Combinatorial Structures

Freely using the textbook by Lovász-Pelikán-Vesztergombi

Péter Gács

Computer Science Department
Boston University

Fall 2009

Introduction

For details on the course structure (syllabus, policies, lecture schedule, homework), see the course homepage
`www.cs.bu.edu/~gacs/courses/cs131` .

The course introduces some general techniques of mathematical reasoning used in computer science. You will get most benefit from it as a freshman, but I hope it is not completely useless for those taking it later just to satisfy the requirement.

The material is sets, functions, relations, counting, graphs. Much emphasis will be on methods of sound reasoning and proof. (We will practice rigorous reasoning, but not learn any rigid formats for doing proofs!)

On our books

We use:

- LPV (Lovász-Pelikán-Vesztergombi)

Homework

- Start early, so that you have time to ask questions. Do not skimp on time: many of the problems will be deliberately such that they cannot be solved in a snap. In my experience, this is necessary for real learning.
- Work neatly. First, your grader is not obliged to spend extra time trying to decipher what you were trying to do. Second, sorting out things on paper helps sorting out your own ideas. Do not skimp on paper: start new line, new paragraph, new sheet of paper frequently.

Grading

- Homework: The purpose of homework grade is to give you some incentive to work and to provide feedback. But the percentage contribution of homework to your final grade is low, so you do not gain much by plagiarizing the work of others. Also it is not worth wasting your and my time coming to office hours and haggling on homework partial credit: come only if you think there is real misunderstanding or mistake by the grader.
- Exams: I do not give partial credit easily, and give it only if I see some real understanding. Even a lot of writing will not get credit if the reasoning is wrong. Be careful about how you argue over a grade. I am frequently amused over students who do it even before they tried to understand what they did wrong.

Counting

Some examples

Names: Alice, Bob, Carl, Diane, Eve, Frank, George.

- How many handshakes among these 7 people?
  - $6+5+\cdots+2+1=\frac{6\cdot7}{2}$ (arithmetic series).
  - $\frac{7\cdot6}{2}$ since everybody shakes with everybody else, and here we counted each shake twice, from both sides.

  The two solutions, for the general case (n people) provide a new proof for the sum formula of the arithmetic series.
- How many ways to seat around the table, with Alice's place fixed? $6\cdot5\cdots2\cdot1=6!$ (everybody seems to be familiar with the factorial notation).
- How many boy-girl pairs can be formed for dancing (4 boys, 3 girls)?

- How many ways to fill out a lottery ticket (90 numbers, 5 must be crossed out)?
  Interesting side result: $90 \cdot 89 \cdot 88 \cdot 87 \cdot 86$ is divisible by $5 \cdot 4 \cdot 3 \cdot 2$. (The earlier side result, that $7 \cdot 6$ is divisible by 2, is less interesting.)
- Bridge: how likely is it that I will get the same hand next time? (A hand is 13 cards, out of the possible 52.)

Matchings

Matching up 6 people at 3 boards to play chess. How many ways?
Discussing the interpretation of the question:

- Do we distinguish the 3 boards (say, by how close they are to the refreshments)?
- Do we consider which player has whites?

Assume that none of those distinctions are made.
In general, before you can solve a practical problem by applying mathematics to it, you must clarify carefully the assumptions, and decide which aspects of the situation you can abstract away from.

Ways to count:

- There are $\frac{6\cdot5}{2}$ choices for the first board, $\frac{4\cdot3}{2}$ for the second board (and just one for the third board). But the order of the boards does not matter, we must divide by 3!:

$$\frac{6\cdot5\cdot4\cdot3}{2^2\cdot3!}.$$

- There are 6! ways to sit on the chairs, divided by: 3! ways to reshuffle the tables, and 2^3 ways to reshuffle the people within the pairs:

$$\frac{6!}{3!\cdot2^3}=\frac{6\cdot5\cdot4}{2^3}.$$

- The youngest chooses first, then the youngest among the remaining people: $5\cdot3$.

Equality obtained: $\frac{6 \cdot 5 \cdot 4}{2^3} = 5 \cdot 3$. More generally $\frac{2n(2n-1)\cdots(n+1)}{2^n} = (2n-1)(2n-3)\cdots 3$. Can you show this without referring to counting?

$$\frac{2n(2n-1)\cdots(n+1)}{2^n} = \frac{(2n)!}{n! \cdot 2^n} = \frac{1 \cdot 2 \cdots 2n}{2 \cdot 4 \cdots 2n}$$
$$= 1 \cdot 3 \cdots (2n-1).$$

Sum and product notation

We will write

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n.$$

For example

$$\sum_{i=1}^{n} i = 1 + 2 + \cdots + n.$$

The variable i here is a bound variable: its meaning is restricted to inside the sum. We could use any other variable in its place, (but of course, not n or another variable in use):

$$\sum_{i=1}^{n} a_i = \sum_{p=1}^{n} a_p.$$

Note that

$$g(j) = \sum_{i=0}^{4} f(i,j) = f(0,j) + f(1,j) + f(2,j) + f(3,j) + f(4,j)$$

depends on j, but

$$g = \sum_{j=0}^{4} f(j,j) = f(0,0) + f(1,0) + f(2,2) + f(3,3) + f(4,4)$$

does not.

There is a corresponding notation for products: The number of pairings among $2n$ people was found to be

$$\frac{\prod_{i=n+1}^{2n} i}{2^n} = \prod_{i=1}^{n}(2i-1) = 1 \cdot 3 \cdots (2n-1) = \prod_{\substack{1 \leqslant i \leqslant 2n \\ i \text{ odd}}} i.$$

Note that i has a completely different meaning in each of the formulas. You can add conditions to the subscript, as in the last formula.

If you learned calculus, you have seen bound variables already. A definite integral is like a sum. In

$$\int_1^{15} \sin x \, dx = \int_1^{15} \sin y \, dy,$$

the variable x is a the bound variable, we could use y instead.

Sets

- Curly bracket notation: $\{2,3,5\}$. The party set :

$$P = \{\text{Alice}, \text{Bob}, \text{Carl}, \text{Diane}, \text{Eve}, \text{Frank}, \text{George}\}.$$

  Order does not matter:

$$\{\text{Alice}, \text{Bob}, \text{George}\} = \{\text{Bob}, \text{Alice}, \text{George}\},$$

- Element relation: $\text{Frank} \in P$.
- Number of elements (cardinality) of a set A: $|A|$. For example, $|\{\text{Alice}, \text{Bob}, \text{George}\}| = 3$, $|\{1,2,3,\dots\}| = \infty$.

- Set notation using conditions:

$$G = \{x \in P : x \text{ is a girl}\} = \{\text{Alice}, \text{Diane}, \text{Eve}\},$$
$$D = \{y \in P : y \text{ is over 21 years old}\} = \{\text{Alice}, \text{Carl}, \text{Frank}\}.$$

The x or y in this notation is a bound variable: its meaning is unrelated to everything outside the braces.

$$\{x \in \mathbb{Z} : 3|x\} = \{3x : x \in \mathbb{Z}\}.$$

Note that x has a different role on the left-hand side and on the right-hand side.

- The subset relation $A \subseteq B$. $A \subset B$ means proper subset. So both $G \subseteq P$ and $G \subset P$ are true.
- The empty set $\emptyset = \{\}$.
- Some important sets: $\emptyset \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.
  Nonnegative integers $\mathbb{Z}_+$.
  Positive integers $\mathbb{N}$ (in the notation of the book LPV, see remark below!).

- Set operations: $A \cup B$, $A \cap B$, $A \setminus B$, $A \Delta B$. For example:

$$\begin{aligned} G \cap D &= \{\text{Alice}\}, \\ G \cup D &= \{\text{Alice}, \text{Carl}, \text{Diane}, \text{Eve}, \text{Frank}\}, \\ G \setminus D &= \{\text{Diane}, \text{Eve}\}. \end{aligned}$$

- Disjoint sets: $A \cap B = \emptyset$. For example, $\{\text{Alice}, \text{George}\}$ is disjoint from $\{\text{Carl}, \text{Frank}\}$.
  One frequently writes for a sequence $A_1, \dots, A_n$ of sets either the statement that they are pairwise disjoint or, equivalently, that $A_i \cap A_j = \emptyset$ for all $i \neq j$.
- Warning: Do not confuse a one-element set with its element! For example, if $C = \{\text{Alice}, \text{George}\}$, then $|C| = 2$, but $|\{C\}| = 1$.

- Many ways to express the same thing, for example

$$A \subseteq B \Leftrightarrow A \cap B = A \Leftrightarrow A \cup B = B.$$

- For addition or subtraction of an element using set operations, we need the one-element set:

$$\{1,2,3\} \cup \{4\} = \{1,2,3,4\},$$
$$\{1,2,3,4\} \setminus \{4\} = \{1,2,3\}.$$

- One more example:

$$C = \{\text{Alice}, \text{George}\} \cup \emptyset, \qquad |C| = 2,$$
$$D = \{\text{Alice}, \text{George}\} \cup \{\emptyset\}, \qquad |D| = 3.$$

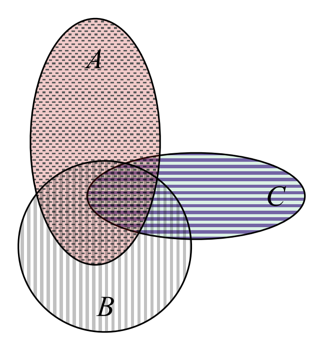
- There are many identities, for example

$$
\begin{aligned}
(A \cup B) \cap A = A = A \cup (B \cap A),& \\
A \cap (B \cup C) = (A \cap B) \cup (A \cap C) & \qquad \text{(distributivity of } \cap\text{)}, \\
A \cup (B \cap C) = (A \cup B) \cap (A \cup C) & \qquad \text{(distributivity of } \cup\text{)}.
\end{aligned}
$$

Let us prove the distributivity of $\cap$.. In proving an equality, it is frequently helpful to break it up into two inequalities, that is we will prove $\subseteq$ and $\supseteq$ separately.

- We prove $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$, that is that $x \in A \cap (B \cup C)$ implies $x \in (A \cap B) \cup (A \cap C)$. If $x \in A \cap (B \cup C)$ then $x \in A$ and either $x \in B$ or $x \in C$.
- Suppose that for example $x \in B$. Then $x \in A \cap B$, hence also $x \in (A \cap B) \cup (A \cap C)$ since $P \subseteq P \cup Q$ in general.
  The case $x \in C$ is handled similarly.
- We still need to prove $A \cap (B \cup C) \supseteq (A \cap B) \cup (A \cap C)$, this is left as an exercise.

- Illustration by Venn diagrams.



- The set of all subsets of a set A is denoted by 2^A.

On notation

- Some people write "|" in place of ":", as in $\{3x \mid x \in \mathbb{Z}\}$.
- The LPV book denotes by $\mathbb{N}$ the positive integers. In computer science (and logic), in general $\mathbb{N} = \mathbb{Z}_+$.
- Many people understand $A \subset B$ to mean the same as $A \subseteq B$, so it is better to be explicit, and write $A \subsetneq B$ for proper subset, if there is any chance of misunderstanding.

In general, mathematics (or computer science) is not about notation! Notation is important to communicate the ideas, but it is your responsibility to make sure that in each case, people understand what you mean: if there is a chance of ambiguity, you must state your conventions explicitly.

- The $\exists$ notation: we will return to it yet in more examples. "x divides y" means: $x|y \Leftrightarrow \exists z \in \mathbb{Z}\, x \cdot z = y$.

Example

Composite positive numbers:

$$\{x \cdot y : x, y \in \mathbb{N} \setminus \{1\}\} = \{n \in \mathbb{N} : \exists m \in \mathbb{N} \setminus \{1, n\}\, m|n\}.$$

- The $\forall$ notation:
  $A \subseteq B$ is the same as saying $\forall x$, $x \in A$ implies $x \in B$.
  Prime positive numbers:

$$\{n \in \mathbb{N} : \forall m \in \mathbb{N} \text{ if } m|n \text{ then } m \in \{1, n\}\}.$$

Logic

Natural language is ambiguous. There are some terms that, when we use them in mathematics, we give them a meaning that is always the same, even if sometimes this meaning seems sometimes strange.

Or

When we say "P or Q", $(P \vee Q)$ we always mean "P or Q or both ". For example, $x \in A \cup B$ is defined meaning $x \in A$ or $x \in B$, and this allows for $x \in A \cap B$.
On the other hand, $x \in A \Delta B$ means $x \in A$ or $x \in B$ but not in both (exclusive "or").

Implication

When we say "P implies Q", or "if P then Q" ($P \Rightarrow Q$), then we just mean that if P is true then Q is also true. This statement is false only if P is true and Q is false. If P is false, the implication is always true.

For example, $A \subseteq B$ means that for all x, $x \in A$ implies $x \in B$. This is true if A is empty, since in this case $x \in A$ is never true.

To emphasize this special case, sometimes if $P \Rightarrow Q$ is true just since P is false, we say that P is vacuously true.

Big unions and intersections

We have notation similar to big sums and products also for big unions and intersections:

$$\bigcup_{i=2}^{n-1} A_i = A_2 \cup \cdots \cup A_{n-1},$$
$$\bigcap_{j=1}^{m} B_j = B_1 \cap \cdots \cap B_m.$$

Complement

Frequently, all the sets we are considering are subsets of one set, called the universal set. For example, if we talk about sets of integers, we can take the set $\mathbb{Z}$ of all integers as the universal set. Let us denote the universal set by X. Then we will write

$$\overline{A} = X \setminus A.$$

The notation is useful because with it, we can write for example:

$$A \setminus B = A \cap \overline{B},$$

which makes some of the properties easier to understand. Also, now union and intersection are connected by the De Morgan rules:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \quad \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

Systematic enumeration

The number of subsets

We may want to not just know the subsets, but also to list them is some order. How to make sure we list all of them once and do not leave out anything?
Various ways to compute: we will learn from all.

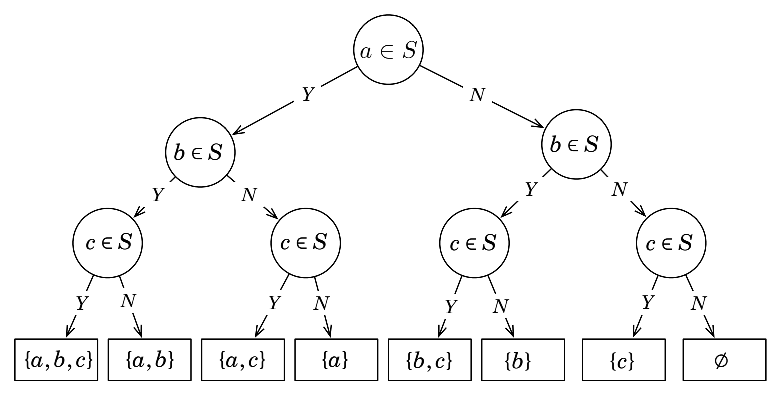
$$\emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}.$$

First the one-element subsets, then the two-element subsets, and so on. Easy for small sets, but not so easy to do systematically for larger ones. Phonebook ordering?

$$\emptyset, a, ab, abc, ac, b, bc, c.$$

Not very practical for enumerating subsets: which is the 233th subset here?

Decision tree



2^n leaves.

For a useful numbering, encode subsets of $\{a, b, c\}$ into 0-1 sequences of length 3:

- If $a \in S$ we write a 1 in position 1, otherwise a 0.
- If $b \in S$ we write a 1 in position 2, otherwise a 0.
- And so on.

Example: $\{a, c\} \to 101$. We represented every subset of a set of size n by a binary string.
We set up a one-to-one correspondence (bijection) between subsets of a set and binary strings.

Binary representation of integers

Recall the binary (base 2) representation of integers, for example

$$5 = 101_2 = 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0.$$

How do we find the binary representation $(b_n b_{n-1} \cdots b_1 b_0)$ of a natural number x? Here is a way, starting from the least significant digit.

1. If $x = 0$ return $(0)_2$. Else let $x_0 = x$.
2. While $x_i \neq 0$ do:
   $a_i :=$ the remainder of x_i after division by 2
   ($a_i = x_i \bmod 2$).
   $x_{i+1} := (x_i - a_i)/2$.
   $i := i + 1$.

And here is a way, starting from the most significant digit.

1. If $x = 0$ return $(0)_2$. Else let n be the largest such that $2^n \leqslant x$. Set $a_n = 1$, $x_{n-1} = x - 2^n$.
2. For $i = n-1$ downto 0 do:

   $a_i := 1$ if $2^i < x_i$, and 0 otherwise.

   $x_{i-1} := x_i - a_i \cdot 2^i$.

To make binary integers all the same length n, pad them by 0's in front. This is a bijection, a one-to-one correspondence between numbers $0, \dots, 2^{n-1}$ and binary strings of length n.
Combining the two bijections:

$$0 \leftrightarrow 000 \leftrightarrow \emptyset \qquad 4 \leftrightarrow 100 \leftrightarrow \{a\}$$
$$1 \leftrightarrow 001 \leftrightarrow \{c\} \qquad 5 \leftrightarrow 101 \leftrightarrow \{a,c\}$$
$$2 \leftrightarrow 010 \leftrightarrow \{b\} \qquad 6 \leftrightarrow 110 \leftrightarrow \{a,b\}$$
$$3 \leftrightarrow 011 \leftrightarrow \{b,c\} \qquad 7 \leftrightarrow 111 \leftrightarrow \{a,b,c\}$$

Now what is the 233th subset of a 10-element set?
Why two proofs? We learned something from each: decision trees, bijections.

Notation

The set of all subsets of a set A is denoted by 2^A. For example,

$$2^{\{a,b,c\}} = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\}.$$

Approximate number of subsets

How large is 2^n?

$$2^3 = 8 < 10, \quad 2^{99} < 10^{33}, \quad 2^{100} < 2 \cdot 10^{33}.$$

$$2^{10} = 1024 > 1000 = 10^3, \quad 2^{100} > 10^{30}.$$

(Note that “kilobyte” means 1024 bytes, not 1000 bytes.) So 2^{100} has between 31 and 34 digits. More precisely, we want to know the k for which

$$10^{k-1} \leqslant 2^{100} < 10^k.$$

Using $x = \log_{10} 2^{100} = 100 \log_{10} 2$, the number of digits is

$$k = \lfloor x \rfloor + 1 = \lfloor 100 \log_{10} 2 \rfloor + 1.$$

Since $\log_{10} 2 = 0.30103$, we get $k = 31$.

Sequences

A string, sequence: obtained by putting things one after the other: first, second, and so on. When elements of the string are coming from a set (an alphabet), it is assumed that each element can be used any number of times:

$$aabacb.$$

Theorem

The number of strings of length n composed of k given elements is k^n.

When there are k_1 choices for the first element, k_2 choices for the second one, and so on, then the number of strings of length n is $k_1 \cdot k_2 \cdots k_n$.

Example

How many nonnegative integers have exactly length n in decimal? $9 \cdot 10^{n-1}$.

Cartesian product

Ordered pair (x, y), unordered pair $\{x, y\}$. Ordered tuple: (a, b, c).
The Cartesian product of sets

$$A \times B \times C = \{(a, b, c) : a \in A, b \in B, c \in C\}.$$

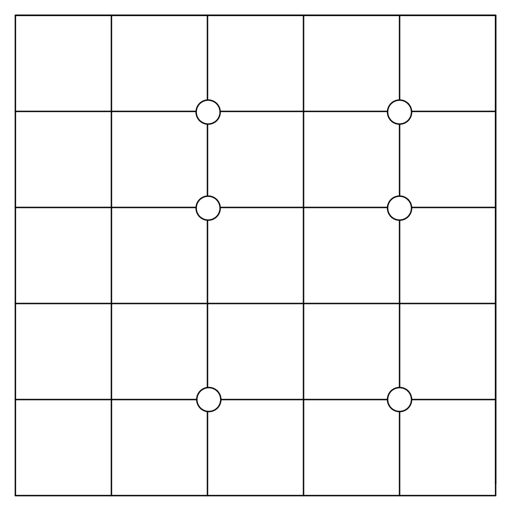
For example

$$\{1,2,3\}^2 = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3))\}.$$

We have $|A \times B \times C| = |A| \times |B| \times |C|$.
In particular, $|A^3| = |A|^3$.

Notation

The (x, y) notation conflicts with the same notation for open intervals. So, sometimes $\langle x, y\rangle$ is used for tuples or the scary notation $]x, y[$ for an open interval.



The Cartesian product $\{2,4\} \times \{1,3,4\}$.

Permutations, ordered subsets

An ordered subset of set A is a sequence of elements of A in which no two elements are the same.
We could use a decision tree again to illustrate the counting of ordered subsets of size k of a set of element n:

$$n(n-1)\cdots(n-k+2)(n-k+1).$$

.

Subsets of given size

Binomial coefficient

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

Values of $\binom{0}{0}$, $\binom{n}{1}$, $\binom{n}{n}$.

Theorem

Identities for binomial coefficients:

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

For $n, k > 0$*:*

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

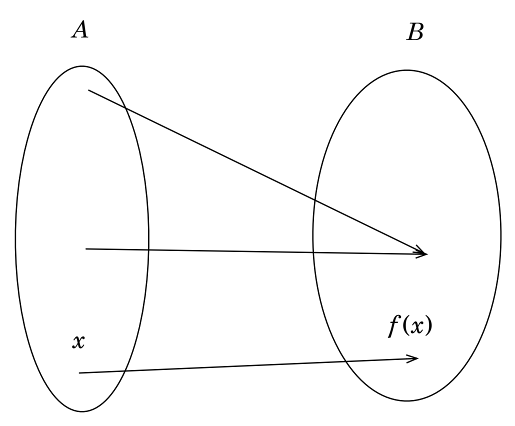
$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n-1} + \binom{n}{n} = 2^n.$$

Pascal triangle

						1						
					1		1					
				1		2		1				
			1		3		3		1			
		1		4		6		4		1		
	1		5		10		10		5		1	
1		6		15		20		15		6		1

Functions

Notation $f : A \to B$. We say that f is a function, or a mapping from A into B. A function from A to A is also called a transformation. We call the value $f(x) \in B$ also the image of the point $x \in A$ under the function f.



- When $f : A \to B$ then A is called the domain of f, and B (less frequently) the codomain. In the expression $f(x)$, we call x the argument (sometimes the parameter), or input, and $f(x)$ the output or value.
- A function is frequently defined with the help of a table: in the top line are listed all the elements $a_1, a_2, \ldots$ of the domain, in the bottom line the values $f(a_1), f(a_2), \ldots$. For example, for $A = \{1,2,3,4,5\}$, where $f(x) = x + 2 \bmod 5$, the table is

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

- A function $f(x, y)$ of two arguments $x \in A$, $y \in B$ with values $f(x, y) \in C$ can be viewed as a one-argument function from $A \times B$ to C, and this is how we denote it:

$$f : A \times B \to C.$$

Example

$g(x) = \frac{1}{x^2-1}$. It maps from $\mathbb{R} \setminus \{-1,1\}$, to $\mathbb{R}$, so

$$g : \mathbb{R} \setminus \{-1,1\} \to \mathbb{R}.$$

$\mathrm{Domain}(g) = \mathbb{R} \setminus \{-1,1\}$.

In general,

$$\mathrm{Range}(f) = \{ f(x) : x \in \mathrm{Domain}(f) \}.$$

In the example,

$$\mathrm{Range}(g) = (-\infty, -1] \cup (0, \infty) = \mathbb{R} \setminus (-1, 0].$$

Note that $(0, \infty)$ is an open interval.
Sometimes we will use the notation

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

to define a function like $g(x)$.

Another frequent way to define a function is case-by-case. For example, in grading an exam I first assign scores to each student. Then I define a function $G : \{1, 2, \ldots, 100\} \to \{A, B, C, F\}$ assigning grades to scores as follows:

$$G(x) = \begin{cases} A & \text{if } x \geqslant 85, \\ B & \text{if } 85 > x \geqslant 70, \\ C & \text{if } 70 > x \geqslant 50, \\ F & \text{otherwise.} \end{cases}$$

(This is just an example. I actually use all the grades $A, A-, B+, B, B-, C+, C, C-, D, F$.)

Indicator function

Sets can also be described by functions. Let X be some set (our universal set) and $A \subseteq X$. We define the indicator function $I_A : X \to \{0, 1\}$ of the set A by the fomula

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases}$$

The indicator function has a nice relation to set operations:

$$I_{A \cap B}(x) = I_A(x) \cdot I_B(x), \quad I_{\overline{A}}(x) = 1 - I_A(x),$$
$$|A| = \sum_{x \in X} I_A(x).$$

Using this and De Morgan's rule we can conclude

$$I_{A \cup B}(x) = 1 - (1 - I_A(x))(1 - I_B(x)) = I_A(x) + I_B(x) - I_A(x) \cdot I_B(x).$$

Inverse image

Whether a function $f : A \to B$ is invertible or not, for an arbitrary subset $D \subseteq B$ we will write

$$f^{-1}(D) = \{ x : f(x) \in D \}.$$

Note that $f^{-1}(D)$ is always a set, and it may be empty. So if $f : A \to B$ then

$$f^{-1} : 2^B \to 2^A$$

where 2^B denotes the set of subsets of B.

Example

If $f : \mathbb{Z} \to \mathbb{Z}$ is the function with $f(x) = 2\lfloor x/2 \rfloor$ then we have

$$f^{-1}(\{0\}) = \{0,1\}, f^{-1}(\{1\}) = \emptyset = \{\}, f^{-1}(\{2\}) = \{2,3\}, f^{-1}(\{3\}) = \emptyset, \dots.$$

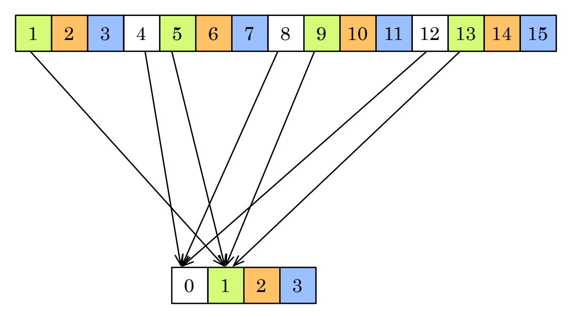
Partitions

An ordered partition of a set A is a finite sequence $(A_1, \dots, A_n)$ of pairwise disjoint subsets of A such that $A_1 \cup \cdots \cup A_n = A$. Given any function $f : A \to \{1, \dots, n\}$, it gives rise to an ordered partition $(f^{-1}(\{1\}), \dots, f^{-1}(\{n\}))$. And every ordered partition defines such a function.

An unordered partition, or simply partition, is just a set $\{A_1, \dots, A_n\}$ of dijoint subsets of A, whose union is A.

Example

The subdivision of 6 people into 3 chess-playing pairs is an unordered partition into sets of size 2.



The function $g : \{1, \dots, 15\} \to \{0, \dots, 4\}$ defined by

$$g(x) = x \bmod 4 \qquad = \text{the remainder of } x \text{ after division by } 4.$$

The partition into inverse images is

$$\begin{aligned}\{1, \dots, 15\} &= g^{-1}(\{0\}) \cup g^{-1}(\{1\}) \cup g^{-1}(\{2\}) \cup g^{-1}(\{3\}) \\ &= \{4, 8, 12\} \cup \{1, 5, 9, 13\} \cup \{2, 6, 10, 14\} \cup \{3, 7, 11, 15\}.\end{aligned}$$

Surjective (onto) property

Sometimes for a function $f : A \to B$, and a set $C \subseteq A$ we will write

$$f(C) = \{ f(x) : x \in A \}.$$

For example, $\mathrm{Range}(f) = f(A)$. Example: $2\mathbb{Z}$ is the set of even numbers.

A function $f : A \to B$ is called onto (surjective), that is a mapping from A onto B, if $\mathrm{Range}(f) = B$.

Example

The function $g : \mathbb{R} \setminus \{-1, 1\} \to \mathbb{R}$ defined in the above example as $g(x) = 1/(x^2 - 1)$ is not surjective, its range is $\mathbb{R} \setminus (-1, 0]$. It becomes surjective if we define it as $g : \mathbb{R} \setminus \{-1, 1\} \to \mathbb{R} \setminus (-1, 0]$.

Injective (one-to-one) property

A function is one-to-one, injective, or 1-1 if $x \neq y$ implies $f(x) \neq f(y)$ (or equivalently, $f(x) = f(y)$ implies $x = y$ for all x, y).

Example

An ordered subset of size 4 of a set A is an injective mapping from the set $\{1, 2, 3, 4\}$ to A.

Example

A one-to-one function that is not onto: the function $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = 2x$.
An onto function $g : \mathbb{Z}_+ \to \mathbb{Z}_+$ that is not one-to-one:

$$g(x) = \begin{cases} x - 1 & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Invertible functions

A function is called invertible, or a bijection, if it is onto and one-to-one. For an invertible function $f : A \to B$, the inverse function $f^{-1} : B \to A$ is always defined uniquely: $f^{-1}(b) = a$ if and only if $f(a) = b$.

An invertible function is also called a one-to-one correspondence.

We have used this notion already several times in counting: if there is a one-to-one correspondence between two finite sets A and B then, of course, $|A| = |B|$. (For infinite sets A, B, this is taken as the definition of the relation $|A| = |B|$.)

An invertible function $f : A \to A$ is also called a permutation.

Examples

1. For the function $s : \mathbb{R} \to \mathbb{R}$, $s(x) = x^2$, we have $\text{Range}(s) = \mathbb{R}_+$, since squares are nonnegative. To make it surjective, we must therefore decrease its codomain to $\mathbb{R}_+$, $s : \mathbb{R} \to \mathbb{R}_+$. But this function is still not injective, since $s(-x) = s(x)$. To make it injective we can restrict it to the domain $\mathbb{R}_+$: the function

$$s : \mathbb{R}_+ \to \mathbb{R}_+, \quad s(x) = x^2$$

   is bijective. Its inverse is $s^{-1}(x) = \sqrt{x}$. Indeed, $y = x^2$, $x \geqslant 0$ means $x = \sqrt{y}$.

2. For the function $E : \mathbb{R} \to \mathbb{R}_+$, $E(x) = 10^x$, we have $\text{Range}(s) = \mathbb{R}_+ \setminus \{0\} = (0, \infty)$, since powers of 10 are positive. The map $E : \mathbb{R} \to (0, \infty)$ is bijective, its inverse is is $E^{-1}(x) = \log_{10} x$. Indeed, $y = 10^x$ means $x = \log_{10} y$.

Example

I am in Manhattan, at the corner of the Eighth Avenue and 25th Street, and want to get to the corner of the Second Avenue and 80th Street. How many different shortest paths do I have?

Each shortest path makes 6 moves eastward and and 55 moves north, so it corresponds to a sequence of the sort $enneennnnnnn\cdots n$ of length 61 with 6 occurrences of e and 55 occurrences of n. This correspondence between the set of shortest paths and the set of such sequences is 1-1.

Similarly, each such sequence corresponds to a subset of size 6 of the set $\{1,2,\ldots,61\}$, namely to the set of positions in which the letter is e. This correspondence is also 1-1. We learned that the number of subsets of size 6 of a set of size 61 is $\binom{61}{6}$.

The discovery of the two 1-1 correspondences helped reduce the original problem to a problem whose solution we already know.

Theorem

Let $A = \{a_1, \dots, a_m\}$, $B = \{b_1, \dots, b_n\}$ be finite sets. For a function $f : A \to B$ the following holds.

- (a) *Suppose that f is one-to-one (injective). Then $m \leqslant n$, and $m = n$ implies that f is onto (surjective).*
- (b) *Suppose that f is onto (surjective). Then $m \geqslant n$, and $m = n$ implies that f is one-to-one (injective).*

It follows that if $m = n$ then f is injective if and only if it is surjective.

- The contrapositive of (a) says that if $m > n$ then f is not one-to-one: this is called the pigeonhole principle: If you put m pigeons into fewer holes, one hole contains more than one pigeon.
- As the earlier examples show, the theorem is false for infinite A.

Proof. Point (a) follows since each a_i gets a distinct image in $f(a_i)$. Listing the elements of B we can start with $f(a_1), \dots, f(a_m)$, so we cannot get more than n. If $m = n$ then we listed all elements of B as some $f(a_i)$, so f is surjective.

To see point (b), look at the partition of A into inverse images $f^{-1}(\{b_j\})$ of elements of B. Pick an element $a'_j \in f^{-1}(\{b_j\})$ for each j. Listing the elements of A we can start with $a'_1, a'_2, \dots, a'_m$, so we cannot get fewer than m. If $m = n$ then each set $f^{-1}(\{b_j\})$ contains only one element, so f is injective. □

An application

Let us apply the above theorem to show the following.

Theorem

Let $A = \{1, \dots, 16\}$. *For every pair of integers* $x, y \in A$ *there is an integer* $z \in A$ *such that* $x \cdot z \bmod 17 = y$.

Proof. We will use the fact that 17 is a prime number. Let us fix x and look at the map g defined by $g(z) = x \cdot z \bmod 17$. We know that $x \cdot z \bmod 17 \in A$: indeed, since 17 does not divide x, z it does not divide $x \cdot z$ either. So g is a map from A to A. The theorem says that g is surjective.

By the previous theorem, it is sufficient to show that g is injective. Assume $g(u) = g(v)$, we will show $u = v$. Now if $x \cdot u \bmod 17 = x \cdot v \bmod 17$ then 17 divides $x \cdot u - x \cdot v = x \cdot (u - v)$. The primality of 17 implies that then 17 divides $u - v$, which can happen only if $u = v$. □

Non-constructive proof

Note that this proof did not give any method for computing the number z whose existence is claimed in the theorem. Such proofs are called existential proofs. Of course, trying out all candidates $1, \ldots, 16$ for z is a method. But the theorem is true for all prime numbers p in place of 17, and the exhaustive search becomes too costly for a p with, say, 100 digits.

Composition of functions

If $f : A \to B$ and $g : B \to C$ are functions then we can always define the composition $h : A \to C$, written as $h = g \circ f$ by

$$(g \circ f)(x) = h(x) = g(f(x)).$$

Example

With $f, g : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x + 1$, $g(y) = 3y$, we have

$$(g \circ f)(x) = 3x + 3, \quad (f \circ g)(x) = 3x + 1.$$

Permutations

It is very frequent to consider compositions of functions of the form $f : A \to A$ (called transformations), especially if they are invertible: they are permutations. If σ, τ are permutations, we frequently write the composition $\sigma \circ \tau$ as just $\sigma\tau$. When a permutation is defined by a table, say $\sigma, \tau : \{1,2,3,4,5\} \to \{1,2,3,4,5\}$ as

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{pmatrix}, \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 2 & 3 & 4 \end{pmatrix}$$

then its inverse can be computed: for each i, just find the j at the top with $\sigma(j) = i$, and set $\sigma^{-1}(i) = j$:

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 5 & 3 & 4 \end{pmatrix}.$$

The product is found by just following the tables: $1 \xrightarrow{\tau} 5 \xrightarrow{\sigma} 3$, $2 \xrightarrow{\tau} 1 \xrightarrow{\sigma} 2$, and so on.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 2 & 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 1 & 4 & 5 \end{pmatrix}.$$

Relations

A binary relation is a set $R \subseteq A \times B$. We will write $(x, y) \in R$ also as $R(x, y)$ (with Boolean value). Thus

$$R(x, y) \Leftrightarrow (x, y) \in R.$$

We sometimes call A the domain and B the codomain of the relation.

Examples

- $L \subseteq \mathbb{R}^2$, $L(x, y) \Leftrightarrow x < y$. Relations are frequently written with the infix notation, like here: thus, "$x < y$" also expresses the relation $<$, we may even write $< \subseteq \mathbb{R} \times \mathbb{R}$.
- Let $G = \{\text{Alice}, \text{Bob}, \text{Carl}, \text{Diana}, \text{Eve}, \text{Frank}, \text{George}\}$. $S \subseteq G^2$, where $S(x, y)$ means that x, y are siblings.
- Let $H \subseteq G^2$ where $H(x, y)$ means that x is husband of y. Of course, we could have defined $H \subseteq G_M \times G_F$ where $G_M = \{\text{Bob}, \text{Carl}, \text{Frank}, \text{George}\}$, $G_F = \{\text{Alice}, \text{Diana}, \text{Eve}\}$.

Ternary relation: $R \subseteq A \times B \times C$.

Example

$M \subseteq \mathbb{Z}^3$, where $M(x, y, z) \Leftrightarrow z | y - x$. This relation is sometimes written as

$$x \equiv y \pmod{z},$$

and is equivalent to $x \bmod z = y \bmod z$.

Let $f : A \to B$ be a function, then we can define the relation $G_f \subseteq A \times B$ as

$$G_f(x, y) \Leftrightarrow y = f(x).$$

This relation is called the graph of function f.

Example

Recall the function $\mathbb{R} \setminus \{-1, 1\} \to \mathbb{R}$ defined by $g(x) = \frac{1}{x^2-1}$. Its graph in the usual sense is the set of points in the plane defined by

$$G_g = \left\{ \left(x, \frac{1}{x^2 - 1} \right) : x \in \mathbb{R} \setminus \{-1, 1\} \right\}.$$

Question

When is a relation $R \subseteq A \times B$ the graph of a function?

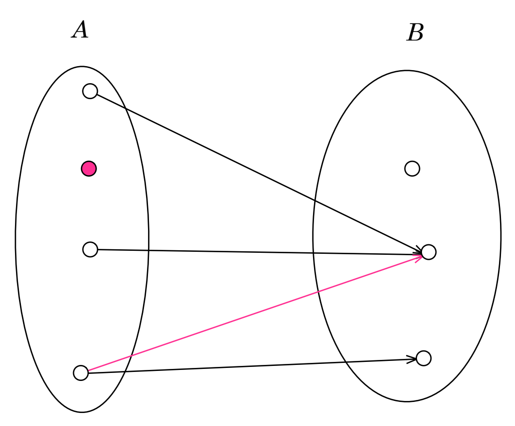
When the following two properties hold:

- $\forall x \in A \, \exists y \in B \, R(x, y)$.
- $\forall x \in A \, \forall y, z \in B \, R(x, y) \wedge R(x, z) \Rightarrow y = z$.

In words, if for all $x \in A$ there is a unique y with $R(x, y)$. The expression “there is a unique y” is sometimes denoted by $\exists! y$:

$$\exists! x \, P(x) \Leftrightarrow \exists x \, P(x) \wedge \forall x, y \, (P(x) \wedge P(y) \Rightarrow x = y).$$

Thus, R is a function iff $\forall x \in A \, \exists! y \in B \, R(x, y)$.



The arrow diagram of a relation R. The red parts show how it may differ from the arrow diagram of a function $A \to B$:

- Some elements of A are not related to any element of B.
- Some elements of A are related to more than one element of B.

Some frequent properties of binary relations

Reflexivity

Relation $R \subseteq A^2$ is reflexive if $R(x,x)$ always holds.

Examples

Let A be the set of cities in Massachusetts.

- The relation

  $$C = \{(x,y) \in A^2 : x \text{ is closer than 10 miles to } y\}$$

  is reflexive.
- The relation

  $$F = \{(x,y) \in A^2 : x \text{ is farther than 10 miles to } y\}$$

  is not reflexive.

Symmetry

Relation $R \subseteq A^2$ is symmetric if $R(x,y)$ implies $R(y,x)$.
Relation it is antisymmetric if $R(x,y), R(y,x)$ implies $x = y$.

Examples

- Let A be the set of cities in Massachusetts. Both of the above relations C,F are symmetric.
- The relation $x \leqslant y$ among real numbers is antisymmetric.
- In the group of people {Alice, ..., George}, the relation $H(x,y)$ expressing that x is the husband of y is antisymmetric (in an uninteresting way).
- In the set $\mathbb{Z}$ the relation $x|y$ is not symmetric, but not antisymmetric either. Indeed, $3|6$ but $6 \nmid 3$. On the other hand, $3 \mid -3$ and $-3 \mid 3$.

Transitivity

Relation $R \subseteq A^2$ is transitive if $R(x,y)$ and $R(y,z)$ implies $R(y,z)$.

Examples

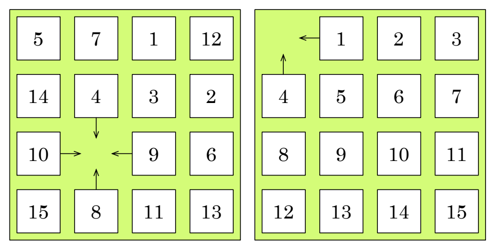
- Let $B = 2^A$ for a set A. The relation $X \subseteq Y$ for $X, Y \in B$ (that is $X, Y \subseteq A$) is transitive.
- Let P the set of all people. The relation $S \subseteq P^2$ where $S(x,y)$ holds if x is a sibling of y, is transitive. The relation $S' \subseteq P^2$ where $S'(x,y)$ holds if x is a half-sibling of y, is not transitive.

Equivalence relation

A relation $R \subseteq A^2$ is calle an equivalence relation if it is reflexive, symmetric and transitive.

Examples

- For a function $f : A \to B$, let $R(x, y) \Leftrightarrow f(x) = f(y)$.
- For $x, y \in \mathbb{Z}$ let $x \sim y$ if $3 | x - y$. We will denote this also as $x \equiv y \pmod 3$. Special case of the previous example, since $x \sim y \Leftrightarrow x \bmod 3 = y \bmod 3$.
- Let N be the set of necklaces of size 10, made up of 2 red beads and 8 blue beads. We say $x \sim y$ for $x, y \in N$ if x can be obtained by a rotation from y.
- The 15-puzzle. For two arrangements we write $x \sim y$ if one can be transformed into the other using shifts, without taking out any pieces.



If you want to play the puzzle without having a physical copy, go for example to
`http://www.cut-the-knot.org/pythagoras/fifteen.shtml`.

Equivalence under a set of permutations

More generally, let P be a set of permutations of a set A such that if $p \in P$ then $p^{-1} \in P$. Write $x \sim_P y$ if there is a sequence $p_1, p_2, \dots, p_n \in P$ with $y = p_n(p_{n-1}(\cdots p_1(x) \cdots))$.

Example

Set of permutations $Q = \{\sigma, \varrho\}$ of $\mathbb{Z}$ where $\sigma(x) = x + 3$, $\varrho(x) = x - 3$. Then $x \equiv y \pmod 3$ iff $x \equiv_Q y$.

We call P a group of permutations if also for all $p, q \in P$ we have $p \circ q \in P$ and $p^{-1} \in P$.

Proposition

If P is a group then $x \sim_P y$ iff $\exists p \in P\ y = p(x)$.

Example

Necklaces: the combination of any two rotations is a rotation.

Rays

Let $A = \mathbb{R}^2 \setminus \{(0,0)\}$, and define $T \subseteq A$ as follows: We say $T((x_1, y_1), (x_2, y_2))$ if $x_1 y_2 = x_2 y_1$. (We want to write $\frac{x_1}{y_1} = \frac{x_2}{y_2}$ but cannot since y_1 or y_2 may be 0.) We will show that T is an equivalence.

For every $\alpha \in \mathbb{R}$, let $p_\alpha : A \to A$ be the mapping defined by

$$p_\alpha((x, y)) = (\alpha x, \alpha y).$$

Whenever $\alpha \neq 0$, this is a permutation. Let $P = \{p_\alpha : \alpha \neq 0\}$. Note that P is a group: if $\alpha, \beta \neq 0$ then $p_\alpha \circ p_\beta = p_{\alpha\beta}$, and $p_\alpha^{-1} = p_{\alpha^{-1}}$.
The fact that T is an equivalence relation follows from the following characterization:

Proposition

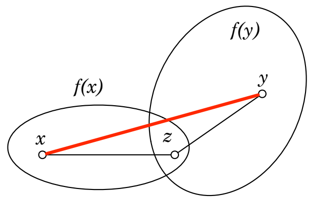
We have $T((x_1, y_1), (x_2, y_2)) \Leftrightarrow \exists \alpha \neq 0 \; p_\alpha((x_1, y_1)) = (x_2, y_2)$.

Theorem

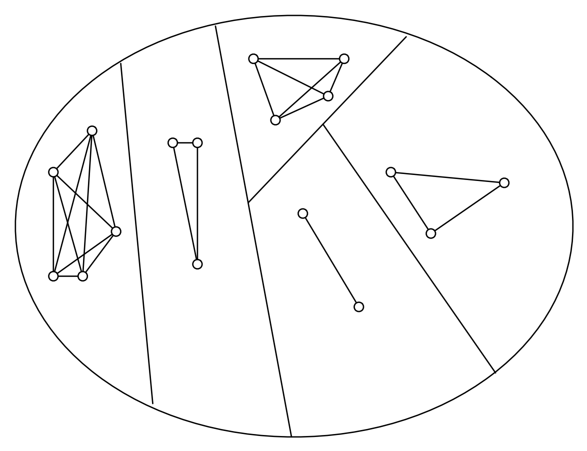
A relation $R \subset A \times A$ is an equivalence relation if and only if there is a partition $\mathscr{P}$ of the set A into nonempty subsets such that $R(x,y) \Leftrightarrow \exists B \in \mathscr{P}\,(x, y \in B)$.

Proof. If R is defined by a partition as in the theorem, it is easy to check that the three properties hold.
Suppose the three properties hold, we define a function $f : A \to 2^A$ as follows: $f(x) = \{y \in A : R(x,y)\}$. We will show that the range of f is the desired partition.

- If $f(x) \cap f(y) \neq \emptyset$ then $x \in f(y)$, $y \in f(x)$, $f(x) = f(y)$.
- The sets $f(x)$ for $x \in A$ form a partition of A.
- $R(x,y) \Leftrightarrow f(x) = f(y)$.

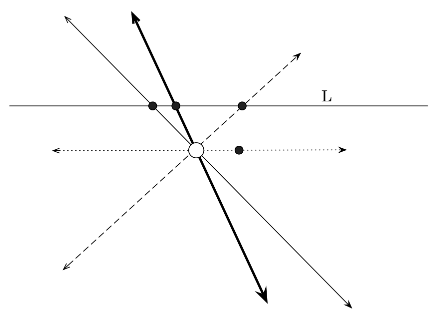


□



Elements (the individual sets) of the partition obtained from the equvalence relation are called its equivalence classes.

Example application: look at the example of the relation T defined above, on $\mathbb{R}^2 \setminus \{(0,0)\}$. The equivalence classes of this relation are called rays.



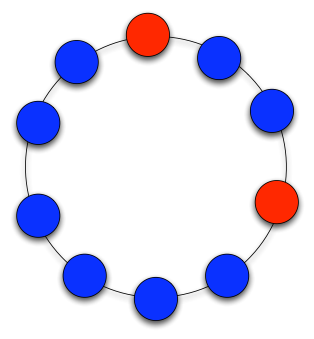
In a partition, we frequently pick a representative in each class (black points on the figure). For a ray (x,y) if $y \neq 0$ we can pick $\left(\frac{x}{y}, 1\right)$ (intersection with horizontal at height 1). If $y = 0$ pick, say, $(1,0)$.

Example

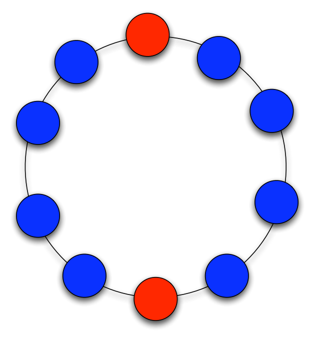
Let C be the set of cities in Massachusetts. We say $T(x, y)$ if x and y are closer than 10 miles. Relation T is reflexive, symmetric, but not transitive.

The non-transitivity of this relation is an important obstacle in many problems faced in science. Suppose we have measurement data, each of them a triple of numbers (x, y, z), say voter opinon on birth control, taxes and health-care reform. All our measurements are representable as a “cloud” in 3-dimensional space. Looking at it, the cloud seems to break up into “clusters”, clumps of data that seem to denote important groups: for example, important groups of voters to be addressed differently. But “closeness” of the points is not an equivalence relation allowing to break up the cloud into clusters. Other ideas are needed, none of them very compelling.

Recall the necklaces of size 10, with 2 red and 8 blue beads.



Necklace with class size 10.



Necklace with class size 5.

Examples

- The number of equivalence classes of the 15-puzzle is 2, and they both have the same size. (I will give homework problems which will help seeing this.) So if you spill out the puzzle and put it back randomly, there is a 50% chance that it will not be solvable.
- The corresponding number of equivalence classes for Rubik's Cube is 12, and they all have the same size. So if you take apart Rubik's Cube and put it together randomly, there is only a 1/12 chance to obtain a solvable cube.

Preorder, partial order

Preorder $\leqslant$: reflexive, transitive.

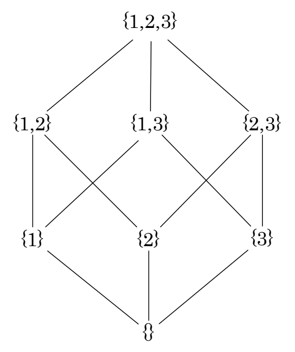
Example

For $x, y \in \mathbb{Z}$ the relation $x|y$.

A preorder is a partial order if it is antisymmetric.

Examples

- $\leqslant$ among real numbers.
- $\subseteq$ among subsets of a set.



Proposition

In a preorder, we can introduce a relation $\sim$*:* $x \sim y$ *if* $x \leqslant y$ *and* $y \leqslant x$*. This is an equivalence relation, and the relation induced by* $\leqslant$ *on the equivalence classes is a partial order.*

Example

The equivalence classes of the preorder $x|y$ among integers are the sets $\{x, -x\}$, for $x \in \mathbb{Z}$.

A partially ordered set is a pair $(A, \preceq)$, where A is a set and $\preceq$ is a partial order defined on it. Element x is minimal if $y \preceq x$ implies $y = x$ for all y.

Examples

- Nonempty subsets of a set A, ordered by inclusion. Minimal elements: the one-element subsets.
- Integers > 1, ordered by $x|y$. Minimal elements: prime numbers.

A partial order $\preceq$ is an order if for all x, y we have $x \preceq y$ or $y \preceq x$. We say that a relation R' extends a relation R if $R \subseteq R'$, that is $R(x, y) \Rightarrow R'(x, y)$.

Theorem

Let A be finite set, with a partial order $\preceq$ defined on it. Then $\preceq$ can always be extended to a complete order $\preceq'$.

Proof. Take a minimal element x_1 (in a finite partially ordered set, there is always one). Set $x_1 \preceq' y$ for all $y \in A$. Let $A_1 = A \setminus \{x_1\}$. Let x_2 be a minimal element of A_1. Set $x_2 \preceq' y$ for all $y \in A_1$. And so on. □

Inclusion-exclusion

In a class of 40 students (set X), say

- 18 have a picture of the Beatles (set A)
- 16 have a picture of the Rolling Stones (set B)
- 12 have a picture of Elvis Presley (set C)
- 7 have a picture of the Beatles and the Rolling Stones
- 5 have a picture of the Beatles and Elvis Presley
- 3 have a picture of the Rolling Stones and Elvis Presley
- 2 have all these pictures

How many students have no picture of any of these? Answer:

$$|X| - (|A| + |B| + |C|) + (|A \cap B| + |A \cap C| + |B \cap C|) - |A \cap B \cap C|$$
$$= 40 - (18 + 16 + 12) + (7 + 5 + 3) - 2 = 7.$$

Naive explanation via repeated corrections.

Let us deduce the formula using indicator functions (recall!). By the De Morgan rule:

$$\overline{(A \cup B \cup C)} = \overline{A} \cap \overline{B} \cap \overline{C}.$$

The indicator function of this set is

$$\begin{aligned}
(1 - I_A(x))(1 - I_B(x))(1 - I_C(x)) &= 1 - (I_A(x) + I_B(x) + I_C(x)) \\
&+ (I_A(x)I_B(x) + I_A(x)I_C(x) + I_B(x)I_C(x)) - I_A(x)I_B(x)I_C(x) \\
&= 1 - (I_A(x) + I_B(x) + I_C(x)) + (I_{A\cap B}(x) + I_{A\cap C}(x) + I_{B\cap C}(x)) \\
&\quad - I_{A\cap B\cap C}(x).
\end{aligned}$$

Summing up by x we get the inclusion-exclusion formula.

Example

In how many ways can we color n cards in red, green, blue, if we have to use all three colors?

Let S be the set of all colorings, S_R the set of colorings not using red, similarly for green and blue. Let S_{RG} be the set of colorings not using either red or green (so, using only blue), and so on. Notice $S_R \cap S_G = S_{RG}$, and so on. We want to know $|S \setminus (S_R \cup S_G \cup S_B)|$. By inclusion-exclusion, it is

$$\begin{aligned} &|S_{RGB}| - (|S_R| + |S_G| + |S_B|) + (|S_{RG}| + |S_{RB}| + |S_{GB}|) \\ &= 3^n - 3 \cdot 2^n + 3. \end{aligned}$$

Derangements

We call a permutation σ of $\{1,2,\dots,n\}$ a derangement if $\sigma(x) \neq x$ is true for all x. For example from the permutations below:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$$

the first one is a derangement, the second one is not. How many derangements of $\{1,\dots,n\}$ are there? We will compute this using inclusion-exclusion. Let P be the set of all permutations. For any subset $S \subseteq \{1,\dots,n\}$ of size k, let P_S be the set of permutations that leave all elements of S fixed. We are interested in the size of the set

$$D = P \setminus (P_{\{1\}} \cup P_{\{2\}} \cup \dots \cup P_{\{n\}}).$$

The inclusion-exclusion formula needs the size of intersections. Note $P_{\{1\}} \cap P_{\{2\}} = P_{\{1,2\}}$ and, in general, if $S = \{s_1,\dots,s_k\}$ then

$$P_S = P_{\{s_1\}} \cap \dots \cap P_{\{s_k\}}.$$

The inclusion-exclusion formula gives

$$|D| = |P| - \sum_i |P_{\{i\}}| + \sum_{i<j} |P_{\{i,j\}}| - \cdots ,$$

the general term is $\sum_{|S|=k} |P_S|$. There are $\binom{n}{k}$ terms in this sum, each has size $|P_S| = (n-k)!$, so the value of the sum is $(n-k)!\frac{n!}{k!(n-k)!} = \frac{n!}{k!}$. This gives

$$|D| = n! - \frac{n!}{1!} + \frac{n!}{2!} - \frac{n!}{3!} + \cdots \pm 1 = n!\left(\frac{1}{2!} - \frac{1}{3!} + \cdots \pm \frac{1}{n!}\right),$$

where the $\pm$ is $+$ or $-$ depending on whether n is even or odd. Recall from analysis:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots ,$$

$$e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} \cdots .$$

For large n this gives $|D| \approx n!/e$.

Mathematical induction

Sometimes we can guess a result from examples, but proving it still seems complicated. In many of these cases, a method called mathematical induction helps.

Example

You may notice

$$1+3=4, \quad 1+3+5=9, \quad 1+3+5+7=16, \dots.$$

This suggests the identity $1+3+\cdots+(2n-1)=n^2$.

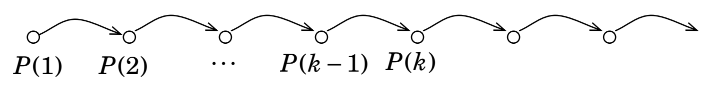
How to prove this?

Theorem

Let $P(k)$ be a predicate on positive integers with the following properties:

a. *$P(1)$ holds. (This is called the base case.)*
b. *For all k, if $P(1), P(2), \dots, P(k-1)$ holds then also $P(k)$ holds. (This is called the induction step.)*

Then $P(n)$ is true for all n.



This theorem is also sometimes called strong induction, since we assumed not only $P(k-1)$ but all of $P(1) \wedge \cdots \wedge P(k-1)$. But we can indeed assume all of that, so I will not distinguish between these two kinds of induction.

Application to the example: Here, $P(n)$ asserts $\sum_{i=1}^{n}(2i-1) = n^2$.
Base case: The statement $P(1)$ just says $1 = 1$, so it is true.
Induction step: Assume $P(k)$, we will prove that then $P(k+1)$. So we know

$$1 + 3 + \cdots + (2k-1) = k^2.$$

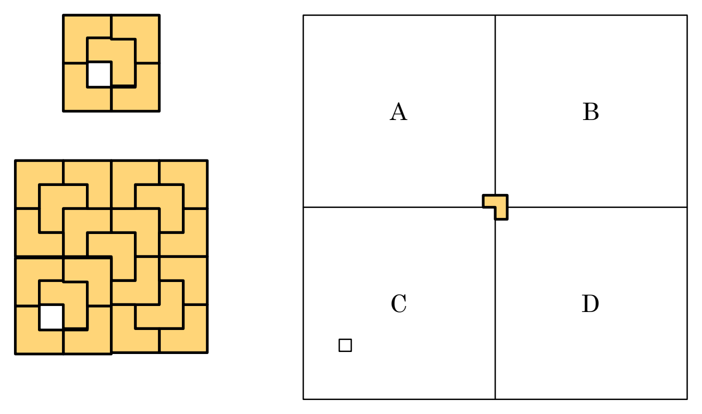
Adding $2(k+1) - 1 = 2k+1$ to both sides:

$$1 + 3 + \cdots + (2(k+1) - 1) = k^2 + 2k + 1 = (k+1)^2.$$

But this is the statement $P(k+1)$. So assuming $P(k)$ we proved $P(k+1)$. The example shows the usefulness of mathematical induction: we can assume additional things in the proof, making it frequently much easier to carry out.
(Recall that looking for proof by contradiction had a similar advantage.)

Theorem

Consider square of size 2^n, subdivided into $2^n \times 2^n$ unit squares, from which one unit square has been removed. The remaining area can be covered by L-shaped figures consisting of 3 unit squares each.



Greatest common divisor

Here is the most important theorem of number theory.

Theorem

Let $a, b \geqslant 0$ be integers, $a + b > 0$. There are integers u, v such that $d = u \cdot a + v \cdot b > 0$ divides both a and b.

Clearly, every common divisor of a, b divides d: therefore d is called the greatest common divisor.

Proof. Mathematical induction on $a + b$. Assume $a \geqslant b$. If $b = 0$ then $d = a = 1 \cdot a + 0 \cdot b$. Otherwise let $a' = a - b$. The pairs (a', b) and (a, b) have the same common divisors. By the inductive assumption, they have a greatest common divisor d with

$$d = u' \cdot a' + v' \cdot b = u'(a - b) + v' \cdot b = u' \cdot a + (v' - u') \cdot b.$$

□

Here is an application of the above theorem. We prove something that you may think obvious, but in fact it is not obvious at all: it is why the prime decomposition is unique.

Theorem

If p is a prime number and $p|a \cdot b$ then either $p|a$ or $p|b$.

Proof. Suppose p does not divide a. Let d be the greatest common divisor of a and p: $d = u \cdot a + v \cdot p$. As a divisor of p it is 1 or p. Since p does not divide a, it is 1. Then

$$1 = d = u \cdot a + v \cdot p,$$
$$b = u \cdot ab + v \cdot p \cdot b.$$

Since p divides both terms on the right-hand side, it divides the left-hand side, too. □

Towers of Hanoi

This game is described in almost every text on recursive programs. Let $f(n)$ be the minimum number of moves to move a tower of size n from one pin to another, using one more pin as a helper. We claim

$$f(n) \leqslant 2^n - 1.$$

The statement is true for $n = 1$. Assume it is true for n, we will prove it for $n+1$. The crucial observation is

$$f(n+1) \leqslant 2f(n) + 1.$$

Using it: $f(n+1) \leqslant 2f(n) + 1 \leqslant 2(2^n - 1) + 1 = 2^{n+1} - 1$.
Can you prove the equality $f(n) = 2^n - 1$?

Frequently, it helps to generalize a statement in order to prove it by induction. Number the disks of the towers of Hanoi from bottom to top by $1,2,3,\dots,n$.

Claim

Consider the configuration in which pin 1 has the disks $1,4,7,11,\dots$, *pin 2 has the disks* $2,5,8,\dots$, *and pin 3 the disks* $3,6,9,\dots$. *There is a sequence of legal moves from here to the initial configuration, with disks* $1,2,3,\dots,n$ *on pin 1.*

We will prove a generalization: for an arbitrary legal configuration of disks, there is a legal sequence of moves back to the initial one.

Mathematical induction on the number of disks. With just one disk, the statement is obvious. Assume that the task is solvable with fewer than n disks, we show a solution for n disks. By the assumption, all disks but the largest one can be moved to pin 2. Now move the largest disk to pin 1 and then, as usual, move the rest to pin 1.

Ramsey numbers

Often, to prove an assertion by induction, we have to generalize it first, in order to obtain a sufficiently strong assumption for the induction step. Of course, after assuming more, you also must prove more in the induction step!

Theorem

Let $s \geqslant 0$ be an integer. In a company with $\binom{2n-2}{n-1}$ people there are either n people who all know each other, or n people who do not know each other.

For example, in a company with $6 = \binom{4}{2}$ people, either there are $3 = 2 + 1$ people who all know each other, or 3 people who do not know each other.

In order to prove this theorem by induction, we generalize it:

Theorem

Let $s, t \geqslant 0$ be integers. In a company with

$$\binom{s+t}{s}$$

people, there are either $s+1$ people who all know each other, or $t+1$ people who do not know each other.

With $s = t = n-1$, this theorem generalizes the previous one.
It is clearly true for $s = 0$ or $t = 0$. We will prove it by induction on $s + t$.

Note

$$\binom{s+t}{s} = \binom{s+t-1}{s-1} + \binom{s+t-1}{s} = \binom{s+t-1}{s-1} + \binom{s+t-1}{t-1}.$$

Consider a company of $\binom{s+t}{s}$ people. Pick a person x, and let K be the set of those he knows, D the set he does not know. Then we have $|K| + |D| = \binom{s+t}{s} - 1$. One cannot have $|K| < \binom{s+t-1}{s-1}$ and $|D| < \binom{s+t-1}{t-1}$ since this would imply $|K| + |D| < \binom{s+t}{s} - 1$. So for example $|K| \geqslant \binom{s+t-1}{s-1}$. By the inductive assumption, K either contains a set D' of $t+1$ people who don't know each other or a set K' of s people who do. In the latter case, $\{x\} \cup K'$ is a set of $s+1$ people who know each other.
The case $|D| \geqslant \binom{s+t-1}{t-1}$ is similar.

Winning strategy in a game

Look at a typical game of strategy, say the Nim game.

- There are two players, Alice and Bob, and Alice starts.
- Players take turns, each making a move.
- Start with 3 piles of pennies, of sizes 10, 10, 10.
- A move means taking off some pennies.
- The player having to take off the last penny loses.

A strategy of a player is a a function $S : \mathbb{N}^3 \to \{1,2,3\} \times \mathbb{N}$. $S(n_1, n_2, n_3) = (i, k)$ says that if it is your turn and the piles have sizes n_1, n_2, n_3 then take off k from pile i. A strategy is winning if it leads to winning no matter what the other player does.

Theorem

In this game, either Alice has a winning strategy or Bob has one.

To prove this theorem, we generalize it to the set of all possible games in which the initial piles have sizes n_1, n_2, n_3, the starting player is X (may be Bob, too), the other player is Y.

Proposition

In the generalized game, either Alice has a winning strategy or Bob has one.

Let us prove the proposition by mathematical induction on $n = n_1 + n_2 + n_3$.

Base case: For $n = 1$, the starting player loses.

Induction step: Suppose that the proposition is true for all games where the sum of piles is $< k$, we will prove that it is also true for all games with the sum of piles equal to $k_1 + k_2 + k_3 = k$.

Ways to take off some coins: $m_1, \dots, m_m$. For example, move m_{15} says take off 5 from pile 1, resulting in $k_1, k_2 - 5, k_3$. New game, with these starting piles, starting player Y. Since here the sum is smaller, we already know that one player has a winning strategy. If move m_i gives a winning strategy for X then write $f(m_i) = X$, otherwise $f(m_i) = Y$.

Now if there is an i with $f(m_i) = X$ then X has a winning strategy: choose move m_i and follow that winning strategy from there. Otherwise Y has a winning strategy no matter what X does: to move m_i, just answer with the winning strategy of Y for the resulting new game.

Asymptotic analysis

We have learned some counting formulas, but in order to have a useful understanding of them, we should learn to estimate how they relate to each other. For this in many cases, we will need an approximate, simplified classification of functions according to how fast they grow.

- Compare n and $\binom{n}{2}$.
- Compare n^2 and 2^n.
- Compare 2^n and $n!$.
- Stirling's formula for $n!$ (without proof):

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

Later we will see how to find easily a weaker version of this.

The birthday (twin) paradox

There are 50 students in a class. What is the probability that two of them have the same birthday?

Assume a fixed (say alphabetic) order of the students. There are 365^{50} possible arrangements of birthdays (ignore the problem of February 29). It is reasonable to assume that these are all equally probable.

There are $365 \cdot 364 \cdots 316$ possible arrangements with no two equal birthdays. So the probability is

$$\frac{365 \cdot 364 \cdots 316}{365^{50}}.$$

It sounds daunting to compute this exactly, though nowadays the the program Mathematica spits back the answer 0.0296264 in no time:

```
In[6]:= Binomial[365, 50] * 50! / 365^50

Out[6]= 216450947969980945018737813684477840905760489196842\
        126408358251528094692173081574234555525510294790\
        233562316563021824 /
       7306010813549515310358093277059651246342214174497\
        508156711617142094873581852472030624097938198246\
        993124485015869140625

In[7]:= % // N

Out[7]= 0.0296264
```

An ordinary program will also compute it well, since the round-offs in the floating-point operations behave well here. But we want more insight than what is given by just a number.

More generally, we want to approximate

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

A useful trick when estimating products: take logarithm, then we will work with sums:

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

In analysis, it is always more practical to use natural logarithm ln, that is logarithm with base $e = 2.718\ldots$: $\ln x = \log_e x$.

Later we will prove the estimate: $\frac{x}{1+x} \leqslant \ln(1+x) \leqslant x$. Applying it here:

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

The other side, using $\frac{-i/n}{1-i/n} = \frac{-i}{n-i}$:

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

So we have, with $n = 365$, $k = 50$:

$$0.0207215 \approx e^{-\frac{k(k-1)}{2(n-k+1)}} \leqslant p \leqslant e^{-\frac{k(k-1)}{2n}} \approx 0.0348687.$$

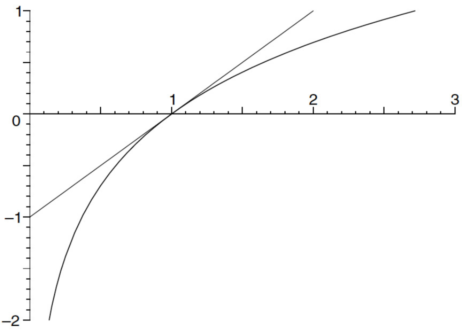
By this approximation, the probability of not having a common birthday is at most 3.5%.
This form is much more useful than the exact formula: it shows that the probability becomes $\approx 1/e$ when

$$k \approx \sqrt{n}.$$

So if there are n days in a year then among $\sqrt{n}$ people it is already likely to have a common birthday.

Estimating the logarithm

The inequality $\ln(1+x) \leqslant x$ is very important and comes from the concavity of the logarithm function:



This same inequality can be used to get a bound from the other side:

$$
\begin{aligned}
-\ln(1+x) = \ln\frac{1}{1+x} &= \ln\left(1-\frac{x}{1+x}\right) \leqslant -\frac{x}{1+x},\\
\ln(1+x) &\geqslant \frac{x}{1+x}.
\end{aligned}
$$

Combining the two estimates:

$$
\frac{x}{1+x} \leqslant \ln(1+x) \leqslant x.
$$

Strong and weak domination

Rough comparison of functions.
$f(n) \ll g(n)$ means $\lim_{n\to\infty} f(n)/g(n) = 0$: in words, $g(n)$ grows faster than $f(n)$. Other notation:

$$f(n) = o(g(n)) \Leftrightarrow f(n) \ll g(n).$$

Example: $n - 4 \gg 116\sqrt{n} + 80$. We may also write

$$116\sqrt{n} + 80 = o(n).$$

Generally, when we write $f(n) = o(g(n))$ then $g(n)$ has a simpler form than $f(n)$ (this is the point of the notation).

$f(n) \overset{*}{<} g(n)$ means $\sup_n f(n)/g(n) < \infty$, that is $f(n) \leqslant c \cdot g(n)$ for some constant c. Other (the common) notation:

$$f(n) = O(g(n)) \Leftrightarrow f(n) \overset{*}{<} g(n).$$

(The notation $\overset{*}{<}$ is mine, you will not find it in your books.)
This is a preorder. If $f \overset{*}{<} g$ and $g \overset{*}{<} f$ then we write $f \overset{*}{=} g$, $f = \Theta(g)$, and say that f and g have the same rate of growth.
Example: $n^2 - 5n$ and $100n(n+2)$ have the same rate of growth.
We can also write

$$100n(n+2) = O(n^2), \quad 100n(n+2) = \Theta(n^2).$$

On the other hand, $n + \sqrt{n} = O(n^2)$ but not $\Theta(n^2)$.

Important special cases:

- $O(1)$ denotes any function that is bounded by a constant, for example $(1+1/n)^n = O(1)$.
- $o(1)$ denotes any function that is converging to 0 as $n \to \infty$. For example, another way of writing Stirling's formula is

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}(1+o(n)).$$

Not all pairs of functions are comparable

Here are two functions that are not comparable. Let $f(n) = n^2$, and for $k = 0, 1, 2, \ldots$, we define $g(n)$ recursively as follows. Let n_k be the sequence defined by $n_0 = 1$, $n_{k+1} = 2^{n_k}$. So, $n_1 = 2$, $n_2 = 4$, $n_3 = 16$, and so on. For $k = 1, 2, \ldots$ let

$$g(n) = n_{k+1} \text{ if } n_k < n \leqslant n_{k+1}.$$

So, $g(n_k + 1) = n_{k+1} = 2^{n_k} = g(n_k + 1) = g(n_k + 2) = \cdots = g(n_{k+1})$. This gives $g(n) = 2^{n-1}$ for $n = n_k + 1$ and $g(n) = n$ for $n = n_{k+1}$. Function $g(n)$ is sometimes much bigger than $f(n) = n^2$ and sometimes much smaller: these functions are incomparable for $\ll, \overset{*}{<}$.

Some function classes

Important classes of increasing functions of n:

- Linear functions: (bounded by) $c \cdot n$ for arbitrary constant c.
- Polynomial functions: (bounded by) n^c for some constant $c > 0$, for $n \geqslant 2$.
- Exponential functions: those (bounded by) c^n for some constant $c > 1$.
- Logarithmic functions: (bounded by) $c \cdot \log n$ for arbitrary constant c. Note: If a function is logarithmic with $\log_2$ then it is also logarithmic with $\log_b$ for any b, since

$$\log_b x = \frac{\log_2 x}{\log_2 b} = (\log_2 x)(\log_b 2).$$

These are all equivalence classes under $\stackrel{*}{=}$.

Some simplification rules

- Addition: take the maximum, that is if $f = O(g)$ then $f + g = O(g)$. Do this always to simplify expressions. Warning: do it only if the number of terms is constant! This is wrong: $n + n + \cdots (n \text{ times}) \cdots + n \neq O(n)$.
- $f(n)^{g(n)}$ is generally worth rewriting as $2^{g(n)\log f(n)}$. For example, $n^{\log n} = 2^{(\log n)\cdot(\log n)} = 2^{\log^2 n}$.
- But sometimes we make the reverse transformation:

$$3^{\log n} = 2^{(\log n)\cdot(\log 3)} = (2^{\log n})^{\log 3} = n^{\log 3}.$$

The last form is the most meaningful, showing that this is a polynomial function.

Examples

$$n/\log\log n + \log^2 n \overset{*}{=} n/\log\log n.$$

Indeed, $\log\log n \ll \log n \ll n^{1/2}$, hence $n/\log\log n \gg n^{1/2} \gg \log^2 n$.

Order the following functions by growth rate:

$$
\begin{aligned}
&n^2 - 3\log\log n && \stackrel{*}{=} n^2,\\
&\log n/n,\\
&\log\log n,\\
&n\log^2 n,\\
&3 + 1/n && \stackrel{*}{=} 1,\\
&\sqrt{5n}/2^n,\\
&(1.2)^{n-1} + \sqrt{n} + \log n && \stackrel{*}{=} (1.2)^n.
\end{aligned}
$$

Solution:

$$
\begin{aligned}
\sqrt{5n}/2^n &\ll \log n/n \ll 1 \ll \log\log n\\
&\ll n/\log\log n \ll n\log^2 n \ll n^2 \ll (1.2)^n.
\end{aligned}
$$

Sums of series

You must know the following three sums:

Arithmetic series $1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$.

Geometric series $1 + q + q^2 + \cdots + q^{n-1} = \frac{1-q^n}{1-q}$.

Infinite geometric series If $|q| < 1$ then $1 + q + q^2 + \cdots = \frac{1}{1-q}$.

Simplification of sums

For rates of growth, the following is more important:

Geometric series grows as fast as its largest element:

$$6 + 18 + \cdots + 2 \cdot 3^n \stackrel{*}{=} 3^n$$

Even more true of series growing faster, say,

$$1! + 2! + \cdots + n! \stackrel{*}{=} n!.$$

Sum of n^c (for example arithmetic series) For rate of growth, replace each term with the maximal one:

$$2^2 + 5^2 + 8^2 + \cdots + (2 + 3n)^2 \stackrel{*}{=} (n + 1)(2 + 3n)^2 \stackrel{*}{=} n^3.$$

Even more true of a series growing slower:

$$\log n! = \log 2 + \log 3 + \cdots + \log n \stackrel{*}{=} n \log n.$$

Let us derive formally, say $1^2 + 2^2 + \cdots + n^2 \overset{*}{=} n^3$. The upper bound is easy.
Lower bound, with $k = \lceil n/2 \rceil$:

$$\begin{aligned} 1^2 + \cdots + n^2 &\geqslant k^2 + (k+1)^2 + \cdots + n^2 \\ &\geqslant (n/2 - 1)(n/2)^2 \overset{*}{=} n^3. \end{aligned}$$

Infinite series

We will prove the following, via rough estimates:

$$1/3 + 2/3^2 + 3/3^3 + 4/3^4 + \cdots < \infty.$$

Since any exponentially growing function grows faster than the linear function, we know $n \overset{*}{<} 3^{n/2}$. Therefore $n \cdot 3^{-n} \overset{*}{<} 3^{n/2} \cdot 3^{-n} = 3^{-n/2}$, and the whole sum is

$$\overset{*}{<} 1 + q + q^2 + \cdots = \frac{1}{1-q}$$

where $q = 3^{-1/2}$.

Another example:

$$1 + 1/2 + 1/3 + \cdots + 1/n = \Theta(\log n).$$

Indeed, for $n = 2^{k-1}$, upper bound:

$$\begin{aligned}&1 + 1/2 + 1/2 + 1/4 + 1/4 + 1/4 + 1/4 + 1/8 + \cdots \\ &\qquad = 1 + 1 + \cdots + 1 \text{ (} k \text{ times).}\end{aligned}$$

Lower bound:

$$\begin{aligned}&1/2 + 1/4 + 1/4 + 1/8 + 1/8 + 1/8 + 1/8 + 1/16 + \cdots \\ &\qquad = 1/2 + 1/2 + \cdots + 1/2 \text{ (} k \text{ times).}\end{aligned}$$

Distributing presents

There are k children and n presents. We give n_1 presents to the first child, n_2 to the second one, and so on. How many ways?

$$\frac{n!}{n_1!n_2!\cdots n_k!}.$$

Interesting special cases:

- $n = k$, $n_1 = n_2 = \cdots = n_k = 1$.
- $n_1 = n_2 = \cdots = n_{k-1} = 1$, $n_k = n - k + 1$.
- $k = 2$.
- $n = 2k$, $n_1 = n_k = \cdots = n_k = 2$.

An equivalent problem: anagrams.

Distributing money

Distribute m pennies to k children, each must get at least 1.
Solution:

$$\underbrace{\text{¢ ¢ ¢}}_{\text{Alice}} \,|\, \underbrace{\text{¢ ¢}}_{\text{Bob}} \,|\, \underbrace{\text{¢ ¢}}_{\text{Carl}} \,|\, \underbrace{\text{¢ ¢ ¢ ¢ ¢}}_{\text{Diane}}$$

Dividing lines show which pennies go to which children: we give m_i presents to child i.

$$\binom{m-1}{k-1}$$

ways to place the lines.

A different problem: there are n children and k presents. Some children are allowed to get nothing.

- Lend them 1 each and take it back at the end. This reduces the problem to the previous one with $m = n + k$:

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

  More detail: a distribution $(n_1, n_2, \dots, n_k)$ with no restriction is in 1-1 correspondence with distribution

$$(m_1, m_2, \dots, m_k) = (n_1 + 1, n_2 + 1, \dots, n_k + 1)$$

  restricted to $m_i \geqslant 1$.
- Another way: the n pennies and $k - 1$ dividing lines come in arbitrary order, so there are $\binom{n+k-1}{k-1}$ possibilities.

Binomial coefficients

The binomial theorem

Let us see some more uses and properties of the binomial coefficients.

The binomial theorem (say, for the case of power 5):

$$
\begin{aligned}
(x+y)^4 &= (x+y)(x+y)(x+y)(x+y) \\
&= \binom{4}{0}x^4 + \binom{4}{1}x^3y + \binom{4}{2}x^2y^2 + \binom{4}{3}x^1y^3 + \binom{4}{4}y^4.
\end{aligned}
$$

In $(x+y)^n$, each term $x^{n-k}y^k$ corresponds to a set $A \subseteq \{1,\dots,n\}$ of size k in which we choose y from the ith bracket if $i \in A$. This is a one-to-one correspondence, so there are $\binom{n}{k}$ such terms.

Some uses:

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

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

Identities in Pascal's Triangle

New proof of $\binom{n}{0} - \binom{n}{1} + \binom{n}{2} + \dots = 0$ reveals more. Represent each term as the sum of the two terms above it in the triangle:

$$\begin{aligned} &\binom{n}{0} && -\binom{n}{1} && +\binom{n}{2} && -\cdots \\ =&\left(0+\binom{n-1}{0}\right) && -\left(\binom{n-1}{0}+\binom{n-1}{1}\right) && +\left(\binom{n-1}{1}+\binom{n-1}{2}\right) && -\cdots \end{aligned}$$

This gives more:

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

Another interesting indentity:

$$\binom{n}{0}^2 + \binom{n}{1}^2 + \cdots + \binom{n}{n}^2 = \binom{2n}{n}.$$

Combinatorial interpretation easier if writing it as

$$\binom{n}{0}\binom{n}{n} + \binom{n}{1}\binom{n}{n-1} + \binom{n}{2}\binom{n}{n-2} + \cdots = \binom{2n}{n}.$$

Let $A = \{1, \dots, n\}$, $B = \{n+1, \dots, 2n\}$. Then a subset C of size n of $A \cup B$ can be written as the disjoint union $C = (C \cap A) \cup (C \cap B)$. For each $0 \leqslant k \leqslant n$, there are $\binom{n}{k}$ ways to choose C with $|C \cap A| = k$ and $\binom{n}{n-k}$ ways still to choose $C \cap B$.

Multinomial theorem

There is a multinomial theorem, analogous to the binomial theorem, an expression for $(x_1 + x_2 + \cdots + x_k)^n$.

- How many terms does this have after expansion?
- What does each term look like?

Example:

$$(x_1 + x_2 + x_3)^n = \sum_{n_1+n_2+n_3=n} \frac{n!}{n_1!n_2!n_3!} x_1^{n_1} x_2^{n_2} x_3^{n_3}.$$

There are $\binom{n+k-1}{k-1} = \binom{n+2}{2}$ terms. Another way of writing the sum is as

$$\sum_{n_1=0}^{n} \sum_{n_2=0}^{n-n_1} \frac{n!}{n_1!n_2!(n-n_1-n_2)!} x_1^{n_1} x_2^{n_2} x_3^{n-n_1-n_2}.$$

Diagonal sums

The elements of each diagonal are the sums of the elements of the previous diagonal. For example:

$$\binom{3}{3} + \binom{4}{3} + \cdots + \binom{n}{3} = \binom{n+1}{4},$$

$$1\cdot 2\cdot 3 + 2\cdot 3\cdot 4 + \cdots + (n-2)(n-1)n = \frac{(n-2)(n-1)n(n+1)}{4}.$$

These are the left-sloping diagonals. The right-sloping diagonals give other, also interesting, identities.

Recursive equations

Leonardo of Pisa ("Fibonacci", 13th century):

> *A farmer raises rabbits. (We only count female rabbits, just assuming that enough males exist.) Each rabbit gives birth to one rabbit when she turns 2 months old, and then to one more each month thereafter. How many rabbits will the farmer have in the nth month if he starts with one newborn?*

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

If there are F_n rabbits at month n, then we get

$$F_1 = F_2 = 1, \tag{1}$$

$$F_{n+1} = F_n + F_{n-1}, \tag{2}$$

This is a recursive definition, or recurrence, an algorithm for computing F_n, but not a simple formula. Equations (1) give the initial conditions.

Other problem leading to the same recursive equation:

> *A staircase has n steps. You walk up taking one or two steps at a time. How many ways can you go up?*

Let J_n be the number of ways. We have

$$J_1 = 1, \quad J_2 = 2,$$
$$J_{n+1} = J_n + J_{n-1}.$$

The recursive part is the same, the initial conditions are slightly different, we get

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

so $J_n = F_{n+1}$.

Defining $F_0 = 0$ keeps the equation valid. Experimentation discovers the relation:

$$F_0 + F_1 + \cdots + F_n = F_{n+2} - 1.$$

Once we discovered it, proving by induction is not hard. A more complicated case is the following pair of equations:

$$\begin{aligned} F_n^2 + F_{n-1}^2 &= F_{2n-1}, \\ F_{n+1}F_n + F_nF_{n-1} &= F_{2n}. \end{aligned}$$

Each by itself is difficult to prove by induction, but we can prove the two simultaneously.

Other initial conditions

Look at the sequence

$$E_0 = A, \quad E_1 = B,$$
$$E_{n+1} = E_n + E_{n-1}. \tag{3}$$

We can guess and prove by induction the formula

$$E_n = F_{n-1}A + F_nB.$$

We will see an easier way to prove this formula based on linearity. But first, a beautiful consequence, if we substitute $A = F_a$, $B = F_{a+1}$:

$$F_{a+b+1} = F_aF_b + F_{a+1}F_{b+1}.$$

Solving the recurrence

Experimentation suggests that F_n grows exponentially, moreover, F_{n+1}/F_n converges to a limit.
Idea: find a geometric progression satisfying the same recurrence:

$$c \cdot \varphi^{n+1} = c \cdot \varphi^n + c \cdot \varphi^{n-1},$$
$$\varphi^2 = \varphi + 1.$$

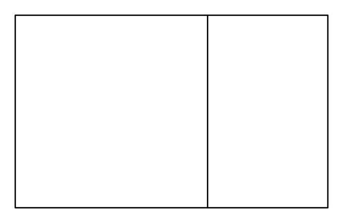
Solution: $\varphi_1 = \frac{1+\sqrt{5}}{2} = 1.618034$, $\varphi_2 = \frac{1-\sqrt{5}}{2} = -0.618034$. The equation can also be written as

$$\varphi = 1 + 1/\varphi.$$

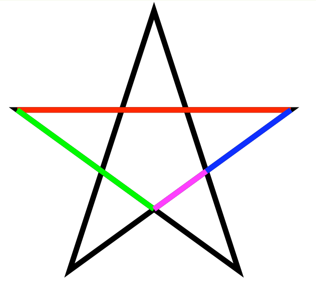
In this form, it is known as the equation of the golden ratio, a proportion with special significance for geometry, art and even natural history, since classic Greek times.

From Wikipedia

Euclid: "A straight line is said to have been cut in extreme and mean ratio when, as the whole line is to the greater segment, so is the greater to the less."



The line segments of various colors in the figure below are related by the golden ratio.



We have found many solutions to the recurrence: $c_1\varphi_1^n$, and $c_2\varphi_2^n$, for arbitrary c_1, c_2. But notice that the recurrence equation

$$F_{n+1} = F_n + F_{n-1}$$

is linear: if $X_1, X_2, \dots$ is a solution and $Y_1, Y_2, \dots$ is a solution then $X_1 + Y_1, X_2 + Y_2, \dots$ is also a solution. So we can look for a solution in form of

$$c_1\varphi_1^n + c_2\varphi_2^n.$$

The initial conditions require $c_1 + c_2 = 0$, $c_1\varphi_1 + c_2\varphi_2 = 1$. The first one gives $c_2 = -c_1$. Using it for the second one:

$$c_1(\varphi_1 - \varphi_2) = c_1\sqrt{5} = 1.$$

So, $c_1 = 5^{-1/2}$, giving the formula

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

Graphs

A graph from now on is similar to the diagram of a relation $E \subseteq V \times V$. (It has nothing to do with the graph of a function.) It is undirected if the relation is symmetric. But it is more convenient to introduce graphs as a new kind of objects. We start with undirected graphs.

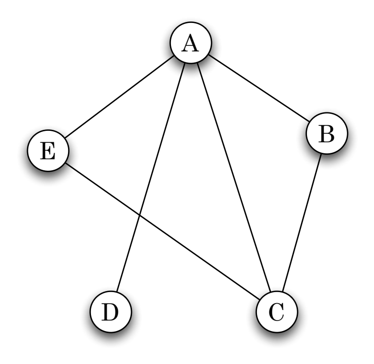
Example

In a group of 51 people, show that there is somebody who know an even number of others. More generally, this is true of any group of an even number of people.

Represent each person by a point (vertex, node), acquaintance between any pair by a line or edge. Graph

$$G = (V, E),$$

where V is the set of vertices, E is the set of edges, our (symmetric) relation. So $\{u, v\} \in E$ if persons u, v are acquainted, if there is an edge between u and v.



(On the drawing, the crossing of two edges is not a node if not marked as such.)

The degree $d(v)$ of a node v is the numbef of edges leaving (or entering, this is the same now) a node. So A has degree 4, C has degree 3, as Alice knows 4 people, Carl knows 3.

- An edge entering (leaving) a node is said to be incident on the node.
- Two nodes are adjacent, or neighbors if they are connected by an edge.
- A loop edge is an edge from a node to itself.
- Parallel edges are several edges going between the same pair of nodes.
- If we do not allow loop edges or parallel edges, we speak of a simple graph, otherwise of a multigraph.

Going back to the party problem, let us add up the degrees of all the nodes.

Theorem

In a graph with vertex set V, the sum of all degrees $\sum_{v \in V} d(v)$ is twice the number of edges.

Proof. Each edge contributes 2 to this sum, at its two ends. □

Corollary

In a graph, the number of nodes with odd degree is even.

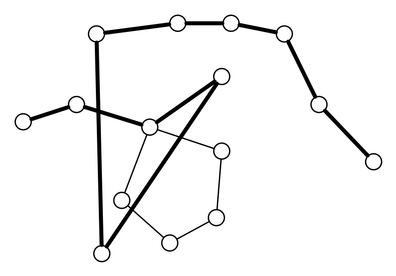
It follows that if the graph has an odd number of nodes, then the set of nodes with even degree has odd size, and so is nonempty. This proves the original statement about the group of 51 people.

Paths, cycles

Some special graphs:

- The complete graph K_n, or clique, the edgeless graph or anticlique.
- The complement $\overline{G}$ of a graph G.
- A star, with $n-1$ edges on n nodes.
- A subgraph $G' = (V', E')$ of a graph has $V' \subseteq V$, $E' \subseteq E$, (of course, all edges of E' are between nodes of V').

- A path, with $n-1$ edges on n nodes, with length $n-1$. Its endpoints.
- A cycle of length k called a k-cycle.
- A walk in G is a sequence of nodes $v_0, v_1, \dots, v_k$ where v_i, v_{i+1} are adjacent for all i. It is like a path or cycle, but may self-intersect. "Cut out the loops" to convert it into a path or cycle with the same endpoints.



Connectivity

An equivalence relation on nodes of a graph G: two points are connected if some walk connects them. (Allow the trivial walk consisting of one point.) We get the same relation requiring that some path connect them.

- Equivalence classes: connected components. The graph is connected if it has only one component.
- If you have a graph and want to show it is connected, typically you need to find a point and show that it has paths to all other points.
- If you have a graph and want to show it is not connected, typically you need to split the graph into two subsets and show that there are no edges between them.

Trees

A tree is a connected graph with no cycles. The following theorem shows that it is also a maximal graph with no cycles and a minimal connected graph.

Theorem

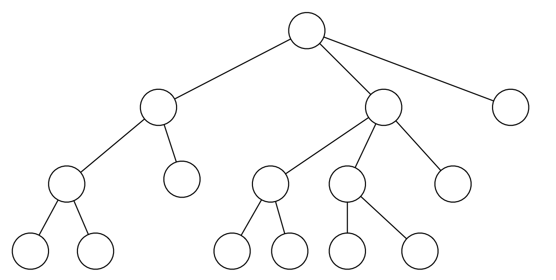
The following are two other ways to characterize trees.

a. *A graph is a tree if and only if it is connected, but deleting any of its edges results in a disconnected graph.*
b. *A graph is a tree if and only if it contains no cycles, but adding any new edge creates a cycle.*

- Spanning trees.
- Cut edges.
- Forests.

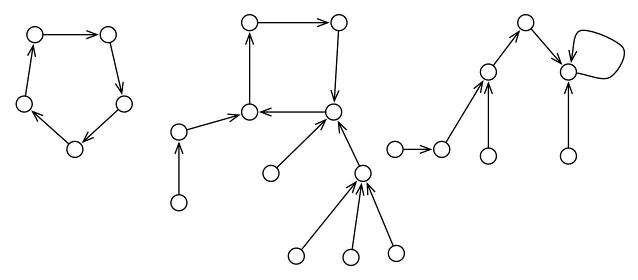
Rooted trees

A rooted tree has a distinguished node, the root. It is generally drawn with the root on top:



Father, sons, internal nodes, leaves.

Every function $f : V \to V$ gives rise to a directed graph, with edges directed from x to $f(x)$.



Directed cycles, with upwards directed trees rooted on them.

Growing trees

Theorem

Every tree with at least 2 nodes has at least two nodes of degree 1.

Tree-growing procedure

- Start with a single node.
- Repeat any number of times: Create a new node and connect it by a new edge to any existing node.

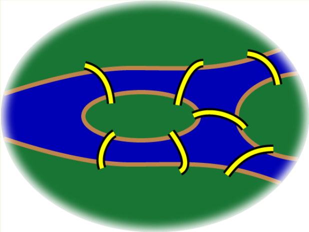
Theorem

Every graph obtained by the Tree-growing Procedure is a tree, and every tree can be obtained this way (thus has $n-1$ edges on n nodes).

Useful for proving facts about trees by induction.

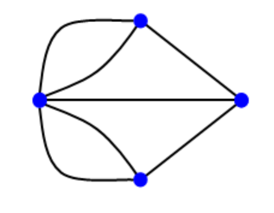
Euler walks

The Königsberg bridges (from Wikipedia):



Is there a walk passing through all the bridges exactly once?

Euler's solution relies on a (multi) graph (without saying so).



He noticed that in the graph of a desired walk all nodes except possibly the start and the end would have even degrees.

Euler walk: a walk passing through each edge of the (multi)graph exactly once.

Theorem

Consider a connected (multi)graph $G = (V, E)$.

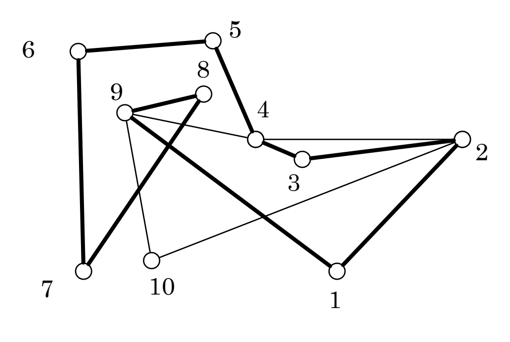
- (a) *More than 2 nodes with odd degree: no Euler walk.*
- (b) *Exactly 2 nodes with odd degree: there is an Euler walk starting at one of these and ending at another.*
- (c) *No nodes with odd degree: there are Euler walks, all these are closed.*

We proved (a). Let us prove (c).

Euler stroll: like a closed Euler walk, but does not have to pass through all edges of the graph.

1. The set of edges of an Euler graph is the disjoint union of some closed Euler strolls. This remains true even if the graph is not connected.
2. Any two strolls C_i, C_j having a common point can be replaced with one stroll covering the same edges.
3. Continue this process of replacement as long as you can. At the end, only a single stroll remains, since the original graph is connected.

Combining two strolls. The first one is $(1, 2, 3, 4, 5, 6, 7, 8, 9, 1)$, the second one is $(10, 2, 4, 9, 10)$.



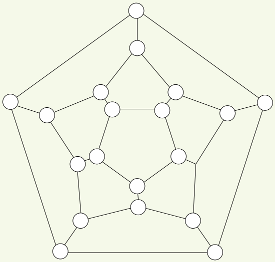
Inserting the second stroll into the first one at the point of first meeting:

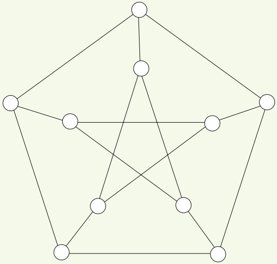
$$(1, 2, 4, 9, 10, 2, 3, 4, 5, 6, 7, 8, 9, 1).$$

Hamilton cycles

A cycle that contains all nodes of a graph.

It is much harder to decide whether a graph has a Hamilton cycle than whether it has an Euler walk. Examples from LPV:





I will not spoil your fun of figuring these out on your own.

Matchings

Example

At a dance party, with 300 students, every boy knows 50 girls and every girl knows 50 boys. Can they all dance simultaneously so that only pairs who know each other dance with each other?

- Bipartite graph: left set A (of girls), right set B (of boys).
- Matching, perfect matching.

Theorem

If every node of a bipartite graph has the same degree $d \geqslant 1$ then it contains a perfect matching.

Examples showing the (local) necessity of the conditions:

- Bipartiteness is necessary, even if all degrees are the same.
- Bipartiteness and positive degrees is insufficient.

Example

6 tribes partition an island into hunting territories of 100 square miles each. 6 species of tortoise, with disjoint habitats of 100 square miles each.
Can each tribe pick a tortoise living on its territory, with different tribes choosing different totems?

Bipartite graph: left set A of tribes, right set B of tortoises. For $S \subseteq A$ let

$$\mathbf{N}(S) \subseteq B$$

be the set of all neighbors of the nodes of A. Special property:
For every $S \subseteq A$ we have $|\mathbf{N}(S)| \geqslant |S|$.
Indeed, the combined hunting area of any k tribes intersects with at least k tortoise habitats.

Example (Workers and jobs)

Suppose that we have n workers and n jobs. Each worker is capable of performing some of the jobs. Is it possible to assign each worker to a different job, so that workers get jobs they can perform?

Theorem (The Marriage Theorem)

A bipartite graph has a perfect matching if and only if $|A| = |B|$ *and for every* $S \subseteq A$ *we have* $|\mathbf{N}(S)| \geqslant |S|$.

The condition is necessary.

Proposition

The condition implies the same condition for all $S \subseteq B$.

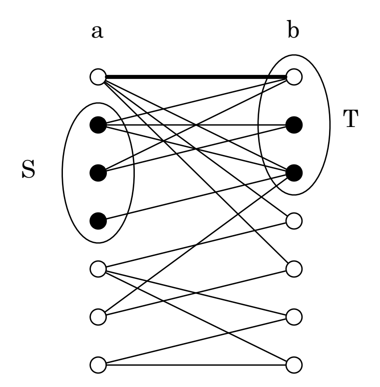
Proof

- Good graph: one that satisfies the conditions.
- A good graph of size 2 clearly has a matching.
- Plan: partition any good graph of size > 2 into two smaller good graphs.
- Try partitioning into an edge (a, b) and the remaining graph on $A \setminus \{a\}$, $B \setminus \{b\}$. If the graph on $(A \setminus \{a\}) \cup (B \setminus \{b\})$ is good, we are done.

- Else there is an

  $$S \subseteq A \setminus \{a\}, \quad b \in \mathbf{N}(S) =: T,$$
  $$|S| = |T|.$$

- Then partition into $S \cup T$, $(A \setminus S) \cup (B \setminus T)$.



Goodness follows from:

- $\forall S' \subseteq S \; |\mathbf{N}(S')| \geqslant |S'|$ (by the goodness of G).
- $\forall U' \subseteq B \setminus T \; |\mathbf{N}(U')| \geqslant |U'|$ (by the Proposition).

Generalization

A matching is any (possibly empty) set of disjoint edges. Let us abandon the condition $|A| = |B|$: we still get a theorem.

Theorem

A bipartite graph has a matching that covers each node of A if and only if for every $S \subseteq A$ we have $|\mathbf{N}(S)| \geqslant |S|$.

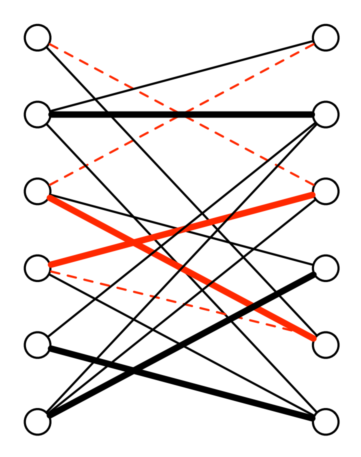
Proof. We will reduce the problem to the original one.
The condition implies $|B| \geqslant |A|$, assume $|B| > |A|$. Let us add $|B| - |A|$ new points to A to get the set A'. Connect each new point to every point of B. The Marriage Theorem implies that the new graph has a matching. Deleting the points of $A' \setminus A$ solves the original problem. □

Finding a perfect matching

- The proof of the Marriage Theorem that we gave is non-constructive. It just shows that if there is no matching, somebody could in principle convince us simply, by showing a set whose shadow is smaller.
- Let us now search for a method to find a perfect matching if it exists. A matching M is any set of disjoint edges.
- Greedy matching method: just keep adding edges to M as long as we can. We may get stuck with a maximal (unextendable) matching that is not perfect, does not have the maximum number of edges.

Augmenting paths

New way to increase the size of a matching M:



- Alternating path: alternates on M and non-M edges.
- Augmenting path: alternating path that starts in A, ends in B, both outside M. To augment, switch the M and non-M edges.

Lemma

If M is not perfect and has no augmenting path, there is no perfect matching.

Proof.

- U := the unmatched points of A.
- Almost augmenting path: alternating path of even size starting in U.
- $S^* \subseteq A$:= the points reachable from U on almost augmenting paths.
- $T^* \subseteq B$:= the points matched to those of S^*.

Then $|S^*| = |T^*| + |U|$, and $T^* = \mathbf{N}(S^*)$. □

Algorithm 12.1: Augment a matching M

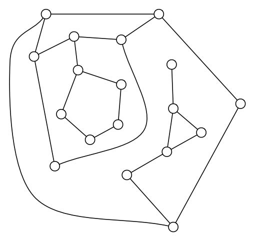
Will gradually build set S reachable on almost augmenting paths, T = the points matched to those of S, and function $f : S \setminus U \to S$ where $f(s)$ = previous point of S (“father”) on the almost augmenting path.

$S \leftarrow U$, $T \leftarrow \emptyset$, $f \leftarrow$ the empty function
while not stopped **do**
    Look for an edge sr between $s \in S$ and $r \in B \setminus T$
    if there is none **then**
        M is a maximum matching, **return**
    else if r is unmatched **then**
        find an augmenting path P from r, s to U using $f(\cdot)$
        apply P to increase M, **return**
    else
        r is matched to a $q \in A$
        $S \leftarrow S \cup \{q\}$, $T \leftarrow T \cup \{r\}$, $f(q) \leftarrow s$

Planar graphs

A graph can sometimes be drawn in the plane, with non-intersecting (possible curved or broken) lines representing the edges. (Consider just connected graphs.) This drawing divides the plane into connected regions that we can call countries and those edges that are between different regions as borders. Using a different terminology, we will sometimes call the regions faces (in analogy with the faces of a polyhedron).

Each country can be described as follows: walk around while having it always on your left, list the edges. (One edge may be listed twice, if passed in different directions.)



Let f = number of countries, e = number of edges, v = number of nodes.

Theorem (Euler)

$f - e + v = 2.$

Proof. View the edges as dams, the infinite country outside as the ocean. Remove, one-by-one, dams connecting dry land with water. Since a country is connected, each dam removal floods one country. We end up with a tree and a single country (water) by the time we removed $f - 1$ edges. The remaining graph is connected since we always removed an edge from some cycle. So it is a tree, with $v - 1$ edges:

$$e = (f - 1) + (v - 1).$$

□

Application

Proposition

In a planar graph of n points, there are at most $3n-6$ edges.

As an application, we get that the graph K_5 is not planar.

Proof. Each country has at least 3 boundary edges, so we have (counting each edge twice)

$$3f \leqslant 2e, \quad f \leqslant 2e/3.$$

Substituting into Euler's formula:

$$2 = v - e + f \leqslant v - e + 2e/3,$$
$$e \leqslant 3v - 6.$$

□

Other ways of showing nonplanarity

Let $A \subseteq \mathbb{R}^2$ be a subset of the plane. We will call two points $p, q \in A$ equivalent if they can be connected by a curve running inside A. (In this class, we will not define the notion of the curve precisely.)

Theorem (Jordan)

Let $\mathbb{R}^2$ be the plane and let $C \subseteq \mathbb{R}^2$ be a simple closed curve. Then $\mathbb{R}^2 \setminus C$ consists of two equivalence classes: a bounded set of points I inside C and the rest: the set of points O outside C.

- We used this theorem implicitly in the notion of a face, which is the inside set of its boundary curve.
- Try to use the theorem to give a direct proof of the fact that K_5 is not planar.

Polyhedra

- Consider polyhedra that can be blown up into a ball:

  tetrahedron, cube, octahedron, dodecahedron, icosahedron, triangular prism, etc.

  These are all convex, but some non-convex ones would also qualify, say a box (remove a smaller cube from one side of a bigger cube). The graph of edges and vertices is always planar, and so each connected part obeys Euler's formula.
- On the other hand, if the polyhedron has a hole passing through (like a window frame) its graph is not planar. (See exercise.)

Regular polyhedra

In the plane, for every n there is a regular n-gon: triangle, square, regular pentagon, hexagon, and so on. In space, we can consider regular polyhedra instead:

- All faces are congruent to each other (in particular, have the same number of edges).
- All corners are congruent to each other (in particular, have the same degree).

Some known regular polyhedra:

tetrahedron, cube, octahedron, dodecahedron, icozahedron.

The ancient Greeks have recognized and proved that there are no other regular polyhedra.
Euler derived this from just the following properties of the surface graph:

- planarity (used only via Euler's formula)
- every face has the same number of edges
- every vertex has the same number of edges (degree).

Counting labeled trees

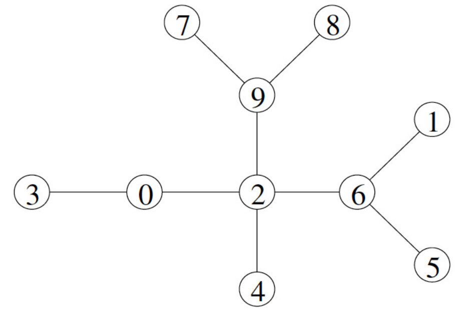
How to represent a tree $T = (V, E)$ on the set $V = \{0, 1, \ldots, n-1\}$, in the computer? Declare point 0 the root. All points $x \neq 0$, have a parent $f(x)$.

One possible representation:

1	2	...	$n-1$
$f(1)$	$f(2)$	...	$f(n-1)$

For example, the tree on the right is represented by

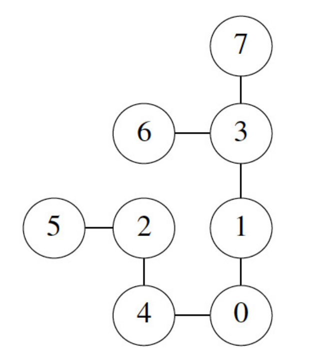
1	2	3	4	5	6	7	8	9
6	0	0	2	6	2	9	9	2



The upper row can be deleted. There are n^{n-1} possible lower rows, this gives an upper bound on the number of trees. But some of these lower rows are illegal. For example $(2, 1)$ is an illegal second row, giving a cycle.
A clever rearrangement of the table, the Prüfer code:

1	3	4	5	6	7	8	9	2
6	0	2	6	2	9	9	2	0

We repeatedly took the leaf with the smallest label for the next column, and deleted it to get a smaller tree. The last column is unnecessary, we know it has 0 at its bottom. It is less obvious that we can figure out the whole top row from the bottom row.



5	2	4	6	7	3	1
2	4	0	3	3	1	0

The first element of the top row is the smallest label missing from the bottom. In general, each element in the top row is the smallest label different from the ones to the left in the top row, and from the ones below it and to the right in the bottom row.

Deleting the top row and the last column, we get the reduced Prüfer code. The number of reduced Prüfer codes is

$$n^{n-2},$$

but then this is also the number of complete Prüfer codes.

Claim

Each Prüfer code belongs to a tree.

Proof. It is sufficient to prove that in the obtained graph, we get to the root from any node by repeating the parent operation. Each parent operation either leads us to the root, or to a nonroot bottom label. A nonroot bottom label has not appeared on the top left yet, so must appear on the top right. Repeating the parent operation this way, we end up at the root. □

We thus proved the famous

Theorem (Cayley)

The number of labeled trees is n^{n-2}.

Unlabeled trees? This is a more complex question, only approximate answers are known.