, \ldots, k_t \in \mathbb{N}$, and assume inductively that the number $R^{p+1}(k'_1, \ldots, k'_t)$ is defined for all $k'_1, \ldots, k'_t \in \mathbb{N}$ such that $k'_1 + \cdots + k'_t < k_1 + \cdots + k_t$.

Suppose first that there exists some $i \in [t]$ such that $k_i = 1$. We then set $N := 1$, and we fix any finite set X such that $|X| \geq N$. Let A_i be any one-element subset of X (so, $A_i \in \binom{X}{k_i}$). Then $\binom{A_i}{p+1} = \emptyset$, and so (vacuously) c assigns color i to each element of $\binom{A_i}{p+1}$. Thus, $R^{p+1}(k_1, \ldots, k_t)$ is defined (and is, in fact, equal to 1). From now on, we assume that $k_1, \ldots, k_t \geq 2$.

To simplify notation, we set $r_i := R^{p+1}(k_1, \ldots, k_{i-1}, k_i - 1, k_{i+1}, \ldots, k_t)$ for all $i \in [t]$ (this is defined by the induction hypothesis for $k_1 + \cdots + k_t$). Further, we set $N := R^p(r_1, \ldots, r_t) + 1$ (this is defined by the induction hypothesis for p).

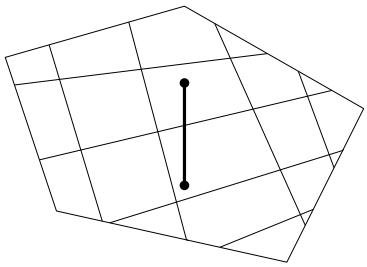
Fix a finite set X such that $|X| \geq N$, and fix a function $c : \binom{X}{p+1} \to [t]$. Set $n := |X|$; we may assume that $X = [n]$.[16] We now define an auxiliary coloring $\widetilde{c} : \binom{[n-1]}{p} \to \{t\}$, as follows: for all $A \in \binom{[n-1]}{p}$, we set $\widetilde{c}(A) = c(A \cup \{n\})$. Since $n - 1 \geq R^p(r_1, \ldots, r_t)$, we know that there exists some $i \in [t]$ and a set $X_i \in \binom{[n-1]}{r_i}$ such that $\widetilde{c}$ assigns color i to each element of $\binom{X_i}{p}$. Finally, since $|X_i| = r_i = R^{p+1}(k_1, \ldots, k_{i-1}, k_i - 1, k_{i+1}, \ldots, k_t)$, we know that there exists some $j \in [t]$ and a set $Y_j \in \binom{X_i}{k'_j}$, where $k'_j = k_j - 1$ if $j = i$ and $k'_j = k_j$ otherwise, such that c assigns color j to each element of $\binom{Y_j}{p+1}$. If $j \neq i$, then we set $A_j = Y_j$, and we observe that $A_j \in \binom{[n]}{k_j}$, and that (by construction) c assigns color j to each element of $\binom{A_j}{p+1}$. Suppose now that $j = i$. Then we set $A_i = Y_i \cup \{n\}$. Once again by construction, we have that $|A_i| = k_i$, and that c assigns color i to each element of $\binom{A_i}{p+1}$.[17] This proves that $R^{p+1}(k_1, \ldots, k_t)$ is defined. □

We now consider a geometric application (see the Erdős-Szekeres theorem

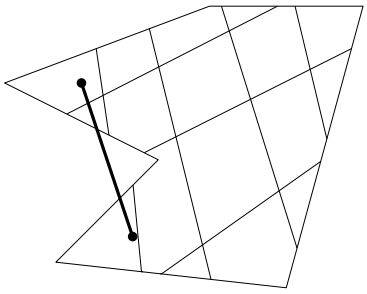
[16] If not, we simply rename the elements of X (via a bijection).

[17] Indeed, fix any $A \in \binom{A_i}{p+1}$. If $n \notin A$, then $A \in \binom{Y_i}{p+1}$, and so $c(A) = i$. On the other hand, if $n \in A$, then $A \setminus \{n\} \in \binom{X_i}{p}$, and we see that $c(A) = \widetilde{c}(A \setminus \{n\}) = i$.

below). A set X of points in the plane is *convex* if for all distinct $x_1, x_2 \in X$, the line segment between x_1 and x_2 lies in X.



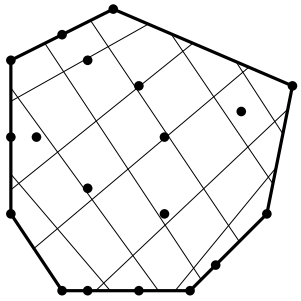
convex



non-convex

The *convex hull* of a non-empty set S of points in the plane is the smallest convex set in the plane that includes S. If S is a non-empty, finite set of points, then the convex hull of S is either a one-point set, a line interval, or a convex polygon (with its interior).

If S is a finite set of points in the plane containing at least three non-collinear points,[18] then the convex hull of S is a convex polygon (with its interior), and the vertices of this polygon are all in S;[19] see the picture below for an example.



Let us say that (pairwise distinct) points $x_1, \ldots, x_t$ ($t \geq 3$) in the plane are in *convex position* if they are the vertices of some convex polygon. Equivalently, (pairwise distinct) points $x_1, \ldots, x_t$ ($t \geq 3$) are in convex position if their convex hull is a convex t-gon whose vertices are precisely $x_1, \ldots, x_t$ (not necessarily in that order).

We now need a geometric lemma.

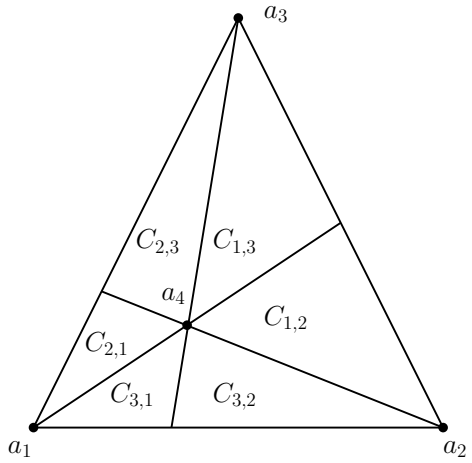
[18] Three or more points are *collinear* if they lie on the same line.

[19] However, not every element of S need be a vertex of the polygon.

Lemma 7.3.1. *Any set of five points in the plane, no three of which are collinear, contains four points in convex position.*

Proof. To simplify notation, for non-collinear points x, y, z in the plane, we denote by Δxyz the triangle with vertices x, y, z.

Let $a_1, \ldots, a_5$ be five point in the plane, no three of which are collinear. We now consider the convex hull of these five points. Since no three of these points are collinear, their convex hull is a convex polygon, and each vertex of the polygon is one of $a_1, \ldots, a_5$.[20] If the polygon is a pentagon, then clearly, any four of our five points are in convex position. If the polygon is a quadrilateral, then its vertices (which are some four of $a_1, \ldots, a_5$) are in convex position. So assume that the polygon is a triangle. By symmetry, we may assume that the vertices of this triangle are a_1, a_2, a_3. Since no three points of $a_1, \ldots, a_5$ are collinear, we see that a_4, a_5 both lie in the interior (and not on any edge) of the triangle $\Delta a_1 a_2 a_3$. Using the fact that a_4 is in the interior of $\Delta a_1 a_2 a_3$, we construct six regions in the interior of $\Delta a_1 a_2 a_3$, as in the picture below (the regions $C_{i,j}$ are the interiors if the triangles in the picture, and in particular, they are disjoint from the lines represented in the picture).



Since no three of $a_1, \ldots, a_5$ are collinear, we see that $a_5 \in C_{1,2} \cup C_{1,3} \cup C_{2,1} \cup C_{2,3} \cup C_{3,1} \cup C_{3,2}$. Now, fix $i, j \in \{1, 2, 3\}$ with $i \neq j$ such that $a_5 \in C_{i,j}$. Then a_i, a_4, a_5, a_j are the vertices of a convex quadrilateral, and we are done. □

The Erdős-Szekeres theorem. *Let $t \geq 4$ be an integer. Any set of at least $R^4(5, t)$ points in the plane, no three of which are collinear, contains t points*

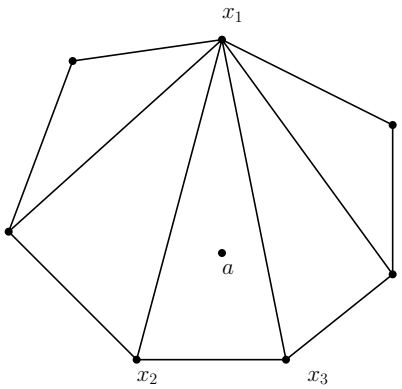
[20] However, not all of $a_1, \ldots, a_5$ need be vertices of the polygon.

in convex position.

Proof. We consider a set S of at least $R^4(5,t)$ points in the plane, and we assume that no three of these points are collinear. We now consider a coloring $c : \binom{S}{4} \to [2]$ defined as follows: for all $X \in \binom{S}{4}$, $c(X) = 1$ if the four points of X are **not** in convex position, and $c(X) = 2$ if they are in convex position. Since $|S| \geq R^4(5,t)$, we know that either there exists some $A_1 \in \binom{S}{5}$ such that c assigns color 1 to all elements of $\binom{A_1}{4}$, or there exists some $A_2 \in \binom{S}{t}$ such that c assigns color 2 to all elements of $\binom{A_2}{4}$.

Suppose that there exists some $A_1 \in \binom{S}{5}$ such that c assigns color 1 to all elements of $\binom{A_1}{4}$. Then A_1 is a set of five points in the plane, no three of which are collinear, and no four of which are in convex position. But this contradicts Lemma 7.3.1.

It now follows that there exists some $A_2 \in \binom{S}{t}$ such that c assigns color 2 to all elements of $\binom{A_2}{4}$. Then A_2 is a set of t points in the plane, no three of which are collinear, and any four of which are in convex position. Let us show that the points in A_2 are in fact in convex position. We consider the convex hull of A_2; this convex hull is a convex polygon, and we let X_2 be the set of vertices of this polygon. Clearly, $X_2 \subseteq A_2$. If $X_2 = A_2$, then we are done. So assume that $X_2 \subsetneq A_2$. Then all points in $X_2 \setminus A_2$ are in the interior of our polygon.[21] We now choose any $a \in A_2 \setminus X_2$. Clearly, there exist three (pairwise distinct) points $x_1, x_2, x_3 \in X_2$ such that a is in the interior of the triangle $\Delta x_1 x_2 x_3$.[22]



But then a, x_1, x_2, x_3 are not in convex position, contrary to the fact that $c(\{a, x_1, x_2, x_3\}) = 2$ (since $\{a, x_1, x_2, x_3\} \in \binom{A_2}{4}$). □

[21] Since no three points in A_2 are collinear, no point of $X_2 \setminus A_2$ is on an edge of the polygon.

[22] Once again, we are using the fact that no three of our points are collinear.

7.4 Ramsey's theorem (infinite version)

For a function $c : A \to B$ and a set $A' \subseteq A$, we denote by $c \restriction A'$ the restriction of c to A'.[23]

Ramsey's theorem (infinite version). *For all $t, p \in \mathbb{N}$, all infinite sets X, and all colorings $c : \binom{X}{p} \to [t]$, there exists an infinite set $A \subseteq X$ such that $c \restriction \binom{A}{p}$ is constant.*[24]

Proof. We fix $t \in \mathbb{N}$, and we proceed by induction on p.

For $p = 1$, we fix an infinite set X and a coloring $c : \binom{X}{1} \to [t]$. For all $i \in [t]$, we set $C_i = \{x \in X \mid c(\{x\}) = i\}$. Then $(C_1, \dots, C_t)$ is a partition of the infinite set X, and consequently, at least one of the sets $C_1, \dots, C_t$, say C_i, is infinite. Furthermore, $c \restriction \binom{C_i}{1}$ is constant (indeed, it assigns color i to each element of $\binom{C_i}{1}$). So, the theorem is true for $p = 1$.

Now, fix $p \in \mathbb{N}$, and assume the theorem is true for p.[25] We must show that it is true for $p+1$. Fix an infinite set X and a coloring $c : \binom{X}{p+1} \to [t]$. Our goal is to recursively construct a sequence $\{X_n\}_{n=1}^{\infty}$ of infinite subsets of X and a sequence $\{x_n\}_{n=1}^{\infty}$ of elements of X with the following three properties:

- $x_n \in X_n$ for all $n \in \mathbb{N}$;
- $X_{n+1} \subseteq X_n \setminus \{x_n\}$ for all $n \in \mathbb{N}$;
- for all $n \in \mathbb{N}$, c assigns the same color to all sets of the form $\{x_n\} \cup Y$, with $Y \in \binom{X_{n+1}}{p}$.

First, we set $X_1 = X$ and we choose $x_1 \in X$ arbitrarily. Now, having constructed $X_1, \dots, X_n$ and $x_1, \dots, x_n$, we construct X_{n+1} and x_{n+1} as follows. We define an auxiliary coloring $c_n : \binom{X_n \setminus \{x_n\}}{p} \to [t]$ by setting $c_n(A) = c(A \cup \{x_n\})$ for all $A \in \binom{X_n \setminus \{x_n\}}{p}$.[26] Since $X_n \setminus \{x_n\}$ is infinite, the induction hypothesis guarantees that there exists some infinite set $X_{n+1} \subseteq X_n \setminus \{x_n\}$ such that $c_n \restriction \binom{X_{n+1}}{p}$ is constant. But now by construction, we have that c assigns the same color to all sets of the form $\{x_n\} \cup Y$, with $Y \in \binom{X_{n+1}}{p}$. Finally, we choose $x_{n+1} \in X_{n+1}$ arbitrarily.

[23] So, $c \restriction A'$ is a function from A' to B, and for all $a \in A'$, we have $(c \restriction A')(a) = c(a)$.

[24] This means that c assigns the same color to all p-element subsets of A.

[25] So, we are assuming that for all infinite sets X, and all colorings $c : \binom{X}{p} \to [t]$, there exists an infinite set $A \subseteq X$ such that $c \restriction \binom{A}{p}$ is constant.

[26] Note that if $A \in \binom{X_n \setminus \{x_n\}}{p}$, then $A \cup \{x_n\} \in \binom{X_n}{p+1} \subseteq \binom{X}{p+1}$, and so $c(A \cup \{x_n\})$ is defined.

We have now constructed our sequences $\{X_n\}_{n=1}^{\infty}$ and $\{x_n\}_{n=1}^{\infty}$. It follows from the construction that for all $n \in \mathbb{N}$, the coloring c assigns the same color to all sets of the form $\{x_n\} \cup \{x_{j_1}, \dots, x_{j_p}\}$, with $n < j_1 < \cdots < j_p$; let us say this color is *associated* with x_n. Now, for all $i \in [t]$, we let $A_i = \{x_n \mid n \in \mathbb{N}, i \text{ is associated with } x_n\}$. Then $(A_1, \dots, A_t)$ is a partition of the infinite set $\{x_1, x_2, x_3, \dots\}$, and we deduce that at least one of the sets $A_1, \dots, A_t$, say A_i, is infinite. But now $c \restriction \binom{A_i}{p+1}$ is constant (it assigns i to all elements of $\binom{A_i}{p+1}$). This completes the induction. □

Note that, to form the sequence $\{x_n\}_{n=1}^{\infty}$ in the proof that we just completed, we made infinitely many "arbitrary choices" (indeed, each x_n was chosen arbitrarily from some specified infinite set). So, we implicitly used the "Axiom of Choice," which allows us to make infinitely many arbitrary choices in this way. It is actually possible to avoid the use of the Axiom of Choice in the proof above, but then the proof would be slightly messier,[27] and we omit the details.

7.5 Kőnig's infinity lemma

An infinite graph (i.e. graph with an infinite vertex set) is *locally finite* if each vertex has finite degree. As in the case of finite graphs, an infinite graph is *connected* if there is a path[28] between any two vertices. An infinite graph is a *forest* if it contains no cycles,[29] and it is a *tree* if it is a connected forest. An *infinite rooted tree* is an ordered pair (T, r) such that T is an infinite tree, and r is some vertex of T, called the *root*.

A *ray* in an infinite graph G is a sequence $x_0, x_1, x_2, x_3, \dots$ of pairwise distinct vertices such that for all integers $n \geq 0$, $x_n x_{n+1}$ is an edge of G.

Kőnig's infinity lemma. *Every infinite, locally finite rooted tree (T, r) contains a ray starting at r (i.e. a ray of the form $r, x_1, x_2, \dots$).*

Proof (outline). Since the tree T is infinite and connected, there are infinitely many paths in T with one endpoint r. Since r has only finitely many neighbors, infinitely many of these paths have the second vertex (say, x_1) in common as well. Since x_1 has only finitely many neighbors, among the infinitely many paths starting with r, x_1, infinitely many have the third

[27] Essentially, we would start with an injection $f : \mathbb{N} \to X$, and then work with $f[\mathbb{N}]$ instead of X. Then, instead of making an arbitary choice, we could choose the $x_n \in X_n$ whose pre-image (via f) is minimum.

[28] The path is supposed to be finite.

[29] Again, cycles are finite.

vertex (say, x_2) in common. We proceed like this, and we obtain an infinite sequence $r, x_1, x_2, x_3, \ldots$. But now $r, x_1, x_2, x_3, \ldots$ is a ray starting at r. $\square$

We remark that the proof of Kőnig's infinity lemma also uses the Axiom of Choice (because at the n-th step, there may be more than one possible choice for x_n, and if so, we choose arbitrarily).

The infinite version of Ramsey's theorem and Kőnig's infinity lemma together imply the hypergraph version of Ramsey's theorem, as we now show.

Ramsey's theorem (hypergraph version). *For all $p, t, k_1, \ldots, k_t \in \mathbb{N}$, the number $R^p(k_1, \ldots, k_t)$ exists.*

Proof. Clearly, it suffices to show that for all $p, t, k \in \mathbb{N}$, the Ramsey number $R^p(\underbrace{k, \ldots, k}_{t})$ exists.[30] Suppose that for some $p, t, k \in \mathbb{N}$, the number $R^p(\underbrace{k, \ldots, k}_{t})$ does not exist. Now, for each integer $n \geq p$, we say that a coloring $c : \binom{[n]}{p} \to [t]$ is *n-bad* if there is no set $A \in \binom{[n]}{k}$ such that $c \restriction \binom{A}{p}$ is constant; a coloring is *bad* if it is n-bad for some integer $n \geq p$. Since $R^p(\underbrace{k, \ldots, k}_{t})$ does not exist, we see that for all integers $n \geq p$, there is at least one n-bad coloring.[31]

Now, let C be the set of all bad colorings, and let T be the graph on the vertex set $C \cup \{r\}$ (where $r \notin C$),[32] with adjacency as follows:

- r is adjacent to all p-bad colorings, and to no other elements of C;
- for all integers $n \geq p$, n-bad colorings are pairwise non-adjacent;
- for all integers $n \geq p$, an n-bad coloring c_n is adjacent to an $(n+1)$-bad coloring c_{n+1} if and only if c_{n+1} is an extension of c_n;[33]
- for all integers $n_1, n_2 \geq p$ such that $|n_1 - n_2| \geq 2$, no n_1-bad coloring is adjacent to any n_2-bad coloring.

[30] Indeed, fix $p, t, k_1, \ldots, k_t \in \mathbb{N}$, and set $k = \max\{k_1, \ldots, k_t\}$. If $R^p(\underbrace{k, \ldots, k}_{t})$ exists, then so does $R^p(k_1, \ldots, k_t)$, and in fact, we have that $R^p(k_1, \ldots, k_t) \leq R^p(\underbrace{k, \ldots, k}_{t})$. (Details?)

[31] Details?

[32] Here, r is simply an artificially added root, which we need in order to make a rooted tree.

[33] This means that $c_{n+1} \restriction \binom{[n]}{p} = c_n$.

Now (T, r) is a rooted tree. Furthermore, for each integer $n \geq p$, the number of n-bad colorings is finite,[34] and so it follows from the construction of T that the T is locally finite. So, by Kőnig's infinity lemma, there is a ray $r, c_p, c_{p+1}, c_{p+2}, \ldots$ in T. Set $c = \bigcup_{n=p}^{\infty} c_n$; then $c : \binom{\mathbb{N}}{p} \to [t]$,[35] and so by the infinite version of Ramsey's theorem, there is an infinite set A such that $c \restriction \binom{A}{p}$ is constant. We now choose any subset $A_k \in \binom{A}{k}$, and we observe that $c \restriction \binom{A_k}{p}$ is constant. Now, A_k is a finite subset of $\mathbb{N}$, and consequently, there exists some $n \in \mathbb{N}$ such that $A_k \subseteq [n]$; we may assume that $n \geq p$.[36] Now $A_k \in \binom{[n]}{k}$, and $c_n \restriction \binom{A_k}{p} = c \restriction \binom{A_k}{p}$ is constant, contrary to the fact that c_n is bad. □

[34] In fact, the number of colorings $c : \binom{[n]}{p} \to [t]$ is finite.

[35] We are using the fact that each coloring in the sequence $c_p, c_{p+1}, c_{p+2}, \ldots$ extends the previous one, and so the union of this sequence is a function (coloring).

[36] Otherwise, we have that $A_k \subseteq [p]$, and we consider p instead of n.

Chapter 8

Error correcting codes

8.1 A motivating example

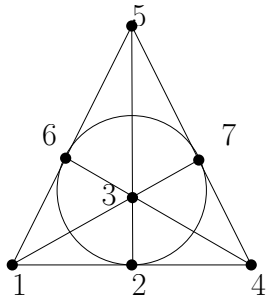
Let us suppose a sender wishes to send a message (say, a sequence of 1's and 0's) to a receiver. If the communication channel is unreliable or noisy, the message may get corrupted. For instance, the sender may send 1011, and the receiver may receive 1001.[1] In this case, the receiver has no chance of spotting and fixing the error.

One way to address this problem might be to agree to triple each bit (i.e. each 1 or 0); so, instead of 1011, we would send 111000111111. Suppose just one error occurred, and the receiver received 111000110111. Because the receiver knows he was supposed to get a sequence of tripled 1's and 0's, he can confidently say that there was an error in the boxed triple: 111000 110 111. The receiver knows that the boxed triple should have been either 000 or 111, and the latter (i.e. 111) is more likely because it is more likely that only one error occurred than that two errors did. So, the receiver guesses that the message sent was 111000111111, which corresponds to 1011. On the other hand, if more than one error occurs in a triple corresponding to one bit, then the receiver will either fail to detect the error or will correct it incorrectly. For instance, if the receiver receives 111000100111, then he will incorrectly guess that the sender sent 111000000111, which corresponds to 1001.

Here is another way to address the same problem. Consider the Fano plane, represented below.[2]

[1] Here, errors are shown in red, to facilitate reading. However, the receiver does not see this: he simply receives a string of 1's and 0's, uncolored.

[2] We saw the Fano plane in chapter 3. Here, points are relabeled (relative to what we had in chapter 3), and the names of lines are omitted. We still have seven lines, represented by the six line segments and one circle. (Each line has exactly three points.)



We now form 16 row vectors of length seven as follows: we take all possible incidence vectors of lines of the Fano plane,[3] the incidence vectors of the complements of the lines of the Fano plane,[4] plus the vectors $(0,0,0,0,0,0,0)$ and $(1,1,1,1,1,1,1)$. Let $\mathcal{H}$ be the set of these 16 vectors. Now, these vectors have the following two properties:

- any two distinct vectors in $\mathcal{H}$ differ in at least three places/coordinates;
- for any vector $\mathbf{w}$ of 1's and 0's of length 7, there exists a unique vector $\mathbf{h} \in \mathcal{H}$ such that $\mathbf{w}$ and $\mathbf{h}$ differ in at most one place/coordinate.

This means that if a sender sends a vector from $\mathcal{H}$, and at most one error is made during transmission, the receiver can correctly guess which vector was sent. Indeed, the receiver simply chooses the unique vector from $\mathcal{H}$ that differs in at most one coordinate from the vector that the receiver received.

How do we use $\mathcal{H}$? First, note that there are precisely 16 strings of 1's and 0's of length four (indeed, these are simply the integers $0, 1, \ldots, 15$ written in binary code). So, we can set up a bijection π between the set of these 16 strings and the set $\mathcal{H}$. Now, suppose we wish to transmit a string of 1's and 0's of length $4n$, for some positive integer n. We divide such a string into n consecutive blocks of length four, and instead of sending these blocks, we send (consecutively) the n vectors from $\mathcal{H}$ that correspond to them. The advantage of this is that if, during transmission, at most one error is made in each vector, the receiver will be able to spot it and correct it, and then to read off (using π^{-1}) the sender's original $4n$-bit message.

Note that, if we use $\mathcal{H}$, then instead of sending $4n$ bits (the number of bits in our original message), we send $7n$ bits. If data is expensive, then this is clearly an improvement over tripling each bit (where we would send $3n$ bits for each n-bit message). We remark that $\mathcal{H}$ is a type of "Hamming

[3]For example, the incidence vector of the line $\{1,2,4\}$ is $(1,1,0,1,0,0,0)$.

[4]For example, the incidence vector of the complement of the line $\{1,2,4\}$ is $(0,0,1,0,1,1,1)$.

code," sometimes called the *Hamming(7,4) code* (because the original 4 bits are converted into 7 bits).

8.2 Basic notions

An *alphabet* is some finite set of symbols $\Sigma = \{s_0, \ldots, s_m\}$. Often, our alphabet is the finite field $\mathbb{F}_q$, where q is a prime power;[5] particularly often, our alphabet is $\mathbb{F}_2 = \mathbb{Z}_2$, which is simply the binary code (and we can perform addition and multiplication modulo 2). A *word* of length n is a string (or row vector) of length n of symbols from our alphabet; Σ^n is the set of all words of length n using symbols from the alphabet Σ. A *code* is a subset C of Σ^n.[6] Elements of the code are *codewords*. Given words $\mathbf{x} = x_1 \ldots x_n$ and $\mathbf{y} = y_1 \ldots y_n$ in Σ^n,[7] the *Hamming distance* between $\mathbf{x}$ and $\mathbf{y}$, denoted by $d(\mathbf{x}, \mathbf{y})$, is the number of places in which $\mathbf{x}$ and $\mathbf{y}$ differ, i.e. $d(\mathbf{x}, \mathbf{y}) = |\{i \in \{1, \ldots, n\} \mid x_i \neq y_i\}|$. It is straightforward to check that the Hamming distance $d(\cdot, \cdot)$ is a "metric" on Σ^n, that is, that is satisfies the following three properties:[8]

- $d(x, y) = 0 \Leftrightarrow x = y$;
- $d(x, y) = d(y, x)$;
- $d(x, z) \leq d(x, y) + d(y, z)$.

The inequality from the third bullet point is referred to as the *triangle inequality*.

Codes are used as follows. A sender would like to send a message to a receiver, and for this, he uses some code $C \subseteq \Sigma^n$, where Σ is some alphabet. There is a bijection π (known to both the sender and the receiver) between all possible messages and the code C. Now, the sender encodes his message (i.e. turns it into a codeword via the bijection) and sends it to the receiver. The

[5] Recall that, for a positive integer q, there is a field of size q if and only if q is a prime power (i.e. $q = p^n$, where p is a prime number and n is a positive integer). Furthermore, all finite fields of the same size are isomorphic. If q a prime power, then $\mathbb{F}_q$ is the unique (up to isomorphism) field of size q. Note that if p is a prime number, then $\mathbb{F}_p = \mathbb{Z}_p$ (but this is only true if p is prime!).

[6] So, in the opening example from section 8.1, we have $\Sigma = \mathbb{F}_2$, $n = 12$ (the original message had four bits, and so after we tripled each bit, we got 12 bits), and $C = \{w_1 \ldots w_{12} \in \Sigma^{12} \mid w_{3k-2} = w_{3k-1} = w_{3k} \ \forall k \in \{1, 2, 3, 4\}\}$.

[7] Here, we treat a string or length n and a row vector of length n as interchangeable. We use one or the other depending on convenience.

[8] Check this!

receiver receives this codeword, but possibly with some errors. (If the sender sends the codeword x and the receiver receives the word $\widetilde{x}$, then $d(x, \widetilde{x})$ is the number of errors created during transmission.) The receiver corrects the errors (this is possible if the number of errors is small enough, where "small enough" depends on the code used), and then recovers the original message using π^{-1}.

In general, there are two competing goals for codes. On the one hand, we wish to send as many different messages as possible, using as few bits as possible. On the other hand, we wish to maximize the number of errors that we can successfully correct.

Now, suppose Σ is an alphabet of size at least two, and $C \subseteq \Sigma^n$ is a code containing at least two codewords. Here are some parameters for the code C:

- the codeword *length* is n;
- the *size* of the alphabet is $q = |\Sigma|$;
- the *dimension* of C is $|C|$, instead of which we often consider the logarithm $k = \log_q |C|$;
- the *minimum distance* in C is $d = \min\{d(x, y) \mid x, y \in C, x \neq y\}$.

A code with these parameters is an $(n, k, d)_q$*-code*. Note that if at most $\lfloor \frac{d-1}{2} \rfloor$ errors are made during the transmission of a codeword, then the receiver can correctly spot and correct the errors by selecting the (unique) codeword with minimum Hamming distance from the word that he received.

8.2.1 Some simple codes

The simplest code is the *total code* Σ^n, where Σ is an alphabet of size $q = |\Sigma| \geq 2$, and n is a positive integer. The total code Σ^n is an $(n, n, 1)_q$ code.[9] If we use this code, we send little data, but we cannot correct even a single error!

The *repetition code* Rep_n of length n over the alphabet Σ (with $q = |\Sigma| \geq 2$) is the code $C = \{\underbrace{x \ldots x}_{n} \mid x \in \Sigma\}$. It is an $(n, 1, n)_q$-code.[10] This code allows us to correct as many as $\lfloor \frac{n-1}{2} \rfloor$ errors, but it uses a lot of data.

[9] Indeed, the size of the alphabet is q, the codeword length is n, and $k = \log_q |\Sigma^n| = \log_q q^n = n$. The minimum distance is $\Delta(\Sigma^n) = 1$ (indeed, recall that $|\Sigma| \geq 2$, and take two symbols $s_1, s_2 \in \Sigma$; then the distance between $s_1 \underbrace{s_1 \ldots s_1}_{n-1}$ and $s_2 \underbrace{s_1 \ldots s_1}_{n-1}$ is 1).

[10] Indeed, the size of the alphabet is q, and the codeword length is n. Further, $|C| = |\Sigma| = q$, and so $k = \log_q |C| = \log_q q = 1$. Finally, the distance between any two distinct words is precisely n.

Another simple example is the *parity code* C of length n (with $n \geq 2$) over the alphabet $\mathbb{F}_2$; it consists of all words of the form $w_1 \dots w_n$ with $w_1, \dots, w_n \in \mathbb{F}_2$ and $\sum_{i=1}^{n} w_i = 0$. Let us check that this is an $(n, n-1, 2)_2$-code. Obviously, the codeword length is n and the size of the alphabet is $q = 2$. Next, $|C| = 2^{n-1}$; this is because the first $n-1$ symbols of a codeword can be chosen arbitrarily (and there are 2^{n-1} ways of doing this), but the n-th symbol is uniquely determined by the previous $n-1$ ones (because the sum must be 0). So, $k = \log_q |C| = \log_2 2^{n-1} = n-1$. Finally, it is obvious that two different words cannot have distance 1, for otherwise, the sum of symbols in one of them would be 1, a contradiction. On the other hand, both $\underbrace{0 \dots 0}_{n-2}00$ and $\underbrace{0 \dots 0}_{n-2}11$ are in our code, and the distance between them is 2. So, the minimum distance in our code is $d = 2$.

8.2.2 The Hadamard code

Given vectors $\mathbf{a} = (a_1, \dots, a_n)^T$ and $\mathbf{b} = (b_1, \dots, b_n)^T$ in $\mathbb{R}^n$, the *standard inner product* (or *dot product*) of $\mathbf{a}$ and $\mathbf{b}$ is $\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i$. Two vectors in $\mathbb{R}^n$ are *orthogonal* with respect to the dot product if their dot product is zero.

A *Hadamard matrix* of order n is an $n \times n$ matrix whose entries are all 1 or -1, and whose columns are pairwise orthogonal (with respect to the dot product). For example, the matrix

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

is Hadamard matrix of order 2. Furthermore, if H is an $n \times n$ Hadamard matrix, then

$$\begin{bmatrix} H & H \\ H & -H \end{bmatrix}$$

is a Hadamard matrix of order $2n$.[11]

Proposition 8.2.1. *Let H be a Hadamard matrix of order n. Then $HH^T = nI_n$.*[12] *Furthermore, H^T is also a Hadamard matrix of order n.*

Proof. Let us show that $H^T H = nI_n$. To simplify notation, set $H = \begin{bmatrix} \mathbf{h}_1 & \dots & \mathbf{h}_n \end{bmatrix}$. For each $i \in \{1, \dots, n\}$, the (i,i)-th entry of $H^T H$ is $\mathbf{h}_i \cdot \mathbf{h}_i$, which is equal to n because all entries of a Hadamard matrix are ± 1.

[11] Check this!

[12] As usual, I_n is the $n \times n$ identity matrix.

On the other hand, for distinct $i, j \in \{1, \ldots, n\}$, the (i, j)-th entry of $H^T H$ is $\mathbf{h}_i \cdot \mathbf{h}_j$, which is equal to 0 since any two distinct columns of a Hadamard matrix are orthogonal. This proves that $H^T H = nI_n$.

Now, since $H^T H = nI_n$, we have that $(\frac{1}{n}H^T)H = I_n$; since $\frac{1}{n}H^T$ and H are square matrices whose product is the identity matrix, we know from Linear Algebra that $\frac{1}{n}H^T$ and H are both invertible and are each other's inverses. Consequently, $H(\frac{1}{n}H^T) = I_n$, and we deduce that $HH^T = nI_n$.

It remains to show that H^T is a Hadamard matrix. Since H is a Hadamard matrix of order n, we know that H^T is an $n \times n$ matrix, and that all entries of H^T are ± 1. It remains to show that the columns of H^T are pairwise orthogonal. To simplify notation, we set $H^T = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix}$; note that this means that $\mathbf{a}_1^T, \ldots, \mathbf{a}_n^T$ are the rows of H (listed from top to bottom). Now, fix distinct $i, j \in \{1, \ldots, n\}$. Then the (i, j)-th entry of HH^T is $\mathbf{a}_i \cdot \mathbf{a}_j$. But we we already showed that $HH^T = nI_n$, and so (since $i \neq j$) the (i, j)-th entry of HH^T is 0. Thus, $\mathbf{a}_i \cdot \mathbf{a}_j = 0$. So, the columns of H^T are pairwise orthogonal, and it follows that H^T is a Hadamard matrix. □

We now construct the Hadamard code as follows. Fix any Hadamard matrix H of order n. Then the Hadamard code associated with H consists of all rows of H and all rows of $-H$. This code has $2n$ codewords.[13] It is easy to check that this is an $(n, 1 + \log_2 n, \frac{n}{2})_2$-code.[14]

8.3 The Singleton, Hamming, and Gilbert-Varshamov bounds

For positive integers n, d, q with $n \geq d$ and $q \geq 2$, let $A_q(n, d)$ be the maximum size of a code (i.e. the maximum possible number of codewords in a code) C with the following parameters:

- the size of the alphabet is q;
- the codeword length is n;
- the minimum distance is at least d.

The Singleton bound. *For all positive integers n, d, q such that $n \geq d$ and $q \geq 2$, we have that $A_q(n, d) \leq q^{n-d+1}$.*

[13] For this, we must check that no two rows of H are the same, and that no row of H is equal to any row of $-H$. But this follows from the fact that, by Proposition 8.2.1, H^T is a Hadamard matrix (details?).

[14] Details?

Proof. We prove this by induction on n, keeping q fixed and allowing d to vary. More precisely, we fix positive integers n, d, q such that $n \geq d$ and $q \geq 2$, and we assume inductively that for all positive integers n', d' with $n' \geq d'$ and $n' < n$, we have that $A_q(n', d') \leq q^{n'-d'+1}$. We must show that $A_q(n, d) \leq q^{n-d+1}$.

Fix a code C over an alphabet Σ with $|\Sigma| = q$, and assume that the codeword length in C is n and that the minimum distance between codewords in C is at least d. We must show that $|C| \leq q^{n-d+1}$. If $d = 1$, then

$$|C| \quad \leq \quad |\Sigma^n| \quad = \quad q^n \quad = \quad q^{n-d+1},$$

and we are done. So, from now on, we assume that $d \geq 2$. This implies that $n - d + 1 < n$; we will apply the induction hypothesis to $n - d + 1$.

We construct the code $\widetilde{C} \subseteq \Sigma^{n-d+1}$ as follows: $\widetilde{C}$ is the set of all words $w_1 \dots w_{n-d+1}$ in Σ^{n-d+1} for which there exist some $w_{n-d+2}, \dots, w_n \in \Sigma$ such that $w_1 \dots w_{n-d+1} w_{n-d+2} \dots w_n \in C$.[15] Let us check that $|\widetilde{C}| = |C|$. We define the function $f : C \to \widetilde{C}$ by setting $f(w_1 \dots w_{n-d+1} w_{n-d+2} \dots w_n) = w_1 \dots w_{n-d+1}$ for all $w_1 \dots w_{n-d+1} w_{n-d+2} \dots w_n \in C$; we will show that f is a bijection. By the construction of $\widetilde{C}$ and f, we have that f is onto $\widetilde{C}$. Now, fix codewords $\mathbf{w} = w_1 \dots w_n$ and $\mathbf{w}' = w'_1 \dots w'_n$ in C such that $f(\mathbf{w}) = f(\mathbf{w}')$; then $w_1 \dots w_{n-d+1} = w'_1 \dots w'_{n-d+1}$, and it follows that the Hamming distance between $\mathbf{w}$ and $\mathbf{w}'$ is at most $d-1$.[16] Since the minimum distance in C is at least d, we conclude that $\mathbf{w} = \mathbf{w}'$, and it follows that f is one-to-one. Thus, $f : C \to \widetilde{C}$ is a bijection, and we deduce that $|\widetilde{C}| = |C|$.

Now, $\widetilde{C}$ is a code over Σ, with $|\Sigma| = q$, the length of codewords in $\widetilde{C}$ is $n - d + 1 < n$,[17] and obviously, the minimum distance in $\widetilde{C}$ is at least 1. So, by the induction hypothesis, we have that

$$|\widetilde{C}| \quad \leq \quad A_q(n - d + 1, 1) \quad \leq \quad q^{(n-d+1)-1+1} \quad = \quad q^{n-d+1}.$$

Since $|\widetilde{C}| = |C|$, we deduce that $|C| \leq q^{n-d+1}$, which is what we needed to show. □

We now need some notation. Suppose n, t, q are positive integers and Σ is an alphabet of size q. For all $\mathbf{w} \in \Sigma^n$, we let $B_t^{\Sigma^n}(\mathbf{w})$ be the "combinatorial ball" of radius t around $\mathbf{w}$, i.e. $B_t^{\Sigma^n}(\mathbf{w})$ is the set of all words in Σ^n whose

[15] So, $\widetilde{C}$ is the set of all words that can be obtained by deleting the last $d-1$ symbols of a codeword in C.

[16] Indeed, $\mathbf{w}$ and $\mathbf{w}'$ are both of length n, and they coincide in their first $n - d + 1$ places. So, they differ in at most $d - 1$ places, i.e. their Hamming distance is at most $d - 1$.

[17] We are using the fact that $d \geq 2$.

Hamming distance from $\mathbf{w}$ is at most t. When no confusion is possible, we write $B_t(\mathbf{w})$ instead of $B_t^{\Sigma^n}(\mathbf{w})$.

Proposition 8.3.1. *Let n, t, q be positive integers such that $n \geq t$ and $q \geq 2$, and let Σ be an alphabet of size q. Then $|B_t(\mathbf{w})| = \sum_{k=0}^{t} \binom{n}{k}(q-1)^k$ for all $\mathbf{w} \in \Sigma^n$.*

Proof. Fix a word $\mathbf{w} \in \Sigma^n$. We must show that the number of words in Σ^n at distance at most t from $\mathbf{w}$ is precisely $\sum_{k=0}^{t} \binom{n}{k}(q-1)^k$. Clearly, it suffices to show that for all $k \in \{0, \ldots, t\}$, the number of words in Σ^n at distance k from $\mathbf{w}$ is precisely $\binom{n}{k}(q-1)^k$. So, fix $k \in \{0, \ldots, t\}$. There are $\binom{n}{k}$ ways to choose the k places in which a word at Hamming distance k from $\mathbf{w}$ differs from $\mathbf{w}$. For each such choice, and for each of the k selected placed, we have $q-1$ ways of altering $\mathbf{w}$ in that place;[18] so, for all k places together, we get $(q-1)^k$ ways of altering $\mathbf{w}$. So, there are precisely $\binom{n}{k}(q-1)^k$ words in Σ^n at distance k from $\mathbf{w}$. □

The Hamming bound. *Let n, d, q be positive integers such that $n \geq d$ and $q \geq 2$, and let $t = \lfloor \frac{d-1}{2} \rfloor$. Then $A_q(n, d) \leq \frac{q^n}{\sum_{k=0}^{t} \binom{n}{k}(q-1)^k}$.*

Proof. Fix a code $C \subseteq \Sigma^n$, where Σ is an alphabet of size q, and assume that the minimum distance between codewords in C is at least d. We must show that $|C| \leq \frac{q^n}{\sum_{k=0}^{t} \binom{n}{k}(q-1)^k}$. Set $m := |C|$ and $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_m\}$. Since the minimum Hamming distance between codewords in C is at least d, and since $t = \lfloor \frac{d-1}{2} \rfloor$, we see that the combinatorial balls $B_t(\mathbf{c}_1), \ldots, B_t(\mathbf{c}_m)$ are

[18] Indeed, we can select any symbol from Σ, except the one that appears in the selected place in the word $\mathbf{w}$ itself. Since $|\Sigma| = q$, we have $q-1$ choices.

pairwise disjoint.[19] We now compute:

$$\begin{aligned}
q^n &= |\Sigma^n| && \text{because } |\Sigma| = q \\
&\geq |\bigcup_{i=1}^{m} B_t(\mathbf{c}_i)| && \\
&= \sum_{i=1}^{m} |B_t(\mathbf{c}_i)| && \text{because } B_t(\mathbf{c}_1), \ldots, B_t(\mathbf{c}_m) \\
& && \text{are pairwise disjoint} \\
&= m \sum_{k=0}^{t} \tbinom{n}{k}(q-1)^k && \text{by Proposition 8.3.1} \\
&= |C| \sum_{k=0}^{t} \tbinom{n}{k}(q-1)^k && \text{because } m = |C|.
\end{aligned}$$

This implies that $|C| \leq \frac{q^n}{\sum_{k=0}^{t}\binom{n}{k}(q-1)^k}$, which is what we needed to show. □

The Gilbert-Varshamov bound. *Let n, d, q be positive integers such that $n \geq d$ and $q \geq 2$. Then $A_q(n,d) \geq \frac{q^n}{\sum_{k=0}^{d-1}\binom{n}{k}(q-1)^k}$.*

Proof. Fix a code $C \subseteq \Sigma^n$, where Σ is some alphabet of size q, with minimum distance between codewords in C at least d, and with $|C| = A_q(n,d)$.[20] We must show that $|C| \geq \frac{q^n}{\sum_{k=0}^{d-1}\binom{n}{k}(q-1)^k}$.

Set $m = |C|$ and $C = \{\mathbf{c}_1, \ldots, \mathbf{c}_m\}$.

Claim. $\Sigma^n = \bigcup_{i=1}^{m} B_{d-1}(\mathbf{c}_i)$.

Proof of the Claim. It is clear that $\bigcup_{i=1}^{m} B_{d-1}(\mathbf{c}_i) \subseteq \Sigma^n$. Suppose that $\bigcup_{i=1}^{m} B_{d-1}(\mathbf{c}_i) \subsetneq \Sigma^n$, and fix some $\mathbf{w} \in \Sigma^n \setminus \left(\bigcup_{i=1}^{m} B_{d-1}(\mathbf{c}_i)\right)$. Then $d(\mathbf{w}, \mathbf{c}_i) \geq d$ for all $i \in \{1, \ldots, m\}$. We now form a new code $\widetilde{C} := C \cup \{\mathbf{w}\}$; obviously, $\widetilde{C} \subseteq \Sigma^n$, with $|\Sigma| = q$, and by construction, the minimum distance in $\widetilde{C}$ is at least d. But now the fact that $|\widetilde{C}| = |C| + 1 = A_q(n,d) + 1$ contradicts the definition of $A_q(n,d)$. This proves the Claim. ♦

[19] Note that we are using the triangle inequality for the Hamming distance here.

[20] Such a code C exists by the definition of $A_q(n,d)$.

We now compute:

$$\begin{aligned} q^n &= |\Sigma^n| && \text{because } |\Sigma| = q \\ &= |\bigcup_{i=1}^{m} B_{d-1}(\mathbf{c}_i)| && \text{by the Claim} \\ &\leq \sum_{i=1}^{m} |B_{d-1}(\mathbf{c}_i)| \\ &= m \sum_{k=0}^{d-1} \tbinom{n}{k}(q-1)^k && \text{by Proposition 8.3.1} \\ &= |C| \sum_{k=0}^{d-1} \tbinom{n}{k}(q-1)^k && \text{because } m = |C|. \end{aligned}$$

It follows that $|C| \geq \frac{q^n}{\sum_{k=0}^{d-1}\binom{n}{k}(q-1)^k}$, which is what we needed to show. □

8.4 Some Linear Algebra preliminaries for linear codes

In what follows, for a field $\mathbb{F}$ and a positive integer n, we denote by $\mathbb{F}^n$ the set of all row vectors of length n whose entries are all in $\mathbb{F}$. For vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ in $\mathbb{F}^n$, we define $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i$, where the summation and multiplication denote the operations from the field $\mathbb{F}$; note that $\langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{F}$. If $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, then $\mathbf{x}$ and $\mathbf{y}$ are said to be *orthogonal*.

Instead of multiplying matrices by column vectors on the right ($A\mathbf{x}$), we will multiply matrices by row vectors on the left ($\mathbf{x}A$). If A is an $n \times m$ matrix with entries in $\mathbb{F}$, and $\mathbf{x} \in \mathbb{F}^n$,[21] then we can think of $\mathbf{x}$ as a $1 \times n$ matrix, and we can compute $\mathbf{x}A$ according to the usual rules of matrix multiplication.[22] Note that if $\mathbf{x} = (x_1, \ldots, x_n)$ and $A = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_n \end{bmatrix}$ (i.e. $\mathbf{r}_1, \ldots, \mathbf{r}_n$ are the rows of A, from top to bottom), then $\mathbf{x}A = \sum_{i=1}^{n} x_i \mathbf{r}_i$. Furthermore, if $\mathbf{e}_i$ is the i-th standard basis vector of $\mathbb{F}^n$, i.e. the row vector whose i-th entry is 1, and all of whose other entries are 0, then $\mathbf{e}_i A$ is equal to the i-th row of A.

[21] So, A has n rows and m columns, and $\mathbf{x}$ is a row vector of length n.

[22] Indeed, we multiply a $1 \times n$ matrix by an $n \times m$ matrix, and we obtain a $1 \times m$ matrix, i.e. a row vector of length m.

With these adjustments, all familiar theorems of Linear Algebra still hold, but with rows and columns reversed. For instance, Gaussian elimination is performed on columns, not rows.[23] The set-up that we just described is customary in the study of linear codes (see section 8.5).

For a field $\mathbb{F}$ and a linear subspace C of $\mathbb{F}^n$, we define $C^\perp = \{\mathbf{y} \in \mathbb{F}^n \mid \langle \mathbf{x}, \mathbf{y} \rangle = 0 \text{ for all } \mathbf{x} \in C\}$. It is easy to check that $C^\perp$ is a linear subspace of $\mathbb{F}^n$.[24]

Theorem 8.4.1. *Let $\mathbb{F}$ be a field, and let C be a linear subspace of $\mathbb{F}^n$. Then $\dim C + \dim C^\perp = n$.*

Proof. Set $k := \dim C$; we must show that $\dim C^\perp = n - k$. If $k = 0$, then $C = \{\mathbf{0}\}$ and $C^\perp = \mathbb{F}^n$, and it follows that $\dim C^\perp = n = n - k$. From now on, we assume that $k \geq 1$. Let $\{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ be some basis for C, and let $G = \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_k \end{bmatrix}$. Then $C^\perp = \{\mathbf{y} \in \mathbb{F}^n \mid \mathbf{y}G^T = \mathbf{0}\} = \mathrm{Ker}(G^T)$.[25] By the Rank-nullity theorem, we have that $\mathrm{rank}(G^T) + \dim \mathrm{Ker}(G^T) = n$. But $\mathrm{rank}(G^T) = \mathrm{rank}(G) = k$ (because G has k rows, and they are linearly independent), and as we saw, $C^\perp = \mathrm{Ker}(G^T)$. It follows that $k + \dim C^\perp = n$, i.e. $\dim C^\perp = n - k$. □

Proposition 8.4.2. *Let $\mathbb{F}$ be a field, and let C be a linear subspace of $\mathbb{F}^n$. Then $(C^\perp)^\perp = C$.*

Proof. Obviously, $C \subseteq (C^\perp)^\perp$;[26] since C and $(C^\perp)^\perp$ are both linear subspaces of $\mathbb{F}^n$, it follows that C is a linear subspace of $(C^\perp)^\perp$. On the other hand, by Theorem 8.4.1, we have that

$$\dim(C^\perp)^\perp \;=\; n - \dim C^\perp \;=\; n - (n - \dim C) \;=\; \dim C,$$

and we deduce that $C = (C^\perp)^\perp$. □

[23] Alternatively, given a matrix A, we can perform Gaussian elimination as follows: we first form the transpose A^T, then we perform the familiar Gaussian elimination on rows to obtain a matrix B, and then we take the transpose of B. The result is the same as if we performed Gaussian elimination on the columns of A directly.

[24] Check this!

[25] $\mathrm{Ker}(G^T) = \{\mathbf{y} \in \mathbb{F}^n \mid \mathbf{y}G^T = \mathbf{0}\}$ is simply the definition of $\mathrm{Ker}(G^T)$.

[26] Indeed, every vector in C is orthogonal to every vector in $C^\perp$. On the other hand, $(C^\perp)^\perp$ is the set of all vectors in $\mathbb{F}$ that are orthogonal to every vector in $C^\perp$. It follows that $C \subseteq (C^\perp)^\perp$.

8.5 Linear codes

A *linear code* is a linear subspace C of a vector space $\mathbb{F}_q^n$, where $\mathbb{F}_q$ is a finite field of size q (here, q is a prime power).[27] Note that every linear code contains the zero vector.

Notationally, if a linear code C is an $(n, k, d)_q$-code, then we write that C is an $[n, k, d]_q$-code (here, square brackets indicate that C is a linear code). Clearly, an $[n, k, d]_q$-code is a linear subspace of $\mathbb{F}_q^n$.[28] Furthermore, as our next proposition shows, the (vector space) dimension of an $[n, k, d]_q$-code is k.

Proposition 8.5.1. *Let C be an $[n, k, d]_q$-code. Then $\dim C = k$, i.e. the dimension of C as a vector space is k.*

Proof. Since C is an $[n, k, d]_q$-code, we know that C is a linear subspace of $\mathbb{F}_q^n$; set $\ell := \dim C$. We must show that $\ell = k$. Let $\{\mathbf{c}_1, \ldots, \mathbf{c}_\ell\}$ be a basis for C. Then C is the set of all vectors of the form $\sum_{i=1}^{\ell} \alpha_i \mathbf{c}_i$, where $\alpha_1, \ldots, \alpha_\ell \in \mathbb{F}_q$. There are q choices for each α_i,[29] and so there are q^ℓ choices for the ℓ-tuple $(\alpha_1, \ldots, \alpha_\ell)$. On the other hand, since $\{\mathbf{c}_1, \ldots, \mathbf{c}_\ell\}$ is linearly independent (because it is a basis), we know that $\sum_{i=1}^{\ell} \alpha_i \mathbf{c}_i = \sum_{i=1}^{\ell} \beta_i \mathbf{c}_i$ (where $\alpha_1, \ldots, \alpha_\ell, \beta_1, \ldots, \beta_\ell \in \mathbb{F}_q$) if and only if $(\alpha_1, \ldots, \alpha_\ell) = (\beta_1, \ldots, \beta_\ell)$. It follows that $|C| = q^\ell$, and consequently, $\ell = \log_q q^\ell = \log_q |C| = k$, which is what we needed to show. □

Now, suppose that $C \subseteq \mathbb{F}_q^n$ is an $[n, k, d]_q$-code, with $0 < k < n$. By Proposition 8.5.1, we have that $\dim C = k$, and so C is a non-trivial proper linear subspace of $\mathbb{F}_q^n$. Let G be any matrix whose rows form a basis for C (in particular, $G \in \mathbb{F}_q^{k \times n}$); then G is called the *generator matrix* of the linear code C. Note that this implies that $C^\perp = \{\mathbf{y} \in \mathbb{F}_q^n \mid \mathbf{y}G^T = \mathbf{0}\}$. Next, suppose H is any matrix such that the rows of H^T form a basis for $C^\perp$ (so, H^T is a generator matrix for $C^\perp$). The matrix H is called a *parity check matrix* for C, and by Proposition 8.4.2, it satisfies $C = \{\mathbf{x} \in \mathbb{F}_q^n \mid \mathbf{x}H = \mathbf{0}\}$,[30] i.e. $C = \mathrm{Ker}(H)$. Note that the parity check matrix H can be used to check

[27] So, elements of $\mathbb{F}_q^n$ are row vectors of length n, all of whose entries are in the field $\mathbb{F}_q$.

[28] This is because the alphabet over which C is a code must be of size q, and since C is a linear code, it is a linear subspace of $\mathbb{F}^n$, where $\mathbb{F}$ is some finite field. So, $\mathbb{F}$ is a field of size q, and so it is equal (technically, isomorphic) to $\mathbb{F}_q$ (because all finite fields of the same size are isomorphic).

[29] This is because $|\mathbb{F}_q| = q$.

[30] Let us check this. Clearly, $(C^\perp)^\perp = \{\mathbf{x} \in \mathbb{F}_q^n \mid \mathbf{x}(H^T)^T = \mathbf{0}\} = \{\mathbf{x} \in \mathbb{F}_q^n \mid \mathbf{x}H = \mathbf{0}\}$. Since $(C^\perp)^\perp = C$ (by Proposition 8.4.2), it follows that $C = \{\mathbf{x} \in \mathbb{F}_q^n \mid \mathbf{x}H = \mathbf{0}\}$.

whether a vector $\mathbf{x} \in \mathbb{F}_q^n$ is a codeword of C. Indeed, if $\mathbf{x}H = \mathbf{0}$, then $\mathbf{x} \in C$, and otherwise, $\mathbf{x} \notin C$. Note that, given a generator matrix for C, one can easily compute a parity check matrix for C, and vice versa.

Given a vector $\mathbf{x} \in \mathbb{F}_q^n$, the *Hamming weight* of $\mathbf{x}$, denoted by $\mathrm{wt}(\mathbf{x})$, is the number of non-zero coordinates in $\mathbf{x}$.

Proposition 8.5.2. *Let $C \subsetneq \mathbb{F}_q^n$ be an $[n, k, d]_q$-code, with $0 < k < n$. Then $d = \min\{wt(\mathbf{x}) \mid \mathbf{x} \in C, \mathbf{x} \neq \mathbf{0}\}$.*

Proof. Fix $\mathbf{x} \in C \setminus \{\mathbf{0}\}$ with minimum Hamming weight. We must show that $d = \mathrm{wt}(\mathbf{x})$.

First, since C is a linear code, we know that $\mathbf{0} \in C$, and so (since $\mathbf{x}$ and $\mathbf{0}$ are distinct codewords in C) we have that $d(\mathbf{x}, \mathbf{0}) \geq d$. But obviously, $d(\mathbf{x}, \mathbf{0}) = \mathrm{wt}(\mathbf{x})$, and it follows that $\mathrm{wt}(\mathbf{x}) \geq d$.

It remains to show that $\mathrm{wt}(\mathbf{x}) \leq d$. Fix distinct $\mathbf{y}, \mathbf{z} \in C$ such that $d(\mathbf{y}, \mathbf{z}) = d$.[31] Since C is a vector space, we know that $\mathbf{y} - \mathbf{z} \in C$, and so by the choice of $\mathbf{x}$, we have that $\mathrm{wt}(\mathbf{x}) \leq \mathrm{wt}(\mathbf{y} - \mathbf{z})$.[32] But now

$$d \;=\; d(\mathbf{y}, \mathbf{z}) \;=\; \mathrm{wt}(\mathbf{y} - \mathbf{z}) \;\geq\; \mathrm{wt}(\mathbf{x}),$$

which is what we needed to show. □

8.6 Hamming codes

Fix an integer $\ell \geq 2$, and set $n = 2^\ell - 1$, $k = 2^\ell - \ell - 1$, and $d = 3$. Our goal in this section is to construct an $[n, k, d]_2$-code, called a *Hamming code*.[33] We do this by constructing its parity check matrix H; then the code in question will simply be the linear subspace $C = \{\mathbf{x} \in \mathbb{F}_2^n \mid \mathbf{x}H = \mathbf{0}\}$ of $\mathbb{F}_2^n$.

Note that the binary representation of the integer $n = 2^\ell - 1$ is $\underbrace{1 \ldots 1}_{\ell}$. More generally, the binary representation of any integer in $\{1, \ldots, n\}$ has at most ℓ digits. Now, for all $i \in \{1, \ldots, n\}$, let $\mathbf{h}_i \in \mathbb{F}_2^\ell$ be the vector giving the binary representation of i, with zeros added to the front if necessary (so

[31] The minimum distance between codewords in C is d. So, there exists distinct vectors in C (say, $\mathbf{y}$ and $\mathbf{z}$) whose distance is precisely d.

[32] We are also using the fact that $\mathbf{y} \neq \mathbf{z}$, and so $\mathbf{y} - \mathbf{z} \neq \mathbf{0}$.

[33] It is also possible to construct "q-ary Hamming codes," which are over the (more general) field $\mathbb{F}_q$. For the sake of simplicity, though, we consider only binary Hamming codes, i.e. those over the field $\mathbb{F}_2$.

that the length of the representation is ℓ).[34] Let

$$H \quad := \quad \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_n \end{bmatrix}.$$

Note that $H \in \mathbb{F}_2^{n\times\ell}$. We now define the code C by setting

$$C \quad := \quad \{\mathbf{x} \in \mathbb{F}_2^n \mid \mathbf{x}H = \mathbf{0}\}.$$

Let us show that C is an $[n, k, d]_2$-code. Obviously, C is a linear subspace of $\mathbb{F}_2^n$.[35] Let us show that $\dim C = k$.[36] As usual, for all $i \in \{1, \ldots, \ell\}$, let $\mathbf{e}_i^\ell$ be the vector in $\mathbb{F}_2^\ell$ whose i-th coordinate is 1, and all of whose other coordinates are 0. Then each of $\mathbf{e}_1^\ell, \ldots, \mathbf{e}_\ell^\ell$ is a row of H, and furthermore, the set $\{\mathbf{e}_1^\ell, \ldots, \mathbf{e}_\ell^\ell\}$ is a basis for $\mathbb{F}_2^\ell$; so, $\operatorname{rank}(H) = \ell$. The Rank-nullity theorem guarantees that $\operatorname{rank}(H) + \dim \operatorname{Ker}(H) = n$, and we deduce that $\dim \operatorname{Ker}(H) = n - \ell = k$. But $C = \operatorname{Ker}(H)$, and so $\dim C = k$.

It remains to show that the minimum distance of words in C is $d = 3$. We will use Proposition 8.5.2. As usual, for all $i \in \{1, \ldots, n\}$, let $\mathbf{e}_i^n$ be the vector in $\mathbb{F}_2^n$ whose i-th coordinate is 1, and all of whose other coordinates are 0. Note that the vectors of $\mathbb{F}_2^n$ of Hamming weight 1 are precisely the vectors $\mathbf{e}_1^n, \ldots, \mathbf{e}_n^n$. But note that, for all $i \in \{1, \ldots, n\}$, we have that $\mathbf{e}_i^n H = \mathbf{h}_i \neq \mathbf{0}$, and so $\mathbf{e}_i^n \notin C$. Next, vectors of $\mathbb{F}_2^n$ of Hamming weight 2 are precisely the vectors of the form $\mathbf{e}_i^n + \mathbf{e}_j^n$, with $i \neq j$. Now, for distinct $i, j \in \{1, \ldots, n\}$, we have that $(\mathbf{e}_i^n + \mathbf{e}_j^n)H = \mathbf{h}_i + \mathbf{h}_j$; since $\mathbf{h}_i \neq \mathbf{h}_j$ (and our field is $\mathbb{F}_2$), we have that $\mathbf{h}_i + \mathbf{h}_j \neq \mathbf{0}$, and it follows that $\mathbf{e}_i^n + \mathbf{e}_j^n \notin C$. We have now shown that C does not contain any non-zero vectors of Hamming weight at most two. On the other hand, C does contain a vector of Hamming weight at most three, e.g. the vector $\mathbf{e}_1^n + \mathbf{e}_2^n + \mathbf{e}_3^n$.[37] So, $\min\{\operatorname{wt}(\mathbf{x}) \mid \mathbf{x} \in C, \mathbf{x} \neq \mathbf{0}\} = 3 = d$, and so by Proposition 8.5.2, we see that the minimum distance in C is d.

[34] For example, if $\ell = 2$, then $n = 3$, and we have that $\mathbf{h}_1 = (0, 1)$, $\mathbf{h}_2 = (1, 0)$, and $\mathbf{h}_3 = (1, 1)$.

[35] So, C is a linear code, and furthermore, the first coordinate (i.e. the n part) and the subscript (i.e. 2) of $[n, k, d]_2$ are correct.

[36] In view of Proposition 8.5.1, this will guarantee that second coordinate (i.e. the k part) of $[n, k, d]_2$ is correct.

[37] Indeed,

$$\begin{aligned}(\mathbf{e}_1^n + \mathbf{e}_2^n + \mathbf{e}_3^n)H &= \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 \\ &= (\underbrace{0, \ldots, 0}_{n-2}, 0, 1) + (\underbrace{0, \ldots, 0}_{n-2}, 0, 1) + (\underbrace{0, \ldots, 0}_{n-2}, 1, 1) \\ &= \mathbf{0},\end{aligned}$$

We have now shown that C is indeed an $[n, k, d]_2$-code, that is, C is a $[2^\ell - 1, 2^\ell - \ell - 1, 3]_2$-code. The code that we just constructed is called a *Hamming code*.

Finally, let us explain how error checking works for the Hamming code C that we just constructed. Suppose $\mathbf{w} \in \mathbb{F}_2^n$. Then by construction, $\mathbf{w} \in C$ if and only if $\mathbf{w}H = \mathbf{0}$. Suppose now that $\mathbf{w}$ differs in exactly one coordinate from some codeword in C, that is, that $\mathbf{w}$ can be obtained from a codeword in C by introducing one error (i.e. by changing exactly one 1 into 0, or vice versa, in some codeword of C). This means that there exist some $\mathbf{x} \in C$ and $i \in \{1, \ldots, n\}$ such that $\mathbf{w} = \mathbf{x} + \mathbf{e}_i^n$, and so

$$\begin{aligned} \mathbf{w}H &= (\mathbf{x} + \mathbf{e}_i^n)H \\ &= \underbrace{\mathbf{x}H}_{=\mathbf{0}} + \underbrace{\mathbf{e}_i^n H}_{=\mathbf{h}_i} \\ &= \mathbf{h}_i. \end{aligned}$$

But $\mathbf{h}_i$ is simply the integer i written in binary code! This means that if $\mathbf{w}$ was obtained from a codeword in C by introducing exactly one error, then the coordinate of that error is the integer whose binary representation is given by the vector $\mathbf{w}H$; we can correct the error by altering the entry (from 1 to 0, or vice versa) in that one coordinate of $\mathbf{w}$.

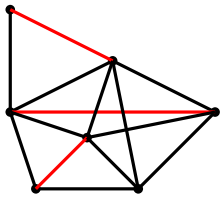
and so $\mathbf{e}_1^n + \mathbf{e}_2^n + \mathbf{e}_3^n \in C$.

Chapter 9

Matchings in general graphs

9.1 Basic notions

Recall that a *matching* in a graph G is a collection of edges of G, no two of which share an endpoint. An example of a matching is shown below (the edges of the matching are in red).



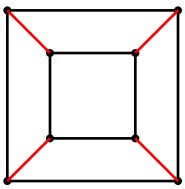
A *maximum matching* of G is a matching M of G such that for all matchings M' of G, we have that $|M'| \leq |M|$. The *matching number* of G, denoted by $\nu(G)$, is the size of a maximum matching (i.e. the number of edges in a maximum matching).[1] Trivially, $\nu(G) \leq \left\lfloor \frac{|V(G)|}{2} \right\rfloor$.

If M is a matching and v is a vertex of a graph G, then we say that v is *saturated* by M (or that M *saturates* v) provided that v is incident with some edge of M. If M does not saturate v, then v is *unsaturated* by M. A set $X \subseteq V(G)$ is *saturated* by M if every vertex in X is saturated by M; if X is saturated by M, then we also say that the matching M is *X-saturating*.

A matching M of a graph G is *perfect* if all vertices of G are saturated by M. Obviously, a graph G has a perfect matching if and only if $\nu(G) = \frac{|V(G)|}{2}$. In particular, every graph that has a perfect matching, has an even number

[1] So, $\nu(G) = \max\{|M| \mid M \text{ is a matching of } G\}$.

of vertices.[2] An example of a perfect matching is shown below (the edges of the matching are in red).



9.2 The Gallai-Edmonds decomposition

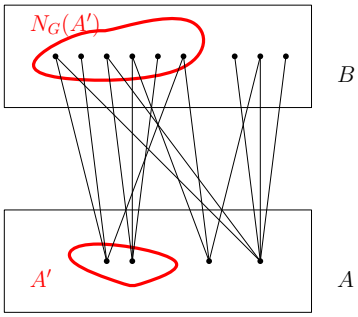
In section 9.3, we will state and prove Tutte's theorem, which gives a necessary and sufficient condition for a graph to have a perfect matching. In fact, we will obtain Tutte's theorem as a corollary of the so called Tutte-Berge formula (also stated and proven in section 9.3). In the present section, we develop some technical tools that we will need in the proof of the Tutte-Berge formula.

We begin by stating Hall's theorem (proven in section 4.4). For a graph G and a set $X \subseteq V(G)$, we denote by $N_G(X)$ the set of all vertices in $V(G) \setminus X$ that have at least one neighbor in X, i.e. $N_G(X) := \{y \in V(G) \setminus X \mid \exists x \in X \text{ s.t. } xy \in E(G)\}$.

Hall's theorem (graph theoretic formulation). *Let G be a bipartite graph with bipartition (A, B). Then the following are equivalent:*

(a) all sets $A' \subseteq A$ satisfy $|A'| \leq |N_G(A')|$;

(b) G has an A-saturating matching.

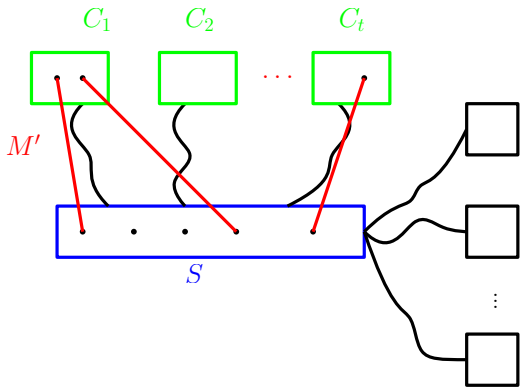


[2] However, there are a great many graphs on an even number of vertices that have no perfect matching. Edgeless graphs are an obvious example, but there are many others.

An *odd component* of a graph G is a (connected) component of G that has an odd number of vertices. We denote by $\mathrm{odd}(G)$ the number of odd components of G.

Remark 9.2.1. *Let G be a graph. Then for all $S \subseteq V(G)$, we have that $\nu(G) \leq \frac{|V(G)|+|S|-odd(G\setminus S)}{2}$.*

Proof. Fix $S \subseteq V(G)$, set $t := \mathrm{odd}(G \setminus S)$, and let $C_1, \ldots, C_t$ be the odd components of $G \setminus S$.

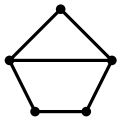


Fix any matching M in G. Let M' be the set of all edges of M that have one endpoint in S and the other one in $V(C_1) \cup \cdots \cup V(C_t)$; obviously, $|M'| \leq |S|$. Next, since the components $C_1, \ldots, C_t$ are all odd, it follows that at least $t - |M'| \geq t - |S|$ of the components $C_1, \ldots, C_t$ have a vertex that is unsaturated by M.[3] So, the total number of vertices of G that are saturated by M is at most $|V(G)| - (t - |S|) = |V(G)| + |S| - t$, and it follows that $|M| \leq \frac{|V(G)|+|S|-t}{2}$. Since the matching M was chosen arbitrarily, we deduce that $\nu(G) \leq \frac{|V(G)|+|S|-t}{2} = \frac{|V(G)|+|S|-\mathrm{odd}(G\setminus S)}{2}$. □

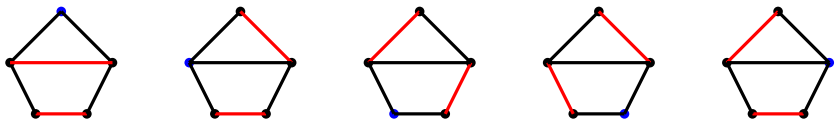
A graph G is *hypomatchable* if it does not have a perfect matching, but for all $v \in V(G)$, the graph $G \setminus v$ does have a perfect matching. Obviously, every hypomatchable graph has an odd number of vertices.[4] For example, the graph below is hypomatchable.

[3] This is because for all odd components C_i, the number of edges of M that have both endpoints in C_i is at most $\left\lfloor \frac{V(C_i)}{2} \right\rfloor = \frac{|V(C_i)|-1}{2}$; if all vertices of C_i are saturated by M, then there must be an edge of M between S and $V(C_i)$. The number of indices i for which such an edge exists is at most $|M'| \leq |S|$. So, at least $t - |S|$ components C_i have a vertex that is unsaturated by M.

[4] But not all graphs with an odd number of vertices are hypomatchable!

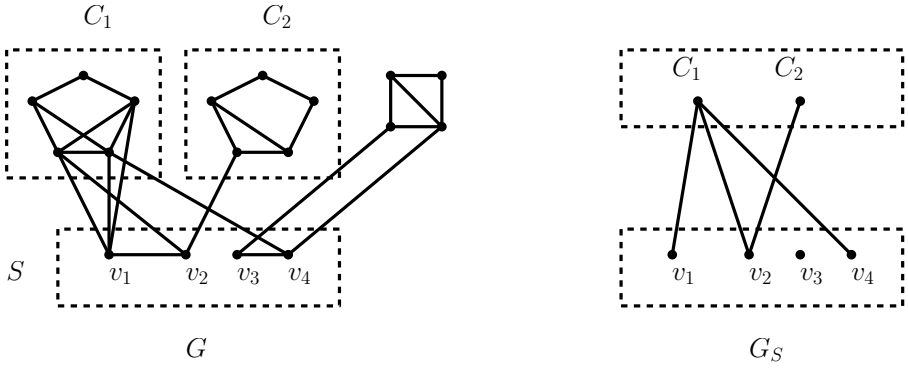


Indeed, deleting any one vertex from the graph above yields a graph that has a perfect matching (as shown below; the vertex that we delete is in blue, and the matching is in red).



A *hypomatchable component* of a graph G is a component of G that is a hypomatchable graph. Obviously, every hypomatchable component of G is odd.

For a graph G and a set $S \subseteq V(G)$, let us denote by G_S the bipartite graph whose one side of the bipartition is S, and whose other side of the bipartition is the collection of all odd components of $G \setminus S$, and in which a vertex $v \in S$ and an odd component C of $G \setminus S$ are adjacent if and only if v has a neighbor in $V(C)$ in G. An example is shown below.



A *Gallai-Edmonds set* in a graph G is a set $S \subseteq V(G)$ that satisfies the following two properties:

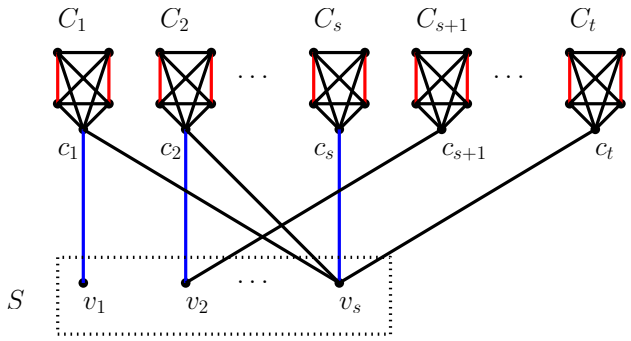
- every component of $G \setminus S$ is hypomatchable (and therefore odd);
- the bipartite graph G_S has an S-saturating matching.

Lemma 9.2.2. *If S is a Gallai-Edmonds set of a graph G, then*

$$\nu(G) \quad = \quad \frac{|V(G)|+|S|-odd(G\setminus S)}{2}.$$

Proof. Let S be a Gallai-Edmonds set of a graph G. By Remark 9.2.1, we have that $\nu(G) \leq \frac{|V(G)|+|S|-\text{odd}(G\setminus S)}{2}$. It remains to show that $\nu(G) \geq \frac{|V(G)|+|S|-\text{odd}(G\setminus S)}{2}$. To simplify notation, set $n := |V(G)|$, $s := |S|$, and $t := \text{odd}(G \setminus S)$. We must show that $\nu(G) \geq \frac{n+s-t}{2}$. We will prove this by exhibiting a matching M in G of size $\frac{n+s-t}{2}$.

Let $C_1, \ldots, C_t$ be the odd components of $G \setminus S$ (since all components of $G \setminus S$ are hypomatchable and therefore odd, we see that $C_1, \ldots, C_t$ are in fact all the components of $G \setminus S$), and set $S = \{v_1, \ldots, v_s\}$.



Since S is a Gallai-Edmonds set, we know that G_S has an S-saturating matching, call it M_S. By symmetry, we may assume that $M_S = \{v_1C_1, \ldots, v_sC_s\}$. For each $i \in \{1, \ldots, s\}$, choose a vertex $c_i \in V(C_i)$ such that $v_ic_i \in E(G)$.[5] For all $i \in \{s+1, \ldots, t\}$, choose any vertex $c_i \in C_i$. Next, since S is a Gallai-Edmonds set, we know that for all $i \in \{1, \ldots, t\}$, C_i is hypomatchable, and in particular, $C_i \setminus c_i$ has a perfect matching, call it M_i; clearly, $|M_i| = \frac{|V(C_i)|-1}{2}$. Now, set $M := \{v_1c_1, \ldots, v_sc_s\} \cup M_1 \cup \cdots \cup M_t$. Then M is a matching in G. Moreover, M saturates all but $t - s$ vertices of G (indeed, the only vertices of G unsaturated by M are $c_{s+1}, \ldots, c_t$), and so $|M| = \frac{n-(t-s)}{2} = \frac{n+s-t}{2}$. □

Lemma 9.2.3. *Every graph has a Gallai-Edmonds set.*

Proof. Let G be a graph, and assume inductively that every graph on fewer than $|V(G)|$ vertices has a Gallai-Edmonds set.

Choose a set $S \subseteq V(G)$ so that $\text{odd}(G \setminus S) - |S|$ is as large as possible, and subject to that, $|S|$ is as large as possible. Our goal is to show that S is a Gallai-Edmonds set.

Claim 1. All components of $G \setminus S$ are odd.

[5] Such a vertex c_i must exist because v_i and C_i are adjacent in G_S.

Proof of Claim 1. Suppose otherwise, and fix a component C of $G \setminus S$ that has an even number of vertices. Fix $v \in V(C)$, and set $S' := S \cup \{v\}$. Since $|V(C)|$ is even, we see that the odd components of $G \setminus S'$ are precisely the odd components of $G \setminus S$, plus the odd components of $C \setminus v$. Furthermore, since $|V(C)|$ is even, we see that $|V(C) \setminus \{v\}|$ is odd, and so $C \setminus v$ has at least one odd component. Thus,

$$\text{odd}(G \setminus S') \;=\; \text{odd}(G \setminus S) + \text{odd}(C \setminus v) \;\geq\; \text{odd}(G \setminus S) + 1,$$

and consequently (since $|S'| = |S| + 1$), we have that

$$\begin{aligned} \text{odd}(G \setminus S') - |S'| \;&\geq\; \Big(\text{odd}(G \setminus S) + 1\Big) - (|S| + 1) \\ &=\; \text{odd}(G \setminus S) - |S|. \end{aligned}$$

Since $|S'| > |S|$, this contradicts the choice of S. This proves Claim 1. ♦

Claim 2. All components of $G \setminus S$ are hypomatchable.

Proof of Claim 2. Suppose otherwise, and fix a component C of $G \setminus S$ and a vertex $v \in V(C)$ such that $C \setminus v$ does not have a perfect matching. By Claim 1, $C \setminus v$ has an even number of vertices; since $C \setminus v$ does not have a perfect matching, it follows that $\nu(C \setminus v) \leq \frac{|V(C)\setminus\{v\}|}{2} - 1 = \frac{|V(C)|-3}{2}$. By the induction hypothesis, $C \setminus v$ has a Gallai-Edmonds set, call it S_C. Thus,

$$\begin{aligned} \tfrac{|V(C)|-3}{2} \;&\geq\; \nu(C \setminus v) \\ &=\; \frac{|V(C\setminus v)|+|S_C|-\text{odd}\Big((C\setminus v)\setminus S_C\Big)}{2} \qquad \text{by Lemma 9.2.2} \\ &=\; \frac{|V(C)|-1+|S_C|-\text{odd}\Big((C\setminus v)\setminus S_C\Big)}{2}, \end{aligned}$$

and consequently,

$$\text{odd}\Big((C \setminus v) \setminus S_C\Big) \;\geq\; |S_C| + 2.$$

Now, set $S' := S \cup \{v\} \cup S_C$. Clearly, the odd components of $G \setminus S'$ are precisely the odd components of $G \setminus S$ other than C, plus the odd components

of $(C \setminus v) \setminus S_C$, and so

$$\begin{aligned} \operatorname{odd}(G \setminus S') &= \operatorname{odd}(G \setminus S) - 1 + \operatorname{odd}\Big((C \setminus v) \setminus S_C\Big) \\ &\geq \operatorname{odd}(G \setminus S) - 1 + (|S_C| + 2) \\ &= \operatorname{odd}(G \setminus S) + |S_C| + 1 \\ &= \operatorname{odd}(G \setminus S) + (|S'| - |S|), \end{aligned}$$

and we deduce that

$$\operatorname{odd}(G \setminus S') - |S'| \geq \operatorname{odd}(G \setminus S) - |S|.$$

Since we also have that $|S'| > |S|$, this contradicts the choice of S. This proves Claim 2. ♦

Claim 3. G_S has an S-saturating matching.

Proof of Claim 3. Suppose otherwise. Then by Hall's theorem, there exists a set $X \subseteq S$ such that $|X| > |N_{G_S}(X)|$. Set $S' := S \setminus X$. Then all odd components of $G \setminus S$ other than the ones in $N_{G_S}(X)$ are still odd components of $G \setminus S'$, and we compute:

$$\begin{aligned} \operatorname{odd}(G \setminus S') &\geq \operatorname{odd}(G \setminus S) - |N_{G_S}(X)| \\ &> \operatorname{odd}(G \setminus S) - |X| \\ &= \operatorname{odd}(G \setminus S) - (|S| - |S'|) \\ &= \operatorname{odd}(G \setminus S) - |S| + |S'|, \end{aligned}$$

and it follows that

$$\operatorname{odd}(G \setminus S') - |S'| > \operatorname{odd}(G \setminus S) - |S|,$$

contrary to the choice of S. This proves Claim 3. ♦

By Claims 2 and 3, we have that S is a Gallai-Edmonds set of G. □

9.3 The Tutte-Berge formula and Tutte's theorem

The Tutte-Berge formula. *Every graph G satisfies*

$$\nu(G) \;=\; \tfrac{1}{2} \min_{U \subseteq V(G)} \Big(|V(G)| + |U| - \mathit{odd}(G \setminus U) \Big).$$

Proof. Fix a graph G. By Lemma 9.2.3, G contains a Gallai-Edmonds set, call it S. Then

$$\begin{aligned} \nu(G) \;&=\; \tfrac{|V(G)|+|S|-\mathrm{odd}(G\setminus S)}{2} && \text{by Lemma 9.2.2} \\ &\geq\; \tfrac{1}{2} \min_{U \subseteq V(G)} \Big(|V(G)| + |U| - \mathrm{odd}(G \setminus U) \Big). \end{aligned}$$

The reverse inequality follows immediately from Remark 9.2.1. □

Tutte's theorem. *A graph G has a perfect matching if and only if every set $S \subseteq V(G)$ satisfies $|S| \geq \mathit{odd}(G \setminus S)$.*

Proof. Fix a graph G. Clearly, the following are equivalent:

(a) every set $S \subseteq V(G)$ satisfies $|S| \geq \mathrm{odd}(G \setminus S)$;

(b) $\min\limits_{U \subseteq V(G)} \Big(|V(G)| + |U| - \mathrm{odd}(G \setminus U) \Big) \geq |V(G)|$.

By the Tutte-Berge formula, (b) is equivalent to

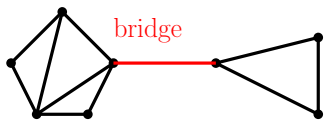
(c) $\nu(G) \geq \frac{|V(G)|}{2}$.

But clearly, (c) holds if and only if G has a perfect matching.[6] So, (a) holds if and only if G has a perfect matching, which is what we needed to show. □

9.4 Petersen's theorem

For a non-negative integer k, a graph G is *k-regular* if all vertices of G are of degree k. A graph is *cubic* if it is 3-regular.

A *bridge* in a graph G is an edge $e \in E(G)$ such that $G - e$ has more components than G. A graph is *bridgeless* if it has no bridge.



[6] Indeed, every graph G satisfies $\nu(G) \leq \frac{|V(G)|}{2}$. So, (c) is in fact equivalent to $\nu(G) = \frac{|V(G)|}{2}$. But $\nu(G) = \frac{|V(G)|}{2}$ if and only if G has a perfect matching.

Petersen's theorem. *Every cubic, bridgeless graph has a perfect matching.*[7]

Proof. Fix a cubic, bridgeless graph G. We will apply Tutte's theorem. Fix $S \subseteq V(G)$; we must show that $|S| \geq \text{odd}(G \setminus S)$.

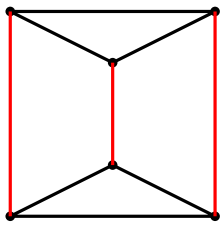
> **Claim.** For all odd components C of $G \setminus S$, there are at least three edges between S and $V(C)$ in G.

Proof of the Claim. Suppose that C is an odd component of $G \setminus S$, and let ℓ be the number of edges between S and $V(C)$. Since G is cubic, we have that $\sum_{v \in V(C)} d_G(v) = 3|V(C)|$; since C is an odd component, we see that $3|V(C)|$ is odd, and consequently, $\sum_{v \in V(C)} d_G(v)$ is odd. On the other hand, every edge incident with a vertex in $V(C)$ either has both its endpoints in $V(C)$, or has one endpoint in $V(C)$ and the other one in S; so, $\sum_{v \in V(C)} d_G(v) = 2|E(G[C])| + \ell$. Since $\sum_{v \in V(C)} d_G(v)$ is odd, we see that ℓ is odd. If $\ell = 1$, then the unique edge between S and $V(C)$ is a bridge in G, contrary to the fact that G is bridgeless. So, $\ell \geq 3$. This proves the Claim. ♦

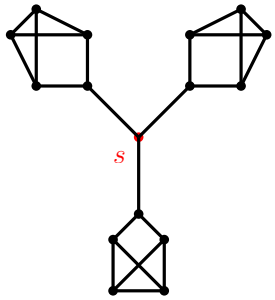
Set $t := \text{odd}(G \setminus S)$. By the Claim, the number of edges between S and $V(G) \setminus S$ is at least $3t$. On the other hand, since G is cubic, the total number of edges incident with at least one vertex of S as at most $3|S|$.[8] Thus, $3t \leq 3|S|$, i.e. $|S| \geq t = \text{odd}(G \setminus S)$. Since $S \subseteq V(G)$ was chosen arbitrarily, Tutte's theorem guarantees that G has a perfect matching. □

The bridgelessness requirement from Petersen's theorem is necessary, as the example below shows.

[7] Here is an example of a cubic, bridgeless graph, with a perfect matching shown in red.



[8] Note that we are double counting edges whose both endpoints are in S. Hence, the number of edges incident with at least one vertex of S is at most $3|S|$, and not necessarily exactly $3|S|$.

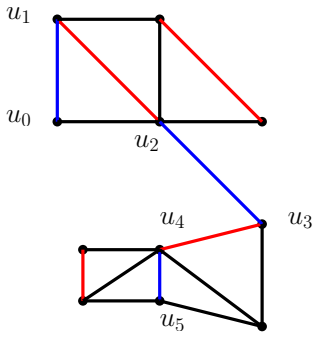


The graph above (call it G) is cubic, but not bridgeless. If we set $S := \{s\}$, then $G \setminus S$ has three odd components, and so $|S| < \text{odd}(G \setminus S)$. Thus, by Tutte's theorem, G does not have a perfect matching.

9.5 M-augmenting paths

Convention: In the remainder of this chapter, in all our figures, edges of the matching in question are in red.

Let M be a matching in a graph G. An *M-alternating path* is a path $u_0, u_1, \ldots, u_t$ in G such that every other edge of the path belongs to M (and the remaining edges do not). An *M-augmenting path* is an M-alternating path $u_0, u_1, \ldots, u_t$ ($t \neq 0$) such that u_0, u_t are both unsaturated by M. For instance, in the picture below, $u_0, u_1, u_2, u_3, u_4, u_5$ is an M-augmenting path (as usual, the edges of the matching M are in red; the edges of the M-augmenting path that do not belong to M are in blue).



We note that if M is a matching of a graph G, and u and v are adjacent vertices of G, both unsaturated by M, then the one-edge path u, v is an M-augmenting path.

Lemma 9.5.1. *Let M be a matching in a graph G, and let $u_0, u_1, \ldots, u_t$ be an M-augmenting path. Then t is odd and*

$$M' \quad := \quad \Big(M \setminus \{u_1u_2, u_3u_4, \ldots, u_{t-2}u_{t-1}\}\Big) \cup \{u_0u_1, u_2u_3, \ldots, u_{t-1}u_t\}$$

is a matching of G satisfying $|M'| = |M| + 1$.

Proof. This follows from the relevant definitions. □

Theorem 9.5.2. *[Berge, 1957] Let M be a matching in a graph G. Then M is a maximum matching of G if and only if G has no M-augmenting path.*

Proof. We will prove that the matching M is **not** maximum if and only if G has an M-augmenting path.

If G has an M-augmenting path, then Lemma 9.5.1 guarantees that M is not a maximum matching of G.

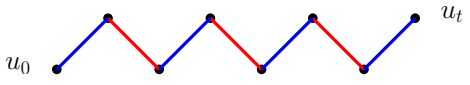
Suppose now that M is not a maximum matching, and let M' be matching of G such that $|M'| > |M|$. Let $F := M \Delta M'$,[9] and let H be the graph with vertex set $V(H) = V(G)$ and edge set $E(H) = F$. Clearly, $\Delta(H) \leq 2$.[10] So, H is the disjoint union of paths and cycles.

Now, since $|M'| > |M|$, some component P of H has more edges of M' than of M. If P is a cycle, then we see that some vertex of P is incident two edges of M', contrary to the fact that M' is a matching. So, P is a path, and it is easy to see that it is in fact an M-augmenting path in G.[11] □

[9] By definition, $M \Delta M' = (M \setminus M') \cup (M' \setminus M)$.

[10] Recall that $\Delta(H)$ is the maximum degree in H, i.e. $\Delta(H) = \max\{d_H(v) \mid v \in V(H)\}$. Let us check that $\Delta(H) \leq 2$. Since M and M' are matchings, we see that every vertex v of G is incident with at most one edge of M and at most one edge of M'. Since $V(H) = V(G)$ and $E(H) \subseteq M \cup M'$, it follows that every vertex of H is incident with at most two edges; thus, $\Delta(H) \leq 2$.

[11] Indeed, let P be of the form $u_0, u_1, \ldots, u_t$. All edges of P are in $M \Delta M'$, and so since M and M' are both matchings, the edges of $M \setminus M'$ and $M' \setminus M$ alternate on P. Since P has more edges of M' than of M, we have that P has an odd number of edges (so, t is odd), and that $u_0u_1, u_2u_3, \ldots, u_{t-1}u_t \in M' \setminus M$ and $u_1u_2, u_3u_4, \ldots, u_{t-2}, u_{t-1} \in M \setminus M'$ (see the picture below; edges of $M \setminus M'$ are in red, and edges of $M' \setminus M$ are in blue).

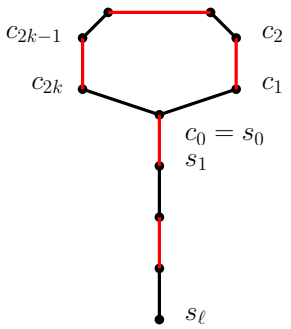


The fact that u_0, u_t are unsaturated by M follows from the construction of H, and from the fact that P is a component of H.

9.6 Blossoms and stems

Our goal is to give a polynomial-time algorithm that finds a maximum matching in a graph. The basic idea is to start with an empty matching, and then repeatedly find augmenting paths and use them to find larger matchings (as in Lemma 9.5.1). We do this until no augmenting path remains, at which point Theorem 9.5.2 guarantees that our matching is maximum. In this section, we describe the basic tools that we need, and in the subsequent section, we describe the algorithm.

We begin with a definition. Suppose that M is a matching in a graph G. A *blossom* is a cycle $c_0, c_1, \ldots, c_{2k}, c_0$ of length $2k+1$ (with $k \geq 1$) in G in which edges $c_1c_2, c_3c_4, \ldots, c_{2k-1}c_{2k}$ belong to M, and the remaining $k+1$ edges do not belong to M. A *stem* for this blossom is an M-alternating path $s_0, \ldots, s_\ell$ of even length[12] such that $s_0 = c_0$ is the unique common vertex of the cycle $c_0, c_1, \ldots, c_{2k}, c_0$ and the path $s_0, \ldots, s_\ell$, and s_ℓ is unsaturated by M.[13] The union of a blossom and a corresponding stem is called a *flower*.[14] An example is shown below.



Note that if M is a matching and F is a flower (with respect to M) in a graph G, then any edge of M that has an endpoint in $V(F)$ is in fact an edge of F.

Next, let G be a graph, and let $C \subseteq V(G)$ and $c \in C$. We say that G' is the graph obtained form G by *contracting* C to c if

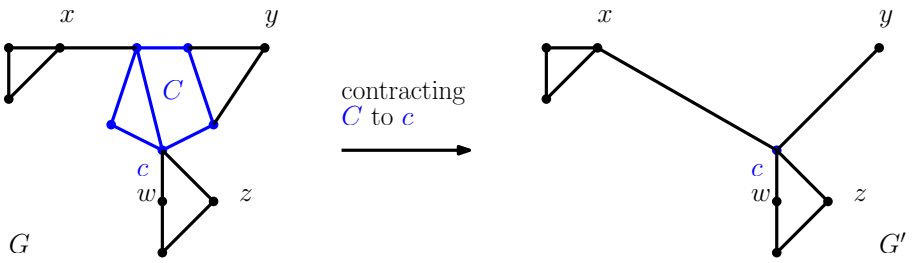
- $V(G') = V(G) \setminus (C \setminus \{c\}) = (V(G) \setminus C) \cup \{c\}$, and

[12] So, the path has an even number of edges, and therefore, ℓ is even.

[13] Note that this implies that either $\ell = 0$ and $c_0 = s_0$ is unsaturated by M, or $\ell \geq 2$ and $s_0s_1 \in M$.

[14] Note that there may be more than one stem for a fixed blossom. Nonetheless, all stems attach to the same vertex of the blossom.

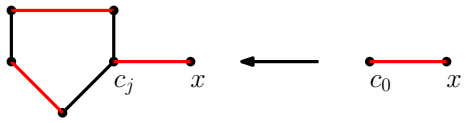
- $E(G') = \left(\binom{V(G)\setminus C}{2} \cap E(G)\right) \cup \left\{xc \mid x \in V(G) \setminus C,\ \exists c' \in C \text{ s.t. } xc' \in E(G)\right\}$.



Lemma 9.6.1. *Let M be a matching in a graph G, and let $C = c_0, \ldots, c_{2k}, c_0$ be a blossom and $S = s_0, \ldots, s_\ell$ a corresponding stem (in particular, $c_0 = s_0$). Let G' be the graph obtained from G by contracting C to c_0,[15] and let $M' = M \setminus E(C)$. Then M' is a matching of G'. Furthermore, M is a maximum matching of G if and only if M' is a maximum matching of G'.*

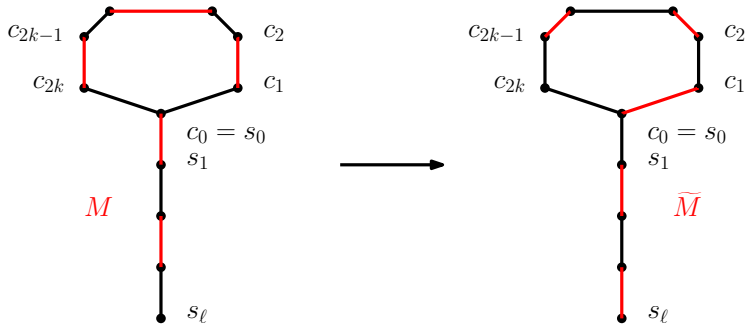
Proof. The fact that M' is a matching of G' follows from the appropriate definitions. Now, we will show that M is **not** a maximum matching in G if and only if M' is **not** a maximum matching in G'.

Suppose first that M' is not a maximum matching of G'; we must show that M is not a maximum matching of G. Let M'' be a matching of G' of size greater than $|M'|$. If c_0 is unsaturated by M'', then $M'' \cup \left(M \cap E(C)\right)$ is a matching of G of size greater than $|M|$. Suppose now that c_0 is saturated by M''. Then there exists some vertex $x \in V(G) \setminus V(C)$ and an index $j \in \{0, \ldots, 2k\}$ such that $xc_j \in E(G)$. But now the matching $(M'' \setminus \{xc_0\}) \cup \{xc_j\} \cup \{c_{j+1}c_{j+2}, c_{j+3}c_{j+4}, \ldots, c_{j+2k-1}c_{j+2k}\}$ is a matching of G of size greater than $|M|$ (see the picture below).



Suppose now that M is not a maximum matching of G; we must show that M' is not a maximum matching of G'. First, let $\widetilde{M} := \left(M \setminus (E(C) \cup E(S)\right) \cup \{c_0c_1, c_2c_3, \ldots, c_{2k-2}c_{2k-1}\} \cup \{s_1s_2, s_3s_4, \ldots, s_{\ell-1}s_\ell\}$ and $\widetilde{M'} = \left(M' \setminus E(S)\right) \cup \{s_1s_2, s_3s_4, \ldots, s_{\ell-1}s_\ell\}$.

[15] Technically, we mean that G is obtained by contracting $V(C)$ to c_0.



Clearly, $\widetilde{M}$ is a matching of G of the same size as M, and $\widetilde{M}'$ is a matching of G' of the same size as M'. Since the matching M of G is not maximum, neither is $\widetilde{M}$; so, by Theorem 9.5.2, there exists an $\widetilde{M}$-augmenting path in G, say $P = p_0, \ldots, p_t$. It now suffices to exhibit an $\widetilde{M}'$-augmenting path in G', for Theorem 9.5.2 will then imply that the matching $\widetilde{M}'$ is not maximum in G', and consequently, that M' is not maximum in G', either.

If $V(P) \cap V(C) = \emptyset$, then P is an $\widetilde{M}'$-augmenting path in G', and we are done. So, we may assume that $V(P) \cap V(C) \neq \emptyset$. First of all, c_{2k} is the only vertex in $V(C)$ that is unsaturated by $\widetilde{M}$; since both p_0, p_t are unsaturated by $\widetilde{M}$, we see that at most one of p_0, p_t belongs to $V(C)$. By symmetry, we may assume that $p_0 \notin V(C)$. Now, set $t_1 := \min\{i \in \{1, \ldots, t\} \mid p_i \in V(C)\}$. But then $p_0, \ldots, p_{t_1-1}, c_0$ is an $\widetilde{M}'$-augmenting path in G',[16] and we are done. □

9.7 Edmonds' Blossom algorithm

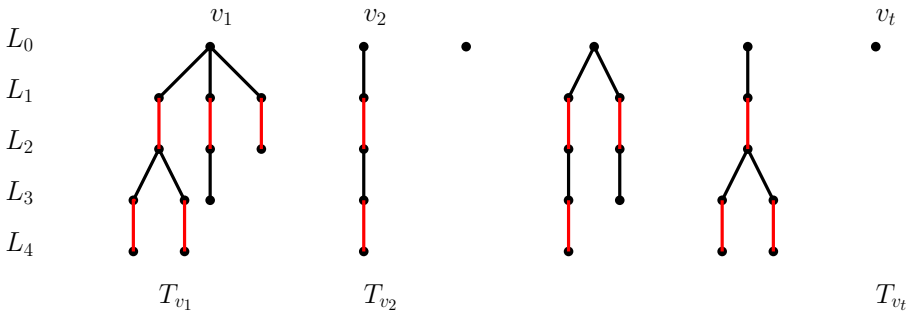
In what follows, we will use the following notation: for a tree T and vertices $x, y \in V(T)$, we denote by $x - T - y$ the unique path between x and y in T.

Let G be an input graph. Initially, we start with the empty matching, and we iteratively increase the size of the matching until this is no longer possible, at which point, our matching is maximum. All we need to do is show how, given a matching M in G, we either produce a larger matching, or determine that no larger matching exists. We proceed as follows.

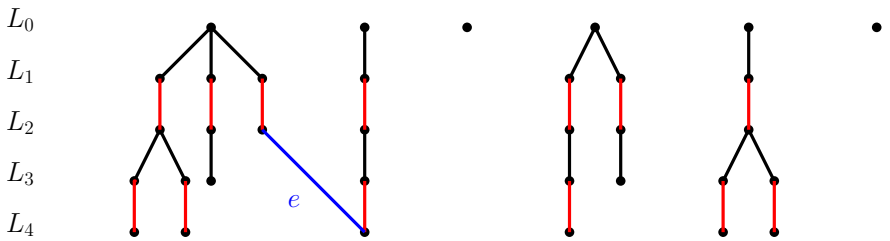
Step 1. First, we form an auxiliary forest F (which is a subgraph of G) as follows. $V(F)$ is partitioned into levels, $L_0, L_1, L_2, \ldots$. Level L_0 consists of all vertices of G that are unsaturated by M. If $L_0 = \emptyset$, then M is a perfect (and therefore maximum) matching of G, and we are done. So, we may assume that $L_0 \neq \emptyset$. Then, using breadth-first-search, we form a (maximal)

[16] We are using the fact that, by construction, c_0 is unsaturated by $\widetilde{M}'$ in G'.

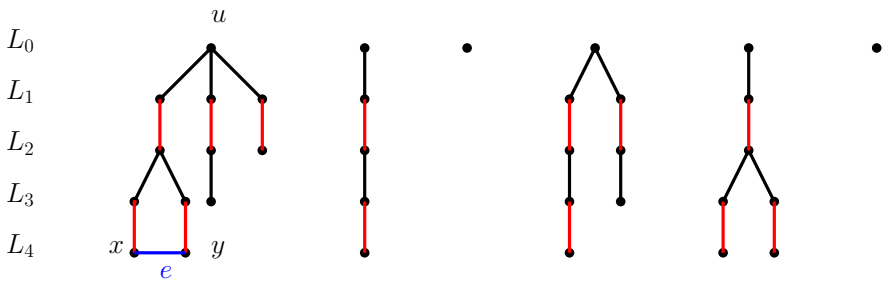
forest F in such a way that, for each integer $k \geq 0$, L_k is the set of all vertices of F that are at distance k from L_0 in F,[17] and moreover, for all even $k \geq 0$, edges between L_k and L_{k+1} in F do not belong to M, and edges between L_{k+1} and L_{k+2} in F do belong to M. For each $v \in L_0$, the unique component of F that contains v is the tree T_v rooted at v.



Step 2. If there exists an edge $e \in E(G)$ between even levels of two distinct trees, we immediately obtain an M-augmenting path,[18] and then we obtain a matching of size $|M| + 1$, as in Lemma 9.5.1.



If there exists an edge $e \in E(G)$ between two vertices, say x and y, belonging to even levels of the same tree T_u, then we can find a flower (i.e. a blossom with a corresponding stem), as follows.



[17] So: distance is counted in the forest F, and not in the whole graph G.

[18] Indeed, suppose that for distinct $u, v \in L_0$, and some even p, q, we have an edge e between a vertex $u' \in V(T_u) \cap L_p$ and a vertex $v' \in V(T_v) \cap L_q$. Then $u - T_u - u' - v' - T_v - v$ is an M-augmenting path in G.

We consider the (unique) path in T_u between x and u in T_u, and the (unique) path in T_u between y and u. The union of these two paths, together with the edge e, is a flower in G, say, with blossom $C = c_0, \ldots, c_{2k}, c_0$ and stem $S = s_0, \ldots, s_\ell$, where $c_0 = s_0$ and $s_\ell \in L_0$.[19] Let G' be the graph obtained from G by contracting C to a vertex c_0, and let $M' = M \setminus E(C)$ (as in Lemma 9.6.1). We now call the algorithm with input G' and M'. Then there are two cases.

- If we obtain the answer that M' is a maximum matching in G', then (by Lemma 9.6.1) M is a maximum matching in G, and we are done.
- Suppose we obtained a matching M'' in G' that is of size greater than $|M'|$. If c_0 is unsaturated by M'', then $(E(C) \cap M) \cup M''$ is a matching in G of size greater than $|M|$, and we are done. Suppose now that c_0 is saturated by M'', and let $x \in V(G) \setminus V(C)$ be such that $xc_0 \in M''$. Let v be some vertex of C such that $xv \in E(G)$, and let M_C be the (unique) matching of size $\frac{|V(C)|-1}{2}$ in C, chosen so that v is M_C-unsaturated. Then $(M'' \setminus \{xc_0\}) \cup \{xv\} \cup M_C$ is a matching in G of size greater than $|M|$.

Next, suppose that some edge $e \in M \setminus E(F)$ has at least one endpoint in $V(F)$. Set $e = xy$. Then e in fact has both its endpoints in $V(F)$, for otherwise, it would have been added to F via our breadth-first-search construction. Moreover, both endpoints of e must belong to odd levels. If both endpoints of e belong to the same tree T_u (for some $u \in L_0$), then similarly to the previous case, we obtain a flower containing e, and we then proceed as in the previous case. So, we may assume that e does not have both its endpoints in the same tree. Then there exist distinct $u, v \in L_0$ such that $x \in V(T_u)$ and $y \in V(T_v)$, and so $u - T_u - x - y - T_v - v$ is an M-augmenting path in G. We can now obtain a matching of size $|M| + 1$, as in Lemma 9.5.1.

From now on, we assume that there are no edges (of G) between vertices in even levels, and moreover, that every edge of M that has an endpoint in $V(F)$ is in fact an edge of F. We now claim that G contains no M-augmenting path, and that M is therefore (by Theorem 9.5.2) a maximum matching in G. Since L_0 is the set of all vertices that are unsaturated by M, it suffices to show that no non-trivial M-alternating path has more than one endpoint in L_0.[20] So, fix an M-alternating path $P = p_0, \ldots, p_t$, with $t \geq 1$. We must show that at most one of p_0, p_t belongs to L_0. If neither p_0 nor

[19] In fact, $s_\ell = u$.

[20] A path is *non-trivial* if it has at least one edge.

p_t belongs to L_0, then we are done. So, by symmetry, we may assume that $p_0 \in L_0$, and we must show that $p_t \notin L_0$.

Claim. For all $i \in \{0, \ldots, t-1\}$, one of the following holds:

(1) $p_i p_{i+1} \in E(F)$, and there exists an integer k such that $p_i \in L_k$ and $p_{i+1} \in L_{k+1}$;

(2) $p_i p_{i+1} \notin E(F)$, p_i belongs to an even level, and p_{i+1} belongs to an odd level.[21]

Proof of the Claim. We proceed by induction on i. First of all, $p_0 \in L_0$. So, if $p_0 p_1 \in E(F)$, then $p_1 \in L_1$, and (1) holds for $i = 0$. So, we may assume that $p_0 p_1 \notin E(F)$. Since vetices of L_0 are unsaturated by M, we know that $p_0 p_1 \notin M$. Now $p_1 \in V(F)$, for otherwise, our breadth-first-search construction of F would have added $p_0 p_1$ to F. Since there are no edges between even levels, we see that p_1 belongs to an odd level, and so (2) holds for $i = 0$.

Now, fix $i \in \{0, \ldots, t-2\}$, and assume that the claim holds for i. We must show it holds for $i+1$.

Suppose first that (1) holds for i, i.e. that $p_i p_{i+1} \in E(F)$, and there exists an integer k such that $p_i \in L_k$ and $p_{i+1} \in L_{k+1}$. If $p_{i+1} p_{i+2} \in E(F)$, then $p_{i+2} \in L_{k+2}$, and (1) holds for $i+1$. So, assume that $p_{i+1} p_{i+2} \notin E(F)$. Then $p_{i+1} p_{i+2} \notin M$,[22] and so since P is M-alternating, we see that $p_i p_{i+1} \in M$. But then k is odd and $k+1$ is even. Note that $p_{i+2} \in V(F)$, for otherwise, our breadth-first-search construction of F would have added $p_{i+1} p_{i+2}$ to F. Since there are no edges between even levels of F, and since p_{i+1} belongs to an even level, it follows that p_{i+2} belongs to an odd level. So, $i+1$ satisfies (2).

Suppose now that (2) holds for i, i.e. that $p_i p_{i+1} \notin E(F)$, p_i belongs to an even level, and p_{i+1} belongs to an odd level. Since $p_i p_{i+1}$ has an endpoint in $V(F)$, but does not belong to $E(F)$, we see that $p_i p_{i+1} \notin M$. Therefore, $p_{i+1} p_{i+2} \in M$, since P is M-alternating. So, $p_{i+1} p_{i+2} \in E(F)$.[23] Since p_{i+1} belongs to an odd level, say L_k, we see that p_{i+2} belongs to the even level L_{k+1}.[24] So, (1) holds for $i+1$. This proves the Claim. ♦

[21] Note that both (1) and (2) imply that $p_i, p_{i+1} \in V(F)$.

[22] Recall that all edges of M that have an endpoint in $V(F)$ are in fact edges of F. So, since $p_{i+1} \in V(F)$, but $p_{i+1} p_{i+2} \notin E(F)$, we have that $p_{i+1} p_{i+2} \notin M$.

[23] Once again, we are using the fact that all edges of M that have an endpoint in $V(F)$ are in fact edges of F. So, since $p_{i+1} \in V(F)$ and $p_{i+1} p_{i+2} \in M$, we see that $p_{i+1} p_{i+2} \in E(F)$.

[24] We are using the fact that $p_{i+1} p_{i+2} \in M$, plus the construction of F.

In view of the Claim, p_0 is the only vertex of P that belongs to L_0.[25] So, $p_t \notin L_0$, and we are done.

Remark: The running time of Edmonds' Blossom algorithm is $O(n^4)$, if the algorithm is implemented in the obvious way. We omit the details.

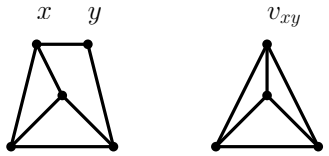
[25]Indeed, fix $i \in \{1, \ldots, t\}$. In view of the Claim, p_i either belongs to an odd level, or it belongs to a level that is one higher than the level that p_{i-1} belongs to. In either case, $p_i \notin L_0$.

Chapter 10

Minors and planar graphs

10.1 3-connected graphs

Given a graph G and an edge $xy \in E(G)$, we denote by G/xy the graph obtained from G by contracting xy to a vertex v_{xy}.



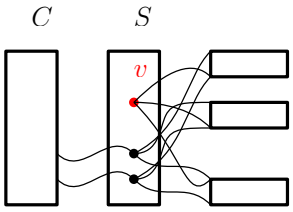
More formally, G/xy is the graph with vertex set $V(G/xy) = (V(G) \setminus \{x, y\}) \cup \{v_{xy}\}$ (where $v_{xy} \notin V(G)$) and edge set $E(G) = \{e \in \binom{V(G)\setminus\{x,y\}}{2} \mid e \in E(G)\} \cup \{vv_{xy} \mid v \in V(G) \setminus \{x, y\}$, and either $vx \in E(G)$ or $vy \in E(G)\}$. If $e = xy$, then we sometimes write G/e instead of G/xy, and v_e instead of v_{xy}.

Recall that for a non-negative integer k, a graph G is *k-connected* if it satisfies the following two conditions:

- $|V(G)| \geq k + 1$;
- for all $S \subseteq V(G)$ such that $|S| \leq k - 1$, the graph $G \setminus S$ is connected.

Proposition 10.1.1. *Let k be a positive integer, let G be a k-connected graph, and let $S \subseteq V(G)$ be such that $|S| = k$. Then every vertex of S has a neighbor in each component of $G \setminus S$.*

Proof. Suppose otherwise, and fix a vertex $v \in S$ and a component C of $G \setminus S$ such that v has no neighbors in $V(C)$. Then $S \setminus \{v\}$ separates v from $V(C)$ in G, and in particular, $G \setminus (S \setminus \{v\})$ is disconnected. But this is impossible since $|S \setminus \{v\}| = k - 1$ and G is k-connected.



□

Lemma 10.1.2. *Let G be a 3-connected graph on more than four vertices. Then G has an edge e such that G/e is 3-connected.*

Proof.

> **Claim.** For all $xy \in E(G)$, either G/xy is 3-connected, or there exists a vertex $z \in V(G) \setminus \{x, y\}$ such that $G \setminus \{x, y, z\}$ is disconnected.

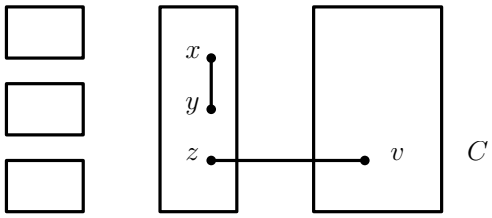
Proof of the Claim. Fix $xy \in E(G)$, and suppose that G/xy is not 3-connected. Clearly, G/xy has at least four vertices,[1] and so there exists some $S \subseteq V(G/xy)$ such that $|S| \leq 2$ and $(G/xy) \setminus S$ is disconnected. If $v_{xy} \notin S$, then it is clear that $G \setminus S$ is disconnected, contrary to the fact that G is 3-connected. So, $v_{xy} \in S$. Now set $S' = (S \setminus \{v_{xy}\}) \cup \{x, y\}$. Then $|S'| = |S| + 1$ and $G \setminus S' = (G/xy) \setminus S$; so, $G \setminus S'$ is disconnected. Since G is 3-connected, it follows that $|S'| \geq 3$; since $|S| \leq 2$, we deduce that $|S'| = 3$, and the result follows.[2] ♦

Since G is 3-connected, it is clear that G has at least one edge. Now, suppose that for all $e \in E(G)$, the graph G/e is not 3-connected. Then using the Claim, we fix an edge $xy \in E(G)$ and a vertex $z \in V(G) \setminus \{x, y\}$ such that $G \setminus \{x, y, z\}$ is disconnected, and we fix a component C of $G \setminus \{x, y, z\}$; we may assume that xy, z, C were chosen so that $|V(C)|$ is minimum.[3]

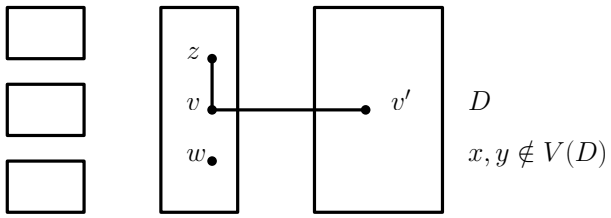
[1] This is because $|V(G)| > 4$, and clearly, $|V(G/xy)| = |V(G)| - 1$.

[2] Indeed, we take z to be the (unique) vertex of $S' \setminus \{x, y\}$.

[3] So, we are assuming that for all edges $x'y' \in E(G)$, all vertices $z' \in V(G) \setminus \{x', y'\}$ such that $\{x', y', z'\}$ is disconnected, and all components C' of $G \setminus \{x', y', z'\}$, we have that $|V(C)| \leq |V(C')|$.



Using Proposition 10.1.1, we let $v \in V(C)$ be a neighbor of z. By our supposition, G/zv is not 3-connected, and so by the Claim, there exists some $w \in V(G) \setminus \{z, v\}$ such that $G \setminus \{z, v, w\}$ is disconnected.[4] Since $xy \in E(G)$, there exists a component D of $G \setminus \{z, v, w\}$ such that $x, y \notin V(D)$; so, D is in fact a component of $G \setminus \{x, y, z, v, w\}$, and in particular, it is a connected induced subgraph of $G \setminus \{x, y, z\}$.



Now, let us show that $V(D) \subsetneqq V(C)$. By Proposition 10.1.1,[5] we know that v has a neighbor v' in $V(D)$. But note that all neighbors of v in G belong to $V(C) \cup \{x, y, z\}$, and so since $x, y, z \notin V(D)$,[6] we have that $v' \in V(D) \cap V(C)$. Since C is a component and D a connected induced subgraph of $G \setminus \{x, y, z\}$, we now deduce that $V(D) \subseteq V(C)$. Since $v \in V(C) \setminus V(D)$, it follows that $V(D) \subsetneqq V(C)$. But this contradicts the minimality of C. □

Proposition 10.1.3. *Let G be a graph, and let $xy \in E(G)$ be such that $d_G(x), d_G(y) \geq 3$. If G/xy is 3-connected, then so is G.*

Proof. To simplify notation, set $G' := G/xy$. Assume that G' is 3-connected. Then by definition, G' has at least four vertices, and consequently, G has at least five vertices.

Now, fix $S \subseteq V(G)$ such that $|S| \leq 2$; we must show that $G \setminus S$ is connected. If $S \cap \{x, y\} = \emptyset$, then $(G \setminus S)/xy = G' \setminus S$; since G' is 3-connected, we see that $G' \setminus S$ is connected, and we deduce that $(G \setminus S)/xy$ is connected. But then clearly, $G \setminus S$ is also connected. Next, if $S = \{x, y\}$,

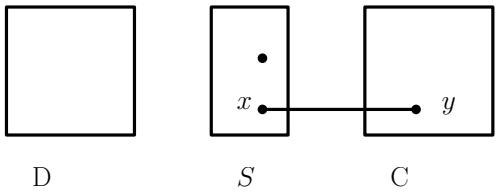
[4] It is possible that $w \in \{x, y\}$.

[5] We are applying Proposition 10.1.1 to G, $k = 3$, and $S = \{z, v, x\}$.

[6] We already saw that $x, y \notin V(D)$. Since D is a component of $G \setminus \{z, v, w\}$, we also have that $z \notin V(D)$. So, $x, y, z \notin V(D)$.

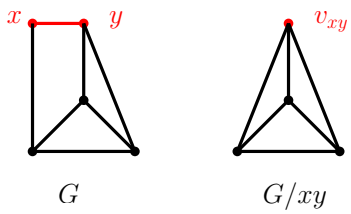
then $G \setminus S = G' \setminus v_{xy}$; since G' is 3-connected, we know that $G' \setminus v_{xy}$ is connected, and so $G \setminus S$ is connected.

It remains to consider the case when S contains exactly one of x, y. By symmetry, we may assume that $x \in S$ and $y \notin S$. Now, suppose that $G \setminus S$ is disconnected. Let C be the component of $G \setminus S$ that contains y, and let D be some other component of $G \setminus S$. Clearly, $N_G(y) \subseteq S \cup (V(C) \setminus \{y\})$; since $d_G(y) \geq 3$ and $|S| \leq 2$, we see that $V(C) \setminus \{y\} \neq \emptyset$. Set $S' := (S \setminus \{x\}) \cup \{v_{xy}\}$, and note that $G \setminus (S \cup \{y\}) = G' \setminus S'$. But now S' separates $V(C) \setminus \{y\} \neq \emptyset$ from $V(D)$ in G', contrary to the fact that G' is 3-connected and $|S'| \leq 2$.



□

Note that in the statement of Proposition 10.1.3, the requirement that $d_G(x), d_G(y) \geq 3$ is necessary, since every 3-connected graph G satisfies $\delta(G) \geq 3$.[7] For a concrete example, see the picture below (G/xy is 3-connected, but G is not).



Theorem 10.1.4 (Tutte, 1961)**.** *A graph G is 3-connected if and only if there exists a sequence $G_0, \ldots, G_n$ of graphs with the following properties:*

(1) $G_0 \cong K_4$ and $G = G_n$;

(2) for all $i \in \{0, \ldots, n-1\}$, G_{i+1} has an edge xy with $d_{G_{i+1}}(x), d_{G_{i+1}}(y) \geq 3$ and $G_i = G_{i+1}/xy$.

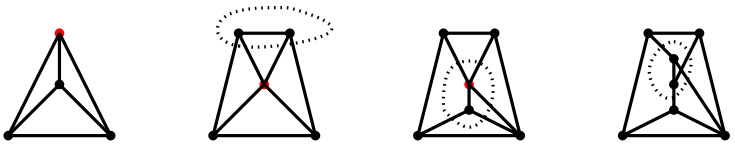
[7] Otherwise, we take a vertex $v \in V(G)$ with $d_G(v) \leq 2$, and we observe that $N_G(v)$ separates v from $V(G) \setminus N_G[v]$ (this is non-empty because $|N_G[v]| \leq 3$, and 3-connected graphs have at least four vertices), contrary to the fact that $|N_G(v)| = d_G(v) \leq 2$ and G is 3-connected.

Proof. Fix a graph G.

By definition, all 3-connected graphs have at least four vertices, and it is easy to see that K_4 is (up to isomorphism) the only 3-connected graph on four vertices. Moreover, by Theorem 5.1.3, the minimum degree of any 3-connected graph is at least three. So, if G is 3-connected, then Lemma 10.1.2 and an easy induction guarantee that there exists a sequence $G_0, \ldots, G_n$, as in the statement of the theorem.

On the other hand, if there exists a sequence $G_0, \ldots, G_n$ as in the statement of the theorem, then Proposition 10.1.3 and an easy induction guarantee that G is 3-connected. □

Note that Theorem 10.1.4 guarantees that every 3-connected graph can be obtained from K_4 by repeatedly "decontracting" vertices into edges, making sure that, at each step, both new vertices have degree at least three. An example is shown below (at each step, the vertex to be "decontracted" is in red, and in the subsequent step, the edge obtained by this "decontraction" is in a dotted bag); each graph in the sequence is 3-connected.



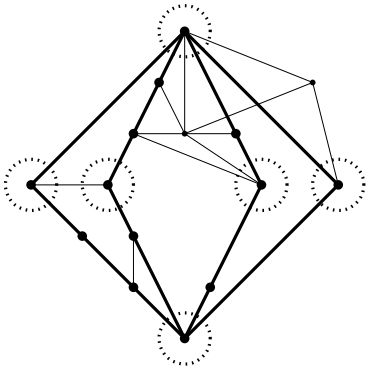
$G_0 \cong K_4$ G_1 G_2 G_3

10.2 Minors and topological minors

A graph H is a *topological minor* of a graph G, and we write $H \preceq_t G$, if G contains some subdivision of H as a subgraph.[8] The vertices of this subdivision that correspond to the vertices of H are called *branch vertices*.[9] For example, the graph below contains $K_{2,4}$ as a topological minor (the branch vertices are in dotted circles).

[8] Every graph is considered to be a subdivision of itself.

[9] If $\delta(H) \geq 3$, then branch vertices are uniquely defined. Otherwise, they need not be uniquely defined.

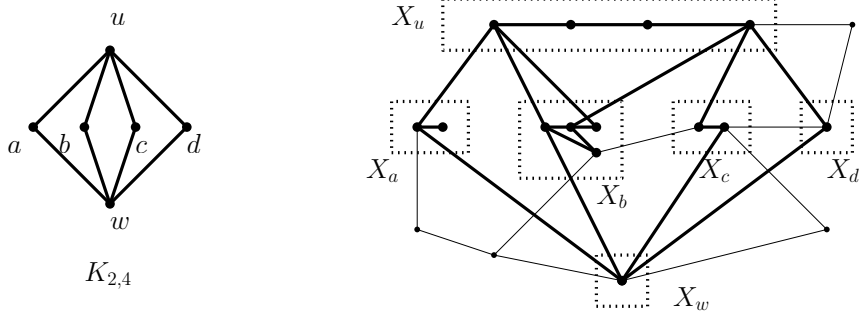


Obviously, the topological minor relation is transitive, that is, for all graphs G_1, G_2, G_3, if $G_1 \preceq_t G_2$ and $G_2 \preceq_t G_3$, then $G_1 \preceq_t G_3$.

A graph H is a *minor* of a graph G, and we write $H \preceq_m G$, if there exists a family $\{X_v\}_{v \in V(H)}$ of pairwise disjoint, non-empty subsets of $V(G)$, called *branch sets*, such that

- $G[X_v]$ is connected for all $v \in V(H)$, and
- for all $uv \in E(H)$, there is an edge between X_u and X_v in G.

For example, the graph below (on the right) contains $K_{2,4}$ as a minor.



Lemma 10.2.1. *For all graphs G and H, the following are equivalent:*

(1) $H \preceq_m G$;

(2) G can be transformed into (an isomorphic copy of) H by a sequence of vertex deletions, edge deletions, and edge contractions;[10]

(3) there exists a subgraph G' of G such that G' can be transformed into (an isomorphic copy) of H by a sequence of edge contractions.[11]

[10] Possibly, $G \cong H$.

[11] Possibly, $G' = G$ or $G' \cong H$.

Proof. Fix graphs G and H. We will show that (1) implies (3), that (3) implies (1), and that (2) and (3) are equivalent.

Suppose first that (1) holds; we must show that (3) holds. Let $\{X_v\}_{v\in V(H)}$ be the family of branch sets of the H minor in G. Let G' be the subgraph of G obtained by first deleting $V(G) \setminus \left(\bigcup_{v\in V(H)} X_v\right)$, and then for all distinct $u, v \in V(H)$ such that $uv \notin E(H)$, deleting all the edges between X_u and X_v. Let G'' be the graph obtained from G' by contracting each X_v into a vertex (we contract the X_v's one edge at a time, in any order). Clearly, $G'' \cong H$. So, (3) holds.

Suppose now that (3) holds; we must show that (1) holds. Let G' be a subgraph of G such that G' can be transformed into (an isomorphic copy) of H by a sequence of edge contractions. Let $G_0, \ldots, G_\ell$ be a sequence of graphs such that $G_0 = G'$, $G_\ell \cong H$, and for all $i \in \{0, \ldots, \ell - 1\}$, G_{i+1} can be obtained from G_i by contracting one edge. We may assume that $G_\ell = H$ (we rename vertices if necessary). For all $v \in V(H)$, we set $X_v^\ell = \{v\}$. Next, for all $i \in \{0, \ldots, \ell - 1\}$, having defined the sets X_v^{i+1}, we define the sets X_v^i as follows. Let $u_1u_2 \in E(G_i)$ be the edge of G_i that was contracted to obtain G_{i+1}, and let u be the vertex formed by contracting that edge.[12] For all $v \in V(H)$, if $u \in X_v^{i+1}$, then we set $X_v^i := (X_v^{i+1} \setminus \{u\}) \cup \{u_1, u_2\}$, and otherwise, we set $X_v^i := X_v^{i+1}$. It then follows by an easy induction that for all $i \in \{0, \ldots, \ell\}$, $\{X_v^i\}_{v\in V(H)}$ is a family of branch sets for the H minor in G_i. In particular, $\{X_v^0\}_{v\in V(H)}$ is a family of branch sets for the H minor in $G_0 = G'$, and therefore (since G' is a subgraph of G) in G as well. So, (1) holds.

It remains to show that (2) and (3) are equivalent. It is clear that (3) implies (2). Let us show that (2) implies (3). It is clear that if a graph G_2 is obtained from a graph G_1 by first contracting an edge and then deleting a vertex or an edge, then we can also obtain G_2 from G_1 by first deleting one or more vertices or edges, and then possibly contracting an edge.[13] Thus, if H can be obtained from G by a sequence of vertex deletions, edge deletions, and edge contractions, then H can be obtained from G by first (possibly)

[12] So, $u = v_{u_1u_2}$.

[13] Let us prove this fully formally. Suppose that G_2 is obtained from G_1 by first contracting an edge xy to a vertex v_{xy}, and then deleting a vertex z. If $z = v_{xy}$, then $G_2 = G_1 \setminus \{x, y\}$; otherwise, G_2 can be obtained from G_1 by first deleting z, and then contracting xy. Suppose now that G_2 is obtained from G_1 by first contracting an edge xy to a vertex v_{xy}, and then deleting an edge e. If v_{xy} is an endpoint of e, say $e = uv_{xy}$, then we can obtain G_2 from G_1 by first deleting all edges between u and $\{x, y\}$ (there is at least one and at most two such edges) and then contracting xy; otherwise, we can obtain G_2 from G_1 by first deleting e and then contracting xy.

deleting some vertices or edges (thus obtaining a subgraph G' of G), and then (possibly) contracting edges of G'. So, (2) implies (3). □

Lemma 10.2.2. *The minor relation is transitive, that is, for all graphs G_1, G_2, G_3, if $G_1 \preceq_m G_2$ and $G_2 \preceq_m G_3$, then $G_1 \preceq_m G_3$.*

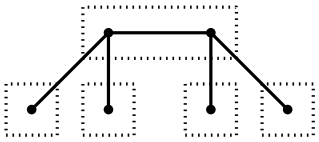
Proof. Fix graphs G_1, G_2, G_3 such that $G_1 \preceq_m G_2$ and $G_2 \preceq_m G_3$. By Lemma 10.2.1, G_1 can be obtained from G_2 by a sequence of vertex deletions, edge deletions, and edge contractions, and G_2 can similarly be obtained from G_3. So, G_1 can be obtained from G_3 by a sequence of vertex deletions, edge deletions, and edge contractions. So, by Lemma 10.2.1, we have that $G_1 \preceq_m G_3$. □

We remark that Lemma 10.2.2 can also be proven directly, using the definition of a minor.[14]

Lemma 10.2.3. *For all graphs G and H, if $H \preceq_t G$, then $H \preceq_m G$.*

Proof. Fix graphs G and H, and assume that $H \preceq_t G$. Then G contains a subgraph G' that is isomorphic to a subdivision of H, and clearly, H can be obtained from the subgraph G' by a sequence of edge contractions. Now Lemma 10.2.1 guarantees that $H \preceq_m G$. □

Note that the converse of Lemma 10.2.3 is false, i.e. it is possible that $H \preceq_m G$, but $H \not\preceq_t G$. For example, the graph below contains $K_{1,4}$ as a minor (the branch sets are in dotted rectangles), but not as a topological minor (this is because $K_{1,4}$ contains a vertex of degree four, whereas the maximum degree in the graph below is three).



We do, however, have the following lemma.

Lemma 10.2.4. *Let G and H be graphs such that $H \preceq_m G$ and $\Delta(H) \leq 3$. Then $H \preceq_t G$.*

Proof. Let G' be a minimal subgraph of G such that $H \preceq_m G'$,[15] and let $\{X_v\}_{v \in V(H)}$ be the corresponding branch sets in G'. Our goal is to show that

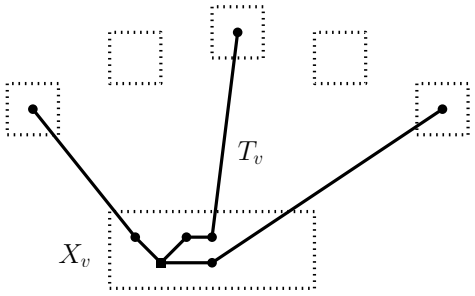
[14] Proof?

[15] So, $H \preceq_m G'$, but for all proper subgraphs G'' of G', we have that $H \not\preceq_m G''$.

G' is itself a subdivision of H; by definition, this will imply that $H \preceq_t G$. By the minimality of G', we know that for all distinct $u, v \in V(H)$, the following hold:

- if $uv \in E(H)$, then there is exactly one edge between X_u and X_v in G',[16]
- if $uv \notin E(H)$, then there are no edges between X_u and X_v.[17]

By the minimality of G', $G'[X_v]$ is a tree.[18] Now, for each $v \in V(H)$, we let T_v be the graph obtained from $G'[X_v]$ by adding to it the edges between X_v and $V(G') \setminus X_v$ (and the endpoints of those edges); see the picture below.



Clearly, for each $v \in V(H)$, the graph T_v is a tree. Since $\Delta(H) \leq 3$, the minimality of G' guarantees that T_v has at most three leaves, and so $\Delta(T_v) \leq 3$. Moreover, T_v has at most one vertex of degree three, and if this vertex exists, then it belongs to X_v. Now, for all $v \in V(H)$, we let v' be the unique vertex of T_v of degree three if such a vertex exists, and otherwise, we let v' be any vertex in X_v. It is now clear that G' is a subdivision of H (vertices v' are the branch vertices), and so $H \preceq_t G$. □

The following lemma will be of use to us in our next section, where we shall study planar graphs and "Kuratowski's theorem."

Lemma 10.2.5. *Let G be a graph. Then the following are equivalent:*

(1) G contains at least one of $K_5, K_{3,3}$ as a topological minor;

[16] By the definition of a minor, there is at least one edge between X_u and X_v. If there is more than one such edge, then we can contradict the minimality of G' by deleting some edge between X_u and X_v.

[17] Otherwise, we can contradict the minimality of G' by deleting an edge between X_u and X_v.

[18] Indeed, $G'[X_v]$ is connected, and therefore has a spanning tree, call it T. If $G'[X_v] \neq T$, then we can contradict the minimality of G' by deleting all edges in $E(G'[X_v]) \setminus E(T)$.

(2) G contains at least one of $K_5, K_{3,3}$ as a minor.

Proof. By Lemma 10.2.3, (1) implies (2). Suppose now that (2) holds; we must show that (1) holds. If $K_{3,3} \preceq_t G$, then Lemma 10.2.4 implies that $K_{3,3} \preceq_t K_{3,3}$, and we are done. Suppose now that $K_5 \preceq_m G$. Our goal is to show that either $K_5 \preceq_t G$ or $K_{3,3} \preceq_m G$.[19]

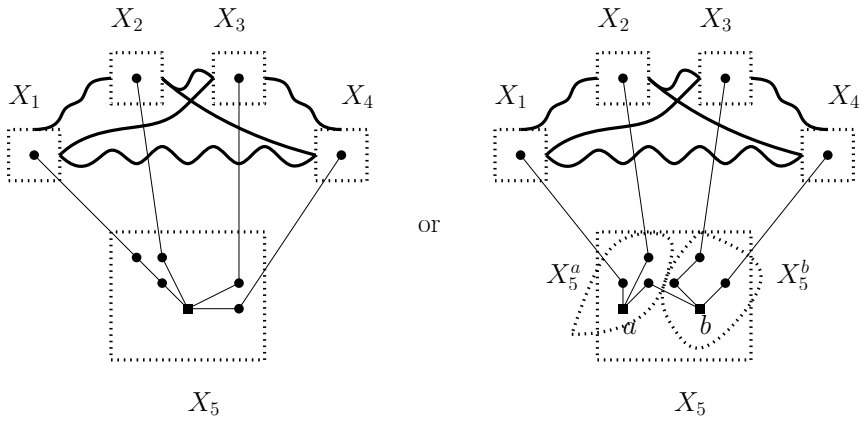
Let G' be a minimal subgraph of G such that $K_5 \preceq_m G'$. Let $X_1, \ldots, X_5$ be the branch sets of the K_5 minor in G'.[20] By the minimality of G', we have that $G'[X_1], \ldots, G'[X_5]$ are all trees, and for all distinct $i, j \in \{1, \ldots, 5\}$, there is exactly one edge between X_i and X_j in G'. For each $i \in \{1, \ldots, 5\}$, let T_i be the graph obtained from $G'[X_i]$ by adding the edges between X_i and $V(G') \setminus X_i$ (and the endpoints of those edges). Then for each $i \in \{1, \ldots, 5\}$, T_i is a tree with exactly four leaves (each one of $X_1, \ldots, X_5$, other than X_i, contains exactly one of those four leaves), and we deduce that T_i is a subdivision of one of the following two trees.

$K_{1,4}$ T

If $T_1, \ldots, T_5$ are all subdivisions of $K_{1,4}$ (see the picture below, on the left), then it is clear that G' is a subdivision of K_5, and it follows that $K_5 \preceq_t G$. Suppose now that at least one of $T_1, \ldots, T_5$ is a subdivision of T (see the picture below, on the right); by symmetry, we may assume that T_5 is a subdivision of T, and we let a, b be the two vertices of T_5 of degree three (note that $a, b \in X_5$). Now let X_5^a be the set of all vertices $v \in X_5$ such that the (unique) path between v and a in the tree T_5 does not contain the vertex b, and let $X_5^b := X_5 \setminus X_5^a$. Then $a \in X_5^a$ and $b \in X_5^b$, and it is easy to see that G contains a $K_{3,3}$ minor with branch sets $X_1, \ldots, X_4, X_5^a, X_5^b$. But now Lemma 10.2.4 implies that $K_{3,3} \preceq G$, and we are done.

[19] Note that this is enough. Indeed, if $K_5 \preceq_t G$, then we are done. And if $K_{3,3} \preceq_m G$, then Lemma 10.2.4 guarantees that $K_{3,3} \preceq_t G$, and again we are done.

[20] So, $G'[X_1], \ldots, G'[X_5]$ are connected, and for all distinct $i, j \in \{1, \ldots, 5\}$, there is an edge between X_i and X_j in G'.



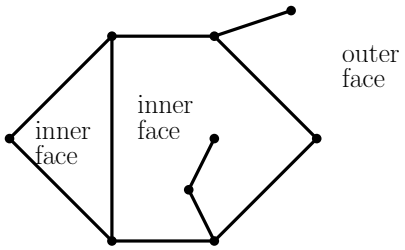
□

10.3 Planar graphs

A graph is *planar* if it can be drawn in the plane without any edge crossings.

Obviously, a graph can be drawn in the plane without any edge crossings if and only if it can be drawn on the sphere without any edge crossings. So, planar graphs are precisely those that can be drawn on the sphere without any edge crossings.

When we draw a graph in the plane without edge crossings, we divide the plane into regions, called *faces*; one of the faces, called the *outer face* is unbounded, and the remaining faces (called *inner faces*) are bounded.



We can define faces on the sphere analogously, but in this case, all faces are bounded, and we get no asymmetry between the inner faces and the outer face. For this reason, for the purposes of proving theorems, it is often more convenient to draw on the sphere than on a plane.

Lemma 10.3.1. *If a graph is planar, then so are all its minors.*

Proof. Clearly, any graph obtained from a planar graph by deleting one vertex, deleting one edge, or contracting one edge is planar. So, by Lemma 10.2.1, all minors of a planar graph are planar. □

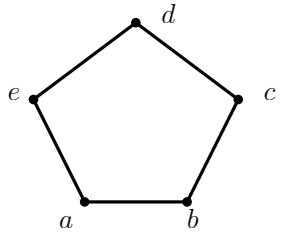
A *homeomorphism* of the sphere is a bijection f from the sphere to itself such that both f and f^{-1} are continuous. Informally, a homeomorphism of the sphere is the result of "stretching" the sphere (and possibly also rotating and taking mirror images).

Two graph drawings on the sphere are *equivalent* if some sphere homeomorphism transforms one drawing into the other.

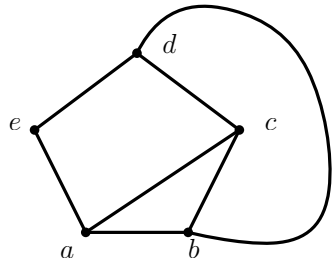
Lemma 10.3.2. *Graphs K_5 and $K_{3,3}$ are not planar. Consequently, no planar graph contains K_5 or $K_{3,3}$ as a minor.*

Proof. In view of Lemma 10.3.1, it suffices to show that K_5 and $K_{3,3}$ are not planar. We will show that K_5 is not planar. The proof is similar for $K_{3,3}$, and we leave it as an exercise.

Suppose that K_5 is planar, so that we can draw it on the sphere without any edge crossings. Let $\{a, b, c, d, e\}$ be the vertex set of the K_5. We first draw the 5-cycle a, b, c, d, e, a on the sphere.

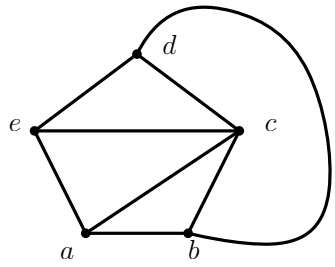


Since edges ac and bd do not cross, we must draw them through distinct faces created by our 5-cycle a, b, c, d, e, a, and we obtain the following.[21]

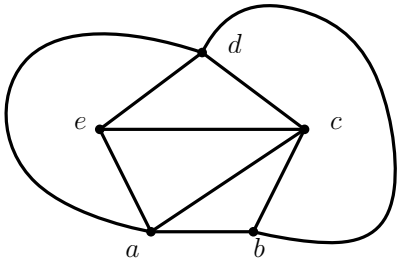


[21] Remember, we are on the sphere! So, we have full symmetry between the two faces produced by the 5-cycle a, b, c, d, e, a.

There is now only one way to add the edge ce to our drawing without creating edge crossings, as shown below.



Further, there is only one way to add the edge ad to our drawing without creating edge crossings, as shown below.



But now it is not possible to add the edge be to our drawing without creating edge crossings. So, K_5 is not planar. □

The following theorem is usually referred to as "Kuratowski's theorem," or sometimes as the "Kuratowski-Wagner theorem."

Theorem 10.3.3 (Kuratowski, 1930; Wagner, 1937)**.** *Let G be a graph. Then the following are equivalent:*

(a) G is planar;

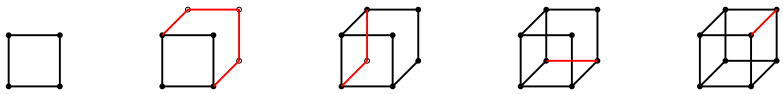
(b) G contains neither K_5 nor $K_{3,3}$ as a minor;

(c) G contains neither K_5 nor $K_{3,3}$ as a topological minor.

We have already proven the "easy" part of Kuratowski's theorem: (a) implies (b) by Lemma 10.3.2, and (b) is equivalent to (c) by Lemma 10.2.5. It remains to prove the "hard" part: (b) implies (a).

A *path addition* (sometimes called *open ear addition*) to a graph H is the addition to H of a path between two distinct vertices of H in such a way that no internal vertex and no edge of the path belongs to H. In the picture

below, we show how the cube graph can be constructed by starting with a cycle of length four and then repeatedly adding paths (the path added at each step is in red).



The following was proven in section 5.3.

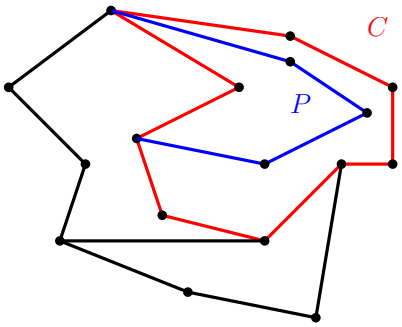
The Ear lemma. *A graph is 2-connected if and only if it is a cycle or can be obtained from a cycle by repeated path addition.*

A *plane drawing* of a planar graph is a drawing of that graph in the plane without any edge crossings.

Lemma 10.3.4. *For any plane drawing of a planar 2-connected graph G, the boundary of each face is a cycle of G.*

Proof. We proceed by induction on the number of edges. Let G be a planar 2-connected graph, and assume inductively that for all planar 2-connected graphs H such that $|E(H)| < |E(G)|$, in any plane drawing of H, the boundary of each face is a cycle of H.

Now, fix a plane drawing of G. If G is a cycle, then the drawing has two faces, and they are both bounded by the cycle G.[22] Suppose now that G is not a cycle. Then the Ear Lemma guarantees that G can be obtained from a 2-connected graph H by adding a path P. If we erase all the edges and all the internal vertices of P from our drawing of G, we obtain a plane drawing of H; by the induction hypothesis, each face of this drawing is bounded by a cycle of H.



[22] Actually, this is somewhat informal. The formal proof requires a theorem from topology called the "Jordan Curve Theorem." We omit the details.

We now put P back into our drawing. The path P must pass through one face of our drawing of H, and it splits this face up into two, each bounded by a cycle of G; the other faces and their boundaries remain unchanged. This completes the proof. □

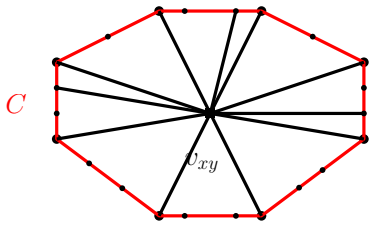
We now prove the "(b) $\Longrightarrow$ (a)" part of Kuratowski's theorem for the case when G is 3-connected. The general case is handled by Lemma 10.3.6, and its proof relies on Lemma 10.3.5.

Lemma 10.3.5. *Let G be a 3-connected graph that contains neither K_5 nor $K_{3,3}$ as a minor. Then G is planar.*

Proof. We may assume inductively that the lemma is true for graphs on fewer than $|V(G)|$ vertices, that is, that for all 3-connected graphs H with $|V(H)| < |V(G)|$ and $K_5, K_{3,3} \not\preceq_m H$, we have that H is planar.

Since G is 3-connected, we know that either $G \cong K_4$ or $|V(G)| > 4$.[23] If $G \cong K_4$, then it is clear that G is planar, and we are done. So assume that $|V(G)| > 4$. Then Lemma 10.1.2 guarantees that G has an edge xy such that $H := G/xy$ is 3-connected. By Lemma 10.2.1, we know that $H \preceq_m G$; since $K_5, K_{3,3} \not\preceq_m G$, Lemma 10.2.2 guarantees that $K_5, K_{3,3} \not\preceq_m H$. Now H is a 3-connected graph on $|V(G)| - 1$ vetrices, with $K_5, K_{3,3} \not\preceq_m H$; so, by the induction hypothesis, H is planar.

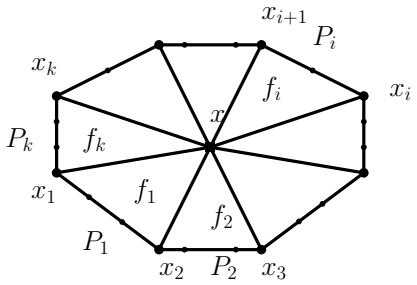
Fix a plane drawing of H. If we erase v_{xy} and all the edges incident in it, we obtain a plane drawing of $H \setminus v_{xy}$. Now, let f be the face of this drawing of $H \setminus v_{xy}$ such that v_{xy} is in the interior of f. Since H is 3-connected, $H \setminus v_{xy}$ is 2-connected; so, by Lemma 10.3.4, the boundary of f is a cycle of $H \setminus v_{xy}$, say C. (Note that C is also a cycle of H and of G.)



Then $N_H(v_{xy}) \subseteq V(C)$, and consequently, $N_G(x) \subseteq \{y\} \cup V(C)$ and $N_G(y) \subseteq \{x\} \cup V(C)$. Since G is 3-connected, Theorem 5.1.3 guarantees that $\delta(G) \geq 3$, and in particular, $d_G(x) \geq 3$; so, since $N_G(x) \subseteq \{y\} \cup V(C)$, x has at least two neighbors in $V(C)$. Let $x_1, \ldots, x_k$ be the neighbors of x in $V(C)$, listed

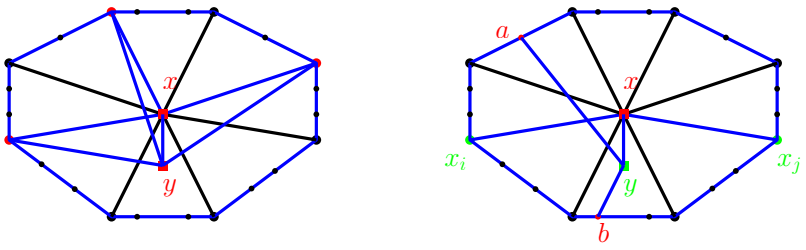
[23] Indeed, since G is 3-connected, we know that $|V(G)| \geq 4$, and clearly, K_4 is (up to isomorphism) the only 3-connected graph on four vertices.

in cyclical order (along the cycle C). For each $i \in \{1, \ldots, k\}$, let P_i be the path from x_i to x_{i+1} (we consider $x_{k+1} = x_1$) along C, as in the picture below.



We now draw $G \setminus y$ in the plane without any edge crossings, as follows. We begin with our drawing of $H = G/xy$, we relabel v_{xy} as x, and we erase the edges between x and $V(C)$ that do not belong to $E(G)$. For each $i \in \{1, \ldots, k\}$, let f_i be the face whose boundary is $x, x_i - P_i - x_{i+1}, x$ and that lies inside f. Our goal is to show that this drawing can be extended to G. If for some $i \in \{1, \ldots, k\}$, we have that $N_G(y) \subseteq \{x\} \cup V(P_i)$, then we simply place the vertex y inside the face f_i, and we draw the edge xy as well as the edges between y and its neighbors in $V(P_i)$, and we obtain a plane drawing of G.

So, suppose that for all $i \in \{1, \ldots, k\}$, we have that $N_G(y) \not\subseteq \{x\} \cup V(P_i)$. Then either x and y have three common neighbors in $V(C)$ (see the picture below, on the left), or y has two neighbors $a, b \in V(C)$ that are separated in C by two neighbors of x, say x_i and x_j (see the picture below, on the right).[24]



In the former case, G contains K_5 as a topological minor (with x, y, and their three common neighbors in C as branch vertices), contrary to the fact that $K_5 \not\preceq_m G$.[25] In the latter case, $G[\{x, y\} \cup V(C)]$ contains $K_{3,3}$ as a

[24] In the second case, it is possible that x is adjacent to one of both of a, b. However, $\{a, b\} \cap \{x_i, x_j\} = \emptyset$.

[25] We are using the fact that, by Lemma 10.2.3, $K_5 \preceq_t G$ implies $K_5 \preceq_m G$.

topological minor, with x, y, a, b, x_i, x_j as the branch vertices,[26] contrary to the fact that $K_{3,3} \not\preceq_m G$.[27] □

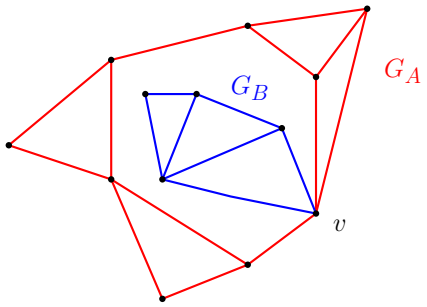
Lemma 10.3.6. *Let G be a graph that contains neither K_5 nor $K_{3,3}$ as a minor. Then G is planar.*

Proof. We may assume inductively that for all graphs H on fewer than $|V(G)|$ vertices, if $K_5, K_{3,3} \not\preceq_m H$, then H is planar.

If $|V(G)| \leq 3$, then it is clear that G is planar. From now on, we assume that $|V(G)| \geq 4$.

Suppose first that G is disconnected, and let $G_1, \ldots, G_t$ be the components of G. Then by the induction hypothesis, $G_1, \ldots, G_t$ are all planar. We obtain a plane drawing of G by drawing $G_1, \ldots, G_t$ in the plane side by side.

Next, suppose that G is connected, but not 2-connected. Then there exists a vertex $v \in V(G)$ such that $G \setminus v$ is disconnected. Let A be the vertex set of one component of $G \setminus v$, and let $B := V(G) \setminus (A \cup \{v\})$. Set $G_A := G[A \cup \{v\}]$ and $G_B := G[B \cup \{v\}]$. By the induction hypothesis, G_A and G_B are both planar. We draw G_A in the plane without any edge crossings, and we let f be some face of this drawing such that v lies on the boundary of f. We then draw G_B inside f, with v coinciding in the drawing of G_A and G_B.[28]



Suppose now that G is 2-connected, but not 3-connected. Since $|V(G)| \geq 4$, the fact that G is not 3-connected guarantees that there is a set $S \subseteq V(G)$ such that $|S| \leq 2$ and $G \setminus S$ is disconnected. Since G is 2-connected, we in fact have that $|S| = 2$; set $S = \{x, y\}$. Let A be the vertex set of some component of $G \setminus S$, and let $B := V(G) \setminus (A \cup S)$. Let $G_A := G[A \cup S] + xy$

[26] Here, $\{x, a, b\}$ and $\{y, x_i, x_j\}$ are the two sides of the bipartition of the subdivided $K_{3,3}$.

[27] We are using the fact that, by Lemma 10.2.3, $K_{3,3} \preceq_t G$ implies $K_{3,3} \preceq_m G$.

[28] This is slightly informal. The point is that we can stretch and shrink our drawing of G_B so that it "fits" inside of f.

and $G_B := G[B \cup S] + xy$.[29] Now, since G is 2-connected, each of $G[A \cup S]$ and $G[B \cup S]$ contains a path between x and y;[30] call these paths P_A and P_B, respectively. Clearly, $G_A \preceq_t G[A \cup S \cup V(P_B)]$ and $G_B \preceq_t G[B \cup S \cup V(P_A)]$; consequently, $G_A, G_B \preceq_t G$, and therefore (by Lemma 10.2.3), $G_A, G_B \preceq_m G$. Since $K_5, K_{3,3} \not\preceq_m G$, Lemma 10.2.2 guarantees that $K_5, K_{3,3} \not\preceq_m G_A$ and $K_5, K_{3,3} \not\preceq_m G_B$. By the induction hypothesis, G_A and G_B are both planar. We now draw G_A in the plane without edge crossings, and we let f be a face of this drawing such that the edge xy lies on the boundary of f. We now draw G_B inside f, with the edge xy coinciding in the drawing of G_A and G_B.[31] This way, we obtain a drawing of $G + xy$ in the plane without any edge crossings;[32] it follows that $G + xy$ is planar, and consequently, G is planar as well.

Finally, if G is 3-connected, then G is planar by Lemma 10.3.5. □

Lemma 10.3.6 proves the "(b) $\Longrightarrow$ (a)" part of Kuratowski's theorem. This completes our proof of Kuratowski's theorem.

10.4 Hajós' Conjecture

In 1961, Hajós conjectured the following.

Hajós' Conjecture. *For every positive integer k, every graph of chromatic number at least k contains K_k as a topological minor.*

Hajós' Conjecture is obviously true for $k = 1$ and $k = 2$. For $k = 3$, we observe that if a graph G satisfies $\chi(G) \geq 3$, then G is not a forest,[33] and in particular, G contains a cycle. Every cycle is a subdivision of K_3, i.e. every cycle contains K_3 as a topological minor. So, if $\chi(G) \geq 3$, then $K_3 \preceq_t G$. Hajós' Conjecture is also true for $k = 4$, as we now show (see Theorem 10.4.2).

A *clique-cutset* of a graph G is a clique $C \subsetneqq V(G)$ of G such that $G \setminus C$ is

[29] So, G_A is the graph with vertex set $A \cup S$ and edge set $E(G[A \cup S]) \cup \{xy\}$; if $xy \in E(G)$, then we simply have $G_A = G[A \cup S]$. Similar remarks apply to G_B.

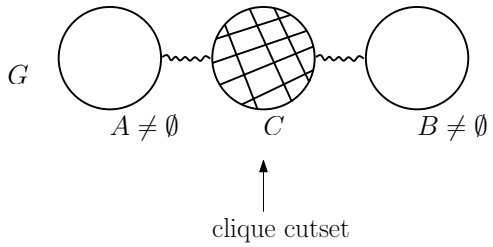
[30] This follows from Proposition 10.1.1. (Details?)

[31] Again, this is slightly informal. The point is that we can stretch and shrink our drawing of G_B so that it "fits" inside of f.

[32] As usual, $G + xy$ is the graph with vertex set $V(G)$ and edge set $E(G) \cup \{xy\}$. If $xy \in E(G)$, then we simply have that $G + xy = G$.

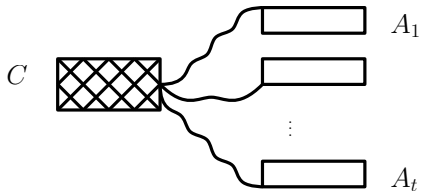
[33] This is because forests are bipartite, and the chromatic number of any bipartite graph is at most two.

disconnected.[34] In particular, if G is disconnected, then $\emptyset$ is a clique-cutset of G.



Lemma 10.4.1. *Let G be a graph, and let C be a clique-cutset of G. Let $A_1, \ldots, A_t$ be the vertex sets of the components of $G \setminus C$. Then $\chi(G) = \max\{\chi(G[A_1 \cup C]), \ldots, \chi(G[A_t \cup C])\}$.*

Proof. To simplify notation, for all $i \in \{1, \ldots, t\}$, set $G_i := G[A_i \cup C]$ and $\chi_i := \chi(G_i)$. We must show that $\chi(G) = \max\{\chi_1, \ldots, \chi_t\}$. It is obvious that $\max\{\chi_1, \ldots, \chi_t\} \leq \chi(G)$. It remains to show that $\chi(G) \leq \max\{\chi_1, \ldots, \chi_t\}$.



For all $i \in \{1, \ldots, t\}$, let $c_i : A_i \cup C \to \{1, \ldots, \chi_i\}$ be a proper coloring of G_i. Since C is a clique of G, we know that for all $i \in \{1, \ldots, t\}$, the coloring c_i assigns distinct colors to all vertices of C. So, after possibly permuting colors, we may assume that $c_1, \ldots, c_t$ all agree on C. But now the union of $c_1, \ldots, c_t$ is a proper coloring of G that uses at most $\max\{\chi_1, \ldots, \chi_t\}$ colors, and we deduce that $\chi(G) \leq \max\{\chi_1, \ldots, \chi_t\}$. □

Theorem 10.4.2 (Dirac, 1952)**.** *Every graph of chromatic number at least* 4 *contains K_4 as a topological minor.*

Proof. Fix a graph G, and assume inductively that for all graphs G' with $|V(G')| < |V(G)|$, if $\chi(G') \geq 4$, then $K_4 \preceq_t G'$. We assume that $\chi(G) \geq 4$, and we show that $K_4 \preceq_t G$. We may assume that all proper induced

[34] In some texts, a *clique-cutset* of G is defined to be a clique $C \subsetneqq V(G)$ of G such that $G \setminus C$ has more components than G. However, the definition that we gave above (requiring only that $G \setminus C$ be disconnected, regardless of the number of components of G) is more convenient for our purposes.

subgraphs of G are 3-colorable,[35] for otherwise, the result follows from the induction hypothesis. In particular, this means that $\chi(G) = 4$.[36]

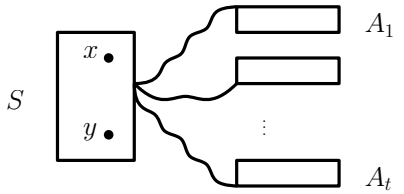
> **Claim 1.** G does not admit a clique-cutset. Furthermore, G is 2-connected.

Proof of Claim 1. The fact that G does not admit a clique-cutset readily follows from Lemma 10.4.1. Indeed, suppose C were a clique-cutset of G, and let $A_1, \ldots, A_t$ be the vertex sets of $G \setminus C$. Then Lemma 10.4.1 guarantees that $\chi(G) = \max\{\chi(G[A_1 \cup C]), \ldots, \chi(G[A_t \cup C])\}$. Since $\chi(G) = 4$, it follows that for some $i \in \{1, \ldots, t\}$, we have that $\chi(G[A_i \cup C]) = 4$, contrary to the fact that all proper induced subgraphs of G are 3-colorable.

Clearly, $|V(G)| \geq \chi(G) = 4$. Furthermore, since G does not admit a clique-cutset, we see that G is connected and has no cut-vertices.[37] So, G is 2-connected. This proves Claim 1. ♦

> **Claim 2.** If G is not 3-connected, then $K_4 \preceq_t G$.

Proof of the Claim. Suppose that G is not 3-connected. Clearly, $|V(G)| \geq \chi(G) = 4$, and so (since G is not 3-connected) there exists a set $S \subseteq V(G)$ such that $|S| \leq 2$ and $G \setminus S$ is disconnected. By Claim 1, we have that $|S| = 2$ (say, $S = \{x, y\}$), and that the two vertices of S are non-adjacent. Let $A_1, \ldots, A_t$ ($t \geq 2$) be the vertex sets of the components of $G \setminus S$, and for each $i \in \{1, \ldots, t\}$, set $G_i := G[A_i \cup S]$. Then $\chi(G_i) \leq 3$ for all $i \in \{1, \ldots, t\}$.[38]



Suppose first that for all $i \in \{1, \ldots, t\}$, there exists a 3-coloring c_i of G_i that assigns distinct colors to x and y.[39] After possibly permuting colors, we may assume that for all $i \in \{1, \ldots, t\}$, we have that $c_i : A_i \cup S \to \{1, 2, 3\}$,

[35] A graph is *k-colorable* if it can be properly colored with at most k colors.

[36] Indeed, if $\chi(G) \geq 5$, then we fix any $v \in V(G)$, and we observe that $\chi(G \setminus v) \geq \chi(G) - 1 \geq 4$, contrary to the fact that all proper induced subgraphs of G are 3-colorable.

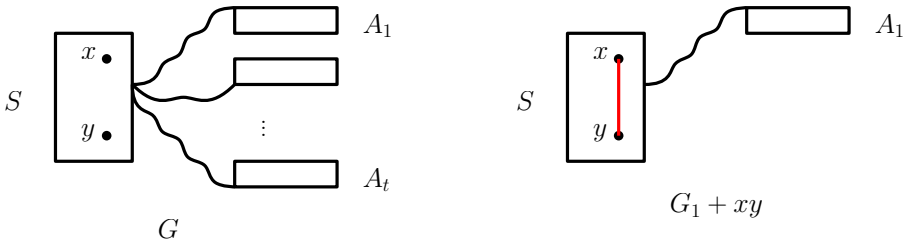
[37] A *cut-vertex* of a connected graph H is a vertex $v \in V(H)$ such that $H \setminus v$ is disconnected. Note that if v is a cut-vertex of H, then $\{v\}$ is a clique-cutset of H.

[38] This is because all proper induced subgraphs of G are 3-colorable.

[39] A *k-coloring* of a graph is a proper coloring of that graph that uses at most k colors.

$c_i(x) = 1$, and $c_i(y) = 2$. But now the union of $c_1, \ldots, c_t$ is a proper coloring of G that uses at most three colors, contrary to the fact that $\chi(G) = 4$.

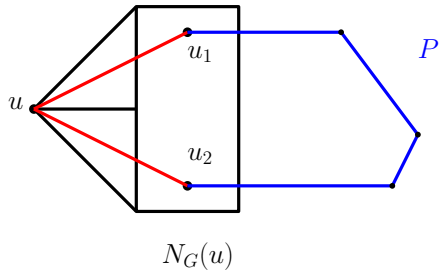
By symmetry, we may now assume that all 3-colorings of G_1 assign the same color to x and y. But then $\chi(G_1 + xy) = 4$.[40] So, by the induction hypothesis, we have that $K_4 \preceq_t G_1 + xy$.



Now, since G is 2-connected, we see that each of x, y has a neighbor in A_2,[41] and so there exists an induced path P in G_2 between x and y. But now $G[A_1 \cup V(P)]$ is a subdivision of $G_1 + xy$, and so $G_1 + xy \preceq_t G$. Since $K_4 \preceq_t G_1 + xy$, we have that $K_4 \preceq_t G$. This proves Claim 2. ♦

> **Claim 3.** If G is 3-connected, then it contains a cycle of length at least four.

Proof of Claim 3. Assume that G is 3-connected. Then Theorem 5.1.3 guarantees that $\delta(G) \geq 3$. Now, fix any vertex u of G; then $d_G(u) \geq \delta(G) \geq 3$. If $N_G(u)$ is a clique, then G contains K_4 as a subgraph,[42] and consequently, G contains a cycle of length four. So, we may assume that some two neighbors of u (call them u_1 and u_2) are non-adjacent.



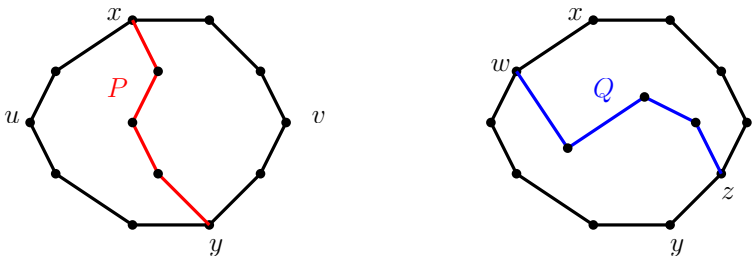
[40] Indeed, since $\chi(G_1) \leq 3$, it is obvious that $\chi(G_1 + xy) \leq 4$. If $\chi(G_1 + xy) \leq 3$, then we fix some 3-coloring of $G_1 + xy$, and we observe that this coloring must assign different colors to x and y (because x and y are adjacent in $G_1 + xy$). But now this coloring is a 3-coloring of G that assigns distinct colors to x and y, a contradiction.

[41] This follows from Proposition 10.1.1.

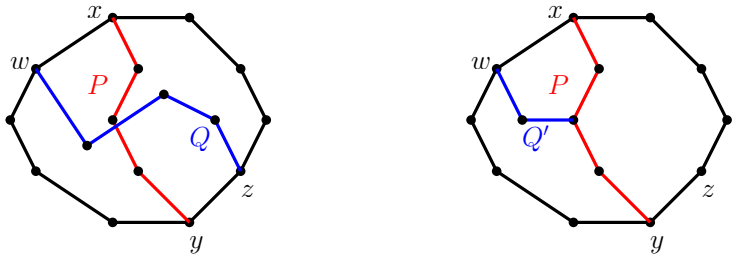
[42] The vertices of our K_4 are u and any three of its neighbors.

Since G is 3-connected, we know that $G \setminus u$ is connected, and consequently, $G \setminus u$ contains a path P between u_1 and u_2. But now $u, u_1 - P - u_2, u$ is a cycle of length at least four in G.[43] This proves Claim 3. ♦

In view of Claim 2, we may assume that G is 3-connected; so, by Claim 3, G contains a cycle C of length at least four. Let u and v be some non-consecutive vertices of C. Since G is 3-connected, we know that $G \setminus \{u, v\}$ is connected; let P be a shortest path in $G \setminus \{u, v\}$ between the two components of $C \setminus \{u, v\}$, and let x and y be the two endpoints of P. (Note that $x, y \in V(C)$, and no internal vertex of P belongs to C. Furthermore, note that x and y are not consecutive vertices of the cycle C.) Since G is 3-connected, $G \setminus \{x, y\}$ is connected; let Q be a shortest path in $G \setminus \{x, y\}$ between the two components of $C \setminus \{x, y\}$, and let w and z be the two endpoints of Q.



Now, if P and Q do not intersect (see the picture below, on the left), then $C \cup P \cup Q$ is a subdivision of K_4,[44] and so $K_4 \preceq_t G$. It remains to consider the case when P and Q do intersect (see the picture below, on the right). Let Q' be the subpath of Q from w to the first intersection point of P and Q. But now $C \cup P \cup Q'$ is a subdivision of K_4, and so $K_4 \preceq_t G$.

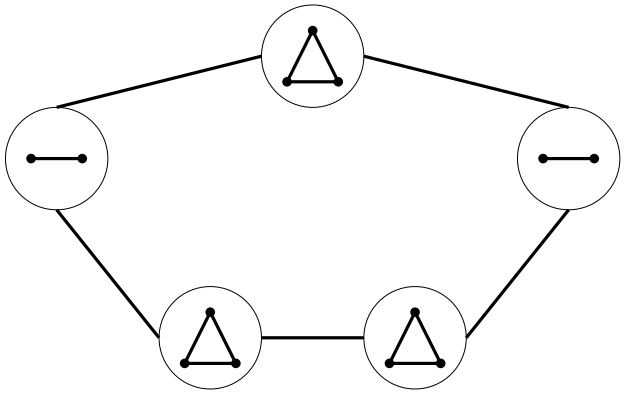


□

[43] We are using the fact that $u_1u_2 \notin E(G)$, and so P has at least one internal vertex.

[44] Here, $C \cup P \cup Q$ is the graph whose vertex set is $V(C) \cup V(P) \cup V(Q)$, and whose edge set is $E(C) \cup E(P) \cup E(Q)$.

In 1979 Catlin proved that Hajós' Conjecture fails for $k \geq 7$, as the example below shows.[45]



Indeed, the graph above has chromatic number 7, and yet it does not contain K_7 as a topological minor.[46] For $k \geq 8$, we can obtain a counterexample to Hajós' Conjecture by adding $k-7$ universal vertices (i.e. vertices adjacent to all other vertices of the graph) to the graph above. Hajós' Conjecture is open for $k = 5$ and $k = 6$.

10.5 Hadwiger's Conjecture

In 1943, Hadwiger conjectured the following.

Hadwiger's Conjecture. *For every positive integer k, every graph of chromatic number at least k contains K_k as a minor.*

Since a topological minor is a special case of a minor (by Lemma 10.2.3), Hadwiger's Conjecture is weaker than Hajós' Conjecture. Thus, since Hajós' Conjecture is true for $k \leq 4$, Hadwiger's conjecture is also true for $k \leq 4$. Hadwiger's Conjecture for $k = 5$ is equivalent to the famous Four Color Theorem (proven by Appel and Haken in 1976), which states that every planar graph is 4-colorable.[47] Further, in 1993, Robertson, Seymour, and Thomas proved that Hadwiger's Conjecture is true for $k = 6$. For $k \geq 7$, the conjecture remains open.

[45] A line between two circles indicates that all vertices inside one of the circles are adjacent to all vertices inside the other circle.

[46] Check this!

[47] The equivalence of Hadwiger's Conjecture for $k = 5$ and the Four Color Theorem is not entirely obvious, though, and we omit the details.

Chapter 11

Graphs on surfaces

11.1 Surfaces

A *surface* is a connected 2-dimensional compact manifold with no boundary. This definition contains several terms we have not defined, and whose formal definition we omit. Here is an intuitive explanation:

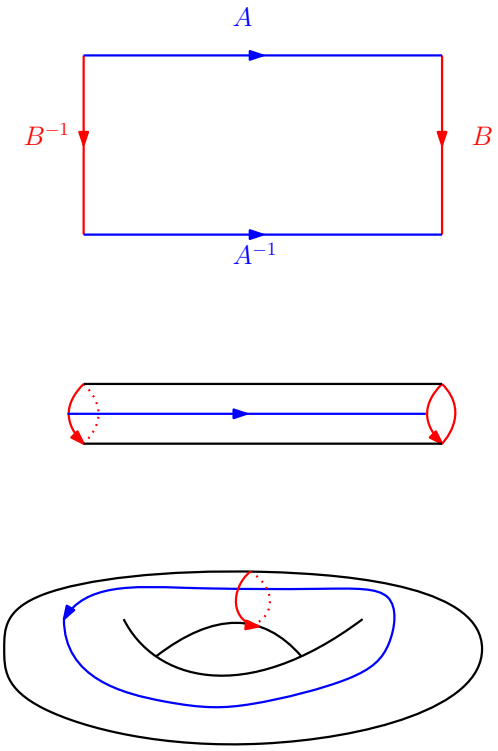
- "2-dimensional manifold with no boundary" means that each point has a neighborhood "homeomorphic" to an open disk (i.e. the neighborhood can be transformed into an open disk by stretching and twisting);
- "compact" means that that the surface admits a triangulation with finitely many triangles;
- "connected" means that there is just one piece.

The sphere and the torus are surfaces. However, the plane is not a surface (because it is not compact). A closed disk is not a surface, either, since it has a boundary.

In what follows, we consider two surfaces to be the "same" if they are "homeomorphic," that is, if there is a bijection f between them such that both f and f^{-1} are continuous. So, if we can obtain one surface from the other by stretching and twisting, then the two surfaces are the same. Thus, a tetrahedron is simply a sphere for our purposes, but a torus is not a sphere.

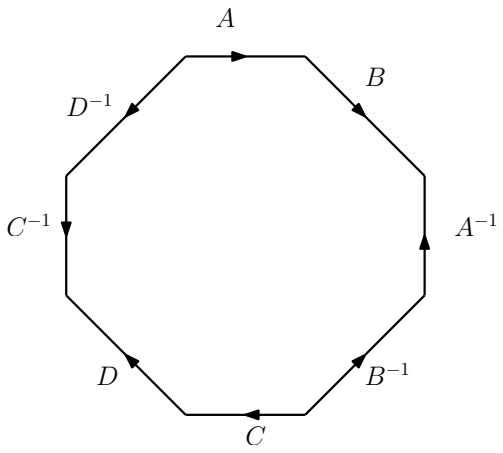
Here is one way of forming a torus: we start with a rectangle (see the picture below), and then we identify the two (directed) blue edges and the two (directed) red edges. Importantly, the corresponding edges must be identified in the direction represented by the arrows. (In the picture below, we first identify the blue edges to get a "tube," and then we identify the two red

edges/circles to get a torus. In our picture, the four vertices of the rectangle all get identified to the same point on the torus.) Note that the blue edges are labeled A (for clockwise direction) and A^{-1} (for counterclockwise direction); a similar labeling applies to B and B^{-1}. Symbolically, the rectangle is represented by the string $ABA^{-1}B^{-1}$.

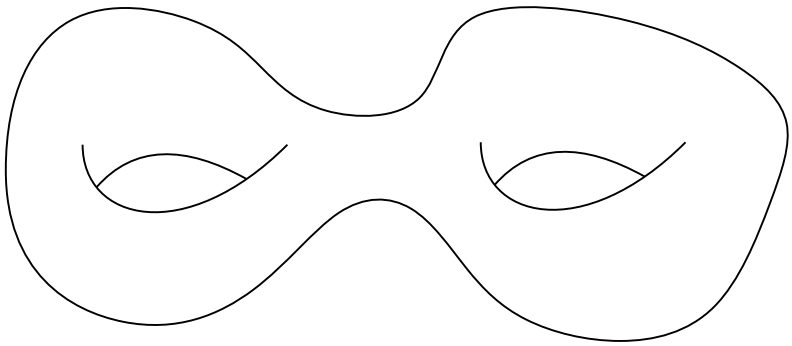


If we identify corresponding edges in the octagon $ABA^{-1}B^{-1}CDC^{-1}D^{-1}$ below, then we get a double torus (also called the "connected sum of two tori"), as you can check.[1]

[1] Alternatively, you can watch this video: `https://www.youtube.com/watch?v=G1yyfPShgqw` (accessed September 2022).

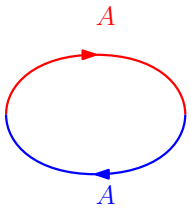


$ABA^{-1}B^{-1}CDC^{-1}D^{-1}$



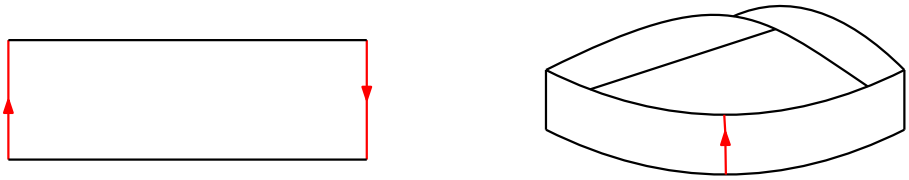
double torus

The *real projective plane* (or simply *projective plane*) is the surface obtained by starting from the sphere, and then identifying each pair of antipodal points. The projective plane has polygonal representation AA (see below). Here, we have two A's (red and blue); they are to be identified in the direction indicated.



Unlike the torus, the projective plane cannot be embedded in $\mathbb{R}^3$. Still, there is a geometric interpretation. Take a rectangle shown below on the left (think

of it as a piece of paper), twist it, and identify the two vertical edges (as shown by the arrows). The result (on the bottom right) is called the *Möbius strip*. Note that the boundary of the Möbius strip consists of just one circle, and the Möbius strip has just one "side."

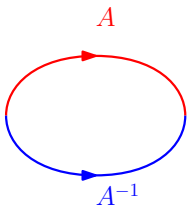


Now, take a sphere, cut out a small disk from it, and then glue the Möbius strip along the boundary obtained by removing the disk. The result is the projective plane (the circle of the Möbius strip corresponds to the edge A from our AA representation of the projective plane). Equivalently, if we cut out a disk from the projective plane, we obtain the Möbius strip.[2] The projective plane is a type of "non-orientable" surface, which roughly means that we cannot set up a left-right distinction. Intuitively, imagine a two-dimensional bug on the surface of the Möbius strip (which is part of the projective plane). If the bug keeps going forward, it will eventually come back to the same place (and facing in the same direction), but with left and right reversed. This sort of thing is impossible on "orientable" surfaces such as the sphere or the torus (or double torus, triple torus, etc.).

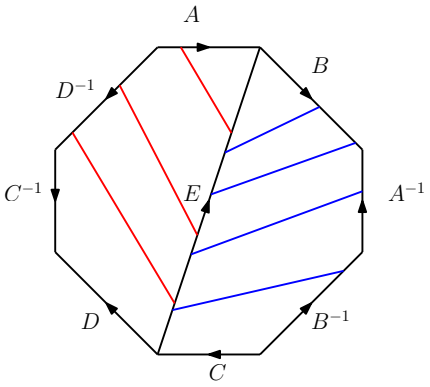
Now, any $2n$-gon, with edges labeled $A_1, \ldots, A_n$ (in any order), with each letter appearing exactly twice on the $2n$-gon, either in the form A (for clockwise direction) or A^{-1} (for counterclockwise direction) can be transformed into a surface via gluing using the rules described above.[3] Some labellings are equivalent. For example, the two octagons below obviously

[2] It is not necessarily obvious that these three descriptions (sphere with antipodal points identified; polygonal AA representation; and Möbius strip with a disk) of the projective plane are equivalent, i.e. that they yield the same surface. For an animation that explains this, see this video: `https://www.youtube.com/watch?v=u0VkikpElMo` (accessed September 2022).

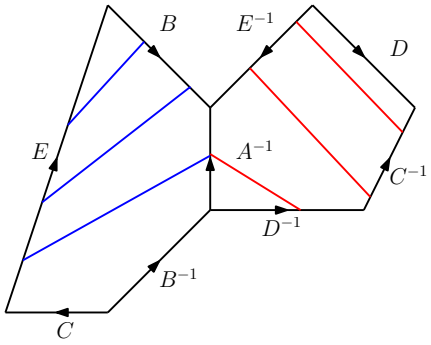
[3] Note: AA^{-1} is simply the sphere.



"encode" the same surface (i.e. after gluing, we get the same surface, in this case, the double torus). We first "cut" and then "glue" the polygon on the left in order to obtain the polygon on the right.[4]

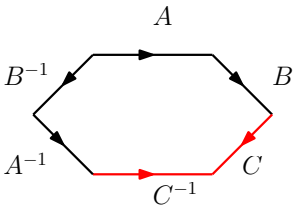


$ABA^{-1}B^{-1}CDC^{-1}D^{-1}$

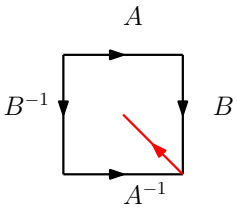


$BE^{-1}DC^{-1}D^{-1}B^{-1}CE$

Sometimes, we have "unnecessary" letters/edges in our polygon, as the picture below shows. Both polygons represent a torus.



$ABCC^{-1}A^{-1}B^{-1}$



$ABA^{-1}B^{-1}$

An argument resembling the one indicated by the pictures above yields the following "classification theorem" (whose proof we omit).

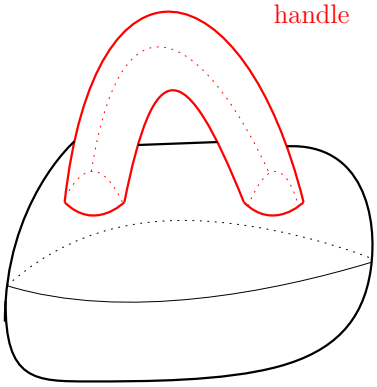
Theorem 11.1.1. *Every surface has a polygonal representation of one of the following forms:*

- AA^{-1};
- $(A_1B_1A_1^{-1}B_1^{-1})(A_2B_2A_2^{-1}B_2^{-1})\dots(A_kB_kA_k^{-1}B_k^{-1})$;
- $(A_1A_1)(A_2A_2)\dots(A_kA_k)$.

[4]Note that the red pentagons have different shapes. This is due to the drawing software, but this is not important: we are allowed to "strech" and "shrink" any way we like.

Importantly, Theorem 11.1.1 does **not** state that each polygonal representation of a surface has one of the forms from the theorem. As a matter of fact, each surface has infinitely many polygonal representations.[5] Theorem 11.1.1 merely states that for each surface S, one of its representations has a "canonical" from (i.e. one of the forms from the theorem).

We remark that the surface with polygonal representation AA^{-1} is simply the sphere. Further, the surface having a polygonal representation $(A_1B_1A_1^{-1}B_1^{-1})(A_2B_2A_2^{-1}B_2^{-1})\dots(A_kB_kA_k^{-1}B_k^{-1})$ is the "connected sum of k tori," i.e. a torus with k holes. This type of torus can be obtained from a sphere by adding k "handles" to a sphere. Adding a handle to a surface we have already constructed consists of removing two small disks from the surface, and then connecting them via a "handle" (a tube). Spheres and connected sums of tori are "orientable surfaces." The *genus* of the sphere is zero, and the *genus* of a connected sum of k tori is k.

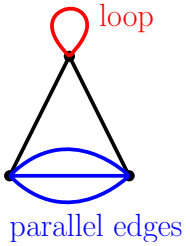


Finally, $(A_1A_1)(A_2A_2)\dots(A_kA_k)$ represents the "connected sum of k real projective planes." We can obtain this surface by starting with a sphere, removing k disks, and then gluing a Möbius strip along the boundary of each removed disk in the sphere. Adding one Möbius strip in this way is called "adding a crosscap." So, $(A_1A_1)(A_2A_2)\dots(A_kA_k)$ is the surface obtained from the sphere by adding k crosscaps. Connected sums of projective planes are "non-orientable surfaces." The *genus* of a connected sum of k real projective planes is k.

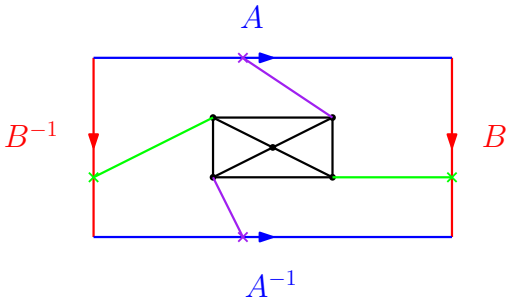
[5] Indeed, for any polygonal representation of a surface, we can obtain another polygonal representation by adding AA^{-1} (where A is a "new" letter) to the end. We can repeat the procedure arbitrarily many times, thus creating infinitely many polygonal representaitons of the same surface.

11.2 Graph drawing on surfaces

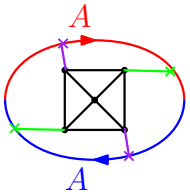
A "multigraph" is a graph thay may possibly have loops and parallel edges.



Just as we can draw graphs (and multigraphs) on the sphere, we can draw them on any other surface. Generally speaking, it is more convenient to use polygonal representation for drawing, than to draw directly on the surface in question. For instance, below is a drawing of K_5 on the torus. (Note how the green edge and the purple edge "wrap around" the rectangle.)



Further, below is a drawing of K_5 on the projective plane (once again, note how the green edge and the purple edge wrap around).



The following was proven in Discrete Math.[6]

[6]Perhaps you saw the proof of the Euler polyhedral formula only for graphs (not multigraphs). But note that any multigraph can be turned into a graph by edge subdivision, which does not alter the expression $V - E + F$ (because subdividing an edge once increases both the number of vertices and the number of edges by one, while leaving the number of faces unchanged). So, we can easily derive the multigraph version of the Euler polyhedral formula from the graph version.

Euler polyhedral formula. *Let G be any connected planar multigraph. Then for any drawing of G on the sphere (without edge crossings), we have that*

$$V - E + F = 2,$$

where V is the number of vertices, E the number of edges, and F the number of faces of the drawing.

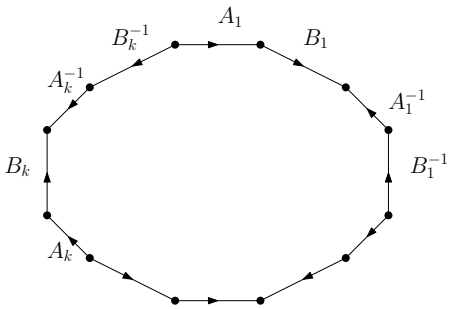
A *net* on a surface is a multigraph drawing on that surface (with no edge crossings) in which every face is homeomorphic to an open disk. Note that the net (or rather, the multigraph whose drawing it is) must be connected. Our next theorem is a generalization of the Euler polyhedral formula for surfaces of arbitrary genus. Before stating and proving the theorem, we make an observation. If G is a net on a surface S, then subdividing an edge ℓ times does not change the value of $V - E + F$ (because both V and E increase by ℓ, and F remains unchanged). Further, adding an edge between two existing vertices and passing through a face of the net does not change $V - E + F$ (because both E and F increase by one, and V remains unchanged; here, we are using the fact that each face is homeomorphic to a disk, and so adding an edge between two existing vertices necessarily "splits" an existing face into two).

Theorem 11.2.1. *Let G be a net on an (orientable or non-orientable) surface S of genus k. Let V be the number of vertices, E the number of edges, and F the number of faces of this net. Then:*

(a) if S is orientable, then $V - E + F = 2 - 2k$;

(b) if S is non-orientable, then $V - E + F = 2 - k$.

Proof. (a) Assume that S is orientable. If $k = 0$, then S is the sphere, and we are done by the Euler polyhedral formula. So, we may assume that $k \geq 1$. Then S is the connected sum of k tori, and it has a polygonal representation $(A_1 B_1 A_1^{-1} B_1^{-1}) \ldots (A_k B_k A_k^{-1} B_k^{-1})$.



Note that the $4k$ vertices of this polygon all correspond to the same point of the surface S; we may assume that this point is a vertex of G (if not, just "move" the net a bit until it is). Next, we will assume that G intersects the boundary of the polygon in only finitely many points;[7] we turn each point of intersection of the net and the polygon boundary into a vertex (this is just edge subdivision, and so $V - E + F$ does not change). Finally, we turn the boundary of the polygon into edges (subdivided according to the vertices that appear on the boundary); this produces $2k$ (potentially subdivided) loops in our net,[8] and it does not alter $V - E + F$. We still call the resulting net G, and we let V be the number of vertices, E the number of edges, and F the number of faces of the net.[9]

Now, our net G on the surface S can be "translated" into a plane drawing in the natural way: we simply place our polygon in the plane. Let V_p be the number of vertices of G on S that lie in the interior of the edges of the polygon (so, in our plane drawing, this turns into $2V_p$ vertices, because each such vertex "doubles"), and let E_p be the number of edges of G on S that lie on the polygon (so, in our plane drawing, this turns into $2E_p$ edges, because each such edge "doubles"). Further, one vertex of G got turned into $4k$ vertices (the vertices of the polygon) in our plane drawing. So, the plane drawing has $4k + 2V_p + (V - 1 - V_p) = V + V_p + 4k - 1$ vertices, $2E_p + (E - E_p) = E + E_p$ edges, and $F + 1$ faces (because of the exterior face). Now, the Euler polyhedral formula gives us

$$(V + V_p + 4k - 1) - (E + E_p) + (F + 1) \quad = \quad 2,$$

and therefore,

$$(V - E + F) + (V_p - E_p) \quad = \quad 2 - 4k.$$

But note that $E_p = V_p + 2k$.[10] So, $V - E + F - 2k = 2 - 4k$, and therefore $V - E + F = 2 - 2k$, which is what we needed.

(b) Assume that S is non-orientable; then S is the connected sum of k projective planes. Let $(A_1A_1)\dots(A_kA_k)$ be the polygonal representation of

[7] This part is a bit informal: a full justification of our assumption requires somewhat sophisticated topology.

[8] We have $4k$ edges on the boundary of the polygon, but after identification, they turn into $2k$ loops.

[9] Technically, we have produced a new net G', with corresponding (new) V', E', and F', and we have that $V' - E' + F' = V - E + F$. However, for the sake of notational simplicity, we just write G, V, E, F instead. Importantly, $V - E + F$ has not changed.

[10] Indeed, the polygon has $4k$ edges, which correspond to $2k$ loops on the surface S. Each time we subdivide an edge, we increase both the number of vertices and the number of edges by one.

the surface S. The proof is now almost identical to that of part (a), except that we have a $2k$-gon, rather than a $4k$-gon, and so the computation yields $V - E + F = 2 - k$.[11] □

Corollary 11.2.2. *Let G be a multigraph drawing (with no edge crossings) on a surface S of genus k.*[12] *Let V be the number of vertices, E the number of edges, and F the number of faces of this net. Then:*

(a) if S is orientable, then $V - E + F \geq 2 - 2k$;

(b) if S is non-orientable, then $V - E + F \geq 2 - k$.

Proof. We keep adding edges (possibly loops) to G without creating edge crossings, until we obtain a net.[13] The result is a net, and so the result follows from Theorem 11.2.1. □

The *Euler characteristic* of a surface S, denoted by $\mathrm{ec}(S)$, is the number $V-E+F$, where V, E, and F are, respectively, the number of vertices, edges, and faces of some net on S.[14] By Theorem 11.2.1, this is well defined, i.e. the number $\mathrm{ec}(S)$ depends only on the surface S, and not on the particular choice of a net. Moreover, Theorem 11.2.1 states that if S is an orientable surface of genus k, then $\mathrm{ec}(S) = 2 - 2k$; on the other hand, if S is a non-orientable surface of genus k, then $\mathrm{ec}(S) = 2 - k$.[15] Moreover, Corollary 11.2.2 implies that if G is a multigraph drawing on a surface S, then $V - E + F \geq \mathrm{ec}(S)$, where V, E, and F are the number of vertices, edges, and faces of the drawing.

Note that the following corollary only holds for graphs (and not for multigraphs). In the remainder of this chapter, for emphasis, we refer to graphs without loops and parallel edges as simple graphs.

Corollary 11.2.3. *Let G be a (simple) graph on at least two edges, drawn on an (orientable or non-orientable) surface S (without edge crossings). Then*

$$|E(G)| \leq 3|V(G)| - 3ec(S).$$

Consequently, the average degree of G is at most $6 - \frac{6ec(S)}{|V(G)|}$.

[11] Check the details!

[12] Note: G need not be a net, that is, it is possible that not all faces are homeomorphic to a disk.

[13] Note that this may possibly decrease the value of $V - E + F$, but it cannot increase it.

[14] The Euler characteristic of a surface S is usually denoted by $\chi(S)$. However, some texts use $\mathrm{ec}(S)$, to avoid confusion with the chromatic number. Here, we will use the notation $\mathrm{ec}(S)$.

[15] So, the Euler characteristic of the sphere is 2, and the Euler characteristic of the torus is 0. The Euler characteristic of the projective plane is 1.

Proof. For each face f, we define $\ell(f)$ to be the number of edges incident with f, with each edge incident with f on both sides counting twice. Since G is simple and $|E(G)| \geq 2$, we see that $\ell(f) \geq 3$ for all faces f. If $F(G)$ is the set of all faces, we get

$$2|E(G)| \;=\; \sum_{f \in F(G)} \ell(f) \;\geq\; 3|F(G)|,$$

and therefore,

$$|F(G)| \;\leq\; \tfrac{2}{3}|E(G)|$$

Now we compute

$$\begin{aligned} |E(G)| &\leq |V(G)| + |F(G)| - \mathrm{ec}(S) \qquad \text{by Corollary 11.2.2} \\ &\leq |V(G)| + \tfrac{2}{3}|E(G)| - \mathrm{ec}(S), \end{aligned}$$

and so

$$|E(G)| \;\leq\; 3|V(G)| - 3\mathrm{ec}(S).$$

Finally, since the average degree of G is $\frac{2|E(G)|}{|V(G)|}$, the inequality above immediately implies that the average degree of G is at most $6 - \frac{6\mathrm{ec}(S)}{|V(G)|}$. □

11.3 The Heawood number

For an integer $c \leq 2$, we define the *Heawood number* as follows:

$$H(c) \;:=\; \left\lfloor \tfrac{7+\sqrt{49-24c}}{2} \right\rfloor.$$

We remark that for the case when S is a sphere, the proof of our next theorem uses the (highly non-trivial) Four Color Theorem, which states that every planar graph is 4-colorable, i.e. has chromatic number at most four. If S is any other surface, then the proof is relatively elementary.

Theorem 11.3.1. *If a (simple) graph G can be drawn without edge crossings on an (orientable or non-orientable) surface S, then $\chi(G) \leq H(ec(S))$.*

Proof. Fix a surface S and a graph G that can be drawn on S without edge crossings. To simplify notation, set $c := \mathrm{ec}(S)$. We must show that $\chi(G) \leq H(c)$.

Suppose first that S is the sphere, so that G is a planar graph and $c = 2$. By the Four Color Theorem, G is 4-colorable. On the other hand, $H(c) = H(2) = 4$. So, $\chi(G) \leq 4 = H(c)$.

From now on, we may assume that S is not a sphere, so that $c \leq 1$. Suppose that there exists a graph G that can be drawn on S without edge crossings, but satisfies $\chi(G) > H(c)$; we may assume that G was chosen with as few vertices as possible. Set $n := |V(G)|$; clearly, $n \geq \chi(G) \geq H(c) + 1$. Moreover, $\delta(G) \geq H(c)$, for otherwise, we fix a vertex $v \in V(G)$ such that $d_G(v) \leq H(c) - 1$, we color $G \setminus v$ with $H(c)$ colors (this is possible by the minimality of n), and then we extend this coloring to a proper coloring of G using at most $H(c)$ colors by assigning to v a color not used on any of its neighbors.[16] On the other hand, by Corollary 11.2.3, the average degree in G is at most $6 - \frac{6c}{n}$.[17] So,

$$H(c) \quad \leq \quad 6 - \tfrac{6c}{n}.$$

Since $H(1) = 6$, the inequality above does not hold if $c = 1$. So, $c \leq 0$. Since $n \geq H(c) + 1 > 0$, it follows that $-\frac{6c}{n} \leq -\frac{6c}{H(c)+1}$, and consequently,

$$H(c) \quad \leq \quad 6 - \tfrac{6c}{H(c)+1}.$$

Since $H(c) > 0$, the above is equivalent to

$$H(c)^2 - 5H(c) + 6(c-1) \quad \leq \quad 0.$$

By solving the corresponding quadratic equation, we now get that

$$\tfrac{5-\sqrt{49-24c}}{2} \quad \leq \quad H(c) \quad \leq \quad \tfrac{5+\sqrt{49-24c}}{2}.$$

But from the definition of $H(c)$, we have that

$$H(c) \quad = \quad \left\lfloor \tfrac{7+\sqrt{49-24c}}{2} \right\rfloor \quad > \quad \tfrac{7+\sqrt{49-24c}}{2} - 1 \quad = \quad \tfrac{5+\sqrt{49-24c}}{2},$$

a contradiction. □

The *Klein bottle* is the surface with polygonal representation $AABB$ or $ABAB^{-1}$.[18] Note that the Klein bottle is non-orientable (and therefore cannot be embedded in $\mathbb{R}^3$);[19] it has genus 2 and Euler characteristic 0.

[16] This contradicts our assumption that $\chi(G) > H(c)$.

[17] Since $\chi(G) \geq H(c)+1 > 2$, we see that G has at least two edges, and so the hypotheses of Corollary 11.2.3 are indeed satisfied.

[18] Check that these are equivalent!

[19] However, a geometric representation of the Klein bottle is possible in $\mathbb{R}^3$, provided we allow the surface to intersect itself. The key is to remember that the intersection is not "really" there, but is simply a feature of our attempt to represent the surface in $\mathbb{R}^3$. For a video, see here: https://www.youtube.com/watch?v=yaeyNjUPVqs (accessed September 2022).

Theorem 11.3.2 (Ringel and Youngs)**.** *If S is a surface other than the Klein bottle, then the complete graph $K_{H(ec(S))}$ can be drawn on S (without edge crossings).*

We omit the proof of Theorem 11.3.2. Note, however, that Theorem 11.3.2 proves that the bound established in Theorem 11.3.1 is best possible, except when S is the Klein bottle. For the Klein bottle, we can get a better bound. Recall that the Euler characteristic of the Klein bottle is 0, and note that $H(0) = 7$. However, as we shall see, the maximum chromatic number of a graph that can be drawn on the Klein bottle without edge crossings is 6 (see Theorem 11.3.4). We begin with the following Lemma, whose proof we omit.

Lemma 11.3.3. *K_6 can be drawn on the Klein bottle (without edge crossings), but K_7 cannot.*

We will also need Brooks' theorem (stated below), whose proof will be given in chapter 12. As usual, $\Delta(G)$ is the maximum degree of a graph G, i.e. $\Delta(G) := \max\{d_G(v) \mid v \in V(G)\}$.

Brooks' theorem. *Let G be a connected graph that is neither a complete graph nor an odd cycle. Then $\chi(G) \leq \Delta(G)$.*

Theorem 11.3.4. *Let G be a graph that can be drawn on the Klein bottle (without edge crossings). Then $\chi(G) \leq 6$.*

Proof. Suppose otherwise, i.e. suppose $\chi(G) \geq 7$. We may assume that, among all graphs that can be drawn on the Klein bottle but are not 6-colorable, G has the smallest possible number of vertices. Note that this means that $\delta(G) \geq 6$.[20] On the other hand, since the Klein bottle has Euler characteristic 0, Corollary 11.2.3 guarantees that G has average degree at most 6. But this is possible only if G is 6-regular. Now, by the minimality of $|V(G)|$, we know that G is connected.[21] Since $\chi(G) \geq 7$, Brooks' theorem guarantees that $G \cong K_7$. But this contradicts Lemma 11.3.3. □

[20] Indeed, suppose G has a vertex v of degree at most five. Then $G \setminus v$ is 6-colorable (by the minimality of $|V(G)|$). We then fix a proper coloring of G with at most six colors, and we extend it to a proper coloring of G with at most six colors by assigning to v a color not used on any of its neighbors. This contradicts our assumption that $\chi(G) \geq 7$.

[21] Otherwise, we take H to be a component of G such that $\chi(H) = \chi(G)$, and we observe that H contradicts the minimality of $|V(G)|$.

Chapter 12

Vertex and edge coloring: Brooks' theorem and Vizing's theorem

12.1 Vertex coloring: Brooks' theorem

Recall that a *proper vertex-coloring* (or simply *proper coloring*) of a graph G is an assignment of colors to the vertices of G in such a way that no two adjacent vertices receive the same color. For an integer k, a *k-vertex-coloring* (or simply *k-coloring*) of G is a proper coloring $c : V(G) \to C$, where C is some set of k colors (typically, $C = \{1, \ldots, k\}$). G is *k-colorable* if it admits a k-coloring. The *chromatic number* of G, denoted by $\chi(G)$, is the smallest integer k such that G is k-colorable. An *optimal vertex-coloring* (or simply *optimal coloring*) of G is a proper coloring of G that uses only $\chi(G)$ colors.

12.1.1 A lower bound for the chromatic number

Recall that a *clique* of G is a set of pairwise adjacent vertices of G, and a *stable set* (or *independent set*) of G is a set of pairwise non-adjacent vertices of G. The *clique number* of G, denoted by $\omega(G)$, is the maximum size of any clique of G. The *stability number* (or *independence number*) of G, denoted by $\alpha(G)$, is the maximum size of a stable set of G.

Note that any proper coloring of a graph G can be thought of as a partition of $V(G)$ into stable sets ("color classes"). Indeed, if $c : V(G) \to \{1, \ldots, k\}$ is a proper coloring of G, then for each $i \in \{1, \ldots, k\}$, we set $S_i := \{v \in V(G) \mid c(v) = i\}$, and we observe that $(S_1, \ldots, S_k)$ is a partition

of $V(G)$ into stable sets. Conversely, any partition $(S_1, \ldots, S_k)$ of $V(G)$ corresponds to a proper coloring of G (indeed, for each $i \in \{1, \ldots, k\}$, we assign color i to all vertices in S_i).

Proposition 12.1.1. *Every graph G satisfies*[1]

$$\chi(G) \geq \max\{\omega(G), \left\lceil \frac{|V(G)|}{\alpha(G)} \right\rceil\}.$$

Proof. Fix a graph G, and set $k := \chi(G)$. Fix an optimal coloring $c : V(G) \to \{1, \ldots, k\}$ of G.

We first show that $\chi(G) \geq \omega(G)$. Fix a clique K of G of size $\omega(G)$. Since K is a clique, c assigns a different color to each vertex of K, and in particular, c uses at least $|K|$ many colors. So, $\chi(G) = k \geq |K| = \omega(G)$.

It remains to show that $\chi(G) \geq \left\lceil \frac{|V(G)|}{\alpha(G)} \right\rceil$. For each $i \in \{1, \ldots, k\}$, set $S_i := \{v \in V(G) \mid c(v) = i\}$. Then $(S_1, \ldots, S_k)$ is a partition of G into stable sets, and it follows that

$$|V(G)| = |S_1 \cup \cdots \cup S_k| = \sum_{i=1}^{k} |S_i| \overset{(*)}{\leq} k\alpha(G) = \chi(G)\alpha(G),$$

where (*) follows from the fact that $S_1, \ldots, S_k$ are all stable sets, and are therefore of size at most $\alpha(G)$. It now follows that $\chi(G) \geq \frac{|V(G)|}{\alpha(G)}$; since $\chi(G)$ is an integer, we deduce that $\chi(G) \geq \left\lceil \frac{|V(G)|}{\alpha(G)} \right\rceil$. □

12.1.2 Greedy coloring and an upper bound for the chromatic number

A *greedy* coloring of a graph G with vertex ordering $V(G) = \{v_1, \ldots, v_n\}$ is a coloring of G obtained as follows: for each $i \in \{1, \ldots, n\}$, we assign to v_i the smallest positive integer that was not used on any smaller-indexed neighbor of v_i.

For example, the greedy coloring applied to the graph below, with the ordering v_1, v_2, v_3, v_4, yields the coloring $c(v_1) = 1$, $c(v_2) = 1$, $c(v_3) = 2$, and $c(v_4) = 3$.

[1] Note that we are implicitly assuming that G is non-null (i.e. has at least one vertex), for otherwise, $\left\lceil \frac{|V(G)|}{\alpha(G)} \right\rceil$ is not defined (we cannot divide by zero).

Note that the greedy coloring of a graph G always produces a proper coloring of G, but the coloring need not be optimal, i.e. it may use more than $\chi(G)$ colors (indeed, this was the case in the example above).

As usual, for a graph G, $\Delta(G)$ is the maximum degree of G, i.e. $\Delta(G) := \max\{d_G(v) \mid v \in V(G)\}$.

Lemma 12.1.2. *Every graph G satisfies $\chi(G) \leq \Delta(G) + 1$.*

Proof. A greedy coloring of a graph G (using any ordering of $V(G)$) produces a proper coloring of G that uses at most $\Delta(G) + 1$ colors; so, $\chi(G) \leq \Delta(G) + 1$. □

If G is a complete graph or an odd cycle, then it is easy to see that $\chi(G) = \Delta(G) + 1$, i.e. the inequality from Lemma 12.1.2 is an equality. However, as we shall see, if G is a connected graph other than a complete graph or odd cycle, then the inequality from Lemma 12.1.2 is strict, i.e. $\chi(G) \leq \Delta(G)$. We prove this in our next section.

12.1.3 Brooks' theorem

We begin with a technical lemma. (Recall that a graph is *regular* if all its vertices are of the same degree.)

Lemma 12.1.3. *If G is connected and not regular, then $\chi(G) \leq \Delta(G)$.*

Proof. Let G be a connected graph that is not regular, and fix a vertex $v \in V(G)$ such that $d_G(v) \leq \Delta(G) - 1$. We order $V(G)$ according to the distance from v, that is, we list v first, then we list all vertices at distance one from v (in any order), then we list all vertices at distance two from v (in any order), etc. Let $v_1, \ldots, v_n$ be the resulting ordering of G. We now color G greedily using the ordering $v_n, \ldots, v_1$.[2] By construction, every vertex in the ordering $v_n, \ldots, v_1$, other than the vertex v_1, has at least one neighbor to the right of it, and therefore at most $\Delta(G) - 1$ neighbors to the left of it. But since $d_G(v) \leq \Delta(G) - 1$, we see that $v_1 = v$ also has at most $\Delta(G) - 1$ neighbors to the left of it in the ordering $v_n, \ldots, v_1$. So, our coloring of G uses at most $\Delta(G)$ colors, and we deduce that $\chi(G) \leq \Delta(G)$. □

Brooks' theorem. *Let G be a connected graph that is neither a complete graph nor an odd cycle. Then $\chi(G) \leq \Delta(G)$.*

[2] Technically, we are applying the greedy coloring algorithm to the graph G with the ordering $u_1, \ldots, u_n$, where $u_i = v_{n-i+1}$ for all $i \in \{1, \ldots, n\}$. So, "smaller indexed" from the description of the greedy coloring algorithm refers to the indices of the u_i's, not the v_i's.

Proof. To simplify notation, we set $\Delta := \Delta(G)$. We must show that $\chi(G) \leq \Delta$. Since G is connected and not complete, we see that $\Delta \geq 2$. Suppose first that $\Delta = 2$. Since G is connected, it follows that G is either a path on at least three vertices or a cycle. But by hypothesis, G is not an odd cycle, and so G is either a path on at least three vertices or an even cycle. It is now obvious that $\chi(G) = 2 = \Delta$.

From now on, we assume that $\Delta \geq 3$. Note that this implies that $|V(G)| \geq 4$.[3] We may further assume that G is regular,[4] for otherwise, we are done by Lemma 12.1.3.

> **Claim 1.** If G has a clique-cutset (i.e. a clique $C \subsetneqq V(G)$ such that $G \setminus C$ is disconnected), then $\chi(G) \leq \Delta$.

Proof of Claim 1. Suppose that G has a clique-cutset, and let C be a minimal clique-cutset of G. Let $A_1, \ldots, A_t$ ($t \geq 2$) be the vertex sets of the components of $G \setminus C$. For all $i \in \{1, \ldots, t\}$, let $G_i := G[A_i \cup C]$. By Lemma 10.4.1, we have that

$$\chi(G) \quad = \quad \max\{\chi(G_1), \ldots, \chi(G_t)\}.$$

Now, since G is connected, we know that C is non-empty. Further, by the minimality of C, we know that each vertex of C has a neighbor in each of $A_1, \ldots, A_t$. This implies that $G_1, \ldots, G_t$ are all connected and not regular.[5] But now Lemma 12.1.3 guarantees that $\chi(G_i) \leq \Delta(G_i) \leq \Delta$ for all $i \in \{1, \ldots, t\}$. Consequently, $\chi(G) \leq \Delta$. This proves Claim 1. ♦

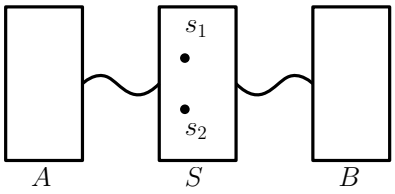
> **Claim 2.** If G is not 3-connected, then $\chi(G) \leq \Delta$.

Proof of Claim 2. Assume that G is not 3-connected; we must show that $\chi(G) \leq \Delta$. Since $|V(G)| \geq 4$, but G is not 3-connected, we see that there exists some $S \subseteq V(G)$ such that $|S| \leq 2$ and $G \setminus S$ is disconnected. By Claim 1, we may assume that G does not admit a clique-cutset. So, S is not a clique, and it follows that $|S| = 2$ and that the two vertices of S (call them s_1 and s_2) are non-adjacent. Let (A, B) be a partition of $V(G) \setminus S$ into non-empty sets such that there are no edges between A and B.

[3] Indeed, consider a vertex of degree Δ, plus all its neighbors.

[4] So, G is Δ-regular, i.e. all vertices of G are of degree Δ.

[5] Indeed, for all $i \in \{1, \ldots, t\}$ and $a_i \in A_i$, we have that $d_{G_i}(a) = d_G(a) = \Delta$, whereas each $c \in C$ has a neighbor in $V(G) \setminus V(G_i)$ and consequently satisfies $d_{G_i}(c) \leq d_G(c) - 1 \leq \Delta - 1$. So, G_i is not regular.

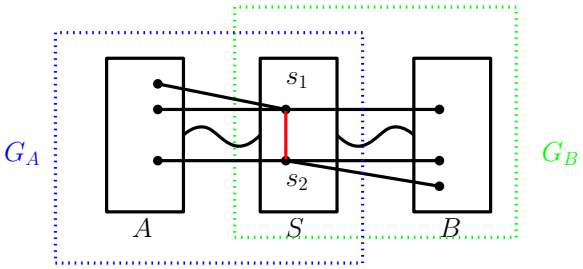


By Proposition 10.1.1, each vertex of S has a neighbor both in A and in B.[6] Furthermore, since $s_1s_2 \notin E(G)$, and since G is Δ-regular, with $\Delta \geq 3$, we see that each of s_1, s_2 has at least two neighbors in at least one of A, B. So, by symmetry, there are two cases to consider:

(i) s_1 has at least two neighbors in A, and s_2 has at least two neighbors in B;

(ii) s_1, s_2 each have exactly one neighbor in A.

Suppose first that (i) holds. Set $G_A := G[A \cup S] + s_1s_2$ and $G_B := G[B \cup S] + s_1s_2$.[7] Then both G_A and G_B are connected,[8] with $\Delta(G_A), \Delta(G_B) = \Delta$.[9] Furthermore, note that $d_{G_A}(s_2) \leq d_G(s_2) - 1 = \Delta - 1$, and so G_A is not regular; thus, Lemma 12.1.3 guarantees that $\chi(G_A) \leq \Delta(G_A) = \Delta$, and similarly, $\chi(G_B) \leq \Delta$. Now, note that S is a clique-cutset of $G + s_1s_2$. Lemma 10.4.1 now implies that $\chi(G + s_1s_2) = \max\{\chi(G_A), \chi(G_B)\} \leq \Delta$,[10] and it follows that $\chi(G) \leq \Delta$.



[6] G is a connected graph on at least four vertices, and it does not admit a clique-cutset. So, G is 2-connected, and we see that the hypotheses of Proposition 10.1.1 are satisfied.

[7] Thus, G_A is obtained from $G[A \cup S]$ by adding an edge between s_1 and s_2. Similarly, G_B is obtained from $G[B \cup S]$ by adding an edge between s_1 and s_2.

[8] Note that we are using Proposition 10.1.1.

[9] Indeed, for any $a \in A$, we have that $d_{G_A}(a) = d_G(a) = \Delta$, and $d_{G_A}(s_1), d_{G_A}(s_2) \leq \Delta$; so, $\Delta(G_A) = \Delta$, and similarly, $\Delta(G_B) = \Delta$.

[10] We are using the fact that A is the union of the vertex sets of some components of $G \setminus S = (G + s_1s_2) \setminus S$, whereas B is the union of the vertex sets of the remaining components of $G \setminus S = (G + s_1s_2) \setminus S$.

Suppose now that (ii) holds. Note that this implies that each of s_1, s_2 has at least two neighbors in B. Let s_1' be the unique neighbor of s_1 in A. Set $S' := \{s_1', s_2\}$, $A' := A \setminus \{s_1'\}$, and $B' := B \cup \{s_1\}$. Since G is Δ-regular, with $\Delta \geq 3$, we know that s_1' has at least three neighbors in G; since all neighbors of s_1' are in $A \cup S$, and $|S| = 2$, we see that s_1' has a neighbor in A. It follows that $A' \neq \emptyset$. Now S' separates $A' \neq \emptyset$ from $B' \neq \emptyset$. Further, if $s_1's_2 \in E(G)$, then S' is a clique-cutset of G, a contradiction. So, we may assume that $s_1's_2 \notin E(G)$. Since s_1' has at least three neighbors, and they all belong to $A \cup S$, we deduce that s_1' in fact has at least two neighbors in A'. But now if we consider S', A', B' instead of S, A, B, we are back in case (i), and so an argument analogous to the one above guarantees that $\chi(G) \leq \Delta$. This proves Claim 2. ♦

In view of Claim 2, we may now assume that G is 3-connected. Since G is connected and not complete, G has two vertices, call them u and v, at distance two from each other; let w be a common neighbor of u and v.[11] Since G is 3-connected, we know that $G' := G \setminus \{u, v\}$ is connected. We now order $V(G')$ according to the distance from w, that is, we list w first, then we list all vertices at distance one from w in G' (in any order), then we list all vertices at distance two from w in G' (in any order), etc. Finally, we list u, v at the end of our list. This produces an ordering $v_1, \ldots, v_n$ of $V(G)$ (with $v_1 = w$, $v_{n-1} = u$, and $v_n = v$). We now color G greedily using the ordering $v_n, \ldots, v_1$. All vertices in the ordering $v_n, \ldots, v_1$ other than the vertex v_1 have at least one neighbor to the right, and therefore at most $\Delta - 1$ neighbors to the left; so, all vertices other than v_1 get a color from the set $\{1, \ldots, \Delta\}$. But v_1 has exactly Δ neighbors, and two of those (namely, $v_{n-1} = u$ and $v_n = v$) got assigned the same color (namely, color 1) by our greedy coloring. So, v_1 also got assigned a color from the color set $\{1, \ldots, \Delta\}$. Thus, our coloring of G uses at most Δ colors, and it follows that $\chi(G) \leq \Delta$. □

12.2 Eulerian graphs

An *Euler circuit* (or *Eulerian circuit*) of a graph G is a walk in the graph that passes through every edge exactly once and comes back to the origin vertex. A graph is *Eulerian* if it has an Eulerian circuit. The following theorem was proven in Discrete Mathematics.

[11] Let us prove that such u, v, w exist. Since G is not complete, it contains a pair of non-adjacent vertices, call them u and u'. Since G is connected, there exists an induced path $P = p_0, \ldots, p_t$, with $p_0 = u$ and $p_t = u'$; since u and u' are non-adjacent, we have that $t \geq 2$. Now set $w := p_1$ and $v := p_2$.

Theorem 12.2.1. *A connected graph is Eulerian if and only if all its vertices are of even degree.*

12.3 Vizing's theorem

A *k-edge-coloring* of a graph G is a mapping $c : E(G) \to C$, with $|C| = k$. Elements of C are called *colors*. An edge-coloring is *proper* if for any two distinct edges e and f that share an endpoint, we have that $c(e) \neq c(f)$.

A graph G is *k-edge-colorable* if it has a proper k-edge-coloring.

The *edge chromatic number* (or *chromatic index*) of a graph G, denoted by $\chi'(G)$, is the minimum integer k such that G is k-edge-colorable.

Clearly, in any proper edge-coloring of a graph G, all edges incident with the same vertex must receive a different color; consequently, $\chi'(G) \geq \Delta(G)$.

Note that any (not necessarily proper) k-edge-coloring $c : E(G) \to \{1, \ldots, k\}$ of a graph G can be represented by a partition $\mathcal{C} = (E_1, \ldots, E_k)$ of $E(G)$, where E_i denotes the subset of $E(G)$ assigned color i. (Sets $E_1, \ldots, E_k$ are called *color classes*.) A proper k-edge-coloring is one where each E_i is a matching.

Lemma 12.3.1. *Every graph G satisfies $\chi'(G)\nu(G) \geq |E(G)|$.*[12] *Consequently, if G has at least one edge, then $\chi'(G) \geq \left\lceil \frac{|E(G)|}{\nu(G)} \right\rceil$.*

Proof. Let G be a graph, and let $k = \chi'(G)$. Let $(E_1, \ldots, E_k)$ be a proper edge-coloring of G. Then

$$\begin{aligned} |E(G)| &= \sum_{i=1}^{k} |E_i| && \text{because } (E_1, \ldots, E_k) \text{ is a partition of } E(G) \\ &\leq \sum_{i=1}^{k} \nu(G) && \text{because } E_1, \ldots, E_k \text{ are matchings of } G \\ &= k\nu(G) \\ &= \chi'(G)\nu(G). \end{aligned}$$

This proves that $\chi'(G)\nu(G) \geq |E(G)|$. If G has at least one edge, then clearly, $\nu(G) \geq 1$, and we deduce that $\chi'(G) \geq \frac{|E(G)|}{\nu(G)}$; since $\chi'(G)$ is an integer, it follows that $\chi'(G) \geq \left\lceil \frac{|E(G)|}{\nu(G)} \right\rceil$. □

[12] Recall that $\nu(G)$ is the matching number of G, i.e. the maximum size of a matching in G.

Given a (not necessarily proper) edge-coloring of a graph G, we say that color i is *represented* at a vertex v of G if some edge incident with v has color i.

Lemma 12.3.2. *Let G be a connected graph that is not an odd cycle. Then G has a (not necessarily proper) 2-edge-coloring in which both colors are represented at each vertex of degree at least 2.*

Proof. We may assume that $\Delta(G) \geq 2$, for otherwise there is nothing to show. By hypothesis, G is connected and not an odd cycle; consequently, if G is 2-regular, then G is an even cycle.

Suppose first that G is Eulerian. Then (by Theorem 12.2.1) all vertices of G are of even degree. If G has a vertex of degree at least four, then let v_0 be such a vertex, and otherwise let v_0 be any vertex. (Note that in the latter case, G is 2-regular, and therefore, by the above, G is an even cycle.) Let $v_0, e_1, v_1, e_2, v_2, \ldots, v_0$ be an Euler circuit of G. Let E_1 be the set of odd-indexed edges, and let E_2 the set of even-indexed edges. If G is an even cycle, then clearly, the edge-coloring (E_1, E_2) satisfies the lemma. Otherwise, v_0 is of degree at least four, and the edge-coloring (E_1, E_2) has the desired property since each vertex of G is an internal vertex of $v_0, e_1, v_1, e_2, v_2, \ldots, v_0$.

So, we may assume that G is not Eulerian. Let G^* be the graph obtained from G by adding a new vertex v^* to G, and joining v^* to each odd-degree vertex of G. Then by Theorem 12.2.1, G^* is Eulerian.[13] Now, let $v_0, e_1, v_1, e_2, v_2, \ldots, v_0$, with $v_0 = v^*$, be an Euler circuit of G^*. Let E_1 be the set of odd-indexed edges, and let E_2 the set of even-indexed edges. Then the edge-coloring $\Big(E_1 \cap E(G), E_2 \cap E(G)\Big)$ of G has the desired property.[14] $\square$

Given a (not necessarily proper) k-edge-coloring $\mathcal{C}$ and a vertex v of G, we denote by $c_{\mathcal{C}}(v)$ the number of distinct colors represented at v. Note that $c_{\mathcal{C}}(v) \leq d_G(v)$ for all $v \in V(G)$. Furthermore, $\mathcal{C}$ is a proper k-edge-coloring of G if and only if $c_{\mathcal{C}}(v) = d_G(v)$ for every vertex $v \in V(G)$. A k-edge-coloring $\mathcal{C}'$ of G is an *improvement* of $\mathcal{C}$ if

$$\sum_{v \in V(G)} c_{\mathcal{C}'}(v) \quad > \quad \sum_{v \in V(G)} c_{\mathcal{C}}(v).$$

[13] Since G is connected and not Eulerian, we know that G has at least one vertex of odd degree. On the other hand, since $\sum_{v \in V(G)} d_G(v) = 2|E(G)|$, we know that $\sum_{v \in V(G)} d_G(v)$ is even, and consequently, G has an even number of vertices of odd degree. So, in G^*, v^* has even degree, strictly greater than zero. We now see that G^* is connected, and that all vertices of G^* have even degree. So, by Theorem 12.2.1, G^* is Eulerian.

[14] Details?

An *unimprovable* k-edge-coloring is one that cannot be improved.[15]

Note that any proper edge-coloring of a graph G is unimprovable, but the converse does not hold in general. We note, however, that for any graph G and positive integer k, there exists at least one unimprovable k-edge-coloring of G.[16]

Lemma 12.3.3. *Let $\mathcal{C} = (E_1, \ldots, E_k)$ be an unimprovable k-edge-coloring of a graph G. If there is a vertex u of G and colors i and j such that i is not represented at u and j is represented at least twice at u, then the component of $G[E_i \cup E_j]$ that contains u is an odd cycle.*[17]

Proof. Let H be the component of $G[E_i \cup E_j]$ that contains u. Suppose that H is not an odd cycle. Then by Lemma 12.3.2, H has a 2-edge-coloring in which both colors are represented at every vertex of degree at least 2 in H. Recolor the edges of H with colors i and j in this way to get a new k-edge-coloring $\mathcal{C}' = (E'_1, \ldots, E'_k)$ of G. To simplify notation, set $c = c_{\mathcal{C}}$ and $c' = c_{\mathcal{C}'}$. By construction, we have that $c(v) \leq c'(v) \leq c(v) + 1$ for all $v \in V(G)$, and that $c'(u) = c(u)+1$. It follows that $\sum_{v \in V(G)} c'(v) > \sum_{v \in V(G)} c(v)$, that is, $\mathcal{C}'$ is an improvement of $\mathcal{C}$. But this contradicts the assumption that $\mathcal{C}$ is unimprovable. □

Theorem 12.3.4. *If G is a bipartite graph, then $\chi'(G) = \Delta(G)$.*

Proof. Let G be a bipartite graph, and let $\Delta := \Delta(G)$. Clearly, $\chi'(G) \geq \Delta$, and we need only show that $\chi'(G) \leq \Delta$. Let $\mathcal{C} = (E_1, \ldots, E_\Delta)$ be an unimprovable Δ-edge-coloring of G. Suppose that $\mathcal{C}$ is not a proper edge-coloring of G. Then there exists a vertex $u \in V(G)$ such that some color j is represented at least twice at u, and (consequently) some color i is not represented at u. But now by Lemma 12.3.3, the component of $G[E_i \cup E_j]$ that contains u is an odd cycle, contrary to the fact that bipartite graphs contain no odd cycles. So, $\mathcal{C}$ is a proper Δ-edge-coloring of G, and it follows that $\chi'(G) \leq \Delta$. □

Vizing's theorem. *Every graph G satisfies $\chi'(G) \leq \Delta(G) + 1$.*[18]

[15] Note that improvability and unimprovability are defined with respect to a fixed k.

[16] Indeed, G has at least one k-edge-coloring (e.g. the one that assigns color 1 to all edges of G), but up to a renaming of colors, G only has finitely many k-edge-colorings. So, some k-edge-coloring of G is unimprovable.

[17] Here, $G[E_i \cup E_j]$ is the graph with vertex set $V(G)$ and edge set $E_i \cup E_j$.

[18] We emphasize that Vizing's theorem holds only for simple graphs (i.e. graphs with no loops and no parallel edges). Vizing's theorem fails for multigraphs!

Proof. Let $\Delta = \Delta(G)$. Suppose that $\chi'(G) > \Delta + 1$. Let $\mathcal{C} = (E_1, \ldots, E_{\Delta+1})$ be an unimprovable $(\Delta+1)$-edge-coloring, and set $c = c_{\mathcal{C}}$. Since no vertex of G has degree greater than Δ, and since we have $\Delta + 1$ colors, we know that for each vertex of G, at least one of our $\Delta + 1$ colors is not represented at that vertex. On the other hand, since $\chi'(G) > \Delta + 1$, we know that $\mathcal{C}$ is not a proper edge-coloring of G, and consequently, at some vertex of G, some color is represented at least twice.

Let vertex $u \in V(G)$ and colors $i_0, i_1 \in \{1, \ldots, \Delta + 1\}$ be such that i_0 is not represented at u, and i_1 is represented at least twice at u. Let uv_1 have color i_1, and let i_2 be a color that is not represented at v_1. (Clearly, $i_1 \neq i_2$.) Color i_2 must be represented at u, since otherwise, recoloring uv_1 with i_2 would yield an improvement of $\mathcal{C}$. So some edge uv_2 has color i_2; let i_3 be a color that is not represented at v_2. (Clearly, $i_2 \neq i_3$.) Color i_3 must be represented at u, since otherwise recoloring uv_1 with i_2 and uv_2 with i_3 would yield an improvement of $\mathcal{C}$. So some edge uv_3 has color i_3. Now, we have only a finite number of colors at our disposal, and so continuing in this way, we eventually start to repeat colors. More formally, we can construct a sequence $v_1, v_2, \ldots, v_\ell$ of vertices and a sequence $i_1, i_2, \ldots, i_\ell, i_{\ell+1}$ of colors such that all the following are satisfied:

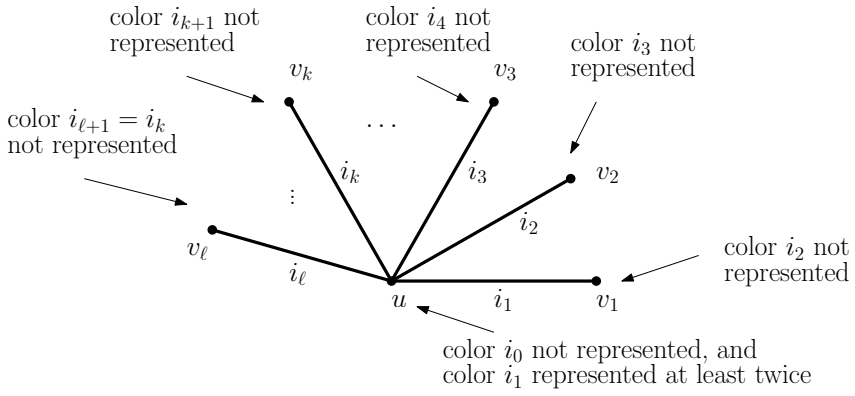
(a) color i_1 is represented at least twice at u;

(b) for all $j \in \{1, \ldots, \ell\}$, edge uv_j has color i_j;

(c) for all $j \in \{1, \ldots, \ell\}$, color i_{j+1} is not represented at v_j;

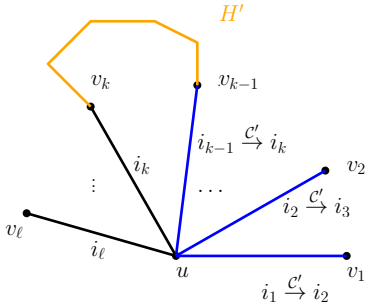
(d) colors $i_1, \ldots, i_\ell$ are pairwise distinct;

(e) there exists some $k \in \{1, \ldots, \ell\}$ such that $i_k = i_{\ell+1}$.

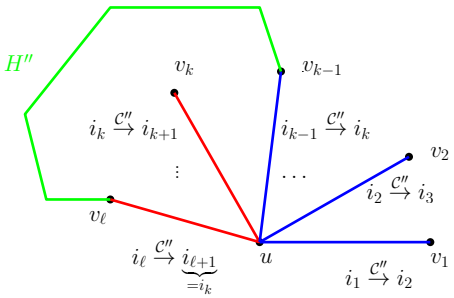


Note that (b) and (c) imply that $i_j \neq i_{j+1}$ for all $j \in \{1, \dots, \ell\}$; in particular, $k \leq \ell - 1$. Further, (b) and (d) imply that vertices $v_1, \dots, v_\ell$ are pairwise distinct.

Let $\mathcal{C}' = (E'_1, \dots, E'_{\Delta+1})$ be the following recoloring of G: for $j = 1, \dots, k-1$, recolor uv_j with i_{j+1}.[19] Set $c' = c_{\mathcal{C}'}$. Then $c'(v) \geq c(v)$ for every $v \in V(G)$; thus, since $\mathcal{C}$ is an unimprovable $(\Delta+1)$-edge-coloring of G, so is $\mathcal{C}'$. Further, by construction, under the coloring $\mathcal{C}'$, color i_0 is not represented at u, and color i_k is represented at least twice at u. (Note that if $k = 1$, then $\mathcal{C}' = \mathcal{C}$ and $i_k = i_1$. In this case, $i_k = i_1$ is still represented twice at u, by the choice of i_1.) Let H' be the component of $G[E'_{i_0} \cup E'_{i_k}]$ that contains u. By Lemma 12.3.3, H' is an odd cycle.



Let $\mathcal{C}'' = (E''_1, \dots, E''_{\Delta+1})$ be the following recoloring of G: for $j = 1, \dots, \ell$, recolor uv_j with i_{j+1}; since $i_{\ell+1} = i_k$, we see that uv_ℓ was recolored with i_k. Set $c'' = c_{\mathcal{C}''}$. Then $c''(v) \geq c(v)$ for every $v \in V(G)$; thus, since $\mathcal{C}$ is an unimprovable $(\Delta+1)$-edge-coloring of G, so is $\mathcal{C}''$. Further, under the coloring $\mathcal{C}'$, color i_0 is not represented at u, and color i_k is represented at least twice at u. Let H'' be the component of $G[E''_{i_0} \cup E''_{i_k}]$ that contains u. By Lemma 12.3.3, H'' is an odd cycle.



Note that the colorings $\mathcal{C}'$ and $\mathcal{C}''$ disagree only on edges $uv_k, \dots, uv_{\ell-1}, uv_\ell$. Further, exactly one edge (namely, uv_k) from $uv_k, \dots, uv_{\ell-1}, uv_\ell$ belongs to

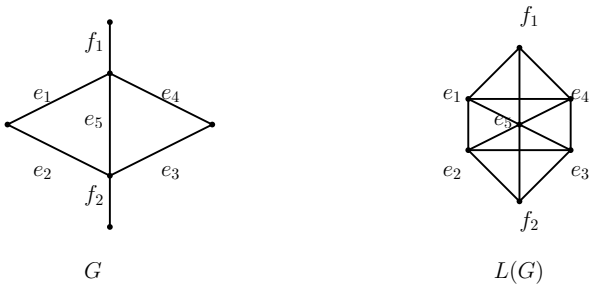
[19] If $k = 1$, then $\mathcal{C}' = \mathcal{C}$.

the cycle H', and exactly one edge (namely, uv_ℓ) from $uv_k, \ldots, uv_{\ell-1}, uv_\ell$ belongs to the cycle H''. It now follows that $H' - uv_k = H'' - uv_\ell$, which is impossible, since two cycles cannot differ in exactly one edge. □

Corollary 12.3.5. *Every graph G satisfies $\Delta(G) \leq \chi'(G) \leq \Delta(G) + 1$.*

We note that it is NP-complete to decide whether $\chi' = \Delta$ (even when $\Delta = 3$). We omit the details.

Finally, we remark that there is a relationship between vertex coloring and edge-coloring, as follows. Given a graph G, the *line graph* of G, denoted by $L(G)$, is the graph with vertex set $E(G)$, in which distinct $e, f \in E(G)$ are adjacent if and only if they share an endpoint in G. An example is shown below.



Obviously, $\chi(L(G)) = \chi'(G)$.

Recall that for a graph G, the *clique number* of G, denoted by $\omega(G)$, is the maximum size of a clique in G.

Lemma 12.3.6. *Every graph G satisfies $\chi(L(G)) \leq \omega(L(G)) + 1$.*

Proof. Let G be a graph. Then clearly, $\chi(L(G)) = \chi'(G)$. Furthermore, for any vertex v, the set of all edges incident with v in G is a clique of size $d_G(v)$ in $L(G)$; consequently, $\omega(L(G)) \geq \Delta(G)$. But now

$$\begin{aligned} \chi(L(G)) &= \chi'(G) \\ &\leq \Delta(G) + 1 \qquad \text{by Vizing's theorem} \\ &\leq \omega(L(G)) + 1. \end{aligned}$$

□

Chapter 13

Chordal graphs

13.1 Triangle-free graphs of arbitrarily large chromatic number

Clearly, every graph G satisfies $\chi(G) \geq \omega(G)$.[1] So, the simplest way to construct a graph of large chromatic number is to construct a graph that has a large clique number. However, as we shall see, it is possible to construct graphs of small clique number and large chromatic number.

A *triangle* in a graph G is a clique of size three. A graph is *triangle-free* if it contains no triangles. So, a graph is triangle-free if and only if its clique number is at most two. Our goal in this section is to construct a family of triangle-free graphs of arbitrarily large chromatic number. There are several known constructions; here, we give the one due to Mycielski (1955).

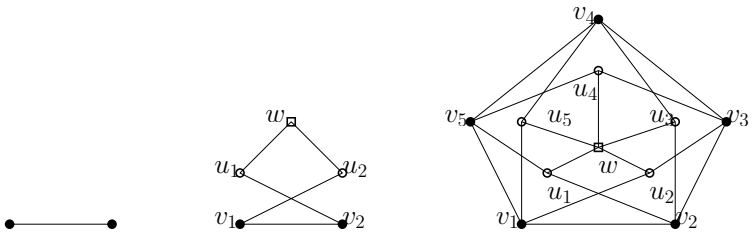
The Mycielski graphs $\{M_k\}_{k=2}^{\infty}$ are defined recursively, as follows. First, let $M_2 := K_2$. Next, fix an integer k, and suppose M_k has been constructed. We construct M_{k+1} as follows. Let $V = \{v_1, \ldots, v_n\}$ be the vertex set of M_k. Let $U = \{u_1, \ldots, u_n\}$ (where the u_i's are "new" vertices; we think of u_i as a "duplicate" of v_i), and let w be another "new" vertex. Let M_{k+1} have vertex set $V \cup U \cup \{w\}$ and adjacency as follows:

- adjacency between the v_i's is inherited from M_k, that is, $M_{k+1}[V] = M_k$;
- for all $i \in \{1, \ldots, n\}$, u_i is non-adjacent to v_i;
- for all distinct $i, j \in \{1, \ldots, n\}$, u_i is adjacent to v_j in M_{k+1} if and only if v_i is adjacent to v_j in M_k;

[1] As usual, $\omega(G)$ is the *clique number* of G, i.e. the maximum size of a clique in G.

- U is a stable set in M_{k+1};
- w is adjacent to all vertices in U and non-adjacent to all vertices in V.

The first three Mycielski graphs are represented below.



M_2 M_3 M_4

Lemma 13.1.1. *For all integers $k \geq 2$, M_k satisfies $\omega(M_k) = 2$ and $\chi(M_k) = k$.*

Proof. We proceed by induction on k. Clearly, $\omega(M_2) = 2$ and $\chi(M_2) = 2$. Next, fix an integer $k \geq 2$, and assume inductively that $\omega(M_k) = 2$ and $\chi(M_k) = k$. We must show that $\omega(M_{k+1}) = 2$ and $\chi(M_{k+1}) = k + 1$. Let $V = \{v_1, \ldots, v_n\}$, $U = \{u_1, \ldots, u_n\}$, and w be as in the definition of M_{k+1}.

We first show that $\omega(M_{k+1}) = 2$. Since $\omega(M_k) = 2$, and M_k is a subgraph of M_{k+1}, it is clear that $\omega(M_{k+1}) \geq 2$. It remains to show that M_{k+1} is triangle-free. Suppose otherwise, and let T be a triangle in M_{k+1}. Since U is a stable set of G, we see that $|T \cap U| \leq 1$. Since $N_{M_{k+1}}(w) = U$, and since U is a stable set, we further see that $w \notin T$. Finally, since $M_{k+1}[V] = M_k$, and since M_k is triangle-free (by the induction hypothesis), we see that $T \not\subseteq V$. It now follows that $|T \cap U| = 1$ and $|T \cap V| = 2$. Let $p, q, r \in \{1, \ldots, k\}$ (with $q \neq r$) be such that $T = \{u_p, v_q, v_r\}$. By the construction of M_{k+1}, $u_p v_p \notin E(M_{k+1})$; since T is a triangle, it follows that $p \notin \{q, r\}$. Since $u_p v_q \in E(M_{k+1})$, it follows from the construction of M_{k+1} that $v_p v_q \in E(M_k)$; similarly, $v_p v_r \in E(M_k)$. But now $\{v_p, v_q, v_r\}$ is a triangle in M_k, a contradiction. So, M_{k+1} is triangle-free, and we deduce that $\omega(M_{k+1}) = 2$.

We now show that $\chi(M_{k+1}) = k + 1$. Let us first show that $\chi(M_{k+1}) \leq k + 1$. First, we properly color M_k with colors $1, \ldots, k$ (this is possible because $\chi(M_k) = k$). Next, for each $i \in \{1, \ldots, n\}$, we assign to u_i the same color as to v_i. Finally, we assign color $k + 1$ to w. Clearly, this is a proper coloring of M_{k+1}, and it follows that $\chi(M_{k+1}) \leq k + 1$.

Finally, we show that $\chi(M_{k+1}) \geq k + 1$. Suppose otherwise, that is, suppose that $\chi(M_{k+1}) \leq k$. Fix a proper coloring $c : V(M_{k+1}) \to \{1, \ldots, k\}$

of M_{k+1}. We will use the coloring c of M_{k+1} to construct a proper $(k-1)$-coloring of M_k, which will contradict the fact that $\chi(M_k) = k$. By symmetry, we may assume that $c(w) = k$. Since w is adjacent to every vertex in U, it follows that c does not assign color k to any vertex in U. Now, let V_k be the set of all vertices in V to which c assigns color k. Since c is a proper coloring of M_{k+1}, we know that V_k is a stable set in M_{k+1} (and therefore, in M_k as well). Now, define $c' : V \to \{1, \ldots, k-1\}$ as follows:

- $c' \restriction (V \setminus V_k) = c \restriction (V \setminus V_k)$;[2]
- for all $v_i \in V_{k+1}$, set $c'(v_i) = c(u_i)$.

Let us check that c' is a proper coloring of M_k. Fix distinct $i, j \in \{1, \ldots, n\}$, and suppose that v_i is adjacent to v_j in M_k. We must show that $c'(v_i) \neq c'(v_j)$. Since V_k is a stable set, we know that at most one of v_i, v_j belongs to V_k. If $v_i, v_j \in V \setminus V_k$, then it follows from the construction of c', and from the fact that c is a proper coloring of M_{k+1}, that $c'(v_i) = c(v_i) \neq c(v_j) = c'(v_j)$. It remains to consider the case when exactly one of v_i, v_j belongs to V_k; by symmetry, we may assume that $v_i \in V_k$ and $v_j \in V \setminus V_k$. By the construction of M_{k+1}, u_i is adjacent to v_j in M_{k+1}, and so $c(u_i) \neq c(v_j)$. But now by the construction of c', we have that $c'(v_i) = c(u_i) \neq c(v_j) = c'(v_j)$, which is what we needed to show. Thus, c' is a proper $(k-1)$-coloring of M_k, contrary to the fact that $\chi(M_k) = k$. □

As an immediate corollary of Lemma 13.1.1, we get the following.

Theorem 13.1.2. *There exist triangle-free graphs of arbitrarily large chromatic number. More precisely, for every positive integer k, there exists a graph G such that $\omega(G) = 2$ and $\chi(G) \geq k$.*

Proof. This follows from Lemma 13.1.1. □

We remark that Erdős (1961) applied the probabilistic method to demonstrate the existence of graphs of arbitrarily high girth and chromatic number (the *girth* of a graph G that has at least one cycle is the length of the shortest cycle in G). Graphs of high girth are triangle-free, and so this result of Erdős is stronger than Theorem 13.1.2.

[2] This means that $c'(v_i) = c(v_i)$ for all $v_i \in V \setminus V_k$.

13.2 Perfect graphs

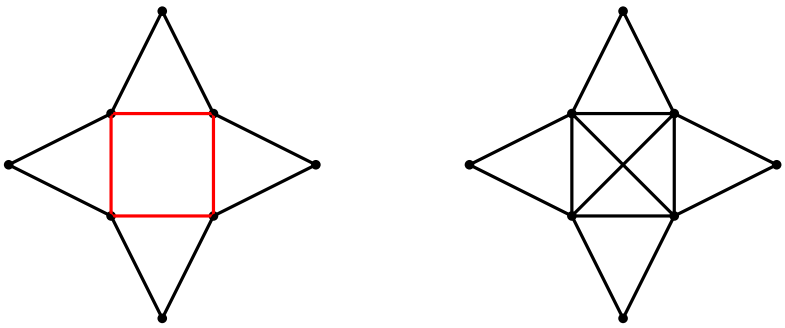
In the previous section, we saw that there exist graphs of small clique number, but large chromatic number. At the other extreme, we might consider graphs for which $\chi = \omega$. This, however, turns out not to be a very interesting question. Indeed, suppose H is any graph at all, and let G be the disjoint union of H and $K_{\chi(H)}$; then $\chi(G) = \omega(G)$, but we can say very little about the structure of G (since G was built starting from an arbitrary graph H).

Here is a more interesting definition. A graph is *perfect* if all its induced subgraphs H satisfy $\chi(H) = \omega(H)$.[3]

Since every graph is an induced subgraph of itself, we see that every perfect graph G satisfies $\chi(G) = \omega(G)$. Importantly, though, in a perfect graph, $\chi = \omega$ should hold not only for the graph itself, but also for all its induced subgraphs.

13.3 Chordal graphs

In this section, we consider a particular subclass of perfect graphs, called "chordal" graphs. A graph is *chordal* (or *triangulated*) if every cycle of length strictly greater than three has a chord (a *chord* of a cycle is an edge joining two non-consecutive vertices of the cycle). In other words, a graph is *chordal* if it contains no induced cycles of length at least four. For example, in the picture below, the graph on the left is not chordal (because it contains an induced cycle of length four, in red), whereas the one on the right is chordal (this graph contains a cycle of length four, but the cycle is not induced).



not chordal chordal

[3] A graph H is an *induced subgraph* of a graph G if $V(H) \subseteq V(G)$, and for all distinct $u, v \in V(H)$, we have that $uv \in E(H)$ if and only if $uv \in E(G)$.

Note that all induced subgraphs of a chordal graph are chordal.

Chordal graphs were one of the first classes of graphs to be recognized as perfect; the study of chordal graphs can be seen as the beginning of the theory of perfect graphs. As we shall see, there are efficient algorithms for recognizing chordal graphs and for solving the vertex coloring and related optimization problems on chordal graphs. In many practical applications of vertex-coloring, the graphs that appear are actually chordal.

In this section, a *cutset* of a graph is a set of vertices whose deletion yields a disconnected graph. More precisely, a *cutset* of a graph G is a (possibly empty) set $S \subsetneq V(G)$ such that $G \setminus S$ is disconnected.[4] A *clique-cutset* is a cutset that is a clique, that is, a *clique-cutset* of a graph G is a clique $C \subsetneq V(G)$ of G such that $G \setminus C$ is disconnected.[5]

For a graph G and non-adjacent vertices $x, y \in V(G)$, an (x, y)-*separator* of G is a set $S \subseteq V(G) \setminus \{x, y\}$ such that x and y belong to distinct components of $G \setminus S$; an (x, y)-separator S of G is *minimal* if no proper subset of S is an (x, y)-separator of G. Note that any (x, y)-separator of G is a cutset of G.

13.3.1 Characterizing chordal graphs

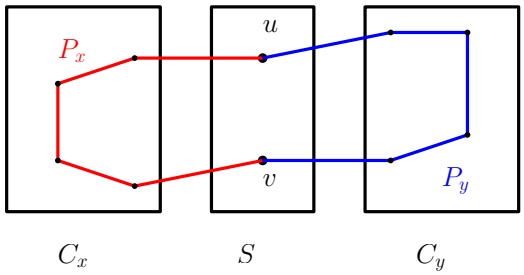
Lemma 13.3.1. *Let G be a chordal graph that is not complete, let x and y be non-adjacent vertices of G, and let S be a minimal (x, y)-separator of G. Then S is a clique of G.*

Proof. Let C_x be the component of $G \setminus S$ that contains x, and let C_y be the component of $G \setminus S$ that contains y. By the minimality of S, every vertex of S has a neighbor both in C_x and in C_y. Now, suppose that S is not a clique, and fix distinct, non-adjacent vertices $u, v \in S$. Let P_x be a minimum-length path between u and v in $G[V(C_x) \cup \{u, v\}]$, and let P_y be a minimum-length path between u and v in $G[V(C_y) \cup \{u, v\}]$.[6]

[4] Sometimes, a cutset is defined to be a set of vertices whose deletion increases the number of components, but that definition is inconvenient in this context.

[5] In particular, if G is disconnected, then $\emptyset$ is a clique-cutset of G.

[6] $G[V(C_x) \cup \{u, v\}]$ is connected because C_x is connected, and both u and v have a neighbor in C_x. Similarly, $G[V(C_y) \cup \{u, v\}]$ is connected. So, P_x and P_y exist.



By the minimality of P_x and P_y, we see that both P_x and P_y are induced paths of G, and since u and v are non-adjacent, we see that each of them has at least two edges. Since the interior of P_x belongs to C_x, and the interior of P_y belongs to C_y, we see that there are no edges between the interiors of P_x and P_y. Thus, $P_x \cup P_y$ is an induced cycle of length at least four in G, a contradiction. □

Theorem 13.3.2. *If G is a chordal graph, then either G is a complete graph or G admits a clique-cutset.*

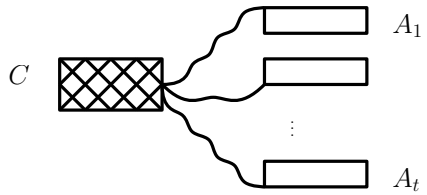
Proof. Let G be a chordal graph that is not complete. Let x and y be non-adjacent vertices of G, and let S be a minimal (x, y)-separator of G.[7] By Lemma 13.3.1, S is a clique. So, S is a clique-cutset of G. □

Corollary 13.3.3. *Chordal graphs are perfect.*

Proof. Since every induced subgraph of a chordal graph is chordal, it is enough to show that every chordal graph G satisfies $\chi(G) = \omega(G)$.[8] So, fix a chordal graph G, and assume inductively that all chordal graphs G' on fewer than $|V(G)|$ vertices satisfy $\chi(G') = \omega(G')$. We must show that $\chi(G) = \omega(G)$. If G is a complete graph, then it is clear that $\chi(G) = \omega(G)$. So, assume that G is not complete. Then by Theorem 13.3.2, G admits a clique-cutset, call it C. Let $A_1, \ldots, A_t$ $(t \geq 2)$ be the vertex sets of the components of $G \setminus C$.

[7] To see that S exists, we first observe that $V(G) \setminus \{x, y\}$ is an (x, y)-separator of G, and in particular, an (x, y)-separator of G exists. Of all (x, y)-separators of G, we can choose S to be one of minimum size.

[8] Indeed, suppose we have shown that all chordal graphs G satisfy $\chi(G) = \omega(G)$. Now, fix a chordal graph G, and let H be an induced subgraph of G. Then H is chordal, and so $\chi(H) = \omega(H)$. So, G is perfect.



For all $i \in \{1, \ldots, t\}$, let $G_i := G[A_i \cup C]$. Note that every clique of G is in fact a clique of one of $G_1, \ldots, G_t$,[9] and it follows that $\omega(G) = \max\{\omega(G_1), \ldots, \omega(G_t)\}$. On the other hand, by Lemma 10.4.1, we have that $\chi(G) = \max\{\chi(G_1), \ldots, \chi(G_t)\}$. Finally, for all $i \in \{1, \ldots, t\}$, the induction hypothesis guarantees that $\chi(G_i) = \omega(G_i)$. So,

$$\begin{aligned} \chi(G) &= \max\{\chi(G_1), \ldots, \chi(G_t)\} \\ &= \max\{\omega(G_1), \ldots, \omega(G_t)\} \\ &= \omega(G), \end{aligned}$$

which is what we needed to show. □

13.3.2 Simplicial vertices

A vertex x of a graph G is *simplicial* if $N_G(x)$ is a clique of G.[10]

Theorem 13.3.4 (Dirac, 1961)**.** *Every chordal graph has a simplicial vertex. Moreover, every chordal graph that is not complete has (at least) two non-adjacent simplicial vertices.*

Proof. We proceed by induction on the number of vertices. Let G be a chordal graph, and assume inductively that the claim holds for chordal graphs on fewer than $|V(G)|$ vertices.[11] We must show that the claim holds for G. If G is a complete graph, then clearly, any vertex of G is simplicial. So assume that G is not complete (and in particular, $|V(G)| \geq 2$). By Theorem 13.3.2, G admits a clique-cutset, call it C. Let A and B be the vertex sets of two distinct components of G, and set $G_A = G[A \cup C]$ and $G_B = G[B \cup C]$.

[9] This is because there are no edges between any two of the sets $A_1, \ldots, A_t$, and so no clique of G intersects more than one of $A_1, \ldots, A_t$.

[10] $N_G(x)$ may possibly be empty, i.e. isolated vertices are simplicial.

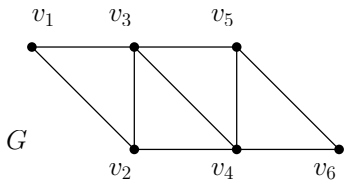
[11] More precisely, we assume inductively that for every chordal graph G' such that $|V(G')| < |V(G)|$, G' has a simplicial vertex, and furthermore, if G' is not a complete graph, then G' has two non-adjacent simplicial vertices.

> **Claim.** *A contains a vertex that is simplicial in G_A, and B contains a vertex that is simplicial in G_B.*

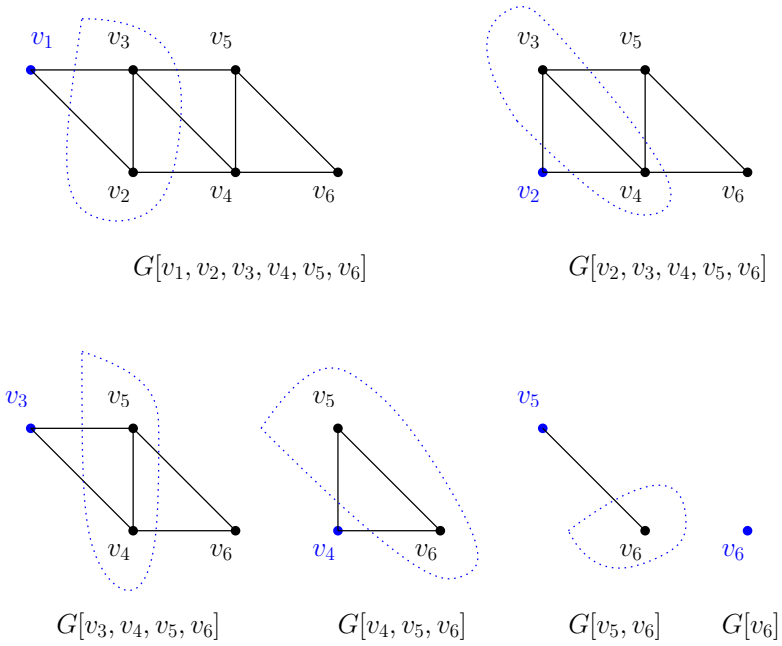
Proof of the Claim. By symmetry, it suffices to show this for A. If G_A is complete, then any vertex in A is simplicial in G_A. Otherwise, by the induction hypothesis, G_A contains two non-adjacent simplicial vertices; since C is a clique, C may contain at most one of these two vertices, and consequently, A contains the other (possibly, A contains both of them). This proves the Claim. ⧫

Now, using the Claim, we let $a \in A$ be a simplicial vertex of G_A, and we let $b \in B$ be a simplicial vertex of G_B. Clearly, a and b are non-adjacent. Furthermore, we have that $N_G(a) = N_{G_A}(a)$ and $N_G(b) = N_{G_B}(b)$, and we deduce that a and b are simplicial vertices of G. □

A *simplicial elimination ordering* (sometimes also called a *perfect elimination ordering*) of a graph G is an ordering $v_1, \ldots, v_n$ of its vertices such that for all $i \in \{1, \ldots, n\}$, v_i is simplicial in the graph $G[v_i, \ldots, v_n]$. For instance, $v_1, \ldots, v_6$ is a simplicial elimination ordering of the graph G in the picture below.



Indeed, consider the picture below. Clearly, for each $i \in \{1, \ldots, 6\}$, v_i is simplicial in $G[v_i, \ldots, v_6]$.



Theorem 13.3.5 (Fulkerson and Gross, 1965)**.** *For a graph G, the following statements are equivalent:*

(i) G is chordal;

(ii) G has a simplicial elimination ordering;

(iii) for all non-adjacent vertices x and y of G, every minimal (x, y)-separator of G is a clique.

Proof. **(i) ⇒ (iii):** This follows from Lemma 13.3.1.

(iii) ⇒ (i): We prove the contrapositive: if (i) if false, then (iii) is false. So, assume that (i) is false, that is, that G is not chordal. Let C be an induced cycle of length at least four in G, let x and y be non-adjacent vertices of C, and let P_1 and P_2 be the two paths between x and y in C; clearly, each of P_1, P_2 has at least two edges, and in particular, $V(P_1) \setminus \{x, y\}$ and $V(P_2) \setminus \{x, y\}$ are non-empty. Let S be a minimal (x, y)-separator of G. Clearly, S must intersect both $V(P_1) \setminus \{x, y\}$ and $V(P_2) \setminus \{x, y\}$. But since C is an induced cycle, we know that there are no edges between $V(P_1) \setminus \{x, y\}$ and $V(P_2) \setminus \{x, y\}$, and it follows that S is not a clique. Thus, (iii) is false.

(i) ⇒ (ii): We proceed by induction on the number of vertices. Clearly, the claim holds for one-vertex graphs. Now, fix a positive integer n, and assume that the claim holds for all chordal graphs on n vertices. Let H be

a chordal graph on $n+1$ vertices. By Theorem 13.3.4, H has at least one simplicial vertex, call it x_0. Then $H \setminus x_0$ is a chordal graph on n vertices, and so by the induction hypothesis, $H \setminus x_0$ has a simplicial elimination ordering, say $x_1, \ldots, x_n$. But now $x_0, x_1, \ldots, x_n$ is a simplicial elimination ordering of H.

(ii) $\Rightarrow$ (i): Suppose that $v_1, \ldots, v_n$ is a simplicial elimination ordering of G; we claim that G is chordal. Let C be an induced cycle of G; we must show that C is a triangle. Let $x = v_i$ be the lowest-indexed vertex from our simplicial elimination ordering that belongs to the cycle C, and let y, z be the two neighbors of x in C. Since $x = v_i$ is simplicial in $G[v_i, v_{i+1}, \ldots, v_n]$, since y, z are distinct neighbors of x, and since (by the minimality of i) we have that $y, z \in \{v_{i+1}, \ldots, v_n\}$, we see that $yz \in E(G)$. Since C is an induced cycle, it follows that C is a triangle. This proves that G is chordal. □

Note that Theorem 13.3.5 gives an $O(n^4)$ time recognition algorithm for chordal graphs (we repeatedly search for simplicial vertices). In fact, chordal graphs can be recognized in $O(n+m)$ time using the so called Lexicographic breadth-first-search (LexBFS) due to Rose, Tarjan, and Lueker (1976), but we omit the details.

13.3.3 Efficient optimization algorithms for chordal graphs

In this subsection, G is a chordal graph on n vertices, and $v_1, \ldots, v_n$ is a simplicial elimination ordering on G.[12] For each $i \in \{1, \ldots, n\}$, set $X_i := N_G[v_i] \cap \{v_i, \ldots, v_n\}$;[13] so, X_i is the closed neighborhood of v_i in the graph $G[v_i, \ldots, v_n]$.

Lemma 13.3.6. *$X_1, \ldots, X_n$ are all cliques of G. Furthermore, for every maximal clique C of G,[14] there exists some $i \in \{1, \ldots, n\}$ such that $C = X_i$.*[15]

Proof. The fact that the sets X_i are cliques follows immediately from the definition of a simplicial elimination ordering and the construction of the sets X_i. Now, let C be a maximal clique of G. Let $i \in \{1, \ldots, n\}$ be minimal with $v_i \in C$. Then clearly, $C \subseteq X_i$. Since C is a maximal clique, and X_i is a clique, it follows that $C = X_i$. □

[12] By Theorem 13.3.5, every chordal graph has a simplicial elimination ordering, and clearly, we can find such an ordering in polynomial time.

[13] As usual, for a graph G and a vertex $x \in V(G)$, we denote by $N_G(x)$ the set of all neighbors of x in G, and we set $N_G[x] = \{x\} \cup N_G(x)$. So, $N_G(x)$ is the open neighborhood (or simply neighborhood) of x in G, and $N_G[x]$ is the closed neighborhood of x in G.

[14] As usual, "maximal" means "inclusion-wise maximal."

[15] However, not all X_i's need be maximal cliques.

Lemma 13.3.7 (Fulkerson and Gross, 1965)**.** *G has at most n maximal cliques. Furthermore, equality holds if and only if G is edgeless.*

Proof. The fact that G has at most n maximal cliques follows immediately from Lemma 13.3.6. Clearly, if G is edgeless, then G has precisely n maximal cliques (indeed, each one-vertex subset of $V(G)$ is a maximal clique of G). Suppose now that G has at least one edge; let $i \in \{1, \ldots, n\}$ be the largest index such that v_i has a neighbor in G. Let v_j be a neighbor of v_i in G; by the maximality of i, we have that $j < i$. Then $X_i = \{v_i\}$ and $\{v_j, v_i\} \subseteq X_j$, and so $X_i \subsetneqq X_j$. By Lemma 13.3.6, both X_i and X_j are cliques. So, X_i is **not** a maximal clique of G, and Lemma 13.3.6 implies that G has fewer than n maximal cliques. □

A *clique cover* of a graph H is a partition of $V(H)$ into cliques. The *clique cover number* of H, denoted by $\overline{\chi}(H)$, is the smallest size of a clique cover of H; a *minimum clique cover* of H is a clique cover of size precisely $\overline{\chi}(H)$. Note that every graph H satisfies $\alpha(H) \leq \overline{\chi}(H)$.[16] Moreover, since proper colorings correspond to partitions of the vertex set into stable sets (color classes), it is clear that every graph H satisfies $\overline{\chi}(H) = \chi(\overline{H})$.

We define a (finite) sequence $i_1, \ldots, i_t$ as follows. First, let $i_1 := 1$. Once $i_1, \ldots, i_{j-1}$ have been defined, we either terminate or extend the sequence, as follows. If $V(G) = X_{i_1} \cup \cdots \cup X_{i_{j-1}}$, then we set $t = j - 1$, and we terminate the sequence; otherwise, we let $i_j \in \{1, \ldots, n\}$ be the smallest index such that $v_{i_j} \notin X_{i_1} \cup \cdots \cup X_{i_{j-1}}$. Set $Y_1 := X_{i_1}$, and for all $j \in \{2, \ldots, t\}$, set $Y_j := X_{i_j} \setminus (Y_1 \cup \cdots \cup Y_{j-1})$.

Theorem 13.3.8 (Gavril, 1972)**.** *The set $\{v_{i_1}, \ldots, v_{i_t}\}$ is a maximum stable set of G, and $(Y_1, \ldots, Y_t)$ is a minimum clique cover of G.*

Proof. First of all, note that $i_1 < \cdots < i_t$. Suppose that $v_p v_q \in E(G)$ for some $p, q \in \{i_1, \ldots, i_j\}$, with $p < q$; then $v_q \in X_p$, contrary to the choice of q. Thus, $\{v_{i_1}, \ldots, v_{i_t}\}$ is a stable set of size t, and we deduce that $t \leq \alpha(G)$.

Further, it is clear that $Y_1, \ldots, Y_t$ are pairwise disjoint cliques.[17] It is also clear that $V(G) = Y_1 \cup \cdots \cup Y_t$, for otherwise, we could extend the sequence $i_1, \ldots, i_t$.[18] Thus, $(Y_1, \ldots, Y_t)$ is a clique cover of G, and it follows that $\overline{\chi}(G) \leq t$.

[16] This is because a clique and a stable set can intersect in at most one vertex. So, if the vertex set of a graph can be partitioned into k cliques, then no stable set of that graph has more than k vertices.

[17] Indeed, for all $j \in \{1, \ldots, t\}$, we have that $Y_j \subseteq X_{i_j}$, and by Lemma 13.3.6, X_{i_j} is a clique. The fact that $Y_1, \ldots, Y_t$ are pairwise disjoint follows from the construction of $Y_1, \ldots, Y_t$.

[18] We are using the fact that $Y_1 \cup \cdots \cup Y_t = X_{i_1} \cup \cdots \cup X_{i_t}$.

We now have that $t \leq \alpha(G) \leq \overline{\chi}(G) \leq t$, and it follows that $\alpha(G) = \overline{\chi}(G) = t$. Thus, $\{y_1, \ldots, y_t\}$ is a maximum stable set of G, and $(Y_1, \ldots, Y_t)$ is a minimum clique cover of G. □

We remind the reader that the greedy coloring algorithm was discussed in section 12.1.

Lemma 13.3.9. *G can be optimally colored (i.e. properly colored using precisely $\chi(G)$ colors) by applying the greedy coloring algorithm to G with the ordering $v_n, \ldots, v_1$.*[19]

Proof. Clearly, the greedy coloring algorithm produces a proper coloring of G. If we apply the greedy coloring algorithm to G with the ordering $v_n, \ldots, v_1$, then when we reach a vertex v_i, the neighbors of v_i that have already been colored are precisely those from the clique $X_i \setminus \{v_i\}$, and consequently, at most $\omega(G) - 1$ neighbors of v_i have already been colored.[20] Thus, the greedy coloring algorithm applied to G with this ordering uses no more than $\omega(G)$ colors. Since every graph H satisfies $\chi(H) \geq \omega(H)$, it follows that the greedy coloring algorithm used precisely $\omega(G)$ colors, and that the coloring that it produced is optimal. □

Clearly, Lemma 13.3.6, Theorem 13.3.8, and Lemma 13.3.9 yield polynomial time algorithms for finding a maximum clique, a maximum stable set, a minimum clique-cover, and an optimal coloring of a chordal graph.

[19] So, we are using the reverse of our simplicial elimination ordering.

[20] Indeed, X_i is a clique, and the size of this clique is at most $\omega(G)$. So, $|X_i \setminus \{v_i\}| \leq \omega(G) - 1$.

Chapter 14

Perfect graphs

Remark: Recall that "maximal" means "inclusion-wise maximal," and "maximum" means "of maximum possible cardinality." This applies (for example) to cliques, stable sets, chains, and antichains.[1]

14.1 The Perfect Graph Theorem

Recall that a graph H is an induced subgraph of a graph G if $V(H) \subseteq V(G)$ and for all distinct $u, v \in V(H)$, we have that $uv \in E(H)$ if and only if $uv \in E(G)$.

Recall that a graph is *perfect* if all its induced subgraphs H satisfy $\chi(H) = \omega(H)$. A graph is *imperfect* if it is not perfect.

We note that in the study of perfect graphs, it is often useful to think of proper colorings as partitions of the vertex set into stable sets (see the discussion in subsection 12.1.1).

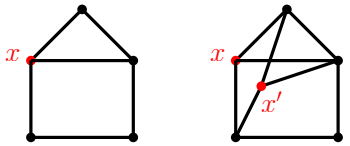
In 1961, Berge conjectured that a graph is perfect if and only if its complement is perfect (this conjecture is known as the "Weak Perfect Graph Conjecture").[2] In 1972, Lovász proved the conjecture, which is now known as the Perfect Graph Theorem. Our goal in this section is to prove this theorem.

Duplicating a vertex x of a graph G produces a supergraph $G \circ x$ by adding to G a vertex x' and making it adjacent to all the neighbors of x in

[1] Chains and antichains are defined in section 14.2.

[2] Recall that for a graph G, the *complement* of G, denoted by $\overline{G}$, is the graph whose vertex set is $V(G)$, and in which any two distinct vertices are adjacent if and only if they are non-adjacent in G.

G, and to no other vertices of G (in particular, x and x' are non-adjacent in $G \circ x$). An example is shown below.



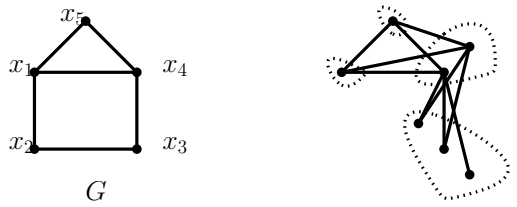
G $\quad$ $G \circ x$

Note that if G' is a proper induced subgraph of $G \circ x$ (i.e. an induced subgraph of $G \circ x$ that is not equal to $G \circ x$), then we have the following:

- if $V(G')$ contains at most one of x, x', then G' is isomorphic to an induced subgraph of G;
- if $V(G')$ contains both x and x', then there exists a proper induced subgraph H of G such that $G' = H \circ x$.

The above observation is key for inductive proofs involving $G \circ x$ (see the proofs of Claims 1 and 2 of Lemma 14.1.1).

Vertex multiplication of a graph G with vertex set $V(G) = \{x_1, \ldots, x_n\}$ by a non-negative integer vector $h = (h_1, \ldots, h_n)$ is the graph $G \circ h$ having h_i pairwise non-adjacent copies of x_i, such that copies of x_i and x_j are adjacent in $G \circ x$ if and only if $x_i x_j \in E(G)$. An example is shown below.



G

$h = (1, 0, 3, 2, 1)$ $\quad$ $G \circ h$

Recall that a *clique cover* of a graph G is a partition of $V(G)$ into cliques. The *clique cover number* of G, denoted by $\overline{\chi}(G)$, is the smallest size of a clique cover of G; a *minimum clique cover* of G is a clique cover of size precisely $\overline{\chi}(G)$. Clearly, $\overline{\chi}(G) = \chi(\overline{G})$ and $\alpha(G) \leq \overline{\chi}(G)$.

Initially, Berge defined two types of perfection, called "χ-perfection" and "α-perfection."[3]

[3]By the Perfect Graph Theorem, χ-perfection and α-perfection are equivalent. However, we have not proven this yet.

- A graph G is *χ-perfect* if every induced subgraph H of G satisfies $\chi(H) = \omega(H)$.[4]
- A graph G is *α-perfect* if every induced subgraph H of G satisfies $\overline{\chi}(H) = \alpha(H)$.

Obviously, a graph is χ-perfect (i.e. perfect) if and only if its complement is α-perfect.

Lemma 14.1.1. *[Berge, 1961] Vertex multiplication preserves χ-perfection and α-perfection.*[5]

Proof. We begin by proving a couple of claims, which (as we shall see) easily imply the lemma.

Claim 1. Vertex duplication preserves χ-perfection.

Proof of Claim 1. Let G be a χ-perfect graph, and assume inductively that any graph obtained by duplicating one vertex of a χ-perfect graph on fewer than $|V(G)|$ vertices is χ-perfect. Let $x \in V(G)$; we must show that $G \circ x$ is χ-perfect. Let x' be the "duplicate" of x in $G \circ x$. It suffices to show that $\chi(G \circ x) = \omega(G \circ x)$, for the rest follows from the induction hypothesis. Clearly, we can extend an optimal coloring of G to a proper coloring of $G \circ x$, by giving x' the same color as x. So, $\chi(G \circ x) = \chi(G)$. Further, no clique contains both x and x', and it readily follows that $\omega(G \circ x) = \omega(G)$. Since G is χ-perfect, we have that $\chi(G) = \omega(G)$, and we now see that $\chi(G \circ x) = \chi(G) = \omega(G) = \omega(G \circ x)$. This proves Claim 1. ♦

Claim 2. Vertex duplication preserves α-perfection.

Proof of Claim 2. Let G be an α-perfect graph, and assume inductively that any graph obtained by duplicating one vertex of an α-perfect graph on fewer than $|V(G)|$ vertices is α-perfect. Let $x \in V(G)$; we must show that $G \circ x$ is α-perfect. Let x' be the "duplicate" of x in $G \circ x$. It suffices to show that $\overline{\chi}(G \circ x) = \alpha(G \circ x)$, for the rest follows from the induction hypothesis.

[4] In other words, χ-perfection is, by definition, the same as perfection.

[5] This means that for every graph G with vertex set $V(G) = \{x_1, \ldots, x_n\}$, and every non-negative integer vector $h = (h_1, \ldots, h_n)$, we have the following:

- if G is χ-perfect, then so is $G \circ h$;
- if G is α-perfect, then so is $G \circ h$.

Suppose first that x belongs to a maximum stable set of G. Then $\alpha(G \circ x) = \alpha(G) + 1$.[6] Since $\overline{\chi}(G) = \alpha(G)$ (because G is α-perfect), we can obtain a clique cover of size $\alpha(G) + 1$ by adding $\{x'\}$ as a one-vertex clique to some set of $\overline{\chi}(G)$ cliques covering G. This is enough because now we have that $\overline{\chi}(G) + 1 = \alpha(G) + 1 = \alpha(G \circ x) \leq \overline{\chi}(G \circ x) \leq \overline{\chi}(G) + 1$, and so $\overline{\chi}(G \circ x) = \alpha(G \circ x)$.

We may now assume that x does not belong to any maximum stable set of G. Then $\alpha(G \circ x) = \alpha(G)$. Let Q be the clique containing x in a minimum clique cover of G. Since $\overline{\chi}(G) = \alpha(G)$, Q intersects every maximum stable set of G.[7] Since x belongs to no maximum stable set, $Q' = Q \setminus \{x\}$ also intersects every maximum stable set, and hence $\alpha(G \setminus Q') = \alpha(G) - 1$. Since G is α-perfect, $\overline{\chi}(G \setminus Q') = \alpha(G \setminus Q')$. To a set of $\alpha(G) - 1$ many cliques covering $G \setminus Q'$, add the clique $Q' \cup \{x'\}$ to obtain a set of $\alpha(G) = \alpha(G \circ x)$ many cliques covering $G \circ x$; we now have that $\overline{\chi}(G \circ x) = \alpha(G \circ x)$. This proves Claim 2. ♦

Let G be a graph with vertex set $V(G) = \{x_1, \ldots, x_n\}$, and let $h = (h_1, \ldots, h_n)$ be a non-negative integer vector. Let A be the set of vertices x_i for which $h_i > 0$. Clearly, if G is χ-perfect (resp. α-perfect), then $G[A]$ is also χ-perfect (resp. α-perfect). Now, $G \circ h$ can be obtained from $G[A]$ by a sequence of vertex duplications: if every h_i is 0 or 1 then $G \circ h = G[A]$, and otherwise, $G \circ h$ can be obtained from $G[A]$ by repeatedly duplicating vertices until there are h_i copies of each x_i. Since vertex duplication preserves χ-perfection and α-perfection (by Claims 1 and 2), an easy induction now guarantees that if G is χ-perfect (resp. α-perfect), then $G \circ h$ is also χ-perfect (resp. α-perfect). This completes the argument. □

The Perfect Graph Theorem (Lovász, 1972)**.** *A graph is perfect if and only if its complement is perfect.*

Proof. Obviously, it is enough to prove that if a graph is perfect, then so is its complement. For this, it is in fact enough to prove that every α-perfect graph is χ-perfect, for then we will have the following sequence of implications for

[6] Indeed, if S is a stable set of G of size $\alpha(G)$, and such that $x \in S$, then $S \cup \{x'\}$ is a stable set of size $\alpha(G) + 1$ in $G \circ x$.

[7] Let us check this. Let S be a maximum stable set of G, i.e. a stable set of G such that $|S| = \alpha(G)$. Let $\{Q_1, \ldots, Q_t\}$ be any optimal clique cover of G, i.e. $t = \overline{\chi}(G)$. Since a clique and a stable set can have at most one vertex in common, we see that S intersects each of $Q_1, \ldots, Q_t$ in at most one vertex. But since $S \subseteq Q_1 \cup \cdots \cup Q_t$ and $|S| = \alpha(G) = \overline{\chi}(G) = t$, it follows that S intersects each of $Q_1, \ldots, Q_t$ in exactly one vertex. Since Q belongs to some optimal clique-cover of G, we deduce that $|Q \cap S| = 1$.

a graph G:

$$G \text{ is } (\chi\text{-})\text{perfect} \implies \overline{G} \text{ is } \alpha\text{-perfect} \implies \overline{G} \text{ is } (\chi\text{-})\text{perfect},$$

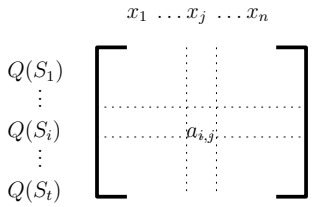
which is what we need.[8]

Now, fix an α-perfect graph G, and assume inductively that all α-perfect graphs on fewer than $|V(G)|$ vertices are χ-perfect. We must show that G is χ-perfect. In view of the induction hypothesis, it suffices to show that $\chi(G) = \omega(G)$.[9] We may assume that $\omega(G) \geq 2$, for otherwise, G is edgeless and the result is immediate.

Suppose first that G has a stable set S that intersects every maximum clique of G. Then by the minimality of G, $\chi(G \setminus S) = \omega(G \setminus S) = \omega(G) - 1$. But now $\chi(G) = \omega(G)$, since we can properly color $G \setminus S$ with $\omega(G) - 1$ colors, and then color all vertices of S with the same new color.

From now on, we assume that every stable set S of G misses (i.e. has an empty intersection with) some maximum clique $Q(S)$; our goal is to derive a contradiction. Set $V(G) = \{x_1, \ldots, x_n\}$, and let $\mathcal{S} = \{S_1, \ldots, S_t\}$ be the set of all maximal stable sets of G. For every vertex x_j, let h_j be the number of stable sets S in $\mathcal{S}$ such that $x_j \in Q(S)$. Set $h := (h_1, \ldots, h_n)$. By Lemma 14.1.1, $H := G \circ h$ is α-perfect, and so $\overline{\chi}(H) = \alpha(H)$.

Let $A = [a_{i,j}]_{t \times n}$ be a 0,1-matrix of the incidence relation between the set of $Q(S)$'s for $S \in \mathcal{S}$ and $V(G)$. So, $a_{i,j} = 1$ if and only if $x_j \in Q(S_i)$.



By construction, h_j is the number of 1's in column j of A, and $|V(H)|$ is the total number of 1's in A.

Since each row contributes $\omega(G)$ many 1's, we have $|V(H)| = \omega(G)|\mathcal{S}|$. Since vertex duplication cannot enlarge cliques, we have $\omega(H) \leq \omega(G)$. Therefore, $\overline{\chi}(H) \geq \frac{|V(H)|}{\omega(H)} \geq \frac{|V(H)|}{\omega(G)} = |\mathcal{S}|$.

Every stable set of H consists of copies of elements in some stable set of G; so, a maximum stable set of H consists of all copies of all vertices in

[8] The first of the implications above ("G is χ-perfect $\implies$ $\overline{G}$ is α-perfect") follows straight from the definition of χ-perfection and α-perfection.

[9] Indeed, suppose that H is a proper induced subgraph of G. Then H is an α-perfect graph on fewer than $|V(G)|$ vertices, and consequently, H is χ-perfect. So, $\chi(H) = \omega(H)$. Thus, to show that G is χ-perfect, it suffices to show that $\chi(G) = \omega(G)$.

some maximal stable set of G. Consequently,

$$\alpha(H) \;=\; \max_{T\in\mathcal{S}} \sum_{j:x_j\in T} h_j.$$

The sum above counts the 1's in A that appear in the columns indexed by the vertices of T. If we count these 1's by rows, we get

$$\alpha(H) \;=\; \max_{T\in\mathcal{S}} \sum_{S\in\mathcal{S}} |T \cap Q(S)|.$$

Since T is a stable set, it has at most one vertex in each chosen clique $Q(S)$. Also, $T \cap Q(T) = \emptyset$. So, $|T \cap Q(S)| \leq 1$ for every $S \in \mathcal{S}$, and $|T \cap Q(T)| = 0$. It follows that $\alpha(H) \leq |\mathcal{S}| - 1$. Therefore, $\alpha(H) < \overline{\chi}(H)$, contrary to the fact that H is α-perfect. □

14.2 Dilworth's theorem and comparability graphs

Recall that a *partial order* of a set X is a binary relation on X that is reflexive, antisymmetric, and transitive.[10] A *partially ordered set* (or *poset*) is an ordered pair $(X, \preceq)$ such that X is a set and $\preceq$ is a partial order on X. A *maximal* element of $(X, \preceq)$ is $x \in X$ such that no $y \in X \setminus \{x\}$ satisfies $x \preceq y$.[11] Similarly, a *minimal* element of $(X, \preceq)$ is $x \in X$ such that no $y \in X \setminus \{x\}$ satisfies $y \preceq x$.[12] We say that $x, y \in X$ are *comparable* with respect to $\preceq$ if either $x \preceq y$ or $y \preceq x$; two elements of X are *incomparable* with respect to $\preceq$ if they are not comparable with respect to $\preceq$. A *chain* in $(X, \preceq)$ is a set $C \subseteq X$ such that any two elements of C are comparable with respect to $\preceq$. A *chain decomposition* of $(X, \preceq)$ is a partition of X into chains of $(X, \preceq)$. An *antichain* in $(X, \preceq)$ is a set $A \subseteq X$ such that no two elements of A are comparable with respect to $\preceq$.

[10] A binary relation $\preceq$ on a set X is

- *reflexive* if for all $x \in X$, we have that $x \preceq x$;
- *antisymmetric* if for all $x, y \in X$, we have that if $x \preceq y$ and $y \preceq x$, then $x = y$;
- *transitive* if for all $x, y, z \in X$, we have that if $x \preceq y$ and $y \preceq z$, then $x \preceq z$.

[11] A partially ordered set may or may not contain a maximal element. Furthermore, if a partially ordered set does contain a maximal element, then this maximal element may or may not be unique.

[12] A partially ordered set may or may not contain a minimal element. Furthermore, if a partially ordered set does contain a minimal element, then this minimal element may or may not be unique.

Dilworth's theorem. *In any finite partially ordered set* $(X, \preceq)$,[13] *the maximum size of an antichain is equal to the minimum size of a chain decomposition of* $(X, \preceq)$.[14]

Proof. Let $(X, \preceq)$ be a finite partially ordered set, and assume inductively that the theorem is true for smaller partially ordered sets.[15] We may assume that $X \neq \emptyset$, for otherwise, the result is immediate. First, it is clear that if $(X, \preceq)$ has an antichain of size k, then no chain decomposition of $(X, \preceq)$ is of size smaller than k (this is because no chain can contain more than one element of an antichain).[16] It remains to exhibit an antichain of $(X, \preceq)$ and a chain decomposition of $(X, \preceq)$ of the same size.[17]

Since $(X, \preceq)$ is a non-empty, finite poset, we see that $(X, \preceq)$ has a maximal element, say x_0. Set $X_0 := X \setminus \{x_0\}$, and let A_0 be a maximum antichain in $(X_0, \preceq)$;[18] set $k := |A_0|$. By the induction hypothesis, $(X_0, \preceq)$ has a chain decomposition of size k, say $\{C_1, \ldots, C_k\}$.

> **Claim 1.** Any antichain of size k in $(X_0, \preceq)$ intersects each of $C_1, \ldots, C_k$ in exactly one element.

Proof of Claim 1. Let B be an antichain of size k in $(X_0, \preceq)$. Since B is an antichain and $C_1, \ldots, C_k$ are chains in $(X_0, \preceq)$, we see that B intersects each of $C_1, \ldots, C_k$ in at most one element. But since $B \subseteq C_1 \cup \cdots \cup C_k$ and $|B| = k$, we see that B intersects each of $C_1, \ldots, C_k$ in exactly one element. This proves Claim 1. ♦

Now, for all $i \in \{1, \ldots, k\}$, let C_i' be the set of all elements of C_i that belong to some antichain of $(X_0, \preceq)$ of size k; then $C_i' \neq \emptyset$ (because, by Claim 1, we have that $C_i \cap A_0 \neq \emptyset$), and we deduce that C_i' has a unique maximal element, call it x_i.[19]

> **Claim 2.** $\{x_1, \ldots, x_k\}$ is an antichain in $(X_0, \preceq)$.[20]

[13] Here, "finite" simply means that X is finite.

[14] The *size* of a chain decomposition is the number of chains in it.

[15] So, we are assuming that for all finite partially ordered sets $(X', \preceq')$ such that $|X'| < |X|$, the maximum size of an antichain is equal to the minimum size of a chain decomposition of $(X', \preceq')$.

[16] Thus, the maximum size of an antichain in $(X, \preceq)$ is no greater than the minimum size of a chain decomposition of $(X, \preceq)$.

[17] This will imply that the maximum size of an antichain in $(X, \preceq)$ is no smaller than the minimum size of a chain decomposition of $(X, \preceq)$.

[18] That is: A_0 is an antichain in $(X_0, \preceq)$ of largest possible cardinality.

[19] Since $C_i' \subseteq C_i$, and C_i is a chain, we know that C_i' is a chain. Furthermore, since X_0 is finite, C_i' is finite, and we already saw that C_i' is non-empty. So, C_i' is a non-empty, finite chain in $(X_0, \preceq)$, and it follows that it has a unique maximal element.

[20] Obviously, this means that $\{x_1, \ldots, x_k\}$ is an antichain in $(X, \preceq)$ as well.

Proof of Claim 2. We may assume that $k \geq 2$, for otherwise, this is immediate. By symmetry, it suffices to show that x_1 and x_2 are incomparable. Let A_1 be an antichain of size k in $(X_0, \preceq)$ such that $x_1 \in A_1 \cap C_1$.[21] By Claim 1, $|A_1 \cap C_2| = 1$; set $A_1 \cap C_2 = \{x_2'\}$. Then $x_2' \in C_2'$, and so (since x_2 is a maximal element of the chain C_2') we have that $x_2' \preceq x_2$.[22] Now, if $x_2 \preceq x_1$, then by the transitivity of $\preceq$, we have that $x_2' \preceq x_1$, which is impossible since x_1 and x_2' are distinct elements of the antichain A_1.[23] So, $x_2 \not\preceq x_1$. An analogous argument establishes that $x_1 \not\preceq x_2$. Thus, x_1 and x_2 are incomparable. This proves Claim 2. ♦

Suppose first that $\{x_0, x_1, \ldots, x_k\}$ is an antichain in $(X, \preceq)$. Then this antichain is of size $k+1$, and $\{C_1, \ldots, C_k, \{x_0\}\}$ is a chain decomposition of $(X, \preceq)$ of size $k+1$, and we are done. So, we may assume that $\{x_0, x_1, \ldots, x_k\}$ is not an antichain in $(X, \preceq)$. By Claim 2, and by symmetry, we may assume that x_0 and x_1 are comparable; since x_0 is a maximal element of $(X, \preceq)$, we see that $x_1 \preceq x_0$. Now, set $D_1 := \{x_0\} \cup \{x \in C_1 \mid x \preceq x_1\}$; since C_1 is a chain, and $x_1 \preceq x_0$, the transitivity of $\preceq$ guarantees that D_1 is a chain in $(X, \preceq)$. Further, by Claim 1, and by the choice of x_1, we know that $(X \setminus D_1, \preceq)$ does not have an antichain of size k. Since $\{x_2, \ldots, x_k\}$ is an antichain of size $k-1$ in $(X \setminus D_1, \preceq)$, we deduce that the maximum size of an antichain in $(X \setminus D_1, \preceq)$ is $k-1$. Then by the induction hypothesis, $(X \setminus D_1, \preceq)$ has a chain decomposition of size $k-1$, say $\{E_1, \ldots, E_{k-1}\}$. But now $\{D_1, E_1, \ldots, E_{k-1}\}$ is a chain decomposition of size k in $(X, \preceq)$, and we are done. □

A *comparability graph* (or a *transitively orientable graph*) is a graph G such that there exists a partial order $\preceq$ on $V(G)$ such that for all distinct $x, y \in V(G)$, we have that $xy \in E(G)$ if and only if x and y are comparable with respect to $\preceq$.[24] Equivalently,[25] G is a comparability graph if there exists an orientation $\vec{G} = (V(G), A(G))$ of G such that for all $\overrightarrow{uv}, \overrightarrow{vw} \in A(G)$, we have that $\overrightarrow{uw} \in A(G)$.

Corollary 14.2.1. *Every comparability graph is perfect. The complement of any comparability graph is perfect.*

[21] Such an A_1 exists because $x_1 \in C_1'$.

[22] Since C_2' is a chain, we know that x_2', x_2 are comparable. Since x_2 is maximal in C_2, we have that $x_2' \preceq x_2$.

[23] The fact that $x_1 \neq x_2'$ follows from the fact that $x_1 \in C_1$, $x_2' \in C_2$, and $C_1 \cap C_2 = \emptyset$.

[24] Note that in a comparability graph, cliques correspond to chains, and stable sets correspond to antichains.

[25] Check that this is really equivalent!

Proof. In view of the Perfect Graph Theorem, it suffices to show that the complement of any comparability graph is perfect. So, fix a comparability graph G, and assume inductively that for all comparability graphs G' on fewer than $|V(G)|$ vertices, the graph $\overline{G'}$ is perfect. We must show that $\overline{G}$ is perfect. Clearly, it suffices to show that $\chi(\overline{G}) = \omega(\overline{G})$, for the rest follows from the induction hypothesis.[26]

Let $\preceq$ be a partial order on $V(G)$ such that for all distinct $x, y \in V(G)$, we have that $xy \in E(G)$ if and only if x and y are comparable with respect to $\preceq$. Let A be a maximum antichain in $(V(G), \preceq)$, and let $(C_1, \ldots, C_k)$ be a chain decomposition of minimum size in $(V(G), \preceq)$. By Dilworth's theorem, we have that $|A| = k$. Now, note that A is a stable set in G, and therefore a clique in $\overline{G}$; so, $\omega(\overline{G}) \geq |A| = k$. On the other hand, $C_1, \ldots, C_k$ are cliques in G, and therefore stable sets in $\overline{G}$; thus, $\{C_1, \ldots, C_k\}$ is a partition of $V(\overline{G})$ into stable sets ("color classes") of $\overline{G}$, and it follows that $\chi(\overline{G}) \leq k$. So, $\chi(\overline{G}) \leq k \leq \omega(\overline{G})$. But obviously, $\chi(\overline{G}) \geq \omega(\overline{G})$, and we deduce that $\chi(\overline{G}) = \omega(\overline{G})$. This completes the argument. □

14.3 Some further examples of perfect graphs

Lemma 14.3.1. *Every bipartite graph is perfect.*

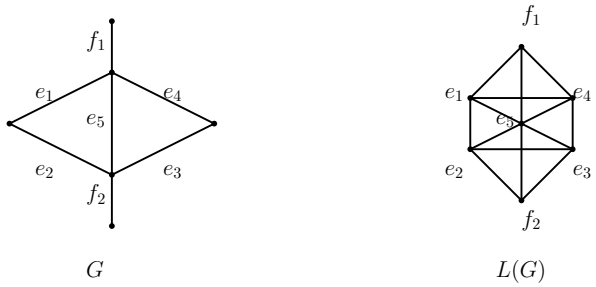
Proof. Since all induced subgraphs of a bipartite graph are bipartite, it suffices to show that every bipartite graph G satisfies $\chi(G) = \omega(G)$. But this is obvious: if G is an edgeless bipartite graph, then $\chi(G) = \omega(G) = 1$, and if G is a bipartite graph that has at least one edge, then $\chi(G) = \omega(G) = 2$. □

Lemma 14.3.2. *The complement of any bipartite graph is perfect.*

Proof. This follows immediately from Lemma 14.3.1 and the Perfect Graph Theorem. □

Recall that the *line graph* of a graph G, denoted by $L(G)$, is the graph with vertex set $E(G)$, in which distinct $e, f \in E(G)$ are adjacent if and only if they share an endpoint in G. An example is shown below.

[26] Indeed, suppose that H is a proper induced subgraph of $\overline{G}$. Then $\overline{H}$ is a comparability graph, and so by the induction hypothesis, $\overline{\overline{H}} = H$ is perfect. Thus, $\chi(H) = \omega(H)$. It only remains to show that $\chi(\overline{G}) = \omega(\overline{G})$.



Lemma 14.3.3. *The line graph of any bipartite graph is perfect.*

Proof. Let G be a bipartite graph, and let $H = L(G)$. We must show that H is perfect. Consider any induced subgraph $H' = H[M]$ of H. So, we have that $M \subseteq V(H)$, and therefore, $M \subseteq E(G)$. Let G' be a subgraph of G with vertex set $V(G)$ and edge set M. Since H' is an induced subgraph of H, it follows that $H' = L(G')$, and consequently, $\chi(H') = \chi'(G')$. On the other hand, since G' is bipartite, Theorem 12.3.4 guarantees that $\chi'(G') = \Delta(G')$. Clearly, $\Delta(G') \leq \omega(H')$, and so it follows that

$$\chi(H') \quad = \quad \chi'(G') \quad = \quad \Delta(G') \quad \leq \quad \omega(H').$$

Since we also know that $\chi(H') \geq \omega(H')$, we deduce that $\chi(H') = \omega(H')$. It follows that H is perfect. □

Lemma 14.3.4. *The complement of the line graph of any bipartite graph is perfect.*

Proof. This follows immediately from Lemma 14.3.3 and the Perfect Graph Theorem. □

14.4 The Strong Perfect Graph Theorem

A *hole* in a graph G is an induced cycle of length at least four.[27] An *antihole* in G is an induced subgraph H of G such that $\overline{H}$ is a hole in $\overline{G}$. An *odd hole* (resp. *odd antihole*) is a hole (resp. antihole) that has an odd number of vertices. Even holes and even antiholes are defined analogously. A graph is *Berge* if it contains no odd holes and no odd antiholes.

The Strong Perfect Graph Theorem (Chudnovsky, Robertson, Seymour, Thomas, 2002)**.** *A graph is perfect if and only if it is Berge.*

[27] Note that this means that chordal graphs are precisely the graphs that contain no holes.

Clearly, a graph is Berge if and only if its complement is Berge. So, the Strong Perfect Graph Theorem immediately implies the Perfect Graph Theorem.

One direction of the Strong Perfect Graph Theorem ("every perfect graph is Berge") is an easy exercise. Indeed, it is easy to check that for each integer $n \geq 2$, we have that

- $\omega(C_{2n+1}) = 2$ and $\chi(C_{2n+1}) = 3$;
- $\omega(\overline{C_{2n+1}}) = n$ and $\chi(\overline{C_{2n+1}}) = n + 1$.

So, odd holes and antiholes are imperfect, and therefore, no perfect graph contains an odd hole or an odd antihole. Thus, every perfect graph is Berge.

What about the other direction ("every Berge graph is perfect")? It relies on a "decomposition theorem" for Berge graphs, which, roughly, states that every Berge graph either is "basic" or admits a "decomposition." (The proof of this decomposition theorem is by far the most complicated part of the proof of the Strong Perfect Graph Theorem, and it is over 100 pages long.) The "basic" graphs are bipartite graphs and their complements, line graphs of bipartite graphs, complements of line graphs of bipartite graphs, and "double split" graphs (we omit the definition). All basic graphs are perfect: we proved this for the first four types of basic graphs, and the proof for double split graphs is easy. There are several "decompositions" (we omit the definitions), and it can be shown that no imperfect Berge graph of minimum possible order admits any of these decompositions. It now follows that all Berge graphs are perfect.

14.5 Algorithmic considerations

In 2005, Chudnovsky, Cornuéjols, Liu, Seymour, and Vušković constructed an $O(n^9)$ time recognition algorithm for Berge graphs. By the Strong Perfect Graph Theorem, it follows that perfect graphs can be recognized in $O(n^9)$ time.

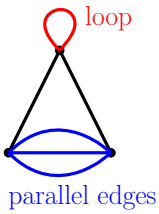
Further, Grötschel, Lovász, and Schrijver (1981) showed that the following optimization problems can be solved in polynomial time for perfect graphs: Maximum Clique, Maximum Stable Set, Graph Coloring (i.e. Vertex Coloring), and Minimum Clique Cover. In fact, weighted versions of these problems can also be solved in polynomial time.

Chapter 15

The Tutte polynomial

15.1 Multigraphs

A *multigraph* is an ordered pair $G = (V(G), E(G))$ such that $V(G)$ and $E(G)$ are finite sets (called the *vertex set* and *edge set*, respectively), and each edge (i.e. element of $E(G)$) is associated with two (possibly identical) vertices (i.e. elements of $V(G)$), called its *endpoints*. If an edge has only one endpoint (i.e. its two endpoints are the same), then this edge is called a *loop*. If two distinct edges have the same endpoints, then those edges are *parallel*. An edge is *incident* with a vertex, if that vertex is an endpoint of the edge. The *degree* of a vertex in a multigraph is the number of edges that it is incident with, with loops counting twice. (In the example below, all vertices are of degree four.) A multigraph is *loopless* if it has no loops.



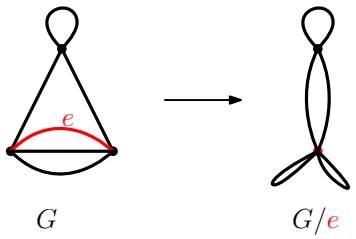
A *proper (vertex) coloring* of a loopless multigraph G is an assignment of colors to the vertices of G in such a way that, whenever two distinct vertices are joined by an edge (i.e. are the endpoints of the same edge), they receive different colors. If a multigraph has a loop, then it has no proper colorings.[1] A *proper k-coloring* (or simply *k-coloring*) of a loopless multigraph G is a proper coloring of G that uses colors $1, \ldots, k$ (not all of these colors need be

[1] Here, the idea is that if e is a loop, then its unique endpoint is adjacent to itself.

used); G is *k-colorable* if it admits a k-coloring. The chromatic number of a loopless multigraph G, denoted by $\chi(G)$, is the smallest integer k such that G is k-colorable.

As usual, for an edge e of a multigraph G, we denote by $G - e$ the multigraph obtained by deleting e from G.

If e is a non-loop edge of a multigraph G, then the multigraph G/e obtained by *contracting* e is the multigraph obtained by first deleting e, and then identifying its endpoints to a single vertex. (Note that edges parallel to e become loops, and it is also possible that new parallel edges are created). An example is shown below.



The topic of this chapter are graph polynomials, or more precisely, multigraph polynomials (for recursive purposes, it is convenient to allow loops and parallel edges). There are a number of such polynomials. Here, we consider two: the chromatic polynomial and the Tutte polynomial.

15.2 The chromatic polynomial

Lemma 15.2.1. *For each multigraph G, there exists a unique polynomial π_G such that for any non-negative integer k, $\pi_G(k)$ is the number of k-colorings of G. Moreover, the (unique) polynomial π_G is of degree at most $|V(G)|$ and has integer coefficients.*

Proof. We proceed by induction on the number of edges. Fix a multigraph G, and assume inductively that the lemma is true for multigrpahs with fewer than $|E(G)|$ edges.[2]

Uniqueness follows immediately from the fact that any two polynomials that agree on infinitely many points are identical. It remains to prove existence. If G is edgeless, then $\pi_G(x) = x^{|V(G)|}$ is the polynomial that we

[2]So, we assume inductively that for all multigraphs G' such that $|E(G')| < |E(G)|$, there exists a unique polynomial $\pi_{G'}$ such that for any non-negative integer k, $\pi_{G'}(k)$ is the number of k-colorings of G', and moreover, the degree of this (unique) polynomial is at most $|V(G')|$ and its coefficients are integers.

need.[3] If G has at least one loop, then $\pi_G(x) = 0$ is the polynomial we need.[4] From now on, we assume that G is loopless and has at least one edge, say e. The induction hypothesis applied to $G - e$ and G/e yields polynomials π_{G-e} and $\pi_{G/e}$ of degree at most $|V(G)|$, and having the desired properties. Set

$$\pi_G \quad := \quad \pi_{G-e} - \pi_{G/e}.$$

By the induction hypothesis, π_G is of degree at most $|V(G)|$ and has integer coefficients. Now, fix a non-negative integer k. We must show that there are precisely $\pi_G(k)$ many k-colorings of G. Clearly, every k-coloring of G is also a proper coloring of $G - e$. On the other hand, a k-coloring of $G - e$ is a k-coloring of G if and only if the two endpoints of e have different colors. Further, k-colorings of $G - e$ in which both endpoints of e receive the same color correspond to k-colorings of G/e in the natural way. So, the number of k-colorings of G is equal to $\pi_{G-e}(k) - \pi_{G/e}(k) = \pi_G(k)$, which is what we needed. □

The *chromatic polynomial* of a multigraph G is the polynomial π_G from the statement of Lemma 15.2.1. Note that the proof of that lemma in fact gives us a recursive formula for π_G, as follows:

- if G is edgeless, then $\pi_G(x) = x^{|V(G)|}$;
- if G has a loop, then $\pi_G(x) = 0$;
- if G is loopless and has at least one edge, say e, then

$$\pi_G(x) \quad = \quad \pi_{G-e}(x) - \pi_{G/e}(x).$$

Note that $G - e$ and G/e have fewer edges than G, and so our formula really is recursive.

We remark that if G is a loopless multigraph, then $\chi(G)$ is equal to the smallest non-negative integer k such that $\pi_G(k) \neq 0$. Note that this implies that computing the chromatic polynomial is NP-hard. However, in some special cases, the chromatic polynomial is easy to compute. For example:

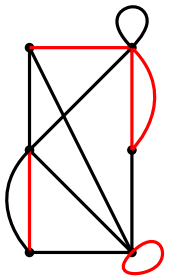
- $\pi_{K_n}(x) = x(x-1)(x-2)\dots(x-n+1)$;
- $\pi_T(x) = x(x-1)^{n-1}$, for any tree T on n vertices.

[3] Indeed, if G is edgeless, then for any non-negative integer k, there are $k^{|V(G)|}$ many k-colorings of G (we simply assign colors from the set $\{1, \dots, k\}$ independently to the vertices of G).

[4] Indeed, if G has at least one loop, then G has no proper colorings.

15.3 The Tutte polynomial

For a multigraph G, let $k(G)$ be the number of components of G; for a set $A \subseteq E(G)$, let $k_G(A)$ be the number of components of the multigraph on vertex set $V(G)$ and edge set A. Note that $k_G(A) \geq \max\{k(G), |V(G)| - |A|\}$,[5] and set $r_G(A) := k_G(A) - k(G)$ and $c_G(A) := k_G(A) + |A| - |V(G)|$.[6] For example, in the multigraph below (with the edges of A in red), we have that $k(G) = 1$, $k_G(A) = 3$, $|A| = 5$, and $|V(G)| = 6$; so, $r_G(A) = 2$ and $c_G(A) = 2$.



Now, the *Tutte polynomial* $T_G(x, y)$ of a multigraph G is defined by

$$T_G(x,y) \quad := \quad \sum_{A \subseteq E(G)} (x-1)^{r_G(A)} (y-1)^{c_G(A)}.$$

[5] Let us check this. Since adding edges to a multigraph cannot increase the number of components, it is clear that $k_G(A) \geq k(G)$. It remains to show that $k_G(A) \geq |V(G)| - |A|$, or equivalently, that $|V(G)| \leq k_G(A) + |A|$. Set $t := k_G(A)$, and let $C_1, \ldots, C_t$ be the components of the multigraph with vertex set $V(G)$ and edge set A. For each $i \in \{1, \ldots, t\}$, let T_i be a spanning tree of C_i, so that $|V(C_i)| = |V(T_i)| = |E(T_i)| + 1$. We now have that

$$\begin{aligned}
|V(G)| &= |V(C_1) \cup \cdots \cup V(C_t)| \\
&= \sum_{i=1}^{t} |V(C_i)| \\
&= \sum_{i=1}^{t} (|E(T_i)| + 1) \\
&= t + \sum_{i=1}^{t} |E(T_i)| \\
&= t + |E(T_1) \cup \cdots \cup E(T_t)| \\
&\leq t + |A|,
\end{aligned}$$

which is what we needed to show.

[6] Note that $r_G(A)$ and $c_G(A)$ are both non-negative.

As we shall see, the Tutte polynomial is more "general" than the chromatic polynomial, i.e. if we know the Tutte polynomial, we can easily compute the chromatic polynomial (see section 15.4 below). Since it is NP-hard to compute the chromatic polynomial, it is NP-hard to compute the Tutte polynomial.

Clearly, if G is edgeless, then $T_G(x, y) = 1$. Otherwise, we can get a recursive formula for $T_G(x, y)$, as follows. (A *bridge* in a multigraph G is an edge e of G such that $G - e$ has more components than G.)

Lemma 15.3.1. *Let e be an edge of a multigraph G. Then*

$$T_G(x,y) = \begin{cases} xT_{G/e}(x,y) & \text{if } e \text{ is a bridge of } G \\ yT_{G-e}(x,y) & \text{if } e \text{ is a loop of } G \\ T_{G-e}(x,y) + T_{G/e}(x,y) & \text{otherwise} \end{cases}$$

Proof.

Claim 1. If e is a bridge of G, then $T_G(x, y) = xT_{G/e}(x, y)$.

Proof of Claim 1. Assume that e is a bridge of G. Then for any $A \subseteq E(G)\backslash\{e\}$, we have the following:

(1) $r_G(A) - 1 \overset{(*)}{=} r_G(A \cup \{e\}) \overset{(**)}{=} r_{G/e}(A)$,

(2) $c_G(A \cup \{e\}) \overset{(*)}{=} c_G(A) \overset{(**)}{=} c_{G/e}(A)$,

where, in both (1) and (2), (*) follows from the fact that e is a bridge of G, and (**) follows from the fact that contracting an edge does not change the

number of components. We now compute:

$$\begin{aligned}
& T_G(x,y) \\
= & \sum_{A\subseteq E(G)} (x-1)^{r_G(A)}(y-1)^{c_G(A)} \\
= & \sum_{A\subseteq E(G)\setminus\{e\}} \Big((x-1)^{r_G(A)}(y-1)^{c_G(A)} + (x-1)^{r_G(A\cup\{e\})}(y-1)^{c_G(A\cup\{e\})}\Big) \\
\overset{(1)\&(2)}{=} & \sum_{A\subseteq E(G/e)} \Big((x-1)^{r_{G/e}(A)+1}(y-1)^{c_{G/e}(A)} + (x-1)^{r_{G/e}(A)}(y-1)^{c_{G/e}(A)}\Big) \\
= & x\sum_{A\subseteq E(G/e)} (x-1)^{r_{G/e}(A)}(y-1)^{c_{G/e}(A)} \\
= & xT_{G/e}(x,y).
\end{aligned}$$

This proves Claim 1. ♦

Claim 2. If e is a loop of G, then $T_G(x,y) = yT_{G-e}(x,y)$.

Proof of Claim 2. Assume that e is a loop of G. Deleting e does not affect the number of components, and so for each $A \subseteq E(G)\setminus\{e\}$, we have the following:

(1) $r_G(A) = r_G(A\cup\{e\}) = r_{G-e}(A)$,

(2) $c_G(A\cup\{e\}) - 1 = c_G(A) = c_{G-e}(A)$.

We now compute:

$$\begin{aligned}
& T_G(x,y) \\
= & \sum_{A\subseteq E(G)} (x-1)^{r_G(A)}(y-1)^{c_G(A)} \\
= & \sum_{A\subseteq E(G)\setminus\{e\}} \Big((x-1)^{r_G(A)}(y-1)^{c_G(A)} + (x-1)^{r_G(A\cup\{e\})}(y-1)^{c_G(A\cup\{e\})}\Big) \\
\overset{(1)\&(2)}{=} & \sum_{A\subseteq E(G-e)} \Big((x-1)^{r_{G-e}(A)}(y-1)^{c_{G-e}(A)} + (x-1)^{r_{G-e}(A)}(y-1)^{c_{G-e}(A)+1}\Big) \\
= & y \sum_{A\subseteq E(G-e)} (x-1)^{r_{G-e}(A)}(y-1)^{c_{G-e}(A)} \\
= & yT_{G-e}(x,y).
\end{aligned}$$

This proves Claim 2. ♦

> **Claim 3.** If e is neither a bridge nor a loop of G, then $T_G(x,y) = T_{G-e}(x,y) + T_{G/e}(x,y)$.

Proof of Claim 3. Assume that e is neither a bridge nor a loop of G. Then $k(G-e) = k(G/e) = k(G)$, and it follows that for all $A \subseteq E(G)\setminus\{e\}$, we have the following:

(1) $r_G(A) = r_{G-e}(A)$,

(2) $r_G(A\cup\{e\}) = r_{G/e}(A)$,

(3) $c_G(A) = c_{G-e}(A)$,

(4) $c_G(A\cup\{e\}) = c_{G/e}(A)$.

We now compute:

$$\begin{aligned}
& T_G(x,y) \\
= & \sum_{A\subseteq E(G)} (x-1)^{r_G(A)}(y-1)^{c_G(A)} \\
= & \sum_{A\subseteq E(G)\setminus\{e\}} \Big((x-1)^{r_G(A)}(y-1)^{c_G(A)} + (x-1)^{r_G(A\cup\{e\})}(y-1)^{c_G(A\cup\{e\})}\Big) \\
\overset{(1)-(4)}{=} & \sum_{A\subseteq E(G)\setminus\{e\}} \Big((x-1)^{r_{G-e}(A)}(y-1)^{c_{G-e}(A)} + (x-1)^{r_{G/e}(A)}(y-1)^{c_{G/e}(A)}\Big) \\
= & T_{G-e}(x,y) + T_{G/e}(x,y).
\end{aligned}$$

This proves Claim 3. ♦

By Claims 1, 2, and 3, we are done. □

Further, it turns out that the Tutte polynomial is "multiplicative" in a certain sense, as the following lemma shows.

Lemma 15.3.2. *If multigraphs G_1 and G_2 have at most one vertex and no edges in common, then $T_{G_1\cup G_2} = T_{G_1}T_{G_2}$.*

Proof. We prove this by induction on the number of edges, using Lemma 15.3.1. So, fix multigraphs G_1 and G_2 that have at most one vertex and no edges in common, set $G := G_1 \cup G_2$, and assume inductively that the lemma is true for multigraphs with fewer than $|E(G)|$ edges.[7] If G is edgeless then so are G_1 and G_2, and we have that $T_G(x,y) = T_{G_1}(x,y) = T_{G_2}(x,y) = 1$, and we are done. So, we may assume that G has at least one edge, say e. By symmetry, we may assume that $e \in E(G_2)$. Note that this means that $G - e = G_1 \cup (G_2 - e)$ and (if e is not a loop) $G/e = G_1 \cup (G_2/e)$.

[7] So, we are assuming inductively that for all multigraphs G_1' and G_2' that have at most one vertex and no edges in common, if the multigraph $G_1' \cup G_2'$ has fewer than $|E(G)|$ edges, then $T_{G_1'\cup G_2'} = T_{G_1'}T_{G_2'}$.

Suppose first that e is a bridge of G (and therefore of G_2 as well). Then

$$\begin{aligned}
T_G(x,y) &= xT_{G/e}(x,y) && \text{by Lemma 15.3.1}\\
&= xT_{G_1\cup(G_2/e)}(x,y)\\
&= xT_{G_1}(x,y)T_{G_2/e}(x,y) && \text{by the induction hypothesis}\\
&= T_{G_1}(x,y)\Big(xT_{G_2/e}(x,y)\Big)\\
&= T_{G_1}(x,y)T_{G_2}(x,y) && \text{by Lemma 15.3.1.}
\end{aligned}$$

Next, suppose that e is a loop of G (and therefore of G_2 as well). Then

$$\begin{aligned}
T_G(x,y) &= yT_{G-e}(x,y) && \text{by Lemma 15.3.1}\\
&= yT_{G_1\cup(G_2-e)}(x,y)\\
&= yT_{G_1}(x,y)T_{G_2-e}(x,y) && \text{by the induction hypothesis}\\
&= T_{G_1}(x,y)\Big(yT_{G_2-e}(x,y)\Big)\\
&= T_{G_1}(x,y)T_{G_2}(x,y) && \text{by Lemma 15.3.1.}
\end{aligned}$$

Finally, suppose that e is neither a bridge nor a loop of G; then e is an edge of G_2 that is neither a bridge nor a loop of G_2. Then

$$\begin{aligned}
T_G(x,y) &= T_{G-e}(x,y)+T_{G/e}(x,y) && \text{by Lemma 15.3.1}\\
&= T_{G_1\cup(G_2-e)}(x,y)+T_{G_1\cup(G_2/e)}(x,y)\\
&= T_{G_1}(x,y)T_{G_2-e}(x,y)+ && \text{by the}\\
&\quad +T_{G_1}(x,y)T_{G_2/e}(x,y) && \text{induction}\\
&&& \text{hypothesis}\\
&= T_{G_1}(x,y)\Big(T_{G_2-e}(x,y)+T_{G/e}(x,y)\Big)\\
&= T_{G_1}(x,y)T_{G_2}(x,y) && \text{by Lemma 15.3.1}
\end{aligned}$$

This completes the argument. □

Note that Lemma 15.3.2 guarantees that the Tutte polynomial of a multigraph G is the product of the Tutte polynomials of its blocks.[8]

15.4 The relationship between the chromatic polynomial and the Tutte polynomial

As our next lemma shows, the Tutte polynomial is more general than the chromatic polynomial, i.e. if we know the Tutte polynomial of a multigraph, we can easily compute the chromatic polynomial of that multigraph.

Lemma 15.4.1. *Every multigraph G satisfies*

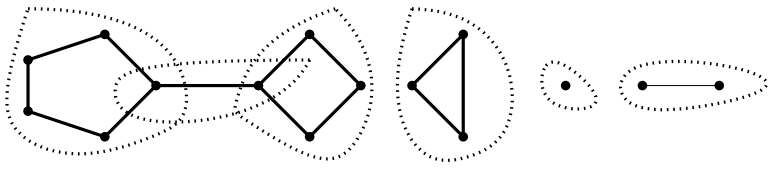
$$\pi_G(x) \quad = \quad (-1)^{|V(G)|-k(G)}\ x^{k(G)}\ T_G(1-x,0).$$

Proof. We proceed by induction on the number of edges. Fix a multigraph G, and assume inductively that the statement is true for all multigraphs on fewer than $|E(G)|$ edges.

Suppose first that G is edgeless. Then by section 15.2, we have that $\pi_G(x) = x^{|V(G)|}$. Further, by the definition of the Tutte polynomial, we have that $T_G(x,y) = 1$, and so $T_G(1-x,0) = 1$. Moreover, $k(G) = |V(G)|$, and so $(-1)^{|V(G)|-k(G)} = 1$. But now it is clear that

$$\begin{aligned}
\pi_G(x) &= x^{|V(G)|} \\
&= (-1)^{|V(G)|-k(G)} \cdot x^{|V(G)|} \cdot 1 && \text{because } (-1)^{|V(G)|-k(G)} = 1 \\
&= (-1)^{|V(G)|-k(G)}\ x^{k(G)}\ T_G(1-x,0) && \text{because } T_G(1-x,0) = 1,
\end{aligned}$$

[8] A *block* of a multigraph G is a maximal connected subgraph of G that has no cut-vertices. (However, not all such subgraphs are blocks! We need maximality.) For example, the (disconnected) graph below has six blocks, in dotted bags.



Note that a (multi)graph can be built from its blocks by repeatedly taking disjoint unions and gluing along single vertices. In the case of graphs (with no loops and no parallel edges), blocks are the maximal 2-connected subgraphs, bridges (with their endpoints), and components on at most two vertices. In the multigraph case, a loop (with its unique endpoint) is considered a block.

which is what we needed.

From now on, we assume that G has at least one edge, say e. We consider three cases: when e is a bridge, when e is a loop, and when e is neither a bridge nor a loop.

Suppose first that e is a bridge of G. Then either $G - e$ and G/e have exactly the same blocks, or $G - e$ can be obtained from G/e by adding an isolated vertex. Since $T_{K_1}(x, y) = 1$, Lemma 15.3.2 now guarantees that $T_{G-e} = T_{G/e}$. We now compute:

$$\begin{aligned}
\pi_G(x) &= \pi_{G-e}(x) - \pi_{G/e}(x) && \text{by section 15.2}\\
&= (-1)^{|V(G-e)|-k(G-e)}\, x^{k(G-e)}\, T_{G-e}(1-x,0) - && \text{by the}\\
&\quad -(-1)^{|V(G/e)|-k(G/e)}\, x^{k(G/e)}\, T_{G/e}(1-x,0) && \text{induction}\\
&&& \text{hypothesis}\\
&= (-1)^{|V(G)|-k(G)-1}\, x^{k(G)+1}\, T_{G-e}(1-x,0) - && \text{since } e \text{ is a}\\
&\quad -(-1)^{|V(G)|-k(G)-1}\, x^{k(G)}\, T_{G/e}(1-x,0) && \text{bridge of } G\\
&= (-1)^{|V(G)|-k(G)-1}\, x^{k(G)}\\
&\quad \Big(xT_{G-e}(1-x,0) - T_{G/e}(1-x,0)\Big)\\
&= (-1)^{|V(G)|-k(G)}\, x^{k(G)}\, (1-x)T_{G/e}(1-x,0) && \text{because}\\
&&& T_{G-e} = T_{G/e}\\
&= (-1)^{|V(G)|-k(G)}\, x^{k(G)}\, T_G(1-x,0) && \text{by Lemma 15.3.1,}
\end{aligned}$$

which is what we needed.

Next, suppose that e is a loop. Then by section 15.2, $\pi_G(x) = 0$. On the other hand, by Lemma 15.3.1, we have that $T_G(x, y) = yT_{G-e}(x, y)$, and consequently, $T_G(1 - x, 0) = 0$. It then immediately follows that

$$\pi_G(x) = (-1)^{|V(G)|-k(G)}\, x^{k(G)}\, T_G(1-x,0).$$

Finally, suppose that e is neither a bridge nor a loop. We then compute:

$$\begin{aligned}
\pi_G(x) &= \pi_{G-e}(x) - \pi_{G/e}(x) && \text{by section 15.2}\\
&= (-1)^{|V(G-e)|-k(G-e)}\ x^{k(G-e)}\ T_{G-e}(1-x,0) - && \text{by the}\\
&\quad -(-1)^{|V(G/e)|-k(G/e)}\ x^{k(G/e)}\ T_{G/e}(1-x,0) && \text{induction}\\
&&& \text{hypothesis}\\
&= (-1)^{|V(G)|-k(G)}\ x^{k(G)}\ T_{G-e}(1-x,0) -\\
&\quad -(-1)^{|V(G)|-k(G)-1}\ x^{k(G)}\ T_{G/e}(1-x,0)\\
&= (-1)^{|V(G)|-k(G)}\ x^{k(G)}\ T_{G-e}(1-x,0) +\\
&\quad +(-1)^{|V(G)|-k(G)}\ x^{k(G)}\ T_{G/e}(1-x,0)\\
&= (-1)^{|V(G)|-k(G)}\ x^{k(G)}\\
&\quad \Big(T_{G-e}(1-x,0) + T_{G/e}(1-x,0)\Big)\\
&= (-1)^{|V(G)|-k(G)}\ x^{k(G)}\ T_G(1-x,0) && \text{by Lemma 15.3.1}
\end{aligned}$$

which is what we needed. This completes the argument. □

15.5 Some special points of the Tutte polynomial

In this section, we give a combinatorial interpretation of the Tutte polynomial evaluated at some special points.

Proposition 15.5.1. *For all multigraphs G, $T_G(2,2) = 2^{|E(G)|}$.*

Proof. By the definition of the Tutte polynomial, we have that

$$T_G(2,2) = \sum_{A\subseteq E(G)} (2-1)^{r_G(A)}(2-1)^{c_G(A)} = \sum_{A\subseteq E(G)} 1.$$

So, $T_G(2,2)$ is equal to the number of subsets A of $E(G)$, which is precisely $2^{|E(G)|}$. □

A *spanning subgraph* of a multigraph G is a multigraph H such that $V(H) = V(G)$ and $E(H) \subseteq E(G)$. A multigraph is *acyclic* if it has no cycles; in particular, acyclic multigraphs have no loops and no parallel edges, and so an acyclic (multi)graph is simply a forest.

Proposition 15.5.2. *For all multigraphs G, $T_G(2,1)$ is the number of acyclic spanning subgraphs of G.*[9]

Proof. By the definition of the Tutte polynomial, we have that

$$T_G(2,1) = \sum_{A\subseteq E(G)} (2-1)^{r_G(A)}(1-1)^{c_G(A)} = \sum_{A\subseteq E(G)} 0^{c_G(A)}$$

Now, $0^{c_G(A)} = 1$ if $c_G(A) = 0$, and $0^{c_G(A)} = 0$ otherwise. So, $T_G(2,1)$ is equal to the number of subsets A of $E(G)$ such that $c_G(A) = 0$, i.e. $k_G(A) + |A| - |V(G)| = 0$, which is equivalent to $k_G(A) = |V(G)| - |A|$. But this last equality holds precisely when the multigraph $(V(G), A)$ is a forest. The result is now immediate. □

Proposition 15.5.3. *If G is a connected multigraph, then $T_G(1,2)$ is the number of connected spanning subgraphs of G.*

Proof. Let G be a connected multigraph. Then by the definition of the Tutte polynomial, we have that

$$T_G(1,2) = \sum_{A\subseteq E(G)} (1-1)^{r_G(A)}(2-1)^{c_G(A)} = \sum_{A\subseteq E(G)} 0^{r_G(A)}$$

Now, $0^{r_G(A)} = 1$ if $r_G(A) = 0$, and $0^{r_G(A)} = 0$ otherwise. So, $T_G(1,2)$ is equal to the number of subsets A of $E(G)$ such that $r_G(A) = 0$, i.e. $k_G(A) - k(G) = 0$. Since G is connected, we have that $k(G) = 1$, and so $T_G(1,2)$ is equal to the number of subsets A of $E(G)$ such that $k_G(A) = 1$, i.e. to the number of connected spanning subgraphs of G. □

Proposition 15.5.4. *If G is a connected multigraph, then $T_G(1,1)$ is the number of spanning trees of G.*

Proof. Let G be a connected multigraph. Then by the definition of the Tutte polynomial, we have that

$$T_G(1,1) = \sum_{A\subseteq E(G)} (1-1)^{r_G(A)}(1-1)^{c_G(A)} = \sum_{A\subseteq E(G)} 0^{r_G(A)+c_G(A)}$$

Now, $0^{r_G(A)+c_G(A)} = 1$ if $r_G(A) + c_G(A) = 0$, and $0^{r_G(A)+c_G(A)} = 0$ otherwise. So, $T_G(1,1)$ is the number of subsets A of $E(G)$ such that $r_G(A) = c_G(A) =$

[9] As a terminological matter, a spanning acyclic subgraph is not quite the same thing as a spanning forest. The term "spanning forest" is generally reserved for forests whose components are spanning trees of the components of the original (multi)graph, which is a more restricted notion. So, $T_G(2,1)$ need **not** be the number of spanning forests of G.

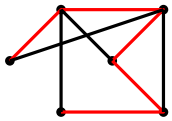
0. But $r_G(A) = 0$ if and only if the multigraph $(V(G), A)$ is connected (as in the proof of Proposition 15.5.3), and $c_G(A) = 0$ if the multigraph $(V(G), A)$ is if and only if acyclic (as in the proof of Proposition 15.5.2). So, $r_G(A) = c_G(A) = 0$ if and only if $(V(G), A)$ is a tree (equivalently: a spanning tree of G). □

Chapter 16

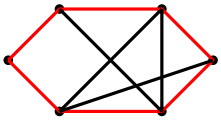
Hamiltonian graphs

16.1 Hamiltonian graphs and t-toughness

A *Hamiltonian path* (or a *Hamilton path*) of a graph G is a path of G that passes through all vertices of G. An example is shown below (the Hamiltonian path is in red.)



A *Hamiltonian cycle* (or a *Hamilton cycle*) of a graph G is a cycle of G that passes through all vertices of G. An example is shown below (the Hamiltonian cycle is in red.)



A graph is *Hamiltonian* if it has a Hamiltonian cycle.

We remark that it is NP-hard to determine whether a graph is Hamiltonian. This is in contrast with Eulerian graphs: to check if a graph is Eulerian, we need only check if it is connected and whether all its vertices are of even degree, which can obviously be done in polynomial time. Nevertheless, there are a number sufficient conditions for Hamiltonicity, which can easily be checked in polynomial time (see section 16.2 below).

For a real number $t > 0$, a graph G is *t-tough* if for every set $S \subsetneq V(G)$, the graph $G \setminus S$ has at most $\max\{1, \frac{|S|}{t}\}$ many components.[1]

Conjecture 16.1.1 (Chvátal)**.** *There exists some $t > 0$ such that every t-tough graph is Hamiltonian.*

The conjecture above remains open. We do have the following simple proposition, though.

Proposition 16.1.2. *Every Hamiltonian graph is 1-tough.*

Proof. Let G be a Hamiltonian graph, and let $S \subsetneq V(G)$. Since G is Hamiltonian, it is connected; so, if $S = \emptyset$, then $G \setminus S = G$ has only one component, and we are done. We may now assume that $S \neq \emptyset$. Let C be a Hamiltonian cycle in G. Clearly, $C \setminus S$ is the disjoint union of at most $|S|$ many paths, and so $C \setminus S$ has at most $|S|$ many components. Since C is a spanning subgraph of G,[2] it is clear that $G \setminus S$ has no more components than $C \setminus S$ does.[3] So, $G \setminus S$ has at most $|S|$ many components, and the result follows. □

16.2 Hamiltonian graphs and vertex degrees

As usual, for non-adjacent vertices x and y of a graph G, we denote by $G+xy$ the graph obtained from G by adding an edge between x and y.

Lemma 16.2.1. *Let G be a graph, and let x and y be distinct, non-adjacent vertices of G that satisfy $d_G(x) + d_G(y) \geq |V(G)|$. Then G is Hamiltonian if and only if $G + xy$ is Hamiltonian.*

Proof. It is clear that if G is Hamiltonian, then so is $G + xy$.[4]

Suppose now that $G + xy$ is Hamiltonian; we must show that G is Hamiltonian. Let C be a Hamiltonian cycle of $G + xy$. If $xy \notin E(C)$, then C is a Hamiltonian cycle of G, and we are done. So, assume that $xy \in E(C)$. Now, consider the path $C - xy = c_1, \ldots, c_n$, with $c_1 = x$ and $c_n = y$.[5] Let $S_x := \{i \mid 1 \leq i \leq n-1,\ xc_{i+1} \in E(G)\}$ and $S_y := \{i \mid 1 \leq$

[1] Equivalently, for a real number $t > 0$, a graph G is *t-tough* if for every set $S \subsetneq V(G)$, the graph $G \setminus S$ either is connected or has at most $\frac{|S|}{t}$ many components.

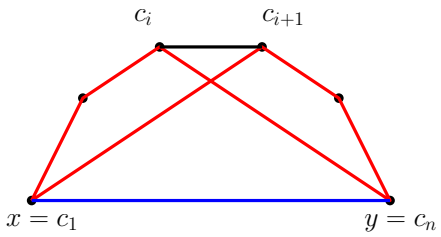
[2] A *spanning subgraph* of a graph G is a subgraph of G that contains all vertices of G.

[3] Indeed, $G \setminus S$ can be obtained from $C \setminus S$ by possibly adding edges, and adding edges cannot increase the number of components.

[4] Indeed, any Hamiltonian cycle of G is also a Hamiltonian cycle of $G + xy$.

[5] Since C is a Hamiltonian cycle of $G + xy$, we have that $V(G) = V(C) = \{c_1, \ldots, c_n\}$.

$i \leq n-1,\ yc_i \in E(G)\}$. Note that $|S_x| + |S_y| = d_G(x) + d_G(y) \geq |V(G)|$, whereas $|S_x \cup S_y| \leq |V(G)| - 1$. So, $S_x \cap S_y \neq \emptyset$. Fix $i \in S_x \cap S_y$. Since $x = c_1$ and $y = c_n$ are non-adjacent in G, we see that $2 \leq i \leq n-2$. But now $\underbrace{x}_{=c_1}, c_2, \ldots, c_i, \underbrace{y}_{=c_n}, c_{n-1}, \ldots, c_{i+1}, \underbrace{x}_{=c_1}$ is a Hamiltonian cycle of G, and so G is Hamiltonian.



□

The *Chvátal closure* of a graph G is the graph obtained by iteratively adding edges between non-adjacent vertices x, y such that $d(x) + d(y) \geq |V(G)|$, until no more such edges can be added. It is easy to see that the Chvátal closure of a graph is uniquely defined (i.e. the order in which edges are added does not matter).[6]

Theorem 16.2.2. *A graph is Hamiltonian if and only if its Chvátal closure is Hamiltonian.*

Proof. This follows from Lemma 16.2.1 by an easy induction. □

Theorem 16.2.3 (Ore)**.** *Let G be a graph on at least three vertices. Assume that for all distinct, non-adjacent vertices x, y of G, we have that $d_G(x) + d_G(y) \geq |V(G)|$. Then G is Hamiltonian.*

Proof. The Chvátal closure of G is the complete graph on $|V(G)|$ vertices, which (since $|V(G)| \geq 3$) is clearly Hamiltonian. So, by Theorem 16.2.2, G is also Hamiltonian. □

Theorem 16.2.4 (Dirac)**.** *Let G be a graph on at least three vertices. If $\delta(G) \geq \frac{|V(G)|}{2}$, then G is Hamiltonian.*

Proof. This is an immediate corollary of Theorem 16.2.3. □

[6] Details?

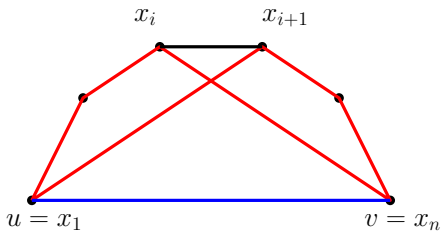
Let $\mathbf{a} = (a_1, \ldots, a_n)$ be a list (vector) of integers such that $0 \leq a_1 \leq \cdots \leq a_n \leq n-1$. A graph G on n vertices *dominates* $\mathbf{a}$ if for some ordering $v_1, \ldots, v_n$ of the vertices of G, we have that $d_G(v_1) \geq a_1, \ldots, d_G(v_n) \geq a_n$. We say that $\mathbf{a}$ is *Hamiltonian* if every n-vertex graph that dominates $\mathbf{a}$ is Hamiltonian.

Theorem 16.2.5. *Let $n \geq 3$ be an integer, and let $\mathbf{a} = (a_1, \ldots, a_n)$ be a sequence of integers such that $0 \leq a_1 \leq \cdots \leq a_n \leq n-1$. Then the following are equivalent:*

(a) for all indices $i < \frac{n}{2}$, if $a_i \leq i$, then $a_{n-i} \geq n-i$;

(b) $\mathbf{a}$ is Hamiltonian.

Proof. Suppose first that (a) holds; we must prove (b). Suppose otherwise. Then there exists a graph on n vertices that dominates $\mathbf{a}$, but is not Hamiltonian; among all such graphs, let G be one with as many edges as possible. Since G has at least three vertices and is not Hamiltonian, we see that G is not complete. Fix distinct, non-adjacent vertices $u, v \in V(G)$ such that $d_G(u)+d_G(v)$ is maximum; by symmetry, we may assume that $d_G(u) \leq d_G(v)$. Then $G + uv$ dominates $\mathbf{a}$ and has more edges than G, and so $G + uv$ is Hamiltonian. Let C be a Hamiltonian cycle in $G + uv$. Then $uv \in E(C)$, for otherwise, C would be a Hamiltonian cycle in G, contrary to the fact that G is not Hamiltonian. We now consider the path $C - uv = x_1, \ldots, x_n$, with $x_1 = u$ and $x_n = v$. Let $S := \{i \mid 1 \leq i \leq n-1,\ ux_{i+1} \in E(G)\}$; clearly, $s := |S| = d_G(u)$. If there exists some $i \in S$ such that $vx_i \in E(G)$,[7] then $\underbrace{x_1}_{=u}, x_2, \ldots, x_i, \underbrace{x_n}_{=v}, x_{n-1}, \ldots, x_{i+1}, \underbrace{x_1}_{=u}$ is a Hamiltonian cycle in G, contrary to the fact that G is not Hamiltonian.



So, no such i exists, and it follows that $d_G(v) \leq n-1-s$. But now $d_G(u) + d_G(v) \leq s + (n-1-s) = n-1$; since $d_G(u) \leq d_G(v)$, we deduce that $d_G(u) < \frac{n}{2}$, and so $s < \frac{n}{2}$. Further, by the maximality of $d_G(u) + d_G(v)$,

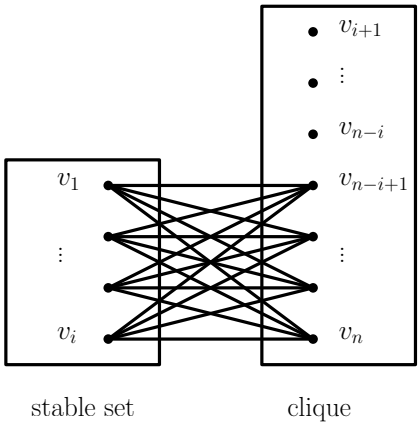
[7] Note that $2 \leq i \leq n-2$, since $uv \notin E(G)$.

we see that for all $i \in S$, we have that $d_G(x_i) \leq d_G(u) = s$.[8] So, at least s vertices of G (i.e. all the x_i's with $i \in S$) have degree at most $s < \frac{n}{2}$ in G, and it follows that $a_1, \ldots, a_s \leq s < \frac{n}{2}$.[9] But since $a_s \leq s < \frac{n}{2}$, (a) guarantees that $a_{n-s} \geq n - s$; but now $n - s \leq a_{n-s} \leq \cdots \leq a_n$, i.e. at least $s + 1$ vertices of G have degree at least $n - s$. Since $d_G(u) = s$, we see that u is non-adjacent to at least one of these $s + 1$ vertices, call it y. But now $d_G(u) + d_G(y) \geq s + (n - s) = n > n - 1 \geq d_G(u) + d_G(v)$, contrary to the maximality of $d_G(u) + d_G(v)$. So, (b) holds.

Suppose now that (a) does not hold; we must show that (b) does not hold either.[10] Since (a) does not hold, there exists some index $i < \frac{n}{2}$ such that $a_i \leq i$ and $a_{n-i} \leq n - i - 1$. Let G be the graph with vertex set $\{v_1, \ldots, v_n\}$, with adjacency as follows:

- $\{v_{i+1}, \ldots, v_n\}$ is a clique;
- $\{v_1, \ldots, v_i\}$ is complete to $\{v_{n-i+1}, \ldots, v_n\}$;[11]
- there are no other edges in G.

The graph G is represented below.



Then

- $d_G(v_1) = \cdots = d_G(v_i) = i \geq a_i \geq \cdots \geq a_1$;
- $d_G(v_{i+1}) = \cdots = d_G(v_{n-i}) = n - i - 1 \geq a_{n-i} \geq \cdots \geq a_{i+1}$;

[8] Here, we are using the fact that v is non-adjacent to all vertices x_i with $i \in S$.

[9] We are using the fact that $a_1 \leq \cdots \leq a_n$, and that G dominates $\mathbf{a}$.

[10] So, we must exhibit an n-vertex graph that dominates $\mathbf{a}$ and is not Hamiltonian.

[11] This means that all possible edges between $\{v_1, \ldots, v_i\}$ and $\{v_{n-i+1}, \ldots, v_n\}$ are present.

- $d_G(v_{n-i+1}) = \cdots = d_G(v_n) = n - 1 \geq a_n \geq \cdots \geq a_{n-i+1}$.

So, G dominates $\mathbf{a}$. On the other hand, $G \setminus \{v_{n-i+1}, \ldots, v_n\}$ has $i + 1$ components, and so G is not 1-tough; so, by Proposition 16.1.2, G is not Hamiltonian, and it follows that (b) does not hold. □

16.3 Number of Hamiltonian cycles

Lemma 16.3.1. *Let G be a graph in which all vertices are of odd degree. Then every edge of G belongs to an even number of Hamiltonian cycles.*[12] *In particular, every edge of G that belongs to a Hamiltonian cycle, belongs to at least two Hamiltonian cycles.*

Proof. Let $e = xy$ be an edge of G; we must show that e belongs to an even number of Hamiltonian cycles of G.

A *lollipop* is a connected subgraph H of G such that $V(H) = V(G)$,[13] $e \in E(H)$, and H satisfies one of the following:

(1) H is a cycle;

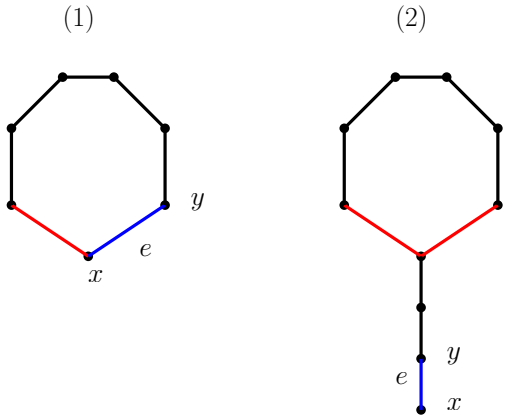
(2) $d_H(x) = 1$, H has one vertex of degree three, and all other vertices of H are of degree two.

Note that lollipops satisfying (1) are precisely the Hamiltonian cycles of G that contain the edge e. On the other hand, in case (2), H consists of a cycle, plus a path that has exactly one vertex in common with the cycle, and furthermore, x is the endpoint of this path that does not belong to the cycle. The two types of lollipop are represented below (the edge $e = xy$ is in blue).[14]

[12] It is possible that an edge of G does not belong to any Hamiltonian cycles of G, and indeed, it is possible that G is not Hamiltonian: zero is an even number.

[13] So, H is a spanning subgraph of G.

[14] In case (2), it is possible that y is in fact the unique vertex of H of degree three.

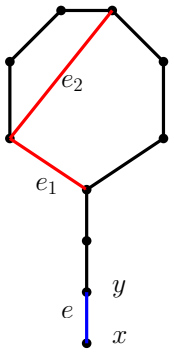


If H is a lollipop that satisfies (1), then H has a unique *tail*, namely the unique edge of H incident with x and distinct from e. On the other hand, if H is a lollipop that satisfies (2), then H has two *tails*, namely, the two edges of the unique cycle of H that are incident with the unique vertex of degree three in H. (In the picture above the tails are in red.)

We now form an auxiliary graph L, as follows. The vertices of L are the lollipops. Two lollipops, H_1 and H_2, are adjacent in L if and only if there exist tails e_1 of H_1 and e_2 of H_2 such that $H_1 - e_1 = H_2 - e_2$.[15]

Our goal is to show that the odd-degree vertices of the auxiliary graph L are precisely the lollipops satisfying (1). This is enough because the number of odd-degree vertices in L is even,[16] and the lollipops satisfying (1) are

[15]For example, in the picture below, if H_i (for $i \in \{1,2\}$) consists of the blue and black edges, plus the red edge e_i (but not the edge e_{3-i}), then lollipops H_1 and H_2 are adjacent in L.



[16]Indeed, every graph has an even number of odd-degree vertices. This follows from the fact that the sum of degrees in any graph is even (because it is equal to twice the number of edges).

precisely the Hamiltonian cycles of G that contain the edge e.

Suppose that $H = x, y, u_1, \ldots, u_t, z, x$ $(t \geq 0)$ is a lollipop satisfying (1), i.e. H is a Hamiltonian cycle of G containing e. Then xz is the unique tail of H, and the neighbors of H in L are precisely the graphs that can be obtained from $H - xz$ by adding an edge between z and a vertex in $N_G(z) \setminus N_H(z)$. So, $d_L(H) = |N_G(z) \setminus N_H(z)| = d_G(z) - 2$; since $d_G(z)$ is odd,[17] so is $d_L(H)$.

Suppose now that H is a lollipop satisfying (2); let $z, u_1, \ldots, u_t, z$ $(t \geq 2)$ be the unique cycle of H, where z is the unique vertex of degree three in H. Then the lollipop H has two tails, namely zu_1 and zu_t, and the neighbors of H in L are precisely the graphs that an be obtained in one of the following two ways:

- by starting with $H - zu_1$, and then adding an edge between u_1 and $N_G(u_1) \setminus \{z, u_2\}$;
- by starting with $H - zu_t$, and then adding an edge between u_t and $N_G(u_t) \setminus \{z, u_{t-1}\}$.

So, $d_L(H) = (d_G(u_1) - 2) + (d_G(u_t) - 2) = d_G(u_1) + d_G(u_t) - 4$. Since all vertices of G have odd degree, we deduce that $d_L(H)$ is even.

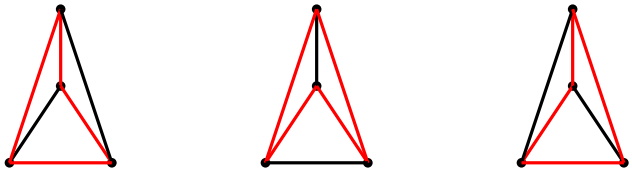
We have now shown that the odd-degree vertices of our auxiliary graph L are precisely the lollipops satisfying (1). This completes the argument. □

Theorem 16.3.2. *Let G be a Hamiltonian graph, all of whose vertices are of odd degree. Then G has at least three Hamiltonian cycles.*

Proof. Let C_1 be a Hamiltonian cycle of G, and let e be some edge of C_1. Then by Lemma 16.3.1, there exists a Hamiltonian cycle $C_2 \neq C_1$ that also contains the edge e. Since C_1, C_2 are distinct Hamiltonian cycles, we see that there exists an edge $e_1 \in E(C_1) \setminus E(C_2)$; but then Lemma 16.3.1 guarantees that there exists a Hamiltonian cycle $C_3 \neq C_1$ that contains e_1. Since $e_1 \in E(C_3) \setminus E(C_2)$, we see that $C_3 \neq C_2$. But now C_1, C_2, C_3 are pairwise distinct Hamiltonian cycles of G. □

We note that the bound from Theorem 16.3.2 is best possible: indeed, K_4 has precisely three Hamiltonian cycles (see the picture below; the Hamiltonian cycles are in red).

[17] By hypothesis, all vertices of G are of odd degree.



Chapter 17

Burnside's lemma and applications

17.1 Groups

A *group* is a set G, together with a binary operation $\circ$, satisfying the following properties:

- $\circ$ is associative, i.e. for all $g_1, g_2, g_3 \in G$, $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$;
- there exists some $e \in G$, called the *identity element*, such that for all $g \in G$, $e \circ g = g \circ e = g$;
- for all $g \in G$, there exists some $g' \in G$, called the *inverse* of g, such that $g \circ g' = g' \circ g = e$.

Usually, for $g_1, g_2 \in G$, we write "$g_1 g_2$" instead of "$g_1 \circ g_2$." It is easy to show that the identity element is unique;[1] typically, this identity element is denoted by 1_G, or simply 1. Furthermore, it can be shown that each element

[1] Indeed, suppose e_1, e_2 are identity elements of G. Then $e_1 e_2 = e_1$ (because e_2 is an identity element), and $e_1 e_2 = e_2$ (because e_1 is an identity element). So, $e_1 = e_2$.

of G has a unique inverse;[2] the unique inverse of an element $g \in G$ is usually denoted by g^{-1}.

For a set X, $\mathrm{Sym}(X)$ is the group of all permutations of X;[3] the group operation is the composition of functions, and the identity element is the identity function on X, denoted by Id_X.[4] For a positive integer n, the group of permutations of the set $\{1, \ldots, n\}$ is denoted by $\mathrm{Sym}(n)$ or Sym_n. A permutation $\pi \in \mathrm{Sym}(n)$ can be denoted by

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ \pi(1) & \pi(2) & \ldots & \pi(n) \end{pmatrix}.$$

Recall that each permutation in $\mathrm{Sym}(n)$ can be represented as a composition of disjoint cycles. For example, the following permutation in Sym(5)

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

can be represented as (143)(25). Cycles of length one are usually omitted (when n is clear from context). For example, in Sym(5), instead of (124)(3)(5), we typically write simply (124).

17.2 Group actions and Burnside's lemma

A *left action* (or simply *action*) of a group G on a set X is a function $a : G \times X \to X$ that satisfies the following two properties:

- for all $x \in X$, $a(1_G, x) = x$.
- for all $g_1, g_2 \in G$ and $x \in X$, $a(g_1, a(g_2, x)) = a(g_1 g_2, x)$.

[2] Indeed, fix $g \in G$, and suppose that $g_1, g_2 \in G$ are inverses of g. Then

$$\begin{aligned} g_1 &= g_1 1_G && \text{because } 1_G \text{ is the identity element} \\ &= g_1(g g_2) && \text{because } g_2 \text{ is an inverse of } g \\ &= (g_1 g) g_2 && \text{because } \circ \text{ is associative} \\ &= 1_G g_2 && \text{because } g_1 \text{ is an inverse of } g \\ &= g_2 && \text{because } 1_G \text{ is an identity element,} \end{aligned}$$

which is what we needed.

[3] A *permutation* of X is a bijection from X to itself.

[4] That is, $\mathrm{Id}_X : X \to X$ satisfies $\mathrm{Id}_X(x) = x$ for all $x \in X$.

Often, instead of $a(g, x)$, we write simply $g \cdot x$. So, using this notation, the axioms above become:

- for all $x \in X$, $1_G \cdot x = x$.
- for all $g_1, g_2 \in G$ and $x \in X$, $g_1 \cdot (g_2 \cdot x) = (g_1 g_2) \cdot x$.

Note that these axioms imply that, for all $g \in G$ and $x, y \in X$, if $g \cdot x = y$, then $g^{-1} \cdot y = x$.[5]

Example 17.2.1. *Any group G acts on itself in a natural way: for all $g \in G$ and $x \in G$,[6] we set $g \cdot x = gx$.*

Given an action $a : G \times X \to X$ of a group G on a set X, and an element $g \in G$, we define a function $a_g : X \to X$ by setting $a_g(x) = a(g, x)$ for all $x \in X$. As our next proposition shows, a_g is simply a permutation of X. So, we can think of group action as a collection of permutations (one permutation of the set X for each member g of the group G), which must satisfy certain additional properties (as in the definition of group action).

Proposition 17.2.2. *Let $a : G \times X \to X$ be an action of a group G on a set X. Then for all $g \in G$, the function a_g is a permutation of X.*

Proof. Fix $g \in G$, and consider its inverse g^{-1}. Then for all $x \in X$, we have that

$$\begin{aligned} a_{g^{-1}} \circ a_g(x) &= a(g^{-1}, a(g, x)) \\ &= a(g^{-1}g, x) \\ &= a(1_G, x) \\ &= x, \end{aligned}$$

and so $a_{g^{-1}} \circ a_g = \mathrm{Id}_X$. A completely analogous argument shows that $a_g \circ a_{g^{-1}} = \mathrm{Id}_X$. So, $a_g : X \to X$ is a bijection with inverse $a_{g^{-1}}$, and the result follows. □

We remark that a converse of sorts of Proposition 17.2.2 also holds: for any set X and any permutation π of X, there is a group G, an action a of G on X, and an element $g \in G$ such that $a_g = \pi$. Indeed, for fixed X and π, we set $G := \mathrm{Sym}(X)$ (the group operation is the composition of functions),

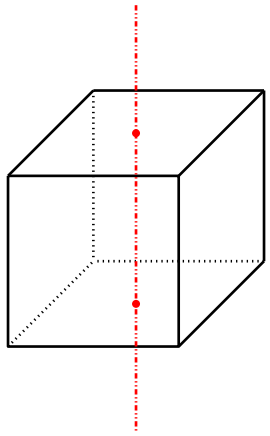
[5] Indeed, if $g \cdot x = y$, then $g^{-1} \cdot y = g^{-1} \cdot (g \cdot x) = (g^{-1}g) \cdot x = 1_G \cdot x = x$.

[6] Here, $X = G$.

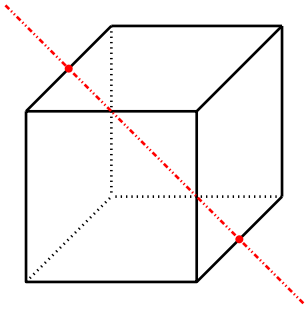
we define $a : G \times X \to X$ by $(\sigma, x) \mapsto \sigma(x)$, and we set $g := \pi$. Then for all $x \in X$, we have that $a_g(x) = a_\pi(x) = a(\pi, x) = \pi(x)$, and so $a_g = \pi$. So, the study of group actions is essentially the same as the study of set permutations.

Example 17.2.3. *Consider a cube in $\mathbb{R}^3$, and let R_{cube} be the group of rotations of $\mathbb{R}^3$ that map this cube to itself. (Here, the group operation is the composition of functions/rotations, and the identity element is the identity function on $\mathbb{R}^3$.) The group R_{cube} acts on the faces of the cube in a natural way: for each rotation $r \in R_{cube}$ and each face f of the cube, $r \cdot f$ is the face of the cube to which the rotation r maps/moves the face f. We note that $|R_{cube}| = 24$. Indeed, the rotations in R_{cube} are as follows:*

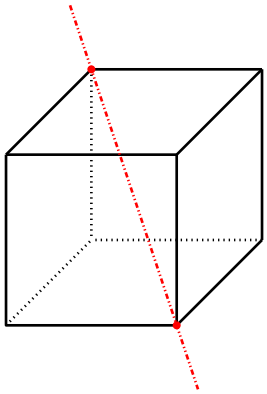
- *the identity function;*
- *nine rotations about an axis passing though centers of opposite faces of the cube (there are three choices of axis, and for each choice, we can rotate by $90°$, $180°$, or $270°$);*



- *six rotations about an axis passing through centers of opposite edges of the cube (there are six choices of axis, and for each choice, we can rotate only by $180°$);*



- *eight rotations around axes passing through opposite vertices of the cube (there are four choices of axis, and in each case, we can rotate by* $120°$ *or* $240°$*).*



We now need a few definitions. Suppose that a is an action of a group G on a set X. Then

- for each $g \in G$, a *fixed point* of g is any $x \in X$ such that $g \cdot x = x$, and we set $X^g := \{x \in X \mid g \cdot x = x\}$;[7]
- for each $x \in X$, we define the *stabilizer* of x to be $G_x := \{g \in G \mid g \cdot x = x\}$;
- for each $x \in X$, we define the *orbit* of x to be $G \cdot x := \{g \cdot x \mid g \in G\}$.[8]

Proposition 17.2.4. *Let a be an action of a group G on a set X. Then*

- *for all $x \in X$, we have that $x \in G \cdot x$;*

[7] So, X^g is the set of all fixed points of g (with respect to the action a).
[8] So, the orbit of x is the set of all elements of x that G can "move" x to.

- *the orbits of the action a form a partition of X.*

Proof. First, since $1_G \cdot x = x$, we see that $x \in G \cdot x$. In particular, each element of X belongs to some orbit. It remains to show that any two distinct orbits are disjoint. So, fix $x_1, x_2 \in X$; we must show that $G \cdot x_1$ and $G \cdot x_2$ are either equal or disjoint. Suppose that $G \cdot x_1$ and $G \cdot x_2$ are not disjoint; we claim that $G \cdot x_1 = G \cdot x_2$. We will show that $G \cdot x_1 \subseteq G \cdot x_2$; the proof of the reverse inclusion is analogous. Fix some $y \in (G \cdot x_1) \cap (G \cdot x_2)$. Then there exist $g_1, g_2 \in G$ such that $y = g_1 \cdot x_1$ and $y = g_2 \cdot x_2$; so, $g_1 \cdot x_1 = g_2 \cdot x_2$. But then $x_1 = (g_1^{-1} g_2) \cdot x_2$.[9] Now, for all $g \in G$, we have that $g \cdot x_1 = g \cdot ((g_1^{-1} g_2) \cdot x_2) = (g g_1^{-1} g_2) \cdot x_2$, and so $g \cdot x_1 \in G \cdot x_2$. Thus, $G \cdot x_1 \subseteq G \cdot x_2$, and we are done. □

Given an action a of a group G on a set X, we denote by X/G the partition of X into orbits of a. So, $|X/G|$ is the number of orbits of a.

Next, given an action a of a group G on a set X, and given $x, y \in X$, we set $M_a(x, y) := \{g \in G \mid g \cdot x = y\}$. Note that $M_a(x, x) = G_x$, and that $M_a(x, y) \neq \emptyset$ if and only if $y \in G \cdot x$.

Lemma 17.2.5. *Let a be an action of a finite group G on a finite set X, and let $x \in X$. Then for all $y \in G \cdot x$, we have that $|M_a(x, y)| = |G_x|$.*

Proof. Fix $y \in G \cdot x$, and fix any $g_y \in G$ such that $g_y \cdot x = y$. We now define a function $f : G \to G$ by setting $f(g) = g_y g$ for all $g \in G$; since G is a group, f is one-to-one. Now, our goal is to show that $f[G_x] = M_a(x, y)$; since f is one-to-one, this will imply that $|M_a(x, y)| = |f[G_x]| = |G_x|$, which is what we need.

First, fix $g \in G_x$. Then $f(g) \cdot x = (g_y g) \cdot x = g_y \cdot (g \cdot x) = g_y \cdot x = y$, and so $f(g) \in M_a(x, y)$. Thus, $f[G_x] \subseteq M_a(x, y)$.

On the other hand, fix any $g' \in M_a(x, y)$. Then $g' \cdot x = y$, and so $(g_y^{-1} g') \cdot x = g_y^{-1} \cdot (g' \cdot x) = g_y^{-1} \cdot y = g_y^{-1} \cdot (g_y \cdot x) = (g_y^{-1} g_y) \cdot x = 1_G \cdot x = x$; consequently, $g_y^{-1} g' \in G_x$. But $f(g_y^{-1} g') = g_y (g_y^{-1} g') = (g_y g_y^{-1}) g' = 1_G g' = g'$, and so $M_a(x, y) \subseteq f[G_x]$.

We have now shown that $f[G_x] = M_a(x, y)$, and we are done. □

As an easy corollary of Lemma 17.2.5, we get the following theorem.

The orbit-stabilizer theorem. *Let a be an action of a finite group G on a finite set X. Then for all $x \in X$, we have that $|G \cdot x| = \frac{|G|}{|G_x|}$.*

[9] Indeed, $x_1 = 1_G \cdot x_1 = (g_1^{-1} g_1) \cdot x_1 = g_1^{-1} \cdot (g_1 \cdot x_1) = g_1^{-1} \cdot (g_2 \cdot x_2) = (g_1^{-1} g_2) \cdot x_2$.

Proof. Fix $x \in X$, and note that sets of the form $M_a(x, y)$, with $y \in G \cdot x$, form a partition of G,[10] and so

$$\begin{aligned} |G| &= |\bigcup_{y \in G \cdot x} M_a(x, y)| \\ &= \sum_{y \in G \cdot x} |M_a(x, y)| \\ &= \sum_{y \in G \cdot x} |G_x| \qquad \text{by Lemma 17.2.5} \\ &= |G \cdot x||G_x|, \end{aligned}$$

and the result follows. □

Lemma 17.2.6. *Let a be an action of a finite group G on a finite set X. Then*

$$|X/G| = \sum_{x \in X} \tfrac{1}{|G \cdot x|}.$$

Proof. Set $t := |X/G|$, and let $O_1, \ldots, O_t$ be the orbits of the action a. Then by Proposition 17.2.4, we have that

- $(O_1, \ldots, O_t)$ is a partition of X;
- for all $i \in \{1, \ldots, t\}$ and $x \in O_i$, $G \cdot x = O_i$.[11]

We now compute

$$\sum_{x \in X} \tfrac{1}{|G \cdot x|} = \sum_{i=1}^{t} \sum_{x \in O_i} \tfrac{1}{|G \cdot x|} = \sum_{i=1}^{t} \sum_{x \in O_i} \tfrac{1}{|O_i|} = t,$$

which is what we needed. □

We are now ready to state and prove Burnside's lemma, which (roughly) states that the number of orbits of an action is equal to the average number of fixed points.

[10] That is: for all distinct $y_1, y_2 \in G \cdot x$, we have that $M_a(x, y_1) \cap M_a(x, y_2) = \emptyset$, and $\bigcup_{y \in G \cdot x} M_a(x, y) = G$.

[11] Indeed, by definition, $G \cdot x$ is equal to one of the orbits $O_1, \ldots, O_t$. Since $(O_1, \ldots, O_t)$ form a partition of X (by Proposition 17.2.4), it suffices to show that $G \cdot x$ and O_i have a non-empty intersection. But by Proposition 17.2.4 and the choice of O_i, we have that $x \in (G \cdot x) \cap O_i$, and so $(G \cdot x) \cap O_i \neq \emptyset$.

Burnside's lemma. *Let a be an action of a finite group G on a finite set X. Then*

$$|X/G| \quad = \quad \tfrac{1}{|G|} \sum_{g \in G} |X^g|.$$

Proof. Let $F := \{(g, x) \in G \times X \mid g \cdot x = x\}$. We will count $|F|$ in two ways.

On the one hand, for all $g \in G$ and $x \in X$, we have that $(g, x) \in F$ if and only if $x \in X^g$; so,

$$|F| \quad = \quad \sum_{g \in G} |X^g|.$$

On the other hand, for all $g \in G$ and $x \in X$, we have that $(g, x) \in F$ if and only if $g \in G_x$, and so

$$\begin{aligned} |F| \quad &= \quad \sum_{x \in X} |G_x| \\ &= \quad \sum_{x \in X} \tfrac{|G|}{|G \cdot x|} \qquad \text{by the orbit-stabilizer theorem} \\ &= \quad |G| \sum_{x \in X} \tfrac{1}{|G \cdot x|}. \end{aligned}$$

But now

$$|G| \sum_{x \in X} \tfrac{1}{|G \cdot x|} \quad = \quad |F| \quad = \quad \sum_{g \in G} |X^g|,$$

and consequently,

$$\sum_{x \in X} \tfrac{1}{|G \cdot x|} \quad = \quad \tfrac{1}{|G|} \sum_{g \in G} |X^g|.$$

But by Lemma 17.2.6, $|X/G| = \sum_{x \in X} \frac{1}{|G \cdot x|}$, and the result follows. □

17.3 Applications of Burnside's lemma

Example 17.3.1. *Let R_{cube} be the group of rotations of the cube, as in Example 17.2.3, and let k be a positive integer. Let B_k be the set of all colorings of the faces of the cube using the color set $\{1, \ldots, k\}$. Then R_{cube} acts on the set B_k in the natural way: a rotation $r \in R_{cube}$ maps each element of B_k to an appropriately rotated coloring. Two colorings of the cube are* equivalent *if one can be transformed into the other by a rotation in R_{cube}. Compute the number of non-equivalent colorings of the cube using the color set $\{1, \ldots, k\}$.*

Solution. Two colorings of the cube using the color set $\{1, \ldots, k\}$ are equivalent if and only if they belong to the same orbit of our group action. So, the number of non-equivalent colorings of the cube using the color set $\{1, \ldots, k\}$ is precisely equal to the number of orbits of our action of R_{cube} on B_k, which we will compute using Burnside's lemma. We know that $|R_{\text{cube}}| = 24$ (see Example 17.2.3), and for each $r \in R_{\text{cube}}$, we compute $|B_k^r|$, the number of fixed points of the rotation r, as follows.

- If r is the identity rotation, then $|B_k^r| = |B_k| = k^6$.
- If r is a rotation by 90° or 270° about an axis passing though the center of opposite faces (there are a total of six such r's), then r fixes precisely the colorings in which the faces not pierced by the axis have the same color. So, we choose one of k colors for one of the faces pieced by the axis, one of k colors for the other face pierced by the axis, and one of k colors for all the remaining four faces. In total, we get $|B_k^r| = k^3$.
- If r is a rotation by 180° about an axis passing though the center of opposite faces (there are a total of three such r's), then r fixes exactly the the colorings for which the opposite faces that are not pierced by the axis have the same color. There are two pairs of opposite faces not pierced by our axis, and it follows that $|B_k^r| = k^4$.
- If r is a rotation by 180° about an axis passing though the center of opposite edges (there are a total of six such r's), then r fixes exactly the colorings for which the two opposite faces not incident with the edges pierced by the axis have the same color, and in which, for each pierced edge, the two faces incident with this edge have the same color. So, $|B_k^r| = k^3$.
- Finally, if r is a rotation by 120° or 240° about an axis passing though opposite vertices (there are a total of eight such r's), then r fixes exactly the colorings for which the three incident faces with each of the pierced vertices have the same color. So, $|B_k^r| = k^2$.

So, by Burnside's lemma and Example 17.2.3, the total number of orbits of our action (and therefore, the total number of non-equivalent colorings) is

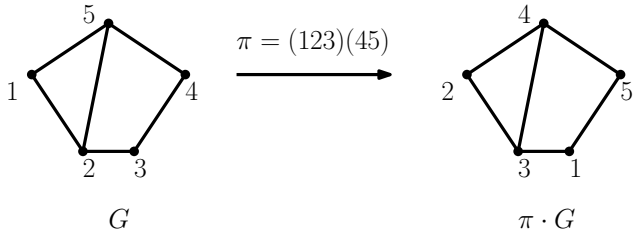
$$\frac{1}{|R_{\text{cube}}|} \sum_{r \in R_{\text{cube}}} |B_k^r| = \frac{k^6+6k^3+3k^4+6k^3+8k^2}{24} = \frac{k^6+3k^4+12k^3+8k^2}{24}.$$

□

We now need a definition. For a graph G, a vertex $x \in V(G)$, and a set $Y \subseteq V(G) \setminus \{x\}$, we say that x is *complete* (resp. *anticomplete*) to Y in G provided that x is adjacent (resp. non-adjacent) to all vertices of Y in G. For a graph G and disjoint sets $X, Y \subseteq V(G)$, we say that X is *complete* (resp. *anticomplete*) to Y in G provided that every vertex of X is complete (resp. anticomplete) to Y in G.[12]

Example 17.3.2. *Find the number of non-isomorphic graphs on five vertices.*

Solution. Let X be the set of all graphs on the vertex set $\{1, \ldots, 5\}$. We let $\mathrm{Sym}(5)$ act on X in the natural way: given a graph $G \in X$ and a permutation $\pi \in \mathrm{Sym}(5)$, we let $\pi \cdot G$ be the graph with vertex set $\{1, \ldots, 5\}$, in which distinct vertices $i, j \in \{1, \ldots, 5\}$ are adjacent if and only if $\pi^{-1}(i)$ and $\pi^{-1}(j)$ are adjacent in G.[13] An example is shown below.



Clearly, two graphs in X are isomorphic if and only if they belong to the same orbit of this action. So, the number of non-isomorphic graphs on five vertices is equal to the number of orbits of our action. We will compute the number of orbits using Burnside's lemma.

Clearly, $|\mathrm{Sym}(5)| = 5!$. We compute the number of fixed points of a permutation $\pi \in \mathrm{Sym}(5)$ according to the cycle structure of π.

- If π is the identity function, then π fixes all elements of X, i.e. $|X^\pi| = |X| = 2^{\binom{5}{2}} = 2^{10}$.
- If $\pi = (ab)$, for distinct $a, b \in \{1, \ldots, 5\}$ (note: there are $\binom{5}{2} = 10$ such π's), then π fixes precisely the graphs $G \in X$ such that $N_G(a) \setminus \{b\} = N_G(b) \setminus \{a\}$. So, we can freely select the neighbors of a (the neighbors of b are then forced), and we can choose adjacency between vertices in $\{1, \ldots, 5\} \setminus \{a, b\}$ arbitrarily. There are 2^4 ways to choose the neighbors

[12] So, if X is complete to Y, then all possible edges between X and Y are present; if X is anticomplete to Y, then there are no edges between X and Y.

[13] Equivalently: $\pi \cdot G$ has the same vertex set as G; each edge ij of G turns into an edge $\pi(i)\pi(j)$ of $\pi \cdot G$; and each non-edge ij of G turns into a non-edge $\pi(i)\pi(j)$ of $\pi \cdot G$.

of a, and there are $2^{\binom{3}{2}} = 2^3$ ways to choose adjacency between vertices in $\{1, \ldots, 5\} \setminus \{a, b\}$. So, $|X^\pi| = 2^4 \cdot 2^3 = 2^7$.

- If $\pi = (ab)(cd)$ for pairwise distinct $a, b, c, d \in \{1, \ldots, 5\}$ (note: there are 15 such π's), then π fixes precisely the graphs $G \in X$ satisfying the following three properties:
  - ac is an edge if and only if bd is an edge,
  - ad is an edge if and only if bc is an edge,
  - the fifth vertex of G (i.e. the unique vertex in $\{1, \ldots, 5\} \setminus \{a, b, c, d\}$) is adjacent to a if and only if it is adjacent to b, and is adjacent to c if and only if it is adjacent to d.

  So, $|X^\pi| = 2^6$.

- If $\pi = (abc)$, for pairwise distinct $a, b, c \in \{1, \ldots, 5\}$ (note: there are 20 such π's), then π fixes precisely the graphs $G \in X$ in which $\{a, b, c\}$ is either a clique or a stable set, and each of the remaining two vertices (i.e. vertices in $\{1, \ldots, 5\} \setminus \{a, b, c\}$) is either complete or anticomplete to $\{a, b, c\}$. So, $|X^\pi| = 2^4$.

- If $\pi = (abc)(de)$, for pairwise distinct $a, b, c, d, e \in \{1, \ldots, 5\}$ (note: there are 20 such π's), then π fixes precisely the graphs $G \in X$ in which $\{a, b, c\}$ is either a clique or a stable set, and $\{a, b, c\}$ is either complete or anticomplete to $\{d, e\}$. So, $|X^\pi| = 2^3$.

- If $\pi = (abcd)$, for pairwise distinct $a, b, c, d \in \{1, \ldots, 5\}$ (note: there are 30 such π's), then π fixes precisely the graphs $G \in X$ in which all the following hold:
  - ab, bc, cd, da are either all edges or all non-edges,
  - ac and bd are either both edges or both non-edges,
  - the fifth vertex of G (i.e. the unique vertex in $\{1, \ldots, 5\} \setminus \{a, b, c, d\}$) is either complete or anticomplete to $\{a, b, c, d\}$.

  So, $|X^\pi| = 2^3$.

- If $\pi = (abcde)$, for pairwise distinct a, b, c, d, e (note: there are 24 such π's), then π fixes precisely the graphs $G \in X$ in which both the following hold:
  - ab, bc, cd, de, ea are either all edges or all non-edges,

– ac, bd, ce, da, eb are either all edges or all non-edges.

So, $|X^\pi| = 2^2$.

Now, by Burnside's lemma, we see that the number of orbits of our action is

$$\begin{aligned} |X/\mathrm{Sym}(5)| &= \tfrac{1}{|\mathrm{Sym}(5)|} \sum_{\pi \in \mathrm{Sym}(5)} |X^\pi| \\ &= \tfrac{1}{5!}\Big(2^{10} + 10 \cdot 2^7 + 15 \cdot 2^6 + 20 \cdot 2^4 + 20 \cdot 2^3 + 30 \cdot 2^3 + 24 \cdot 2^2\Big) \\ &= 34. \end{aligned}$$

So, there are 34 non-isomorphic graphs on five vertices. □

17.4 Pólya enumeration theorem

Our goal in this section is to compute the number of different colorings of certain objects, up to symmetry. The symmetry will be determined by an appropriate group action.

A *subgroup* of a group G is a subset of G that is a group under the operation inherited from G. Note that every group is a subgroup of itself, as is the one-element group consisting only of the identity element.

Let X be a set of size n, and let G be a subgroup of $\mathrm{Sym}(X)$. Each element of G can be represented as a composition of disjoint cycles, the sum of whose lengths is n. Now, for $g \in G$ and $k \in \{1, \ldots, n\}$, we denote by $j_k(g)$ the number of cycles of length k, when g is written as a composition of disjoint cycles.[14] For $g \in G$, we set $x^{\mathrm{cs}(g)} := x_1^{j_1(g)} x_2^{j_2(g)} \ldots x_n^{j_n(g)}$. Finally, the *cycle index* of the group G is

$$\mathcal{Z}_G(x_1, \ldots, x_n) := \tfrac{1}{|G|} \sum_{g \in G} x^{\mathrm{cs}(g)}.$$

Example 17.4.1. *Compute cycle index of the group Sym*(2)*.*

Solution. Here, using the notation from the definition of a cycle index, we have that $X = \{1, 2\}$ and $n = 2$. Moreover, we have that $\mathrm{Sym}(2) = \{(1)(2), (12)\}$, and clearly,

- $x^{\mathrm{cs}\big((1)(2)\big)} = x_1^2 x_2^0 = x_1^2$;

[14] For example, if $n = 7$ and $g = (124)(35)(6)(7)$, then $j_1(g) = 2$, $j_2(g) = 1$, $j_3(g) = 1$, and $j_4(g) = j_5(g) = j_6(g) = j_7(g) = 0$. Do not forget to count cycles of length one!

- $x^{\mathrm{cs}\left((12)\right)} = x_1^0 x_2^1 = x_2$.

So,

$$\mathcal{Z}_{\mathrm{Sym}(2)}(x_1, x_2) \quad = \quad \tfrac{x_1^2+x_2}{2}.$$

□

Example 17.4.2. *Compute cycle index of the group Sym*(3).

Solution. Here, using the notation from the definition of a cycle index, we have that $X = \{1, 2, 3\}$ and $n = 3$. Sym(3) has one element that is a composition of three 1-cycles; it has three elements that are a composition of one 2-cycle and one 1-cycle; and it has two elements that consist of one 3-cycle. So,

$$\mathcal{Z}_{\mathrm{Sym}(3)}(x_1, x_2, x_3) \quad = \quad \tfrac{x_1^3+3x_1x_2+2x_3}{6}.$$

□

Recall that for a set X, $\binom{X}{2}$ is the set of all 2-element subsets of X. For each positive integer n and permutation $\pi \in \mathrm{Sym}(n)$, we define a permutation π' on the set $\binom{\{1,\ldots,n\}}{2}$ by setting $\pi'(\{i, j\}) = \{\pi(i), \pi(j)\}$, and we set $\mathrm{Sym}'(n) = \{\pi' \mid \pi \in \mathrm{Sym}(n)\}$. It is easy to check that $\mathrm{Sym}'(n)$ is a subgroup of $\mathrm{Sym}\left(\binom{\{1,\ldots,n\}}{2}\right)$. In particular, every permutation in $\mathrm{Sym}'(n)$ can be represented as a composition of disjoint cycles, the sum of whose lengths is $\binom{n}{2}$.

Example 17.4.3. *Compute the cycle index of the group Sym*$'(5)$.

Solution. We remark that $\binom{5}{2} = 10$, and so each permutation in $\mathrm{Sym}'(5)$ can be represented as a composition of disjoint cycles, the sum of whose lengths is 10.

We analyze the cycle structure of permutations in Sym(5): given the cycle structure of a permutation $\pi \in \mathrm{Sym}(5)$, we describe the cycle structure of π'. If we, in addition, keep track of the number of permutations of each type in Sym(5), we can easily find the cycle index of $\mathrm{Sym}'(5)$.

- There is one permutation π in Sym(5) (namely, the identity permutation) of the form $(a)(b)(c)(d)(e)$. For such a π, we have that π' is the composition of ten cycles of length one. So, $x^{\mathrm{cs}(\pi')} = x_1^{10}$.

- There are 10 permutations π in Sym(5) of the form $(ab)(c)(d)(e)$. For such a π, we see that π' has three cycles of the length two (these cycles are of the form $(\{a, x\}, \{b, x\})$, with $x \notin \{a, b\}$), and it has four cycles of length one. So, $x^{\mathrm{cs}(\pi')} = x_1^4 x_2^3$.

- There are 15 permutation π in Sym(5) of the form $(ab)(cd)(e)$. For such a π, we see that π' has exactly two cycles of length one (namely, $(\{a,b\})$ and $(\{c,d\})$), and the remaining cycles of π' (four of them) are of length two. So, $x^{\mathrm{cs}(\pi')} = x_1^2 x_2^4$.

- There are 20 permutations π in Sym(5) of the form $(abc)(d)(e)$. For such a π, we see that π' has one cycle of length one (namely, $(\{d,e\})$), and the remaining cycles of π' (three of them) are of length three. So, $x^{\mathrm{cs}(\pi')} = x_1 x_3^3$.

- There are 20 permutations π in Sym(5) of the form $(abc)(de)$. For such a π, we see that π' has one cycle of length one (namely, $(\{d,e\})$), one cycle of length three (namely, $(\{a,b\},\{b,c\},\{c,a\})$), and one cycle of length six (containing all the remaining elements of $\binom{\{1,\ldots,5\}}{2}$). So, $x^{\mathrm{cs}(\pi')} = x_1 x_3 x_6$.

- There are 30 permutations π in Sym(5) of the form $(abcd)(e)$. For such a π, we see that π' has two 4-cycles (namely, $(\{a,e\},\{b,e\},\{c,e\},\{d,e\})$ and $(\{a,b\},\{b,c\},\{c,d\},\{d,a\})$) and one 2-cycle (namely, $(\{a,c\},\{b,d\})$). So, $x^{\mathrm{cs}(\pi')} = x_2 x_4^2$.

- There are 24 permutations π in Sym(5) of the form $(abcde)$. For such a π, we see that π' has two 5-cycles (namely, $(\{a,b\},\{b,c\},\{c,d\},\{d,e\},\{e,a\})$ and $(\{a,c\},\{b,d\},\{c,e\},\{d,a\},\{e,b\})$). So, $x^{\mathrm{cs}(\pi')} = x_5^2$.

Since $|\mathrm{Sym}'(5)| = |\mathrm{Sym}(5)| = 5! = 120$, we now see that

$$\begin{aligned} & \mathcal{Z}_{\mathrm{Sym}'(5)}(x_1,\ldots,x_{10}) \\ = \; & \tfrac{1}{120}\Big(x_1^{10} + 10x_1^4x_2^3 + 15x_1^2x_2^4 + 20x_1x_3^3 + 20x_1x_3x_6 + 30x_2x_4^2 + 24x_5^2\Big). \end{aligned}$$

□

We now need a couple more definitions. Suppose that $C = \{c_1,\ldots,c_k\}$ is some set of colors, and G is a subgroup of $\mathrm{Sym}(X)$ acting on a finite set X in the natural way, i.e. for $\pi \in G$ and $x \in X$, we have $\pi \cdot x = \pi(x)$. Let $\mathcal{C}$ be the set of all colorings of X using the color set C (formally, $\mathcal{C}$ is simply the set of all functions from X to C). Then G acts on $\mathcal{C}$ in the natural way: for all $\pi \in G$, $c \in C$, and $x \in X$, we set $(\pi \cdot c)(x) = c(\pi^{-1} \cdot x)$;[15] the idea is that $\pi \cdot c$ should assign to x the color that c assigned to the element of X

[15]Let us check that this is really a group action. For $c \in \mathcal{C}$ and $x \in X$, we have that $(1_G \cdot c)(x) = c(1_G^{-1} \cdot x) = c(1_G \cdot x) = c(x)$, and it follows that $1_G \cdot c = c$. Further, for

that got "moved" to x via π, i.e. to the element $\pi^{-1} \cdot x$. Two colorings are *equivalent* if one can be transformed into the other via our group action, i.e. if they belong to the same orbit of our action. Now, let $\mathcal{D} \subseteq \mathcal{C}$.

- The *coloring inventory* of $\mathcal{D}$ is a polynomial in $c_1, \ldots, c_k$, which is the sum of terms of the form $c_1^{d_1} \ldots c_k^{d_k}$, and the coefficient in front of the term $c_1^{d_1} \ldots c_k^{d_k}$ is the number of colorings in $\mathcal{D}$ that, for each $i \in \{1, \ldots, k\}$, assign color c_i to precisely d_i elements of X.
- The *pattern inventory* of $\mathcal{D}$ is a polynomial in $c_1, \ldots, c_k$, which is the sum of terms of the form $c_1^{d_1} \ldots c_k^{d_k}$, and the coefficient in front of the term $c_1^{d_1} \ldots c_k^{d_k}$ is the number of **non-equivalent** colorings in $\mathcal{D}$ that, for each $i \in \{1, \ldots, k\}$, assign color c_i to precisely d_i elements of X.

Lemma 17.4.4. *Let $C = \{c_1, \ldots, c_k\}$ be a set of colors, let X be a finite set of size n, and let G be a subgroup of Sym(X), acting on X in the natural way.*[16] *Let $\mathcal{C}$ be the set of all colorings of X with colors from C, and let G act on $\mathcal{C}$ in the natural way.*[17] *Then for all $\pi \in G$, the coloring inventory of $\mathcal{C}^\pi$ (the set of fixed points of π in $\mathcal{C}$) is the polynomial $p_\pi(c_1, \ldots, c_k)$ obtained by substituting $\sum_{i=1}^k c_i^r$ for each x_r in $x^{cs(\pi)}$.*[18]

Proof. We write π as a product of disjoint cycles, and we set up a correspondence between the cycles of π and the terms in the product $x^{\mathrm{cs}}(\pi)$,[19] in such

---

$\pi_1, \pi_2 \in G$, $c \in \mathcal{C}$, and $x \in X$, we have that

$$\begin{aligned}(\pi_1 \cdot (\pi_2 \cdot c))(x) &= (\pi_2 \cdot c)(\pi_1^{-1} \cdot x)\\ &= c(\pi_2^{-1} \cdot (\pi_1^{-1} \cdot x))\\ &= c((\pi_2^{-1}\pi_1^{-1}) \cdot x)\\ &= c((\pi_1\pi_2)^{-1} \cdot x)\\ &= ((\pi_1\pi_2) \cdot c)(x);\end{aligned}$$

so, $\pi_1 \cdot (\pi_2 \cdot c) = (\pi_1\pi_2) \cdot c$. Thus, this is indeed a group action on $\mathcal{C}$.

[16] This means that for all $\pi \in \mathrm{Sym}(X)$ and $x \in X$, we have that $\pi \cdot x = \pi(x)$.

[17] That is, for all $\pi \in G$, $c \in C$, and $x \in X$, we set $(\pi \cdot c)(x) = c(\pi^{-1} \cdot x)$.

[18] For example, if $C = \{c_1, c_2\}$, $X = \{1, \ldots, 7\}$, $G = \mathrm{Sym}(7)$, and $\pi = (125)(36)(47)$, then $x^{\mathrm{cs}(\pi)} = x_2^2 x_3$; if we substitute $\sum_{i=1}^k c_i^r = c_1^r + c_2^r$ for each x_r in $x^{\mathrm{cs}(\pi)}$, then we get $p_\pi(c_1, c_2) = (c_1^2 + c_2^2)^2(c_1^3 + c_2^3) = c_1^7 + 2c_1^5c_2^2 + c_1^4c_2^3 + c_1^3c_2^4 + 2c_1^2c_2^5 + c_2^7$.

[19] Here, $x_i^{d_i}$ is understood as a term of d_i different terms (namely, d_i copies of x_i), and not as a single term.

a way that a cycle of length r corresponds to an x_r term.[20] Then a coloring $c \in \mathcal{C}$ is a fixed point of π if and only if, for each cycle of π, c assigns the same color to each element of X in the cycle. We can choose colors independently for each cycle. Now, if we substitute $\sum_{i=1}^{k} c_i^r$ for each x_r in $x^{\mathrm{cs}(\pi)}$, then each r-cycle of π has a corresponding term of the form $\sum_{i=1}^{k} c_i^r$; selecting color c_i for all elements of the r-cycle is equivalent to choosing the summand c_i^r from the corresponding term $\sum_{i=1}^{k} c_i^r$ in the product defining the polynomial $p_\pi(c_1, \ldots, c_k)$. It follows that the number of ways that we can color X in such a way that π fixes the coloring, and that there are precisely d_i elements of X colored c_i (for each $i \in \{1, \ldots, k\}$) is precisely the coefficient in front of the summand $c_1^{d_1} \ldots c_k^{d_k}$ in the polynomial $p_\pi(c_1, \ldots, c_k)$. The result now follows. □

Pólya enumeration theorem. *Let $C = \{c_1, \ldots, c_k\}$ be a set of colors, let X be a finite set of size n, and let G be a subgroup of Sym(X), acting on X in the natural way.*[21] *Let $\mathcal{C}$ be the set of all colorings of X with colors from C, and let G act on $\mathcal{C}$ in the natural way.*[22] *Then the pattern inventory of $\mathcal{C}$ is $\mathcal{Z}_G(\sum_{i=1}^{k} c_i, \sum_{i=1}^{k} c_i^2, \ldots, \sum_{i=1}^{k} c_i^n)$.*

Proof. Fix a vector $\mathbf{d} = (d_1, \ldots, d_k)$ with non-negative integer entries, and let $\mathcal{C}_\mathbf{d}$ be the set of all colorings in $\mathcal{C}$ in which, for each $i \in \{1, \ldots, n\}$, the number of elements of X receiving color c_i is precisely d_i.[23] Then $\mathcal{C}_\mathbf{d}$ is the union of some orbits of the action of G on $\mathcal{C}$, and so in fact, G acts on $\mathcal{C}_\mathbf{d}$ as well. By Burnside's lemma, we have that

$$|\mathcal{C}_\mathbf{d}/G| \;=\; \tfrac{1}{|G|} \sum_{\pi \in G} |\mathcal{C}_\mathbf{d}^\pi|.$$

and consequently,

$$|\mathcal{C}_\mathbf{d}/G|\; c_1^{d_1} \ldots c_k^{d_k} \;=\; \tfrac{1}{|G|} \sum_{\pi \in G} |\mathcal{C}_\mathbf{d}^\pi|\; c_1^{d_1} \ldots c_k^{d_k}.$$

Now we sum up over all possible choices of the vector $\mathbf{d} = (d_1, \ldots, d_k)$, and we get[24]

$$\sum_\mathbf{d} |\mathcal{C}_\mathbf{d}/G|\; c_1^{d_1} \ldots c_k^{d_k} \;=\; \sum_\mathbf{d} \tfrac{1}{|G|} \sum_{\pi \in G} |\mathcal{C}_\mathbf{d}^\pi|\; c_1^{d_1} \ldots c_k^{d_k},$$

[20] For example, if $\pi = (125)(36)(47)$, then $x^{\mathrm{cs}(\pi)} = x_2^2 x_3$, and we can set up a correspondence $(125) \mapsto x_3$, $(36) \mapsto x_2$, and $(47) \mapsto x_2$. (So, two different cycles of length two get mapped to two "different" x_2's.)

[21] This means that for all $\pi \in \mathrm{Sym}(X)$ and $x \in X$, we have that $\pi \cdot x = \pi(x)$.

[22] That is, for all $\pi \in G$, $c \in C$, and $x \in X$, we set $(\pi \cdot c)(x) = c(\pi^{-1} \cdot x)$.

[23] Note that if $d_1 + \cdots + d_k \neq n$, then $\mathcal{C}_\mathbf{d} = \emptyset$.

[24] Note that our sums are in fact finite because if $d_1 + \cdots + d_k \neq n$, then $\mathcal{C}_\mathbf{d} = \emptyset$.

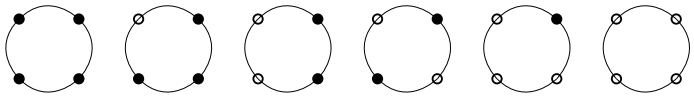
and consequently,

$$\sum_{\mathbf{d}} |\mathcal{C}_{\mathbf{d}}/G|\ c_1^{d_1}\dots c_k^{d_k} \quad = \quad \tfrac{1}{|G|}\sum_{\pi\in G}\sum_{\mathbf{d}} |\mathcal{C}^{\pi}_{\mathbf{d}}|\ c_1^{d_1}\dots c_k^{d_k}.$$

Clearly, the left-hand-side of this last equality is precisely the pattern inventory of $\mathcal{C}$. On the other hand, for each $\pi \in G$, $\sum_{\mathbf{d}} |\mathcal{C}^{\pi}_{\mathbf{d}}|\ c_1^{d_1}\dots c_k^{d_k}$ is precisely the coloring inventory of $\mathcal{C}^{\pi}$, which (by Lemma 17.4.4) is precisely $p_\pi(c_1,\dots,c_k)$, where $p_\pi(c_1,\dots,c_k)$ is the polynomial obtained by substituting $\sum_{i=1}^k c_i^r$ for each x_r in $x^{\mathrm{cs}(\pi)}$. So, the pattern inventory of $\mathcal{C}$ is $\frac{1}{|G|}\sum_{\pi\in G} p_\pi(c_1,\dots,c_k)$, which (by the definition of cycle index) is precisely $\mathcal{Z}_G(\sum_{i=1}^k c_i, \sum_{i=1}^k c_i^2, \dots, \sum_{i=1}^k c_i^n)$. □

Example 17.4.5. *Compute the number of non-equivalent colorings of a bracelet with four beads, using colors black and white for the beads. (Two colorings are equivalent if one can be transformed into the other via a rotation or a reflection.)*

Solution. In this particular case, it is easy to see that there are exactly six non-equivalent colorings, represented below.



However, let us apply the Pólya enumeration theorem in order to illustrate the principle. We label the beads $1, 2, 3, 4$ counterclockwise. The group acting on the beads is simply the dihedral group D_8 (symmetries of the square). The elements of the group are:

- (1)(2)(3)(4) - identity;
- (1234) - rotation by 90° ccw;[25]
- (13)(24) - rotation by 180°;
- (1432) - rotation by 270° ccw;
- (12)(34) - reflection about the axis through the centers of edges $12, 34$;
- (14)(23) - reflection about the axis through the centers of edges $14, 23$;
- (1)(24)(3) - reflection about the axis through vertices/beads $1, 3$;

[25] ccw = counterclockwise

- (13)(2)(4) - reflection about the axis through vertices/beads $2, 4$.

So,

$$\mathcal{Z}_{D_8}(x_1, \ldots, x_4) \;=\; \tfrac{1}{8}(x_1^4 + 2x_1^2x_2 + 3x_2^2 + 2x_4),$$

and we have that

$$\begin{aligned} & \mathcal{Z}_{D_8}(b+w, b^2+w^2, b^3+w^3, b^4+w^4) \\ = \;& \tfrac{1}{8}\Big((b+w)^4 + 2(b+w)^2(b^2+w^2) + 3(b^2+w^2)^2 + 2(b^4+w^4)\Big) \\ = \;& b^4 + b^3w + 2b^2w^2 + bw^3 + w^4. \end{aligned}$$

The total number of colorings is equal to the sum of coefficients of the polynomial above: $1 + 1 + 2 + 1 + 1 = 6$.

We also remark that the polynomial above allows us to do more, namely, to count the number of non-equivalent colorings with a fixed number of black and white beads. So, there are two non-equivalent colorings with two beads colored black and two colored white. For any other (fixed) combination of black and white beads, where the total number of beads adds up to four, we only have one non-equivalent coloring. □

Proposition 17.4.6. *Let $n \geq 2$ and $k \geq 0$ be integers. Then the number of non-isomorphic graphs on n vertices and k edges is equal to the coefficient in front of the term x^k in the polynomial $\mathcal{Z}_{Sym'(n)}(1 + x, 1 + x^2, \ldots, 1 + x^{\binom{n}{2}})$.*

Proof. Let $\mathcal{C}$ be the set of all colorings of the set $\binom{\{1,\ldots,n\}}{2}$ using the color set $\{b, w\}$. We let $\mathrm{Sym}'(n)$ act on $\mathcal{C}$ in the natural way. Now, colorings in $\mathcal{C}$ correspond to n-vertex graphs in the natural way: the vertex-set is $\{1, \ldots, n\}$, and edges are the pairs colored b ("black"), whereas the non-edges are the pairs colored w ("white"). The number of non-isomorphic n-vertex graphs with k edges is precisely the number of non-equivalent colorings in $\mathcal{C}$ (with respect to our group action) in which exactly k elements of $\binom{\{1,\ldots,n\}}{2}$ are colored b (and the remaining $\binom{n}{2} - k$ elements are colored w). By the Pólya enumeration theorem, the latter is precisely the coefficient in front of $b^k w^{\binom{n}{2}-k}$ in the polynomial $\mathcal{Z}_{\mathrm{Sym}'(5)}(b + w, b^2 + w^2, \ldots, b^{\binom{n}{2}} + w^{\binom{n}{2}})$. But this is exactly the coefficient in front of x^k in the polynomial $\mathcal{Z}_{\mathrm{Sym}'(n)}(1 + x, 1 + x^2, \ldots, 1 + x^{\binom{n}{2}})$ (we replace b by x and w by 1). □

Example 17.4.7. *For each non-negative integer k, find the number of non-isomorphic k-edge graphs on five vertices.*

Solution. We apply Proposition 17.4.6. By Example 17.4.3, we know that

$$\begin{aligned} & \mathcal{Z}_{\mathrm{Sym}'(5)}(x_1, \ldots, x_{10}) \\ = \; & \tfrac{1}{120}\Big(x_1^{10} + 10x_1^4x_2^3 + 15x_1^2x_2^4 + 20x_1x_3^3 + 20x_1x_3x_6 + 30x_2x_4^2 + 24x_5^2\Big), \end{aligned}$$

and so

$$\begin{aligned} & \mathcal{Z}_{\mathrm{Sym}'(5)}(1 + x, \ldots, 1 + x^{10}) \\ = \; & \tfrac{1}{120}\Big((1+x)^{10} + 10(1+x)^4(1+x^2)^3 + 15(1+x)^2(1+x^2)^4 + \\ & +20(1+x)(1+x^3)^3 + 20(1+x)(1+x^3)(1+x^6) + \\ & +30(1+x^2)(1+x^4)^2 + 24(1+x^5)^2\Big), \\ = \; & 1 + x + 2x^2 + 4x^3 + 6x^4 + 6x^5 + 6x^6 + 4x^7 + 2x^8 + x^9 + x^{10}. \end{aligned}$$

Thus, up to isomorphism,

- there is one edgeless graph on five vertices;
- there is one graph on five vertices with one edge;
- there are two graphs on five vertices with two edges;
- there are four graphs on five vertices with three edges;
- there are six graphs on five vertices with four edges;
- there are six graphs on five vertices with five edges;
- there are six graphs on five vertices with six edges;
- there are four graphs on five vertices with seven edges;
- there are two graphs on five vertices with eight edges;
- there is one graph on five vertices with nine edges;
- there is one graph on five vertices with ten edges;
- there are no graphs on five vertices with more than ten edges.

□

Chapter 18

Exponential generating functions

18.1 Ordinary and exponential generating functions

Let $\{a_n\}_{n=0}^{\infty}$ be a sequence of real (or complex) numbers. The *ordinary generating function* (abbreviated *ogf*) of $\{a_n\}_{n=0}^{\infty}$ is the function

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots$$

The *exponential generating function* (abbreviated *egf*) of $\{a_n\}_{n=0}^{\infty}$ is the function

$$g(x) = \sum_{n=0}^{\infty} \frac{a_n x^n}{n!} = \frac{a_0}{0!} + \frac{a_1 x}{1!} + \frac{a_2 x^2}{2!} + \frac{a_3 x^3}{3!} + \dots$$

Ordinary generating functions (or simply "generating functions") were studied in chapter 2. Here, we give a brief introduction to exponential generating functions. We begin with a simple example, in which we contrast the use of ogf's and egf's.

Example 18.1.1.

(a) Find the number of ways that three letters from the word SEQUENCE can be selected (order does not matter).[1]

[1]Note that the letter E appears three times, and so we may select between zero and three copies of E. The three E's are considered the same: so, if we select (say) two E's, we do not care which particular two we selected.

(b) Find the number of ways that three letters from the word SEQUENCE can be arranged (order matters).[2]

Solution. (a) The number of ways we can select three letters from the word SEQUENCE is the coefficient in front of x^3 in the polynomial

$$f(x) \quad = \quad (1+x+x^2+x^3)(1+x)^5,$$

which is 26. (Here, the polynomial $1+x+x^2+x^3$ corresponds to the letter E, and the five terms $1+x$ correspond to the remaining five letters of the word SEQUENCE.)

More generally, the coefficient in front of x^k in $f(x)$ is the number of ways we can select k letters from the word SEQUENCE (when order does not matter). So in fact, $f(x)$ is the ogf for the sequence $\{a_k\}_{k=0}^{\infty}$, where a_k is the number of ways of selecting k letters from the word SEQUENCE (when order does not matter).

(b) Here, we use an egf. The number of ways we can arrange three letters from the word SEQUENCE is the coefficient in front of $\frac{x^3}{3!}$ in the polynomial

$$g(x) \quad = \quad (1+x+\tfrac{x^2}{2!}+\tfrac{x^3}{3!})(1+x)^5,$$

which is 136.

Let us explain why this is correct. For each $k \in \{0,1,2,3\}$, we select k E's and $3-k$ of the remaining five letters. The number of ways of selecting those $3-k$ other letters is precisely the coefficient in front of x^{3-k} in $(1+x)^5$, and then the number of ways of arranging our three chosen letters (k E's and $3-k$ other letters) is $\frac{3!}{k!}$. So, the total number of ways of arranging three letters from the word SEQUENCE is precisely the coefficient in front of $\frac{x^3}{3!}$ in $g(x)$.

More generally, the coefficient in front of $\frac{x^k}{k!}$ in $g(x)$ is the number of ways we can arrange k letters from the word SEQUENCE (when order matters). So in fact, $g(x)$ is the egf for the sequence $\{b_k\}_{k=0}^{\infty}$, where b_k is the number of ways of arranging k letters from the word SEQUENCE (when order matters). □

Example 18.1.2. *Find the ogf and egf of the constant sequence* $1,1,1,1,\ldots$.

[2] For example, SEE and ESE count as different. However, the E's are still interchangeable: we do not care which of the three E's from the word SEQUENCE correspond to the two E's from SEE.

Solution. The ogf of the sequence is

$$f(x) \;=\; \sum_{n=0}^{\infty} x^n \;=\; \tfrac{1}{1-x},$$

whereas the egf of the sequence is

$$g(x) \;=\; \sum_{n=0}^{\infty} \tfrac{x^n}{n!} \;=\; e^x.$$

□

For some sequences, it is possible to find a closed formula for the egf, but not for the ogf. For instance, consider the sequence $\{n!\}_{n=0}^{\infty}$. The ogf of this sequence is

$$f(x) \;=\; \sum_{n=0}^{\infty} n!x^n,$$

which has radius of convergence 0,[3] i.e. the series only converges for $x = 0$. On the other hand, the egf of the sequence is

$$g(x) \;=\; \sum_{n=0}^{\infty} \tfrac{n!x^n}{n!} \;=\; \sum_{n=0}^{\infty} x^n \;=\; \tfrac{1}{1-x},$$

with the radius of convergence 1 (the series converges when $|x| < 1$).

The formulas for the basic operations with egf's are as follows. (Here, $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ are sequences, and c is a constant.)

- $\Big(\sum_{n=0}^{\infty} \frac{a_n x^n}{n!}\Big) \pm \Big(\sum_{n=0}^{\infty} \frac{b_n x^n}{n!}\Big) = \sum_{n=0}^{\infty} \frac{(a_n \pm b_n)x^n}{n!}$
- $c\Big(\sum_{n=0}^{\infty} \frac{a_n x^n}{n!}\Big) = \sum_{n=0}^{\infty} \frac{ca_n x^n}{n!}$.
- $\Big(\sum_{n=0}^{\infty} \frac{a_n x^n}{n!}\Big)\Big(\sum_{n=0}^{\infty} \frac{b_n x^n}{n!}\Big) = \sum_{n=0}^{\infty} \big(\sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}\big) \frac{x^n}{n!}$
- $\frac{d}{dx}\Big(\sum_{n=0}^{\infty} \frac{a_n x^n}{n!}\Big) = \sum_{n=0}^{\infty} \frac{a_{n+1} x^n}{n!}$

The first two formulas above are obvious. For the third, we observe that the coefficient in front of x^n is $\sum_{k=0}^{n} \frac{a_k}{k!} \frac{b_{n-k}}{(n-k)!} = \sum_{k=0}^{n} \binom{n}{k} \frac{a_k b_{n-k}}{n!}$, and the formula follows. Finally, for the fourth formula, we compute:

$$a'(x) \;=\; \sum_{n=1}^{\infty} \tfrac{n a_n x^{n-1}}{n!} \;=\; \sum_{n=1}^{\infty} \tfrac{a_n x^{n-1}}{(n-1)!} \;=\; \sum_{n=0}^{\infty} \tfrac{a_{n+1} x^n}{n!}.$$

[3] This can be shown using, for example, the Ratio Test.

Example 18.1.3. *A* derangement *of a set X is a permutation of X that has no fixed points.*[4] *For all integers $n \geq 0$, let d_n be the number of derangements of an n-element set. Find a recursive formula for the sequence $\{d_n\}_{n=0}^{\infty}$.*

Solution. Clearly, $d_0 = 1$ and $d_1 = 0$.[5] Now, fix an integer $n \geq 0$, and let X be a set of size $n+2$. Fix any $a \in X$. Then a derangement of X can map a to any element of $b \in X \setminus \{a\}$ (so, there $n+1$ choices for b). Now, suppose we have chosen b. Then our derangement of X either does or does not map b to a. If it does map b to a, then our derangement swaps a and b and then deranges $X \setminus \{a, b\}$; for fixed b, there are d_n choices for this type of derangement.[6] Suppose now that our derangement π does not map b to a. The number of such derangement is equal to the number of derangements of $X \setminus \{b\}$,[7] which is d_{n+1}. So, $d_{n+2} = (n+1)(d_n + d_{n+1})$.

We have now obtained the desired recursive formula:

- $d_0 = 1$, $d_1 = 0$;
- $d_{n+2} = (n+1)(d_n + d_{n+1})$ for all integers $n \geq 0$.

□

In our next example, we use egf's to find a non-recursive formula for d_n (from Example 18.1.3).

Example 18.1.4. *Let the sequence $\{d_n\}_{n=0}^{\infty}$ be defined recursively as follows:*

- $d_0 = 1$, $d_1 = 0$*;*
- $d_{n+2} = (n+1)(d_n + d_{n+1})$ *for all integers $n \geq 0$.*

Find a closed formula for the egf of the sequences $\{d_n\}_{n=0}^{\infty}$, and then find a non-recursive formula for d_n.

[4] In other words, a *derangement* of X is a permutation π of X such that for all $x \in X$, $\pi(x) \neq x$.

[5] Indeed, the empty function is a permutation of the empty set, and it has no fixed points (so, it is a derangement). On the other hand, any one-element set admits only one permutation (namely, the identity), and this permutation has one fixed point (and so it is not a derangement).

[6] We are using the fact that $|X \setminus \{a, b\}| = n$.

[7] Indeed any derangement π of X such that $\pi(a) = b$ and $\pi(b) \neq a$ corresponds to a derangement of $X \setminus \{b\}$ that maps a to $\pi(b)$.

Solution. Let $d(x) = \sum_{n=0}^{\infty} \frac{d_n x^n}{n!}$ be the egf of the sequence $\{d_n\}_{n=0}^{\infty}$. We first differentiate $d(x)$, and then we apply the recursive formula, as follows.

$$\begin{aligned}
d'(x) &= \sum_{n=0}^{\infty} \tfrac{d_{n+1}x^n}{n!} \\
&= \sum_{n=1}^{\infty} \tfrac{d_{n+1}x^n}{n!} && \text{because } d_1 = 0 \\
&= \Big(\sum_{n=1}^{\infty} \tfrac{nd_{n-1}x^n}{n!}\Big) + \Big(\sum_{n=1}^{\infty} \tfrac{nd_n x^n}{n!}\Big) && \text{by the recursive formula} \\
&= x\Big(\sum_{n=0}^{\infty} \tfrac{d_n x^n}{n!}\Big) + x\Big(\sum_{n=0}^{\infty} \tfrac{d_{n+1}x^n}{n!}\Big) \\
&= xd(x) + xd'(x).
\end{aligned}$$

So, we have obtained a differential equation:

$$d'(x) = xd(x) + xd'(x).$$

The differential equation above is equivalent to $\frac{d'(x)}{d(x)} = \frac{x}{1-x}$, i.e.

$$\tfrac{d'(x)}{d(x)} = \tfrac{1}{1-x} - 1.$$

By integrating both sides, we get

$$\ln(d(x)) = -\ln(1-x) - x + C,$$

and since $d(0) = d_0 = 1$, we have that $C = 0$. So, $\ln(d(x)) = -\ln(1-x) - x$. By exponentiating both sides, we get

$$d(x) = \tfrac{e^{-x}}{1-x}.$$

We have now obtained a closed formula for the exponential generating function $d(x)$. To obtain a formula for d_n, we note that $e^{-x} = \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{n!}$ and $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = \sum_{n=0}^{\infty} \frac{n!x^n}{n!}$. By the formula for the product of egf's, we now have that, for all integers $n \geq 0$,

$$d_n = \sum_{k=0}^{n} \tbinom{n}{k}(-1)^k (n-k)!,$$

and we are done. □

Chapter 19

Extremal combinatorics

19.1 Turán's theorem

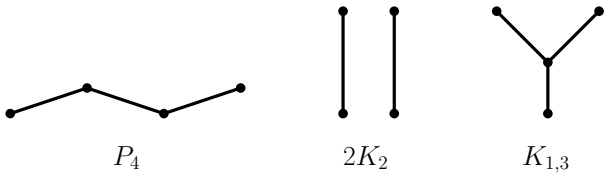
Given graphs H and G, we say that G *contains* H *as a subgraph* provided that some subgraph of G is isomorphic to H. If G does not contain H as a subgraph, we also say that G is *without an* H *subgraph*. G is *edge-maximal* without an H subgraph provided that G does not contain H as a subgraph, but any graph obtained from G by adding at least one edge to G does contain H as a subgraph.

Given a positive integer n and a graph H, an n-vertex graph G without an H subgraph is *extremal* (for the property of not containing H as a subgraph) if it has the largest possible number of edges among all n-vertex graphs without an H subgraph; $\mathrm{ex}(n, H)$ is the number of edges of an extremal n-vertex graph without an H subgraph. In other words, $\mathrm{ex}(n, H)$ is the maximum number of edges that an n-vertex graph that does not contain H as a subgraph can have. Note that if H has at least one edge, then $\mathrm{ex}(n, H)$ is defined;[1] however, if H is edgeless, then $\mathrm{ex}(n, H)$ is undefined whenever $n \geq |V(H)|$.

Obviously, any extremal graph G without an H subgraph is edge-maximal without an H subgraph. The converse, however, does not hold in general: a graph may be edge-maximal without an H subgraph, without being extremal. For example (see the picture below), $2K_2$ is a four-vertex edge-maximal graph without a P_4 subgraph, but it is not extremal: indeed, $K_{1,3}$ also has four vertices and no P_4 subgraph, and it has more edges than $2K_2$.[2]

[1] This is because there exists at least one n-vertex graph (namely, the graph $\overline{K_n}$, the edgeless graph on n vertices) that does not contain H as a subgraph.

[2] Note that we cannot obtain $K_{1,3}$ by adding edges to $2K_2$.

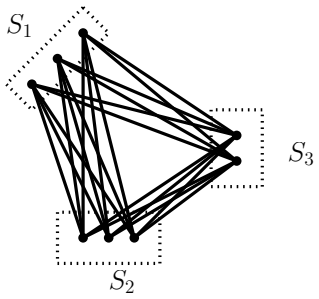


The following was proven in chapter 6.

Mantel's theorem. *For any positive integer n, we have that $ex(n, K_3) = \lfloor \frac{n^2}{4} \rfloor$, and moreover, $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ is an extremal n-vertex graph without a K_3 subgraph.*

Note that the statement of Mantel's theorem above is slightly different from the statement given in chapter 6. However, the two statements are obviously equivalent. Mantel's theorem is a special case of Turán's theorem, to which we now turn.

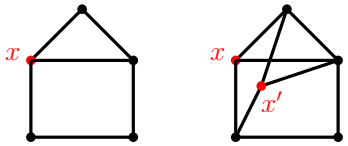
For a positive integer r, a *complete r-partite graph* is a graph G whose vertex set can be partitioned into r (possibly empty) stable sets (called *parts*), pairwise complete to each other (i.e. all possible edges between different parts are present). For example, the graph below is complete 3-partite, with parts S_1, S_2, S_3.



A *complete multipartite graph* is any graph that is complete r-partite for some r.

The *r-partite Turán graph on n vertices*, denoted by $T_r(n)$, is the complete r-partite graph on n vertices, in which the sizes of any two parts differ by at most one (so, each part is of size $\lfloor \frac{n}{r} \rfloor$ or $\lceil \frac{n}{r} \rceil$); $t_r(n)$ is the number of edges of $T_r(n)$. We note that the complete 3-partite graph above is in fact the graph $T_3(8)$.

Recall that *duplicating a vertex x* of a graph G produces a supergraph $G \circ x$ by adding to G a vertex x' and making it adjacent to all the neighbors of x in G, and to no other vertices of G (in particular, x and x' are non-adjacent in $G \circ x$). An example is shown below.



G $G \circ x$

Obviously, $\omega(G \circ x) = \omega(G)$, and so G contains K_{r+1} is a subgraph if and only if $G \circ x$ does.

Turán's theorem. *Let n and r be positive integers. Then $ex(n, K_{r+1}) = t_r(n)$, and furthermore, $T_r(n)$ is the unique (up to isomorphism) extremal n-vertex graph without a K_{r+1} subgraph.*

Proof. We may assume that $r < n$, for otherwise, $T_r(n) \cong K_n$, and the result is immediate. It is clear that $T_r(n)$ is an n-vertex graph without a K_{r+1} subgraph. Now, let G be any n-vertex extremal graph without a K_{r+1} subgraph. We must show that $G \cong T_r(n)$.

Claim. G is a complete multipartite graph.

Proof of the Claim. Suppose otherwise. Then there exist pairwise distinct vertices $y_1, x, y_2 \in V(G)$ such that $y_1x, xy_2 \notin E(G)$, but $y_1y_2 \in E(G)$.[3] If $d_G(y_1) > d_G(x)$, then $G_1 := (G \setminus x) \circ y_1$ is an n-vertex graph that does not contain K_{r+1} as a subgraph, and $|E(G_1)| > |E(G)|$, contrary to the fact that G is extremal.

So, $d_G(y_1) \leq d_G(x)$, and similarly, $d_G(y_2) \leq d_G(x)$. Now, let G' be the graph obtained from $G \setminus \{y_1, y_2\}$ by duplicating x twice. Then G' is an n-vertex graph without a K_{r+1} subgraph, and (since $y_1y_2 \in E(G)$) we have that $|E(G')| = |E(G)| - (d_G(y_1) + d_G(y_2)) + 1 + 2d_G(x) \geq |E(G)| + 1$, contrary to the fact that G is extremal. This proves the Claim. ♦

Now, using the Claim, we fix a partition $(S_1, \ldots, S_k)$ of $V(G)$ into non-empty stable sets ("parts"), pairwise complete to each other. Clearly, G contains K_k is a subgraph,[4] and so $k \leq r$. Suppose that $k < r$. Then since $r < n$, at least one of the sets $S_1, \ldots, S_k$ has more than one vertex; by

[3] Let us justify this. Since G is not complete multipartite, $\overline{G}$ is not the disjoint union of complete graphs. Consequently, some component C of $\overline{G}$ is not a complete graph. Since C is not complete, we see that C contains some two distinct, non-adjacent vertices, call them a and b. Since C is connected, there is an induced path $p_1, \ldots, p_t$ in C, with $p_1 = a$ and $p_t = b$; since a and b are non-adjacent in C, we see that $t \geq 3$. We now set $y_1 := p_1$, $x := p_2$, and $y_2 := p_3$, and we observe that $y_1x, y_2x \notin E(G)$ and $y_1y_2 \in E(G)$.

[4] Indeed, we just take one vertex from each S_i, and we obtain a clique of size k.

symmetry, we may assume that $|S_k| \geq 2$. Fix $a \in S_k$. Then consider the graph G' obtained from G by adding edges between a and all vertices of $S_k \setminus \{a\}$; then G' is a complete $(k+1)$-partite graph, it does not contain K_{r+1} as a subgraph (because $k < r$), and it has more edges than G, contrary to the fact that G is extremal. So, $k = r$.

It remains to show that any two of $S_1, \ldots, S_r$ differ in size by at most one (this will imply that $G \cong T_r(n)$). Suppose otherwise. By symmetry, we may assume that $|S_1| \geq |S_2| + 2$. Fix a vetex $a \in S_1$, and let G' be the graph obtained by first deleting all edges between a and S_2, and then adding all edges between a and $S_1 \setminus \{a\}$. (This effectively "moves" a into S_2.) Now G is still a complete r-partite graph on n vertices, and it does not contain K_{r+1} as a subgraph. Furthermore, since $|S_1| \geq |S_2| + 2$, we see that $|E(G')| \geq |E(G)| + 1$. But this contradicts the fact that G is extremal. □

19.2 The Erdős-Ko-Rado theorem

Suppose we are given positive integers r and n, and we want to select a maximum number of pairwise intersecting r-element subsets of $\{1, \ldots, n\}$. What is this maximum number? For $r > \frac{n}{2}$, any two r-element subsets of $\{1, \ldots, n\}$ intersect, and there are $\binom{n}{r}$ many such subsets. How about if $r \leq \frac{n}{2}$? In that case, we can fix any $x \in \{1, \ldots, n\}$, and consider all r-element subsets of $\{1, \ldots, n\}$ that contain x; there are $\binom{n-1}{r-1}$ many such subsets, and obviously, they pairwise intersect.[5] As the following theorem shows, this is in fact best possible (the proof that we give is due to Katona).

The Erdős-Ko-Rado theorem. *Let r and n be positive integers such that $r \leq \frac{n}{2}$. Then there are at most $\binom{n-1}{r-1}$ many pairwise distinct and pairwise intersecting r-element subsets of $\{1, \ldots, n\}$.*

Proof. Let $A_1, \ldots, A_m$ be pairwise distinct and pairwise intersecting r-element subsets of $\{1, \ldots, n\}$. We must show that $m \leq \binom{n-1}{r-1}$.

Let c be the number of ordered pairs (C, A), where

- C is a directed cycle with vertex set $\{1, \ldots, n\}$;[6]
- A is an r-vertex directed subpath of C;

[5] For the case when $r = \frac{n}{2}$, here is another construction. Fix any $x \in \{1, \ldots, n\}$, and consider all r-element subsets of $\{1, \ldots, n\} \setminus \{x\}$. Since $r = \frac{n}{2}$, all these subsets pairwise intersect, and there are $\binom{n-1}{r} = \binom{n-1}{(n-1)-r} = \binom{n-1}{(n-1)-\frac{n}{2}} = \binom{n-1}{\frac{n}{2}-1} = \binom{n-1}{r-1}$ many of them.

[6] Vertics $1, \ldots, n$ need **not** appear in that order on the cycle.

- $V(A) = A_i$ for some $i \in \{1, \ldots, m\}$.

Now we count in two ways, as follows. On the one hand, we can form an ordered pair (C, A) by first selecting one of the sets $A_1, \ldots, A_m$ (we have m choices), then ordering its vertices to form a directed path (there are $r!$ choices), and then ordering the remaining $n - r$ vertices to complete the cycle C (there are $(n - r)!$ choices). So,

$$c \quad = \quad mr!(n - r)!.$$

We now count in another way. First, there are $(n - 1)!$ ways of ordering $\{1, \ldots, n\}$ to obtain a directed cycle C. Next, we claim that for fixed C, there are at most r directed subpaths of C that correspond to one of the A_i's. Indeed, suppose a subpath $a_1, a_2, \ldots, a_r$ of C corresponds to one of $A_1, \ldots, A_m$. Then for any other subpath of C corresponding to one of $A_1, \ldots, A_m$ (and therefore containing at least one of $a_1, \ldots, a_r$), there exists some $i \in \{1, \ldots, r - 1\}$ such that either a_i is the terminal vertex of the path, or a_{i+1} is the initial vertex of the path; but since $r \leq \frac{n}{2}$, the r-vertex subpath terminating at a_i and the r-vertex subpath starting at a_{i+1} have no vertices in common, and so at most one of them can correspond to one of $A_1, \ldots, A_m$. Thus, in addition to $a_1, \ldots, a_r$, there are at most $r - 1$ subpaths of C corresponding to one of $A_1, \ldots, A_m$; in total, at most r subpaths of C correspond to one of $A_1, \ldots, A_m$. This proves that

$$c \quad \leq \quad (n - 1)!r.$$

We now have that

$$mr!(n - r)! \quad = \quad c \quad \leq (n - 1)!r,$$

and so

$$m \quad \leq \quad \tfrac{(n-1)!r}{r!(n-r)!} \quad = \quad \tbinom{n-1}{r-1},$$

which is what we needed to show. □

19.3 The Sunflower lemma

A *sunflower* is a family (i.e. collection) $\mathscr{S}$ of sets (called *petals*) such that there exists a set S (called a *kernel*) with the property that for all distinct $S_1, S_2 \in \mathscr{S}$, we have that $S_1 \cap S_2 = S$.[7]

[7] It is possible that $S = \emptyset$.

The Sunflower lemma (Erdős-Rado)**.** *Let ℓ and p be positive integers, and let $\mathscr{A}$ be a family of sets such that*

- $|A| \leq \ell$ *for all* $A \in \mathscr{A}$*, and*
- $|\mathscr{A}| > (p-1)^\ell \ell!$.

Then there exists a sunflower $\mathscr{S} \subseteq \mathscr{A}$ with p petals.

Proof. We keep p fixed, and we assume inductively that the lemma is true for smaller values of ℓ. More precisely, we assume that for all positive integers $\ell' < \ell$, and all families $\mathscr{A}'$ of sets such that

- $|A| \leq \ell'$ for all $A \in \mathscr{A}'$, and
- $|\mathscr{A}'| > (p-1)^{\ell'} \ell'!$,

there exists a sunflower $\mathscr{S}' \subseteq \mathscr{A}'$ with p petals. Clearly, we may assume that $\mathscr{A}$ is finite (otherwise, instead of $\mathscr{A}$, we consider any finite subset of $\mathscr{A}$ with more than $(p-1)^\ell \ell!$ elements).

Note that $|\mathscr{A}| \geq p$; so, if $p \leq 2$, then any p elements of $\mathscr{A}$ form a sunflower with p petals, and we are done. So, we may assume that $p \geq 3$. Next, suppose that $\ell = 1$. Then $|A| \leq 1$ for all $A \in \mathscr{A}$ and $|\mathscr{A}| > p-1$. We then take any p elements of $\mathscr{A}$, and we observe that they form a sunflower (with an empty kernel). So, from now on, we assume that $\ell \geq 2$.

Let $\mathscr{D} \subseteq \mathscr{A}$ be a collection of pairwise disjoint sets, chosen so that $|\mathscr{D}|$ is as large as possible. If $|\mathscr{D}| \geq p$, then any p elements of $\mathscr{D}$ form a sunflower (with an empty kernel), and we are done. So assume that $|\mathscr{D}| < p$. Let $D := \bigcup \mathscr{D}$;[8] then $|D| \leq |\mathscr{D}|\ell \leq (p-1)\ell$. Furthermore, since $|\mathscr{A}| \geq \ell \geq 2$, the maximality of $\mathscr{D}$ guarantees that $\mathscr{D}$ contains at least one non-empty set,[9] and so $D \neq \emptyset$.

> **Claim.** There exists some $d \in D$ such that d belongs to more than $(p-1)^{\ell-1}(\ell-1)!$ many elements of $\mathscr{A}$.

Proof of the Claim. We consider two cases: when $\emptyset \notin \mathscr{A}$, and when $\emptyset \in \mathscr{A}$.

Suppose first that $\emptyset \notin \mathscr{A}$ (and consequently, $\emptyset \notin \mathscr{D}$). Then every element of $\mathscr{A}$ intersects D: indeed, since $\emptyset \notin \mathscr{D}$, we know that every element of $\mathscr{D}$

[8] This means that $D = \bigcup_{X \in \mathscr{D}} X$, i.e. D is the union of elements of $\mathscr{D}$.

[9] Let us justify this. Suppose that $\mathscr{D}$ contains no non-empty sets. Then either $\mathscr{D} = \emptyset$ or $\mathscr{D} = \{\emptyset\}$. Since $|\mathscr{A}| \geq 2$, we see that $\mathscr{A}$ contains at least one non-empty set, say A. But then $\mathscr{D} \cup \{A\}$ is a set of pairwise disjoint elements of $\mathscr{A}$, contrary to the maximality of $\mathscr{D}$.

intersects D, and by the maximality of $\mathscr{D}$, every element of $\mathscr{A} \setminus \mathscr{D}$ intersects D. But then by the Pigeonhole Principle, some element of D belongs to at least

$$\left\lceil \tfrac{|\mathscr{A}|}{|D|} \right\rceil \;>\; \tfrac{(p-1)^\ell \ell!}{(p-1)\ell} \;=\; (p-1)^{\ell-1}(\ell-1)!$$

many elements of $\mathscr{A}$, which is what we needed.

Suppose now that $\emptyset \in \mathscr{A}$. Then the maximality of $\mathscr{D}$ guarantees that $\emptyset \in D$. Since $D \neq \emptyset$, it follows that $|\mathscr{D}| \geq 2$. Since $\emptyset \in \mathscr{D}$, we see that $|D| \leq (|\mathscr{D}|-1)\ell \leq (p-2)\ell$. Now by the maximality of $\mathscr{D}$, every element of $\mathscr{A} \setminus \{\emptyset\}$ intersects D. But then by the Pigeonhole Principle, some element of D belongs to at least

$$\begin{aligned} \left\lceil \tfrac{|\mathscr{A}\setminus\{\emptyset\}|}{|D|} \right\rceil &\geq \tfrac{(p-1)^\ell \ell!}{(p-2)\ell} \\ &= (p-1)^{\ell-1}(\ell-1)!\tfrac{p-1}{p-2} \\ &> (p-1)^{\ell-1}(\ell-1)! \end{aligned}$$

many elements of $\mathscr{A}$, which is what we needed. This proves the Claim. ♦

Let $d \in D$ be as in the Claim, and set $\mathscr{A}' := \{A \setminus \{d\} \mid A \in \mathscr{A}, d \in A\}$. Then $|\mathscr{A}'| > (p-1)^{\ell-1}(\ell-1)!$; furthermore, $|A| \leq \ell - 1$ for all $A \in \mathscr{A}'$. So, by the induction hypothesis, there exists a sunflower $\mathscr{S}' \subseteq \mathscr{A}'$ with p petals. Now, set $\mathscr{S} := \{A \cup \{d\} \mid A \in \mathscr{S}'\}$; then $\mathscr{S} \subseteq \mathscr{A}$ is a sunflower with p petals, and we are done. □

Bibliography

[1] M. Balko, *Kombinatorika a grafy I* (course web page, 2019), Charles University, `https://kam.mff.cuni.cz/~balko/kgI1819/KGI.html` (accessed September 22, 2022).

[2] *Blossom algorithm* (In *Wikipedia*), `https://en.wikipedia.org/wiki/Blossom_algorithm` (accessed September 22, 2022).

[3] B. Bollobás, *Modern Graph Theory*, Springer, New York, 1998.

[4] A. Camina and B. Lewis, *An Introduction to Enumeration*, Springer, London, 2011.

[5] R. Diestel, *Graph Theory* (3rd ed.), Springer, Berlin, Heidelberg, 2005.

[6] Z. Dvořák, *Combinatorics and graph theory II* (course web page, 2020), Charles University, `https://iuuk.mff.cuni.cz/~rakdver/index.php?which=uceni&subject=kg2` (accessed September 22, 2022).

[7] *Erdős–Ko–Rado theorem* (In *Wikipedia*), `https://en.wikipedia.org/wiki/Erdos-Ko-Rado_theorem` (accessed September 22, 2022).

[8] J. Gross and J. Yellen, *Graph Theory and Its Applications* (2nd ed.), Chapman and Hall/CRC, Boca Raton, 2006.

[9] V. Jelínek, *Combinatorics and Graph Theory II* (course web page, 2018), Charles University, `https://iuuk.mff.cuni.cz/~jelinek/1819/cag2.html` (accessed September 22, 2022).

[10] T. Kaiser, *Samoopravné kódy* (lecture notes), `http://home.zcu.cz/~kaisert/kody/kody.pdf`

[11] G.O.H. Katona, *A simple proof of the Erdös-Chao Ko-Rado theorem*, Journal of Combinatorial Theory, Series B, 13(2):183–184, 1972.

[12] M. Mareš, *Ramseyovy věty*, `http://mj.ucw.cz/papers/ramsey.pdf` (accessed September 22, 2022).

[13] J. Matoušek and J. Nešetřil, *Kapitoly z diskrétní matematiky*, Karolinum, Prague, 2009.

[14] J. Matoušek and T. Valla, *Kombinatorika a grafy I* (lecture notes, 2005), `https://iuuk.mff.cuni.cz/~valla/kg.pdf` (accessed September 22, 2022).

[15] B.M. Scott, *Sunflower combinatorics* (In *Stack Exchange*), `https://math.stackexchange.com/q/248874` (accessed September 22, 2022).

[16] R. Tarjan, *Sketchy Notes on Edmonds' Incredible Shrinking Blossom Algorithm for General Matching*, `https://www.cs.dartmouth.edu/~ac/Teach/CS105-Winter05/Handouts/tarjan-blossom.pdf` (accessed September 22, 2022).

[17] K. Vušković. *Graph Theory: Structure and Algorithms* (unpublished lecture notes, 2015/16), University of Leeds.

[18] D.B. West, *Introduction to Graph Theory* (2nd ed.), Pearson, 2001.