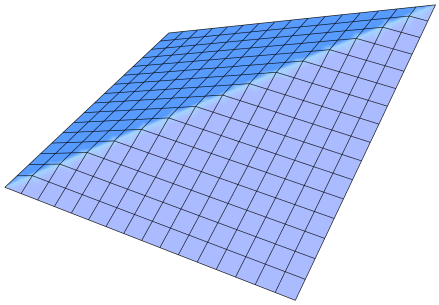


MATH 249: Introduction to Combinatorics (Advanced)

Prof. Kevin Purbhoo • Fall 2015 • University of Waterloo
MWF 1:30–2:20PM in HH 139

Transcribed, LaTeXed, and edited by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is https://rob-zimmerman.github.io/coursenotes.

Contents

Lecture 1 — Monday, September 14, 2015

1. Counting Principles and Binomial Identities

The Twelvefold Way

We begin with the following basic question: how many ways are there to place k balls in n boxes? We can impose restrictions on this problem, such as requiring at most one ball per box or at least one ball per box. We can also decide whether the balls and boxes are distinguishable. These choices lead to the **twelvefold way**.

Each way to place the balls is a function $f : K \to N$, where $|K| = k$ and $|N| = n$. Restrictions on the placement correspond to special conditions on the function:

Condition 1: No restriction on f.

Condition 2: f is injective.

Condition 3: f is surjective.

Distinguishability can be encoded as equivalence relations on functions $f : K \to N$.

Case A: We have $f \sim g$ if and only if $f = g$.

Case B: We have $f \sim g$ if and only if there exists a bijection $\alpha : K \to K$ such that $f = g \circ \alpha$.

Case C: We have $f \sim g$ if and only if there exists a bijection $\beta : N \to N$ such that $f = \beta \circ g$.

Case D: We have $f \sim g$ if and only if there exist bijections $\alpha : K \to K$ and $\beta : N \to N$ such that $f = \beta \circ g \circ \alpha$.

These equivalence relations describe which relabelings of balls and boxes are ignored.

Example 1.1 In case 3A with $k = 3$ and $n = 2$, the possible placements include those shown in Figure 1.

FIGURE 1

1 2 3	
2 3	1
1 3	2
1 2	3

In **Case 1A**, there is no restriction and both the balls and boxes are distinguishable. There are n^k placements, since there are n choices for each of the k balls.

In **Case 2A**, there is at most one ball per box and both the balls and boxes are distinguishable. There are

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

placements, since there are n choices for the first ball, $n-1$ choices for the second ball, and so on.

Definition 1.2

$$x^{\underline{k}} := \prod_{i=0}^{k-1}(x-i) = x(x-1)(x-2)\cdots(x-k+1) \qquad \text{is the } \textbf{falling factorial}.$$

$$x^{\overline{k}} := \prod_{i=0}^{k-1}(x+i) = x(x+1)(x+2)\cdots(x+k-1) \qquad \text{is the } \textbf{rising factorial}.$$

$$k! := k^{\underline{k}} \qquad \text{is just the } \textbf{factorial}.$$

Lecture 2 — Wednesday, September 16, 2015

In **Case 2C**, there is at most one ball per box and the boxes are indistinguishable. The answer is 1 if $k \le n$, and 0 otherwise. This is because the only way to place k balls in n indistinguishable boxes with at most one ball per box is to put one ball in each of k boxes, and this is only possible if $k \le n$.

Case 2D has the same answer as Case 2C: once both the balls and boxes are indistinguishable, an injective placement still consists of one ball in each of k boxes.

Definition 2.1 If k is a nonnegative integer, then the **binomial coefficient** is

$$\binom{x}{k} := x^{\underline{k}}/k!.$$

If $x = n \in \mathbb{N}$ and $0 \le k \le n$, then

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

For $k > n$, the falling-factorial definition gives $\binom{n}{k} = 0$.

In **Case 2B**, there is at most one ball per box and the balls are indistinguishable.

Theorem 2.2 The answer to Case 2B is $\binom{n}{k}$.

Proof. Let N and K be sets with $|N| = n$ and $|K| = k$. Let

$$X = \{f : K \to N : f \text{ is injective}\}.$$

Let $f \sim g$ if there exists a bijection $\alpha : K \to K$ such that $f = g \circ \alpha$. Let Y be the set of equivalence classes of $\sim$. Then $|Y|$ is the answer to the problem (this is the standard way to begin a **combinatorial proof**). From Case 2A, we have $|X| = n^{\underline{k}}$. Also

$$|X| = \sum_{C \in Y} |C|$$

since Y is the set of disjoint classes whose union is X. Now suppose that $C \in Y$ and $g \in C$. Then

$$C = \{f \in X : f = g \circ \alpha \text{ for some bijection } \alpha\} = \{g \circ \alpha : \alpha : K \to K \text{ is bijective}\}.$$

Since g is injective, $g \circ \alpha = g \circ \alpha'$ if and only if $\alpha = \alpha'$. Thus $|C|$ is the number of bijections $K \to K$, which is $k!$. Putting these facts together,

$$n^{\underline{k}} = |X| = \sum_{C \in Y} |C| = k! \cdot |Y|,$$

and therefore

$$|Y| = \frac{n^{\underline{k}}}{k!} = \binom{n}{k},$$

as required. □

The interpretation here is that $\binom{n}{k}$ is the number of k-element subsets of an n-element set.

Example 2.3 If $n = 2$ and $k = 4$, then there are five k-element multisets of $N = \{a, b\}$. They are

$$\{a,a,a,a\}, \quad \{a,a,a,b\}, \quad \{a,a,b,b\}, \quad \{a,b,b,b\}, \quad \text{and} \quad \{b,b,b,b\}.$$

Notation 2.4 Define

$$\left(\binom{n}{k}\right) := \frac{n^{\overline{k}}}{k!} = \binom{n+k-1}{k}.$$

This is occasionally called the **multiset coefficient**.

In **Case 1B**, there is no restriction and the balls are indistinguishable. This is equivalent to

counting k-element multisets from an n-element set.

Theorem 2.5 The answer to Case 1B is $\left(\binom{n}{k}\right)$.

Proof. Assume $N = \{1, 2, \ldots, n\}$. Let X be the set of k-element multisets of N. Let Y be the set of k-element subsets of $\{1, 2, \ldots, n+k-1\}$. Since we know $|Y| = \binom{n+k-1}{k}$, we can prove the result by finding a bijection $f : X \to Y$. If $A \in X$, we can write $A = \{a_1, \ldots, a_k\}$ where

$$a_1 \leq a_2 \leq \cdots \leq a_k.$$

Define $f(A) = \{a_1, a_2 + 1, a_3 + 2, \ldots, a_k + k - 1\}$.

Since $a_i \leq a_{i+1}$, we have

$$a_i + (i-1) < a_{i+1} + i,$$

so $f(A)$ has k distinct elements; also $1 \leq a_1$ and $a_k + k - 1 \leq n + k - 1$, hence $f(A) \in Y$. Define $g : Y \to X$ by

$$g(\{b_1 < \cdots < b_k\}) = \{b_1,\ b_2 - 1,\ \ldots,\ b_k - (k-1)\},$$

interpreted as a multiset. The sequence $b_1,\ b_2 - 1, \ldots, b_k - (k-1)$ is nondecreasing, so $g(Y) \subseteq X$. Directly,

$$g(f(A)) = A \qquad \text{and} \qquad f(g(B)) = B,$$

so f is bijective. □

Remark 2.6 To show that a function is bijective, it is often enough to construct an inverse explicitly and check that the two composites are identity maps.

In **Case 3B**, there is at least one ball per box and the balls are indistinguishable. This is equivalent to counting **compositions** of an integer.

How many ways are there to write $k = c_1 + \cdots + c_n$, where $c_1, \ldots, c_n$ are positive integers? Here c_i is the number of balls in the ith box. Case 1B can be interpreted similarly, but with nonnegative parts.

In **Case 3D**, there is at least one ball per box and both the balls and boxes are indistinguishable. This is equivalent to counting **partitions** of an integer.

How many ways are there to write $k = c_1 + \cdots + c_n$, where $c_1, \ldots, c_n$ are positive integers and

$$c_1 \geq c_2 \geq \cdots \geq c_n?$$

Here c_i is the number of balls in the box with the ith largest number of balls. Case 1D is analogous, with nonnegative parts.

Lecture 3 — Friday, September 18, 2015

In **Case 3A**, there is at least one ball per box and both the balls and boxes are distinguishable. This is equivalent to counting **surjective functions**.

In **Case 3C**, there is at least one ball per box and the boxes are indistinguishable. This is equivalent to counting **set partitions**.

Example 3.1 $\{\{1,4\},\{2,5,6\},\{3,7\}\}$ is a set partition of $\{1,\ldots,7\}$.

In **Case 1C**, there is no restriction, the balls are distinguishable, and the boxes are indistinguishable. A placement is therefore determined by a partition of the balls into at most n nonempty fibres. Equivalently, the answer is the number of set partitions of K into at most n blocks.

Notation 3.2 We use $\mathbb{N} = \mathbb{Z}_{\geq 0}$ for the nonnegative integers, so $0 \in \mathbb{N}$. For $n \in \mathbb{N}$, let $[n] = \{1, 2, \ldots, n\}$.

Recall that $x^{\underline{k}} = x(x-1)(x-2)\cdots(x-k+1)$ is the falling factorial, $x^{\overline{k}} = x(x+1)(x+2)\cdots(x+k-1)$ is the rising factorial, and $k! = k^{\underline{k}}$ is the factorial. The binomial coefficient $\binom{x}{k}$ is defined as $x^{\underline{k}}/k!$ and the multiset coefficient $\left(\!\binom{x}{k}\!\right)$ is defined as $x^{\overline{k}}/k!$.

In degenerate cases, we set $x^{\underline{0}} = x^{\overline{0}} = 1$, $0! = 1$, and $\binom{x}{0} = 1$. Every empty product is 1, every empty sum is 0, and $[0] = \emptyset$. We also use the convention $0^0 = 1$, which is appropriate here because there is exactly one function $\emptyset \to \emptyset$, vacuously. We leave concerns about indeterminate forms to the calculus nerds.

One more convention: if n is a negative integer, $1/(n!) = 0$ and $\binom{x}{n} = 0$. Why? Because it works. Don't ask questions.

Theorem 3.3 (Pascal's identity) For $n, k \in \mathbb{N}$,

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

Combinatorial proof. Assume $1 \leq k \leq n$; the boundary cases are covered by the algebraic proof below. Let $X = \{k\text{-element subsets of } [n]\}$. Let

$$Y = \{k \text{ or } (k-1)\text{-element subsets of } [n-1]\}.$$

Then the left-hand side is $|X|$ and the right-hand side is $|Y|$. We prove the identity by finding a bijection between X and Y. Define $f : X \to Y$ as follows. Given a k-element subset $A \in X$, let $f(A) = A \setminus \{n\}$. Then $f(A) \subseteq [n-1]$ has k or $k-1$ elements, so $f(A) \in Y$. Define $g : Y \to X$ as follows. Given $B \in Y$, let

$$g(B) = \begin{cases} B & \text{if } |B| = k \\ B \cup \{n\} & \text{if } |B| = k-1 \end{cases}.$$

In either case, $|g(B)| = k$ and $g(B) \subseteq [n]$, so $g(B) \in X$.

We check that $g(f(A)) = A$ for $A \in X$. If $n \notin A$, then $f(A) = A \setminus \{n\} = A$, and $|f(A)| = k$, so $g(f(A)) = A$. If $n \in A$, then $|f(A)| = k-1$, so

$$g(f(A)) = (A \setminus \{n\}) \cup \{n\} = A.$$

Similarly, $f(g(B)) = B$ for $B \in Y$. If $|B| = k$, then $g(B) = B$ and $n \notin B$, so $f(g(B)) = f(B) = B$. If $|B| = k-1$, then $g(B) = B \cup \{n\}$, so $f(g(B)) = (B \cup \{n\}) \setminus \{n\} = B$. Thus f and g are mutually inverse bijections. □

Algebraic proof. Recall the binomial theorem: for $|x| < 1$ and $a \in \mathbb{R}$,

$$(1+x)^a = \sum_{j \geq 0} \binom{a}{j} x^j.$$

If

$$A(x) = \sum_{j \geq 0} a_j x^j$$

is a polynomial or power series, then $[x^k]A(x) \coloneqq a_k$, the coefficient of x^k in $A(x)$. Thus the

binomial theorem can be restated as

$$[x^k](1+x)^a = \binom{a}{k}.$$

Now consider

$$(1+x)^{a-1} = \sum_{j\geq 0} \binom{a-1}{j} x^j.$$

Multiplying both sides by $(1+x)$ gives

$$\begin{aligned}
(1+x)^a &= (1+x)\sum_{j\geq 0}\binom{a-1}{j}x^j \\
&= \left[\sum_{j\geq 0}\binom{a-1}{j}x^j\right] + x\left[\sum_{j\geq 0}\binom{a-1}{j}x^j\right] \\
&= \left[\sum_{k\geq 0}\binom{a-1}{k}x^k\right] + \left[\sum_{j\geq 0}\binom{a-1}{j}x^{j+1}\right] \\
&= \left[\sum_{k\geq 0}\binom{a-1}{k}x^k\right] + \left[\sum_{k\geq 1}\binom{a-1}{k-1}x^k\right] \\
&= \left[1+\sum_{k\geq 1}\binom{a-1}{k}x^k\right] + \left[\sum_{k\geq 1}\binom{a-1}{k-1}x^k\right] \\
&= 1+\sum_{k\geq 1}\left[\binom{a-1}{k}+\binom{a-1}{k-1}\right]x^k
\end{aligned}$$

Thus

$$[x^k](1+x)^a = \binom{a-1}{k} + \binom{a-1}{k-1}.$$

Since

$$[x^k](1+x)^a = \binom{a}{k},$$

we obtain

$$\binom{a}{k} = \binom{a-1}{k} + \binom{a-1}{k-1}$$

for $k \geq 1$. The case $k = 0$ also holds. □

Lecture 4 — Monday, September 21, 2015

The idea of the algebraic proof of Pascal's identity is to encode the sequence

$$\binom{a-1}{0}, \binom{a-1}{1}, \binom{a-1}{2}, \binom{a-1}{3}, \dots$$

in the power series

$$A(x) = \binom{a-1}{0}x^0 + \binom{a-1}{1}x^1 + \binom{a-1}{2}x^2 + \binom{a-1}{3}x^3 + \cdots.$$

To add consecutive terms, compare $A(x)$ and $xA(x)$:

$$A(x) = \binom{a-1}{0} + \binom{a-1}{1}x + \binom{a-1}{2}x^2 + \dots$$
$$xA(x) = \binom{a-1}{0}x + \binom{a-1}{1}x^2 + \dots$$

Then

$$[x^k](A(x) + xA(x)) = \binom{a-1}{k} + \binom{a-1}{k-1}.$$

The binomial theorem identifies the same coefficient as $\binom{a}{k}$, which gives Pascal's identity.

Variations on the Binomial Theorem

The plain-Jane binomial theorem says that

$$(1+x)^a = \sum_{k\geq 0} \binom{a}{k} x^k$$

for $a \in \mathbb{R}$ and $|x| < 1$.

For nonnegative integers, the binomial theorem becomes

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

for $n \in \mathbb{N}$. It is often more convenient to write the same identity as an infinite sum, with

the convention that $\binom{n}{k} = 0$ for $k > n$. But why? It is much easier to work with algebraic manipulations; in the finite case, we have to worry about both lower and upper limits of the sum, and the upper limit is a function of n. In the infinite case, we only have to worry about the lower limit, and the upper limit is just ∞. Hurray!

Theorem 4.1 (Negative binomial theorem) For $a \in \mathbb{R}$ and $|x| < 1$,

$$(1-x)^{-a} = \sum_{k\geq 0} \binom{-a}{k} (-x)^k = \sum_{k\geq 0} \left(\binom{a}{k}\right) x^k.$$

Proof. We have

$$(1-x)^{-a} = (1+(-x))^{-a} = \sum_{k\geq 0} \binom{-a}{k} (-x)^k.$$

Now

$$\begin{aligned}\binom{-a}{k} &= \frac{(-a)^{\underline{k}}}{k!} = \frac{(-a)(-a-1)(-a-2)\cdots(-a-k+1)}{k!} \\ &= \frac{(-1)^k a(a+1)(a+2)\cdots(a+k-1)}{k!} = (-1)^k \left(\binom{a}{k}\right).\end{aligned}$$

Thus

$$(1-x)^{-a} = \sum_{k\geq 0} (-1)^k \left(\binom{a}{k}\right) (-x)^k = \sum_{k\geq 0} \left(\binom{a}{k}\right) x^k,$$

as required. □

Example 4.2 The **Fibonacci sequence** $f_0, f_1, f_2, \ldots$ satisfies

$$f_0 = 0, f_1 = 1, f_n = f_{n-1} + f_{n-2}$$

for $n \geq 2$. Consider the related series $F(x) = f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \ldots$. We have

$$\begin{aligned} F(x) &= f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \ldots \\ -xF(x) &= \quad - f_0 x - f_1 x^2 - f_2 x^3 + \ldots \\ -x^2 F(x) &= \qquad\quad - f_0 x^2 - f_1 x^3 + \ldots \end{aligned}$$

Summing these three expressions gives $F(x) - xF(x) - x^2 F(x) = x$, since all other terms

cancel nicely. Therefore,

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

within its radius of convergence. By partial fractions,

$$F(x) = \frac{A}{1 - \frac{1+\sqrt{5}}{2}x} + \frac{B}{1 - \frac{1-\sqrt{5}}{2}x},$$

which gives $A = 1/\sqrt{5}$ and $B = -1/\sqrt{5}$.

Plug in the geometric series:

$$F(x) = \sum_{n \geq 0} \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^n x^n - \sum_{n \geq 0} \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^n x^n.$$

Therefore,

$$f_n = [x^n]F(x) = \left(\frac{1+\sqrt{5}}{2}\right)^n \frac{1}{\sqrt{5}} - \left(\frac{1-\sqrt{5}}{2}\right)^n \frac{1}{\sqrt{5}}.$$

It's Binet's formula!

Definition 4.3 A **lattice path** of length r is a sequence $P_0, P_1, \ldots, P_r \in \mathbb{Z}^2$ of points in the plane such that $P_i - P_{i-1} \in \{(0,1), (1,0)\}$. The step $(0,1)$ is a **North step**, and the step $(1,0)$ is an **East step**. We often represent a path by a string of Ns and Es.

Example 4.4 The lattice path encoded by $ENENNEEN$ is shown in Figure 2.

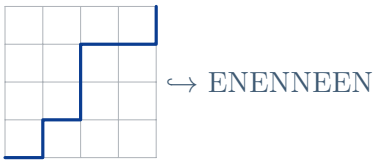


FIGURE 2

The length of a lattice path is the number of steps, *not* the number of points.

Lecture 5 — Wednesday, September 23, 2015

Recall that a lattice path of length r is a sequence $P_0, \ldots, P_r \in \mathbb{Z}^2 \subseteq \mathbb{R}^2$ of points in the plane such that $P_i - P_{i-1} \in \{(0,1),(1,0)\}$. The step $(0,1)$ is a North step, and the step $(1,0)$ is an East step. We represent such a path as a sequence of Ns and Es.

Theorem 5.1 The number of lattice paths from $(0,0)$ to (m,n) is $\binom{m+n}{m} = \binom{m+n}{n}$.

Proof. Each path is represented by a string $s_1 s_2 \ldots s_{m+n}$ with m Es and n Ns. Let $a = \{i \in [m+n] \mid s_i = E\}$. Then a is an m-element subset of $[m+n]$. We show that this defines a bijection between our lattice paths and m-element subsets of $[m+n]$.

Given a lattice path, the set a is uniquely determined, so the map is well-defined. Conversely, given any m-element subset $a \subseteq [m+n]$, define a string $s_1 \ldots s_{m+n}$ by setting $s_i = E$ if $i \in a$ and $s_i = N$ if $i \notin a$. This string has exactly m Es and n Ns, so it determines a lattice path from $(0,0)$ to (m,n). This construction is inverse to the map above, so the correspondence is a bijection.

Therefore, the number of lattice paths from $(0,0)$ to (m,n) equals the number of m-element subsets of $[m+n]$, namely $\binom{m+n}{m}$. Equivalently, by choosing the n positions of N, the same number is $\binom{m+n}{n}$. □

Definition 5.2 A **Dyck path** (or **Catalan path**) is a lattice path $P_0, P_1, \ldots, P_{2n}$, $n \in \mathbb{N}$, from $(0,0)$ to (n,n) with the property that $P_i = (x_i, y_i)$ where $x_i \leq y_i$.

A Dyck path is entirely on or above the line $y = x$, with endpoints on the line, as illustrated in Figure 3.

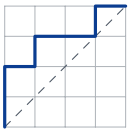


FIGURE 3

Remark 5.3 The trivial path $P_0 = (0, 0)$ is a Dyck path of length 0.

Theorem 5.4 The number of Dyck paths of length $2n$ is

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

These numbers are called the **Catalan numbers**.

We prove this theorem in three steps.

1. *Combinatorial argument.* Let c_n be the number of Dyck paths of length $2n$. Obtain a recurrence relation.
2. *Algebraic formulation.* Consider the series

$$C(x) = \sum_{n \geq 0} c_n x^n.$$

   Convert the recurrence relation to an algebraic equation involving $C(x)$.
3. *Algebraic extraction.* Perform algebraic manipulations on $C(x)$ to compute c_n.

Proof. For (1), consider a Dyck path π of length $2n$, where $n \geq 1$. The first step must be N, and the path must end on the line $y = x$. Suppose the first return to the line $y = x$ occurs after $2k + 2$ steps. Then the $(2k + 2)$th step of π is E. Write

$$\pi = N\pi' E\pi'',$$

where π' has length $2k$ and π'' has length $2(n - k - 1)$. The path π' starts and ends on the line $y = x + 1$ and never goes below it, while π'' starts on $y = x$ and never goes below it. Thus π' and π'' correspond to Dyck paths, and the decomposition is unique. There are c_k possibilities for π' and c_{n-k-1} possibilities for π''. Summing over all possible k gives the recurrence

$$c_n = \sum_{k=0}^{n-1} c_k c_{n-k-1}$$

with $c_0 = 1$. For instance, $c_1 = 1 \cdot 1 = 1$, $c_2 = 1 \cdot 1 + 1 \cdot 1 = 2$, and $c_3 = 1 \cdot 2 + 1 \cdot 1 + 2 \cdot 1 = 5$.

For (2), consider the series

$$C(x) = \sum_{n \geq 0} c_n x^n.$$

We claim that $x(C(x))^2 - C(x) + 1 = 0$. Indeed,

$$\begin{aligned} x(C(x))^2 &= x \left(\sum_{j \geq 0} c_j x^j \right) \left(\sum_{k \geq 0} c_k x^k \right) \\ &= \sum_{j \geq 0} \sum_{k \geq 0} c_j c_k x^{j+k+1} \\ &= \sum_{n \geq 1} \left(\sum_{k=0}^{n-1} c_k c_{n-k-1} \right) x^n \\ &= \sum_{n \geq 1} c_n x^n \\ &= -1 + \sum_{n \geq 0} c_n x^n \\ &= C(x) - 1. \end{aligned}$$

For (3), the quadratic formula gives

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

By the binomial theorem,

$$(1 - 4x)^{1/2} = 1 + \sum_{k \geq 1} \binom{1/2}{k} (-4)^k x^k.$$

For $k \geq 1$,

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

$$= -\frac{2}{k}\binom{2k-2}{k-1}.$$

Lecture 6 — Monday, September 28, 2015

Proof continued. Putting it all together, we get

$$\begin{aligned} C(x) &= \frac{1}{2x}\left[1 \pm \left(1 + \sum_{k\geq 1} -\frac{2}{k}\binom{2k-2}{k-1}x^k\right)\right] \\ &= \frac{1}{2x} \pm \left(\frac{1}{2x} - \sum_{k\geq 1}\frac{1}{k}\binom{2k-2}{k-1}x^{k-1}\right). \end{aligned}$$

The plus sign cannot be correct, since it would give a $1/x$ term in $C(x)$. Therefore, the minus sign must be correct. Thus

$$C(x) = \sum_{k\geq 1}\frac{1}{k}\binom{2k-2}{k-1}x^{k-1} = \sum_{n\geq 0}\frac{1}{n+1}\binom{2n}{n}x^n,$$

and therefore $c_n = \dfrac{1}{n+1}\binom{2n}{n}$. □

We next record two combinatorial proofs of the same result.

Combinatorial proof of Theorem 5.4. Let D_n be the set of Dyck paths of length $2n$, and let P_n be the set of all lattice paths from $(0,0)$ to (n,n). We give a bijection $f : P_n \to D_n \times [n+1]$. Start with a word $s_1 s_2 \dots s_{2n} \in P_n$, and prepend one North step by setting $s_0 = N$. Consider the $2n+1$ cyclic windows of length $2n$:

$$\begin{gathered} s_0 s_1 \dots s_{2n-2} s_{2n-1}, \qquad s_1 s_2 \dots s_{2n-1} s_{2n}, \\ s_2 s_3 \dots s_{2n} s_0, \qquad \dots, \qquad \text{and} \qquad s_{2n} s_0 \dots s_{2n-3} s_{2n-2}. \end{gathered}$$

Each window omits one s_i. Retain precisely those that omit an N; these are the $n+1$ windows in P_n.

Exactly one retained window is a Dyck path. To see why, assign height $+1$ to N and -1 to E. Among the cyclic rotations of the $(2n+1)$-letter word $s_0 s_1 \dots s_{2n}$, exactly one has every partial sum positive: start immediately after the rightmost occurrence of the minimum partial sum. Its first letter is N, and deleting that first letter leaves a Dyck word. Conversely, every retained Dyck window arises this way.

If the unique Dyck window π is obtained by omitting the kth N in cyclic order, let $f(s_1 s_2 \dots s_{2n}) = (\pi, k)$. For example, for the path $NEENNE$, the full list of cyclic windows is

$$NNEENN,\ NEENNE,\ EENNEN,\ ENNENN,\ NNENNE,\ NENNEE,\ ENNEEN.$$

Each word in the list is a substring of the doubled word $NNEENNENNEENNE$; after discarding the three words with unequal numbers of Ns and Es, one retained word is Dyck. The inverse construction prepends N to a Dyck word, marks its kth N, and rotates so that the marked N becomes s_0. Hence f is a bijection. □

Theorem 6.1 For $0 \le a \le b$, the number of lattice paths from $(0,0)$ to (a,b) that are entirely on or above the line $y = x$ is

$$\binom{a+b}{a} - \binom{a+b}{a-1}.$$

Proof. Let X be the set of lattice paths from $(0,0)$ to (a,b) that have at least one point strictly below the line $y = x$. It is enough to show that $|X| = \binom{a+b}{a-1}$. Let Y be the set of lattice paths from $(1,-1)$ to (a,b). We give a bijection between X and Y.

Let $s_1 s_2 \dots s_{a+b} \in X$. Since the path has at least one point below the line $y = x$, it has at least one point on the line $y = x - 1$. Suppose the first such point occurs after $2k-1$ steps. Define

$$f(s_1 s_2 \dots s_{a+b}) = \overline{s}_1 \overline{s}_2 \dots \overline{s}_{2k-1} s_{2k} \dots s_{a+b},$$

where $\overline{s}_i = N$ if $s_i = E$, and $\overline{s}_i = E$ if $s_i = N$. The reflected path begins at $(1,-1)$ and ends at (a,b). Conversely, a path from $(1,-1)$ to (a,b) must meet the line $y = x - 1$; reflecting its initial segment through its first point on that line recovers a unique path in X. Thus f is a bijection. A path in Y uses $a-1$ East steps and $b+1$ North steps, so

$$|X| = |Y| = \binom{a+b}{a-1},$$

as required. □

Lecture 7 — Monday, October 05, 2015

2. Generating Functions and Integer Compositions

Generating Functions

The basic ingredients here are a set $\mathscr{S}$ of combinatorial objects and a function $w : \mathscr{S} \to \mathbb{N}$. A function of this kind is called a **weight function**. For $\sigma \in \mathscr{S}$, the number $w(\sigma)$ is called the **weight** of σ.

The general counting problem is the following: for $n \in \mathbb{N}$, determine the number of objects in $\mathscr{S}$ that have weight n, namely

$$|\{\sigma \in \mathscr{S} : w(\sigma) = n\}|.$$

A weight function is **good** if this number is finite for all $n \in \mathbb{N}$.

The basic strategy is to shove as many of our counting problems as possible into this framework.

Example 7.1 Fix $m \in \mathbb{N}$. Let $\mathscr{S}_m = \{0,1\}^m$, the set of binary strings of length m. If $\sigma \in \mathscr{S}_m$, define the weight of σ to be the number of ones in σ. The counting problem is to determine the number of binary strings of length m that have n ones. The answer is, fairly obviously,

$$\binom{m}{n}.$$

Example 7.2 Let $\mathscr{C}$ be the set of all Dyck paths. If $\pi \in \mathscr{C}$, define the weight of π to be the number of N steps in π. The counting problem is to determine the number of Dyck paths of length $2n$. The answer is

$$\frac{1}{n+1}\binom{2n}{n},$$

by Theorem 5.4.

Definition 7.3 When w is good, the **generating function** for $\mathscr{S}$ with respect to w is the formal power series

$$\Phi_{\mathscr{S}}(x) = \sum_{\sigma \in \mathscr{S}} x^{w(\sigma)}.$$

Proposition 7.4 The answer to the counting problem is $[x^n]\Phi_{\mathscr{S}}(x)$. Equivalently,

$$|\{\sigma \in \mathscr{S} : w(\sigma) = n\}| = [x^n]\Phi_{\mathscr{S}}(x).$$

Proof. We have

$$\begin{aligned}
[x^n]\Phi_{\mathscr{S}}(x) &= [x^n] \sum_{\sigma \in \mathscr{S}} x^{w(\sigma)} \\
&= [x^n] \sum_{k \geq 0} \sum_{\substack{\sigma \in \mathscr{S} \\ w(\sigma)=k}} x^{w(\sigma)} \\
&= [x^n] \sum_{k \geq 0} \left(\sum_{\substack{\sigma \in \mathscr{S} \\ w(\sigma)=k}} 1 \right) x^k \\
&= \sum_{\substack{\sigma \in \mathscr{S} \\ w(\sigma)=n}} 1 \\
&= |\{\sigma \in \mathscr{S} : w(\sigma) = n\}|.
\end{aligned}$$

□

Example 7.5 For Example 7.1, the generating function is

$$\Phi_{\mathscr{S}_m}(x) = \sum_{n=0}^{m} \binom{m}{n} x^n = (1+x)^m.$$

Here we first solved the counting problem and then used that answer to determine the generating function.

Example 7.6 For Example 7.2, the generating function is

$$\Phi_{\mathscr{C}}(x) = \sum_{n\geq 0} \frac{1}{n+1}\binom{2n}{n} x^n = \frac{1-\sqrt{1-4x}}{2x},$$

where the second equality comes from the algebraic manipulations in the proof of Theorem 5.4.

Lemma 7.7 (Sum lemma) Let $\mathscr{S}$ be a set with weight function w, and $\mathscr{A} \cup \mathscr{B} = \mathscr{S}$. Then $\Phi_{\mathscr{S}}(x) = \Phi_{\mathscr{A}}(x) + \Phi_{\mathscr{B}}(x) - \Phi_{\mathscr{A}\cap\mathscr{B}}(x)$.

Proof. For a subset $\mathscr{A} \subseteq \mathscr{S}$, let $\mathbb{1}_{\mathscr{A}} : \mathscr{S} \to \mathbb{Z}$ be the indicator function

$$\mathbb{1}_{\mathscr{A}}(\sigma) = \begin{cases} 1, & \sigma \in \mathscr{A} \\ 0, & \sigma \notin \mathscr{A} \end{cases}.$$

Then $\mathbb{1}_{\mathscr{A}}(\sigma) + \mathbb{1}_{\mathscr{B}}(\sigma) - \mathbb{1}_{\mathscr{A}\cap\mathscr{B}}(\sigma) = 1$ for all $\sigma \in \mathscr{S}$. So

$$\begin{aligned}
\Phi_{\mathscr{A}}(x) + \Phi_{\mathscr{B}}(x) - \Phi_{\mathscr{A}\cap\mathscr{B}}(x) &= \sum_{\sigma\in\mathscr{A}} x^{w(\sigma)} + \sum_{\sigma\in\mathscr{B}} x^{w(\sigma)} - \sum_{\sigma\in\mathscr{A}\cap\mathscr{B}} x^{w(\sigma)} \\
&= \sum_{\sigma\in\mathscr{S}} \mathbb{1}_{\mathscr{A}}(\sigma) x^{w(\sigma)} + \sum_{\sigma\in\mathscr{S}} \mathbb{1}_{\mathscr{B}}(\sigma) x^{w(\sigma)} - \sum_{\sigma\in\mathscr{S}} \mathbb{1}_{\mathscr{A}\cap\mathscr{B}}(\sigma) x^{w(\sigma)} \\
&= \sum_{\sigma\in\mathscr{S}} (\mathbb{1}_{\mathscr{A}}(\sigma) + \mathbb{1}_{\mathscr{B}}(\sigma) - \mathbb{1}_{\mathscr{A}\cap\mathscr{B}}(\sigma)) x^{w(\sigma)} \\
&= \Phi_{\mathscr{S}}(x).
\end{aligned}$$

□

Lemma 7.7 is most commonly used in the case where $\mathscr{A} \cap \mathscr{B} = \emptyset$, in which case $\Phi_{\mathscr{A}\cap\mathscr{B}}(x) = 0$.

Remark 7.8 The same argument gives the following generalization. If $\mathscr{S}$ is a set with weight function w, and $\mathscr{S} = \bigcup_i \mathscr{A}_i$, where $\mathscr{A}_i \cap \mathscr{A}_j = \emptyset$ for $i \neq j$, then

$$\Phi_{\mathscr{S}}(x) = \sum_i \Phi_{\mathscr{A}_i}(x).$$

Example 7.9 Suppose that we have five loonies and three toonies. How many ways are there to make \$7?

We make a table of all possibilities:

	Loonies					
Toonies	\$0	\$1	\$2	\$3	\$4	\$5
\$0	0	1	2	3	4	5
\$2	2	3	4	5	6	(7)
\$4	4	5	6	(7)	8	9
\$6	6	(7)	8	9	10	11

There are three ways to make \$7. There is an actual point to this.

Example 7.10 Compute $[x^7](1+x+x^2+x^3+x^4+x^5)(1+x^2+x^4+x^6)$.

We multiply each pair of possible terms:

	x^0	x^1	x^2	x^3	x^4	x^5
x^0	x^0	x^1	x^2	x^3	x^4	x^5
x^2	x^2	x^3	x^4	x^5	x^6	(x^7)
x^4	x^4	x^5	x^6	(x^7)	x^8	x^9
x^6	x^6	(x^7)	x^8	x^9	x^{10}	x^{11}

Therefore,

$$[x^7](1+x+\cdots+x^5)(1+x^2+x^4+x^6) = 3.$$

Examples 7.9 and 7.10 are two forms of the same question! This is the key.

Lecture 8 — Wednesday, September 30, 2015

Lemma 8.1 (Product lemma) Let $\mathscr{A}$ and $\mathscr{B}$ be sets with good weight functions α and β, respectively. Let w be a weight function on $\mathscr{A} \times \mathscr{B}$ and let $\gamma \in \mathbb{N}$ be constant. If

$$w(a,b) = \alpha(a) + \beta(b) + \gamma$$

for all $(a,b) \in \mathscr{A} \times \mathscr{B}$, then

$$\Phi_{\mathscr{A}\times\mathscr{B}}(x) = x^\gamma \cdot \Phi_{\mathscr{A}}(x) \cdot \Phi_{\mathscr{B}}(x).$$

Proof.

$$\begin{aligned}
\Phi_{\mathscr{A}\times\mathscr{B}}(x) &= \sum_{(a,b)\in\mathscr{A}\times\mathscr{B}} x^{w(a,b)} \\
&= \sum_{a\in\mathscr{A}} \sum_{b\in\mathscr{B}} x^{\alpha(a)+\beta(b)+\gamma} \\
&= x^\gamma \sum_{a\in\mathscr{A}} x^{\alpha(a)} \sum_{b\in\mathscr{B}} x^{\beta(b)} \\
&= x^\gamma \cdot \Phi_{\mathscr{A}}(x) \cdot \Phi_{\mathscr{B}}(x).
\end{aligned}$$

□

Remark 8.2 This generalizes as follows. Let $\mathscr{A}_1, \ldots, \mathscr{A}_k$ be sets with good weight functions $\alpha_1, \ldots, \alpha_k$. Let w be a weight function on $\mathscr{A}_1 \times \cdots \times \mathscr{A}_k$ and let $\gamma \in \mathbb{N}$ be constant. If

$$w(a_1, \ldots, a_k) = \gamma + \sum_{i=1}^{k} \alpha_i(a_i)$$

for all $(a_1, \ldots, a_k) \in \mathscr{A}_1 \times \cdots \times \mathscr{A}_k$, then

$$\Phi_{\mathscr{A}_1\times\cdots\times\mathscr{A}_k}(x) = x^\gamma \Phi_{\mathscr{A}_1}(x) \cdots \Phi_{\mathscr{A}_k}(x).$$

The proof is by induction.

Warning An infinite Cartesian product needs separate treatment. Cardinality is not the decisive issue: what matters is whether each coefficient receives only finitely many contributions and whether the coefficients of the finite partial products eventually stabilize.

Integer Compositions

The motivating problem here is the following:

Example 8.3 For $k \geq 1$, determine the number of tuples $(c_1, \ldots, c_k) \in \mathbb{N}^k$ such that $c_1 + c_2 + \cdots + c_k = n$.

Solution. Our set of objects is $\mathbb{N}^k$. Define a weight function $w : \mathbb{N}^k \to \mathbb{N}$ by $w(c_1, \ldots, c_k) = c_1 + \cdots + c_k$. Define another weight function $\alpha : \mathbb{N} \to \mathbb{N}$ by $\alpha(i) = i$. Then

$$w(c_1, \ldots, c_k) = \sum_{i=1}^{k} \alpha(c_i),$$

so we can apply Lemma 8.1. That is,

$$\Phi_{\mathbb{N}^k}(x) = [\Phi_{\mathbb{N}}(x)]^k = (1 + x + x^2 + \ldots)^k = \left(\frac{1}{1-x}\right)^k.$$

Therefore, the answer is

$$[x^n]\Phi_{\mathbb{N}^k}(x) = [x^n](1-x)^{-k} = \binom{n+k-1}{n} = \binom{n+k-1}{k-1} = \left(\!\binom{k}{n}\!\right).$$

□

Definition 8.4 Let $n, k \in \mathbb{N}$. A **composition of n with k parts** is a k-tuple $(c_1, \ldots, c_k) \in \mathbb{N}_{\geq 1}^k$ such that

$$\sum_{i=1}^{k} c_i = n.$$

The integers $c_1, \ldots, c_k$ are the **parts** of the composition. There is exactly one composition of $n = 0$, and it has zero parts.

Example 8.5 Determine the number of compositions of n with k parts.

Solution. Assume $n, k \geq 1$. The sole composition with zero parts is the empty composition of 0. The previous proof applies with $\mathbb{N}$ replaced by $\mathbb{N}_{\geq 1}$. We have $\Phi_{\mathbb{N}_{\geq 1}}(x) = x(1-x)^{-1}$, so $\Phi_{\mathbb{N}_{\geq 1}^k}(x) = x^k(1-x)^{-k}$. Therefore, the answer is

$$[x^n]\Phi_{\mathbb{N}_{\geq 1}^k} = [x^n]x^k(1-x)^{-k} = [x^{n-k}](1-x)^{-k} = \left(\!\binom{k}{n-k}\!\right) = \binom{n-1}{n-k} = \binom{n-1}{k-1}.$$

□

Alternatively, start with $(1+1+\cdots+1) = n$. A composition with k parts is obtained by changing $k-1$ of the plus signs into commas.

Example 8.6 Determine the number of compositions of n with an even number of parts, where each part is odd.

Solution. Let $\mathscr{S}$ be the set of all compositions with an even number of parts, all of which are odd. Define the weight of a composition to be the sum of its parts. Let $\mathbb{N}_{\text{odd}} = \{1, 3, 5, 7, \dots\}$. Then

$$\begin{aligned}\mathscr{S} &= \mathbb{N}_{\text{odd}}^0 \cup \mathbb{N}_{\text{odd}}^2 \cup \mathbb{N}_{\text{odd}}^4 \cup \dots \\ &= \bigcup_{k=0}^{\infty} \mathbb{N}_{\text{odd}}^{2k}.\end{aligned}$$

Here $\mathbb{N}^0 = \{()\}$ consists of the empty tuple; it is not the empty set. Then

$$\Phi_{\mathscr{S}}(x) = \sum_{k=0}^{\infty} \Phi_{\mathbb{N}_{\text{odd}}^{2k}}(x).$$

By the same reasoning as before, $\Phi_{\mathbb{N}_{\text{odd}}^{2k}}(x) = \Phi_{\mathbb{N}_{\text{odd}}}(x)^{2k} = (x(1-x^2)^{-1})^{2k}$. Then

$$\Phi_{\mathscr{S}}(x) = \sum_{k=0}^{\infty} (x(1-x^2)^{-1})^{2k} = \frac{1}{1 - x^2(1-x^2)^{-2}}.$$

Thus the answer is

$$[x^n]\frac{1}{1-x^2(1-x^2)^{-2}} = [x^n]\frac{(1-x^2)^2}{(1-x^2)^2-x^2} = [x^n]\frac{(1-x^2)^2}{1-3x^2+x^4}.$$

Equivalently, after simplifying,

$$\Phi_{\mathscr{S}}(x) = \frac{(1-x^2)^2}{1-3x^2+x^4}.$$

Hence there are no such compositions when n is odd. When $n = 2m > 0$, the coefficient turns out to be f_{2m}, the $(2m)$th term of the Fibonacci sequence. For $n = 0$, the empty composition gives one composition. □

Lecture 9 — Friday, October 02, 2015

We have been working with generating functions while ignoring issues of convergence. We now formalize that approach.

Definition 9.1

A **variable** is a letter used to represent something that can take on a range of values.

A **constant** is a number, such as e, with one specific value.

A **parameter** can take on different values, but is treated as constant within a given problem.

An **unknown** is the quantity being solved for; it is treated as a variable during the solution.

An **indeterminate** is a symbol that takes on no values at all, but is manipulated according to algebraic rules.

For example, $x^2 + 3$ is a **polynomial**, while $f : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto x^2 + 3$ is a **polynomial function**. In the former, x is a symbol; in the latter, x is a variable. Over $\mathbb{R}$ this distinction is usually harmless, but over a finite field such as $\mathbb{F}_5$, a polynomial contains more information than the function it defines.

We now reinterpret power series so that x is *disallowed* from being a variable and treated as an indeterminate instead. We cure the disease by killing the patient. However, just because we are

reinventing the wheel does not mean that the old wheel was all bad. We will still use analytic power series later, but the formal setting lets us avoid convergence issues when doing algebra.

If $\mathscr{S}$ is a finite set, then $\Phi_{\mathscr{S}}(x)$ is a polynomial, so plugging in numbers is meaningful. For example:

1. $\Phi_{\mathscr{S}}(1) = |\mathscr{S}|$.
2. $$\Phi'_{\mathscr{S}}(1) = \sum_{\sigma \in \mathscr{S}} w(\sigma)$$
   is the total weight of all elements of $\mathscr{S}$.
3. If $\mathscr{S} \neq \emptyset$, then $\Phi'_{\mathscr{S}}(1)/\Phi_{\mathscr{S}}(1)$ is the average weight of an element of $\mathscr{S}$.
4. If $\mathscr{S} \neq \emptyset$, then
   $$\frac{\Phi''_{\mathscr{S}}(1)}{\Phi_{\mathscr{S}}(1)} + \frac{\Phi'_{\mathscr{S}}(1)}{\Phi_{\mathscr{S}}(1)} - \left(\frac{\Phi'_{\mathscr{S}}(1)}{\Phi_{\mathscr{S}}(1)}\right)^2$$
   is the variance of the weight of an element of $\mathscr{S}$.

If $\mathscr{S}$ is nonempty and infinite, and $0 < \Phi_{\mathscr{S}}(p) < \infty$ for some $p > 0$, we can put the probability distribution

$$\mathbb{P}_p(\sigma) = \frac{p^{w(\sigma)}}{\Phi_{\mathscr{S}}(p)}.$$

Then the expected weight is $p\Phi'_{\mathscr{S}}(p)/\Phi_{\mathscr{S}}(p)$. In some problems, however, the radius of convergence is 0. Fortunately, the formal theory does not care.

Our goal is to construct a framework in which we can work with power series without worrying about convergence issues. Our approach will be to redefine everything.

Definition 9.2 A **formal power series** with real coefficients is a power series

$$A(x) = \sum_{n \geq 0} a_n x^n.$$

The notation $\mathbb{R}[[x]]$ denotes the set of formal power series with real coefficients. It is a commutative ring: if

$$A(x) = \sum_{n \geq 0} a_n x^n \qquad \text{and} \qquad B(x) = \sum_{n \geq 0} b_n x^n,$$

the addition is defined by

$$A(x) + B(x) = \sum_{n\geq 0}(a_n + b_n)x^n$$

and multiplication is defined by

$$A(x)B(x) = \sum_{n\geq 0}\left(\sum_{k=0}^{n} a_k b_{n-k}\right)x^n.$$

The additive identity is

$$0 = \sum_{n\geq 0} 0x^n.$$

The multiplicative identity is

$$1 = 1 + \sum_{n\geq 1} 0x^n.$$

There are a few extra operations:

1. **Coefficient extraction:** $[x^n]A(x) = a_n$.
2. **Scalar multiplication:**
$$cA(x) = \sum_{n\geq 0}(ca_n)x^n,$$
where $c \in \mathbb{R}$.
3. **Formal derivative:**
$$\frac{\mathrm{d}}{\mathrm{d}x}A(x) = \sum_{n\geq 1} n a_n x^{n-1} = \sum_{n\geq 0}(n+1)a_{n+1}x^n.$$
4. **Formal integral:**
$$I_x A(x) = \sum_{n\geq 0}\frac{a_n}{n+1}x^{n+1} = \sum_{n\geq 1}\frac{a_{n-1}}{n}x^n$$
This formal integral satisfies the expected formal version of the fundamental theorem of calculus.

Sometimes, we can substitute. There are two cases where $A(B(x))$ is allowed:

1. If
$$A(x) = \sum_{n=0}^{d} a_n x^n$$
is a polynomial of degree d, then we can define
$$A(B(x)) = \sum_{n=0}^{d} a_n B(x)^n$$
which is well-defined because it only requires finitely many operations.

2. If $[x^0]B(x) = 0$, then define
$$A(B(x)) = \sum_{n=0}^{\infty} a_n B(x)^n.$$
The key point is that if we want to know $[x^m]A(B(x))$, then $B(x)^n$ does not contribute for $n > m$, so
$$[x^m]A(B(x)) = [x^m] \sum_{n=0}^{m} a_n B(x)^n.$$
For this coefficient calculation, we may treat $A(x)$ as a polynomial.

In all other cases, $A(B(x))$ is not defined. This includes the case when $B(x) = c \in \mathbb{R} \setminus \{0\}$.

Lecture 10 — Monday, October 05, 2015

Theorem 10.1 A formal power series $A(x) \in \mathbb{R}[[x]]$ has a multiplicative inverse if and only if $[x^0]A(x) \neq 0$.

Proof. First suppose $[x^0]A(x) = 0$. Then $[x^0]A(x)B(x) = 0$ for any $B(x)$, so we cannot have $A(x)B(x) = 1$. Conversely, suppose $[x^0]A(x) = a_0 \neq 0$. Let $f(x) = a_0 - x$. This has multiplicative inverse
$$g(x) = \sum_{n \geq 0} a_0^{-n-1} x^n,$$
that is, $f(x)g(x) = 1$. Let $B(x) = a_0 - A(x)$. Since $[x^0]B(x) = 0$, we can substitute $B(x)$ into the equation above to get $f(B(x)) \cdot g(B(x)) = 1$. But $f(B(x)) = A(x)$, so $g(B(x)) = A(x)^{-1}$. □

Note that in the formal setting, we treat the binomial theorem for $\alpha \in \mathbb{R}$ as a definition. This is consistent with the rules of formal power series.

When we counted Dyck paths in Theorem 5.4, the most mysterious part was the equation $x \cdot C(x)^2 - C(x) + 1 = 0$. It is time to solve the mystery.

Let $\mathscr{C}$ be the set of all Dyck paths, with weight function $w : \mathscr{C} \to \mathbb{N}$ where $w(\pi)$ is the number of N steps in π. Then $C(x) = \Phi_{\mathscr{C}}(x)$. The combinatorial decomposition argument gives a bijection

$$\mathscr{C} \times \mathscr{C} \to \mathscr{C} \setminus \{\varepsilon\}, \quad (\pi', \pi'') \mapsto N\pi' E \pi'',$$

where ε denotes the trivial Dyck path of length 0. Define a weight function W on $\mathscr{C} \times \mathscr{C}$ by

$$W(\pi', \pi'') = w(N\pi' E \pi'').$$

Then, from the bijection, we have

$$\begin{aligned} \Phi_{\mathscr{C}\times\mathscr{C}}(x) &= \sum_{(\pi',\pi'')\in\mathscr{C}\times\mathscr{C}} x^{W(\pi',\pi'')} = \sum_{\pi\in\mathscr{C}\setminus\{\varepsilon\}} x^{w(\pi)} \\ &= C(x) - 1. \end{aligned}$$

On the other hand,

$$W(\pi', \pi'') = w(\pi') + w(\pi'') + 1,$$

so Lemma 8.1 gives

$$\Phi_{\mathscr{C}\times\mathscr{C}}(x) = x \cdot \Phi_{\mathscr{C}}(x)^2 = x \cdot C(x)^2.$$

Putting these together gives

$$x \cdot C(x)^2 = C(x) - 1.$$

Example 10.2 The **crazy dice problem** asks whether it is possible to replace the numbers on a pair of dice with different positive integers without changing the probability table for the sum of the two face-up numbers. We use generating functions to answer this.

Let $\mathscr{S}$ be the set of sides of an ordinary 6-sided die. Define the weight of $\sigma \in \mathscr{S}$ to be the number written on side σ. Then $\Phi_{\mathscr{S}}(x) = x + x^2 + x^3 + x^4 + x^5 + x^6$. The set of ways to roll a pair is $\mathscr{S} \times \mathscr{S}$. Define the weight to be the total of the two numbers rolled. Then $\Phi_{\mathscr{S}\times\mathscr{S}}(x) = \Phi_{\mathscr{S}}(x)^2$. This tells us how many ways to roll each total, and hence determines the probability table.

Now suppose A and B are the sets of sides of two alternative dice, and the weight is the number written on the side. For example, if the first die has $1, 2, 4, 4, 4, 8$ written on the sides, then $\Phi_A(x) = x + x^2 + 3x^4 + x^8$. In this case, the generating function from rolling the pair is $\Phi_A(x)\Phi_B(x)$, and this determines the probability table. Therefore, we want

$$\Phi_A(x)\Phi_B(x) = \Phi_{\mathscr{S}}(x)^2.$$

We now factor $\Phi_{\mathscr{S}}(x)^2$ and redistribute the factors. Because

$$\Phi_{\mathscr{S}}(x) = x(x+1)(x^2+x+1)(x^2-x+1),$$

and both dice have six sides, we need

$$\Phi_A(1) = \Phi_B(1) = \Phi_{\mathscr{S}}(1) = 6.$$

Thus we need a factor of $(x+1)(x^2+x+1)$ in both $\Phi_A(x)$ and $\Phi_B(x)$. Since the numbers on each die must be positive integers, we need a factor of x in both.

The nonordinary possibility, up to swapping the two dice, is

$$\Phi_A(x) = x(x+1)(x^2+x+1) = x + 2x^2 + 2x^3 + x^4$$

and

$$\Phi_B(x) = x(x+1)(x^2+x+1)(x^2-x+1)^2 = x + x^3 + x^4 + x^5 + x^6 + x^8.$$

These polynomials have nonnegative integer coefficients and the correct product. Thus crazy dice exist: the first die has sides $1, 2, 2, 3, 3, 4$, and the second has sides $1, 3, 4, 5, 6, 8$.

The general recipe for solving combinatorial problems using generating functions is as follows:

1. Introduce the set of objects under consideration and a weight function on them. What we remove from the objects is reintroduced by the weight function.
2. Find a bijection or decomposition involving the set, such as the decomposition of a nonempty Dyck path into a pair of Dyck paths.
3. Define weight functions on any new sets so that the weight of a composite object is the sum of the weights of its parts.
4. Write down the corresponding generating function equations.
5. Solve using algebra.

The previous examples all follow this recipe. The hard combinatorial step is usually (2), while (5) is the algebraic step.

Lecture 11 — Wednesday, October 07, 2015

How did we factor $\Phi_{\mathscr{S}}(x)$ in Example 10.2? We used the factorization of $x^6 - 1$. In general, we had

$$\Phi_{\mathscr{S}}(x) = \frac{x(x^6 - 1)}{x - 1}.$$

Thus we need to factor $x^6 - 1$.

Theorem 11.1 There exist polynomials $\phi_n(x) \in \mathbb{Z}[x]$, $n \geq 1$, called the **cyclotomic polynomials**, satisfying the following:

1. $x^n - 1 = \prod_{d|n} \phi_d(x)$;
2. $\phi_n(x) = \prod_{d|n} (x^d - 1)^{\mu(n/d)}$, where $\mu(k)$ is the **Möbius function** defined by

$$\mu(k) = \begin{cases} (-1)^r, & \text{if } k \text{ is squarefree with } r \text{ prime factors,} \\ 0, & \text{otherwise;} \end{cases}$$

3. $$\phi_n(x) = \prod_{\substack{1 \leq k \leq n \\ \gcd(k,n)=1}} (x - e^{2\pi i k/n})$$

   in $\mathbb{C}[x]$. The latter terms $e^{2\pi i k/n}$ are the ***n*th roots of unity**.

Proof sketch.. Use (3) as the definition. Then (1) holds, and by Möbius inversion, so does (2).

It remains to justify the integer coefficients. Proceed by induction on n. From (1),

$$\phi_n(x) = \frac{x^n - 1}{\prod_{\substack{d|n \\ d<n}} \phi_d(x)}.$$

By induction, the denominator is monic and belongs to $\mathbb{Z}[x]$. Polynomial division by a monic polynomial in $\mathbb{Z}[x]$ produces an integer-coefficient quotient and remainder; (3) shows that the remainder is zero. Hence $\phi_n(x) \in \mathbb{Z}[x]$. □

Integer Partitions

Definition 11.2 A **partition** of n with k parts is a k-tuple $\lambda = (\lambda_1, \ldots, \lambda_k) \in (\mathbb{N}_{\geq 1})^k$ of positive integers such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$$

and

$$\sum_{i=1}^{k} \lambda_i = n.$$

In other words, a partition is a composition in which the parts are in weakly decreasing order. The integer n is called the **size** of λ, and we write $|\lambda| = n$ or $\lambda \vdash n$. The integer k is called the **length** of λ; we write $\ell(\lambda) = k$.

For all $m, n \in \mathbb{N}$, we first determine the number of partitions of n in which all parts are at most m. For example, when $n = 4$ and $m = 3$, the partitions are

$$31, \qquad 22, \qquad 211, \qquad \text{and} \qquad 1111,$$

so there are four.

Let $\mathscr{P}_m$ denote the set of all partitions in which all parts are at most m. Define the weight of a partition to be its size: $w(\lambda) = \lambda_1 + \cdots + \lambda_k$. Let $\mathscr{J}_m = \mathbb{N} \times 2\mathbb{N} \times \cdots \times m\mathbb{N}$ with weight function $\hat{w} : \mathscr{J}_m \to \mathbb{N}$ given by $\hat{w}(c_1, \ldots, c_m) = c_1 + c_2 + \cdots + c_m$.

We give a bijection $\phi : \mathscr{P}_m \to \mathscr{J}_m$. If $\lambda = (\lambda_1, \ldots, \lambda_k) \in \mathscr{P}_m$, define $\phi(\lambda) = (c_1, \ldots, c_m)$ where

$$c_i = \sum_{\substack{j: \\ \lambda_j = i}} \lambda_j.$$

The inverse map records c_i/i, the number of parts equal to i in λ. For example, if $m = 7$ and $\lambda = (6, 6, 6, 3, 2, 2)$, then $\phi(\lambda) = (0, 4, 3, 0, 0, 18, 0)$.

Note that $w(\lambda) = \hat{w}(\phi(\lambda))$, so ϕ is a **weight-preserving bijection**. Therefore, $\Phi_{\mathscr{P}_m}(x) = \Phi_{\mathscr{J}_m}(x)$. To compute $\Phi_{\mathscr{J}_m}(x)$, define a weight function $\alpha_i : i\mathbb{N} \to \mathbb{N}$ by $\alpha_i(c) = c$. Then

$\hat{w}(c_1, \ldots, c_m) = \alpha_1(c_1) + \cdots + \alpha_m(c_m)$, so we can use Lemma 8.1. In particular,

$$\Phi_{\mathscr{J}_m}(x) = \prod_{i=1}^{m} \Phi_{i\mathbb{N}}(x) = \prod_{i=1}^{m} (1 - x^i)^{-1}.$$

Therefore, the number of partitions of n with all parts at most m is $[x^n] \prod_{i=1}^{m} (1 - x^i)^{-1}$. There is no simpler closed form for this coefficient in general.

Next, for $m, n \in \mathbb{N}$, we determine the number of partitions of n with length at most m.

Let $\mathscr{P}'_m$ be the set of partitions of length at most m. The weight of a partition is its size. The **Ferrers diagram** is a left-justified array of boxes, with λ_j boxes in row j, for $j \in [k]$. The Ferrers diagram of $\lambda = 6632$ is shown in Figure 4.

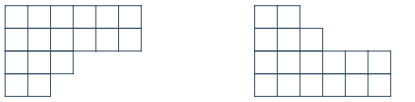


FIGURE 4

The size of λ is the number of boxes, and the length of λ is the number of rows. In the previous problem, we defined a bijection between $\mathscr{P}_m$ and $\mathscr{J}_m$; in terms of the Ferrers diagram, $c_i = i \cdot$ (number of rows of length i). This works because the diagram has at most m columns. In the present problem, we use columns of Ferrers diagrams instead, defining a bijection between $\mathscr{P}'_m$ and $\mathscr{J}_m$ by letting c_i record i times the number of columns of length i. This works because the diagram has at most m rows. Hence the answer is the same as in the previous problem.

Lecture 12 — Friday, October 09, 2015

We now determine the number of partitions of n.

Let $m \geq n$. Then every partition of n has at most n parts, and all parts are at most n, so the answer is $[x^n] \prod_{i=1}^{m} (1 - x^i)^{-1}$ as long as $m \geq n$. We could take $m = n$, but it is more useful to write the answer as

$$[x^n] \prod_{i=1}^{\infty} (1 - x^i)^{-1}.$$

We now make this infinite product precise.

Let $A_1(x), A_2(x), A_3(x), \cdots \in \mathbb{R}[[x]]$. We say that $a_n = [x^n]\prod_{i=1}^{\infty} A_i(x)$ if and only if there exists $N \in \mathbb{N}$ such that for all $m \geq N$,

$$a_n = [x^n]\prod_{i=1}^{m} A_i(x).$$

If such an N exists, we define

$$\prod_{i=1}^{\infty} A_i(x) := \sum_{n\geq 0} a_n x^n.$$

Example 12.1 Compute $f(x) = \prod_{i=1}^{\infty}(1 + x^{2^{i-1}})$.

Solution. The partial product is $f_m(x) = \prod_{i=1}^{m}(1 + x^{2^{i-1}})$.

$$\begin{aligned} f_1(x) &= 1 + x \\ f_2(x) &= (1 + x)(1 + x^2) \\ f_3(x) &= (1 + x)(1 + x^2)(1 + x^4) = 1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7 \end{aligned}$$

For $n = 0$, the coefficient $[x^0]f_m(x)$ is always 1. For $n \geq 1$, binary expansion gives $[x^n]f_m(x) = 1$ whenever $m > \log_2(n)$. Therefore,

$$\prod_{i=1}^{\infty}(1 + x^{2^{i-1}}) = \sum_{n\geq 0} x^n = \frac{1}{1-x}.$$

□

Returning to partitions, each factor has the expansion

$$\frac{1}{1-x^i} = \underbrace{1}_{\substack{0 \text{ parts} \\ \text{of size } i}} + \underbrace{x^i}_{\substack{1 \text{ part} \\ \text{of size } i}} + \underbrace{x^{2i}}_{\substack{2 \text{ parts} \\ \text{of size } i}} + x^{3i} + \cdots.$$

Example 12.2 Show that the number of partitions of n with all parts odd is equal to the number of partitions of n with all parts distinct.

Solution. Let $\mathscr{O}$ be the set of partitions with all parts odd, and $\mathscr{D}$ be the set of partitions with all parts distinct.

We have

$$\Phi_{\mathscr{O}}(x) = \prod_{j=1}^{\infty}(1-x^{2j-1})^{-1} \qquad \text{and} \qquad \Phi_{\mathscr{D}}(x) = \prod_{i=1}^{\infty}(1+x^i),$$

since for each i, we can have either zero or one part of size i in a partition with distinct parts.

We claim that $\Phi_{\mathscr{O}}(x) = \Phi_{\mathscr{D}}(x)$.

To verify the claim, let $A(x) = \prod_{i=1}^{\infty}(1-x^{2i})^{-1}$. Then

$$A(x)\cdot\Phi_{\mathscr{O}}(x) = \prod_{i=1}^{\infty}(1-x^{2i-1})^{-1}(1-x^{2i})^{-1} = \prod_{j=1}^{\infty}(1-x^j)^{-1}.$$

Meanwhile,

$$\begin{aligned} A(x)\cdot\Phi_{\mathscr{D}}(x) &= \prod_{i=1}^{\infty}\frac{1+x^i}{1-x^{2i}} \\ &= \prod_{i=1}^{\infty}\frac{1+x^i}{(1-x^i)(1+x^i)} \\ &= \prod_{i=1}^{\infty}(1-x^i)^{-1} \end{aligned}$$

as required. Hence, $\Phi_{\mathscr{O}}(x) = \Phi_{\mathscr{D}}(x)$ since $\mathbb{R}[[x]]$ is an integral domain. □

If we have two weight functions, then we get a multivariate generating function:

$$\Phi_{\mathscr{S}}(x,y) = \sum_{\sigma\in\mathscr{S}} x^{w_1(\sigma)}y^{w_2(\sigma)}.$$

It is often said that x and y are **marking** the weights w_1 and w_2, respectively: x "marks" w_1 and y "marks" w_2.

Exercise 12.3 For $m, n \in \mathbb{N}$, show that the number of objects $\sigma \in \mathscr{S}$ satisfying $w_1(\sigma) = n$ and $w_2(\sigma) = m$ is $[x^n y^m]\Phi_{\mathscr{S}}(x,y)$.

The sum and product lemmas for multivariate generating functions are essentially the same as

Lemmas 7.7 and 8.1. We define

$$[x^n]A(x,y) = \sum_{m\geq 0}([x^n y^m]A(x,y))y^m,$$

and similarly for coefficient extraction in the other variable. For the product version, we require

$$w_i(a,b) = \alpha_i(a) + \beta_i(b) + \gamma_i$$

for $i = 1, 2$. Here $\gamma_1, \gamma_2 \in \mathbb{N}$. Then $\Phi_{\mathscr{A}\times\mathscr{B}}(x,y) = x^{\gamma_1}y^{\gamma_2}\Phi_{\mathscr{A}}(x,y)\Phi_{\mathscr{B}}(x,y)$.

Lecture 13 — Wednesday, October 14, 2015

Recall that if $\mathscr{S}$ is a set of objects and $w_1, w_2 : \mathscr{S} \to \mathbb{N}$ are weight functions, then the multivariate generating function is

$$\Phi_{\mathscr{S}}(x,y) = \sum_{\sigma\in\mathscr{S}} x^{w_1(\sigma)}y^{w_2(\sigma)}.$$

We have $\Phi_{\mathscr{S}}(1,1) = |\mathscr{S}|$ if $\mathscr{S}$ is finite, and $\Phi_{\mathscr{S}}(x,1) = \Phi_{\mathscr{S}}(x)$ whenever the univariate generating function with respect to w_1 is defined. Garbage in, garbage out: w_1 and w_2 may be perfectly good weight functions when used together, but bad when used separately. For example, if $\mathscr{S} = \mathbb{N}^2$, $w_1(a,b) = a$, and $w_2(a,b) = b$, then

$$\Phi_{\mathscr{S}}(x,y) = \frac{1}{(1-x)(1-y)},$$

but the univariate series for w_1 is undefined because infinitely many pairs have each fixed first coordinate.

Example 13.1 Let $\mathscr{C}$ be the set of all compositions. For $c \in \mathscr{C}$, write $c = (c_1, \ldots, c_k)$ and let

$$w_1(c) = \sum_{i=1}^{k} c_i$$

and let $w_2(c) = k$, the number of parts. Determine $\Phi_{\mathscr{C}}(x,y)$.

Solution. We have $\mathscr{C} = \bigcup_{k\geq 0}\mathbb{N}_{\geq 1}^k$, where $\mathbb{N}_{\geq 1}^0 = \{\varepsilon\}$ contains the empty composition. Its two weights are $w_1(\varepsilon) = w_2(\varepsilon) = 0$. For $k > 0$, the tuple $(c_1, \ldots, c_k)$ has weights $w_1(c_1, \ldots, c_k) = c_1 + \cdots + c_k$ and $w_2(c_1, \ldots, c_k) = k$. Let $\alpha_1 : \mathbb{N}_{\geq 1} \to \mathbb{N}$ be given by $\alpha_1(c) = c$. Then $w_1(c_1, \ldots, c_k) =$

$\alpha_1(c_1) + \cdots + \alpha_1(c_k)$. Let $\alpha_2 : \mathbb{N}_{\geq 1} \to \mathbb{N}$ be given by $\alpha_2(c) = 1$. Then $w_2(c_1, \ldots, c_k) = \alpha_2(c_1) + \cdots + \alpha_2(c_k)$. Also

$$\Phi_{\mathbb{N}_{\geq 1}}(x, y) = xy + x^2y + x^3y + \cdots = \frac{xy}{1-x}.$$

By the two equations for w_1 and w_2, Lemma 8.1 applies. Then

$$\begin{aligned}
\Phi_{\mathscr{C}}(x, y) &= \sum_{k \geq 0} \Phi_{\mathbb{N}_{\geq 1}^k}(x, y) \\
&= \sum_{k \geq 0} \Phi_{\mathbb{N}_{\geq 1}}(x, y)^k \\
&= \sum_{k \geq 0} \left(\frac{xy}{1-x} \right)^k \\
&= \frac{1}{1 - \dfrac{xy}{1-x}}.
\end{aligned}$$

□

Suppose we want to know the number of compositions of n with k parts. Then

$$\begin{aligned}
[x^n y^k] \Phi_{\mathscr{C}}(x, y) &= [x^n y^k] \frac{1}{1 - \dfrac{xy}{1-x}} \\
&= [x^n] \left(\frac{x}{1-x} \right)^k \\
&= [x^n] x^k (1-x)^{-k} \\
&= [x^{n-k}] (1-x)^{-k} \\
&= \left(\binom{k}{n-k} \right).
\end{aligned}$$

The number of compositions of n with any number of parts is obtained by setting $y = 1$:

$$[x^n] \Phi_{\mathscr{C}}(x, 1) = [x^n] \frac{1-x}{1-2x} = \begin{cases} 2^{n-1}, & \text{if } n > 0, \\ 1, & \text{if } n = 0. \end{cases}$$

Remark 13.2 To solve a rational generating function problem by hand, we can reverse engineer a recurrence relation for $p(x)/q(x)$, where $p(x)$ and $q(x)$ are polynomials. This is the same method used for the Fibonacci sequence.

Exercise 13.3 Do the same thing with partitions; that is, show

$$\Phi_{\mathscr{P}_m}(x,y) = \prod_{i=1}^{m}(1-yx^i)^{-1},$$

where $\mathscr{P}_m$ is the set of partitions whose parts are at most m, x marks size, and y marks length.

Strings

Definition 13.4 A **string** is a finite sequence of symbols taken from an alphabet.

Our first examples will use the alphabet $\{0,1\}$. Strings of this type are called **binary strings**.

For example, $a = 11010$ and $b = 0001$ are binary strings of lengths 5 and 4, respectively. The empty string ε has length 0.

Definition 13.5 The **concatenation product** of a, b is the string ab obtained by writing a and b next to each other.

For example, if $a = 11010$ and $b = 0001$, then $ab = 110100001$. The concatenation product is associative, has an identity element, is noncommutative, and admits no inverses. Thus, binary strings with the concatenation product form a **monoid**.

If A and B are sets of strings, define $AB \coloneqq \{ab \mid a \in A, b \in B\}$ and

$$A^* \coloneqq \{\varepsilon\} \cup A \cup AA \cup AAA \cup \cdots = \bigcup_{k\geq 0} A^k.$$

Here A^k is the k-fold concatenation product, and $A^0 = \{\varepsilon\}$. This is different from the Cartesian product $A \times \cdots \times A$.

Definition 13.6 The concatenation product defines a surjective map from $A \times B \to AB$, $(a, b) \mapsto ab$. If this map is a bijection, we say that AB is **unambiguous** (more precisely: the *expression* AB is unambiguous). If $A \cap B = \emptyset$, we say that $A \cup B$ is unambiguous too. For a more complicated expression built out of concatenations/unions, the expression is unambiguous if each individual expression is.

Lecture 14 — Friday, October 16, 2015

Example 14.1 Let $A = \{0, 00\}$, $B = \{1, 11\}$, $C = \{\varepsilon, 0, 1\}$.

1. The product $AB = \{01, 011, 001, 0011\}$ is unambiguous.
2. The product $AC = \{0, 00, 01, 00, 000, 001\}$ is not unambiguous.
3. The union $A \cup B = \{0, 00, 1, 11\}$ is unambiguous.
4. The union $A \cup C = \{0, 00, \varepsilon, 0, 1\}$ is not unambiguous.
5. The set $A^* = \{\varepsilon, 0, 00, 000, 0000, \dots\}$ is not unambiguous.

Unless specified otherwise, the default weight function on a set of strings will be the length of a string.

Using the sets A and B from Example 14.1, we have $\Phi_{AB}(x) = x^2 + 2x^3 + x^4$.

Theorem 14.2 If A, B are sets of strings, then:

1. If AB is unambiguous, then $\Phi_{AB}(x) = \Phi_A(x)\Phi_B(x)$.
2. If $A \cup B$ is unambiguous, then $\Phi_{A \cup B}(x) = \Phi_A(x) + \Phi_B(x)$.
3. If A^* is unambiguous, then $\Phi_{A^*}(x) = (1 - \Phi_A(x))^{-1}$.

Proof. For (1), the expression AB is unambiguous if and only if there is a weight-preserving bijection $A \times B \to AB$, where the weight of $(a, b) \in A \times B$ is defined to be $w(a, b) = w(a) + w(b)$. By Lemma 8.1,

$$\Phi_{A \times B}(x) = \Phi_A(x)\Phi_B(x) = \Phi_{AB}(x).$$

For (2), this is just Lemma 7.7.

For (3), we have $A^* = \bigcup_{k\geq 0} A^k$. Therefore,

$$\Phi_{A^*}(x) = \sum_{k\geq 0} \Phi_{A^k}(x) = \sum_{k\geq 0} [\Phi_A(x)]^k = (1 - \Phi_A(x))^{-1}.$$

□

Example 14.3 The set of all 01-strings is $\{0,1\}^*$. Therefore,

$$\Phi_{\{0,1\}^*}(x) = (1 - \Phi_{\{0,1\}}(x))^{-1} = (1-2x)^{-1} = \frac{1}{1-2x}.$$

Since $[x^n]1/(1-2x) = 2^n$, there are 2^n binary strings of length n. We probably already knew this.

Example 14.4 Let $\mathscr{S}$ be the set of 01-strings that do not have 11 as a substring.

One decomposition is $\mathscr{S} = \{0, 10\}^*\{\varepsilon, 1\}$. Then

$$\Phi_{\mathscr{S}}(x) = (1 - \Phi_{\{0,10\}}(x))^{-1} \cdot \Phi_{\{\varepsilon,1\}}(x) = \left(\frac{1}{1-(x+x^2)}\right)(1+x).$$

This is an example of a **0-decomposition**. The set can also be written as $\mathscr{S} = \{\varepsilon, 1\}\{0\{1\}^*\}^*$. The general 0-decomposition is

$$\{0,1\}^* = (\{1\}^*\{0\})^*\{1\}^* = \{1\}^*(\{0\}\{1\}^*)^*,$$

and both expressions are unambiguous.

We will use the following principle:

Proposition 14.5 If $\mathscr{S}$ is a set of strings, $p(A_1, \ldots, A_k)$ is an unambiguous expression for $\mathscr{S}$ in terms of sets $A_1, \ldots, A_k$, and $B_i \subseteq A_i$ for each i, then $p(B_1, \ldots, B_k)$ is unambiguous too.

We omit the proof, which is straightforward.

For the 0-decomposition, if $A_1 = \{1\}^*$, $A_2 = \{0\}$, $A_3 = \{1\}^*$, then $p(A_1, A_2, A_3) = (A_1A_2)^*A_3$

is unambiguous. If $B_1 = \{\varepsilon, 1\} \subseteq A_1$, $B_2 = \{0\} = A_2$, $B_3 = B_1$, then the principle implies that $(\{\varepsilon, 1\}\{0\})^*\{\varepsilon, 1\}$ is unambiguous.

There is also a **1-decomposition**,

$$\{0,1\}^* = (\{0\}^*\{1\})^*\{0\}^* = \{0\}^*(\{1\}\{0\}^*)^*.$$

The **block decomposition** is

$$\begin{aligned}\{0,1\}^* &= \{0\}^*(\{1\}\{1\}^*\{0\}\{0\}^*)^*\{1\}^* \\ &= \{1\}^*(\{0\}\{0\}^*\{1\}\{1\}^*)^*\{0\}^*.\end{aligned}$$

A set of strings built from finite sets and the operations $\cup$, concatenation, and $*$ is a **regular language**, and such an expression is called a **regular expression**.

Definition 14.6 A **block** is a maximal substring using only one symbol. Here, "maximal" means that it cannot be extended to a longer substring using only that symbol.

Example 14.7 The string 00111101000 is made up of 5 blocks: $\boxed{00}\boxed{1111}\boxed{0}\boxed{1}\boxed{000}$.

In the same string, the substring 111 is not a block since it can be extended to 1111.

Lecture 15 — Monday, October 19, 2015

Recall that the block decomposition is

$$\{0,1\}^* = \{0\}^*\left(\{1\}\{1\}^*\{0\}\{0\}^*\right)^*\{1\}^*.$$

Blocks are nonempty by definition.

Example 15.1 Consider the string $\boxed{000}\boxed{1}\boxed{0}\boxed{1}\boxed{111}\boxed{0}$. We find the generating function for the set $\mathscr{S}$ of $\{0,1\}$ strings in which each block has odd length, with respect to length.

The regular expression for $\mathscr{S}$ is

$$(\{\varepsilon\} \cup \{0\}\{00\}^*)\,(\{1\}\{11\}^*\{0\}\{00\}^*)^*\,(\{\varepsilon\} \cup \{1\}\{11\}^*)\,.$$

Then

$$\Phi_{\mathscr{S}}(x) = \left(1 + \frac{x}{1-x^2}\right)\left(1 - \left(\frac{x}{1-x^2}\right)\left(\frac{x}{1-x^2}\right)\right)^{-1} \cdot \left(1 + \frac{x}{1-x^2}\right).$$

This expression can then be simplified algebraically. The algebra is more afraid of you than you are of it.

You know what? Let's just do it. Let $y = \dfrac{x}{1-x^2}$. Then

$$\begin{aligned}
\Phi_{\mathscr{S}}(x) &= (1+y)(1-y^2)^{-1}(1+y) \\
&= (1+y)^2(1-y^2)^{-1} \\
&= \frac{(1+y)^2}{(1-y)(1+y)} \\
&= \frac{1+y}{1-y} \\
&= \frac{1 + \dfrac{x}{1-x^2}}{1 - \dfrac{x}{1-x^2}} \\
&= \frac{1+x-x^2}{1-x-x^2}.
\end{aligned}$$

Partitions as Strings

First we need a version of Lemma 8.1 for infinite products.

Definition 15.2 Let $A_1, A_2, A_3, \ldots$ be sets of strings, and let $\varepsilon \in A_i$ for all i. Define the infinite concatenation product as

$$A_1 A_2 \cdots = \bigcup_{m \geq 1} A_1 A_2 \cdots A_m.$$

We have $A_1 \subseteq A_1A_2 \subseteq \cdots$. In this context, we say that $A_1A_2\cdots$ is **unambiguous** if and only if $A_1A_2\cdots A_m$ is unambiguous for every $m \geq 1$.

Theorem 15.3 If an infinite concatenation product $A_1A_2\cdots$ is unambiguous and carries a good additive weight function, then

$$\Phi_{A_1A_2\ldots}(x) = \prod_{i=1}^{\infty} \Phi_{A_i}(x).$$

Proof. Let $n \in \mathbb{N}$. Since there are only finitely many strings of weight n in $A_1A_2\cdots$, there exists $N \in \mathbb{N}$ such that all of them lie in $A_1A_2\cdots A_N$. Then, for all $m \geq N$,

$$[x^n]\Phi_{A_1A_2\ldots}(x) = [x^n]\Phi_{A_1\ldots A_m}(x) = [x^n]\prod_{i=1}^{m} \Phi_{A_i}(x).$$

The result follows from the definition of the infinite product. □

The set of partitions is in bijection with $\{1\}^*\{2\}^*\{3\}^*\cdots$. Define the weight of $ii\cdots i$, with k copies of i, to be ki. By Theorem 15.3,

$$\Phi_{\mathscr{P}}(x) = \prod_{i=1}^{\infty} \Phi_{\{i\}^*}(x) = \prod_{i=1}^{\infty} \frac{1}{1-x^i}$$

which is the same generating function as before.

We can also use more than one weight function. For a 01-string σ, let $w_0(\sigma)$ be the number of 0s in σ, and let $w_1(\sigma)$ be the number of 1s in σ. For instance, if $\sigma = 011000$, then $w_0(\sigma) = 4$ and $w_1(\sigma) = 2$. The same product principles from Theorems 14.2 and 15.3 still apply.

Example 15.4 Determine $\Phi_{\{0,1\}^*}(x, y)$ with respect to w_0 and w_1, and then determine $\Phi^{\text{length}}_{\{0,1\}^*}(t)$.

Solution. We have the unambiguous decomposition

$$\{0,1\}^* = \{\varepsilon\} \cup \{0,1\}^*\{0,1\}.$$

Therefore,

$$\Phi_{\{0,1\}^*}(x,y) = 1 + \Phi_{\{0,1\}^*}(x,y)\Phi_{\{0,1\}}(x,y).$$

Solving for $\Phi_{\{0,1\}^*}(x,y)$ gives

$$\Phi_{\{0,1\}^*}(x,y) = \frac{1}{1-\Phi_{\{0,1\}}(x,y)} = \frac{1}{1-(x+y)}.$$

Since $\text{length}(\sigma) = w_0(\sigma) + w_1(\sigma)$, we have

$$\Phi^{\text{length}}_{\{0,1\}^*}(t) = \Phi^{w_0,w_1}_{\{0,1\}^*}(t,t) = \frac{1}{1-2t}.$$

□

Example 15.5 Let $\mathscr{S}$ be the set of 01-strings that do not have 011 as a substring. Determine $\Phi_{\mathscr{S}}(x,y)$.

Solution 1 (brute force). One direct decomposition is $\mathscr{S} = \{1\}^* (\{0\}\{0\}^*\{1\})^* \{0\}^*$. Then

$$\begin{aligned}\Phi_{\mathscr{S}}(x,y) &= \left(\frac{1}{1-y}\right)\left(\frac{1}{1-\left(\frac{x}{1-x}\right)y}\right)\left(\frac{1}{1-x}\right) \\ &= \frac{1}{1-x-y+xy^2}.\end{aligned}$$

□

Solution 2 (trickery). Another decomposition is

$$\{0,1\}^* = \mathscr{S}(\{011\}\mathscr{S})^*.$$

This decomposition is unambiguous because the final occurrence of 011 cannot overlap the preceding avoided string. This method does not work for every excluded substring; for example, the substring 000 can overlap itself. The decomposition gives

$$\frac{1}{1-(x+y)} = \Phi_{\mathscr{S}}(x,y)\frac{1}{1-xy^2\Phi_{\mathscr{S}}(x,y)}.$$

Solving for $\Phi_{\mathscr{S}}(x, y)$ gives

$$\Phi_{\mathscr{S}}(x, y) = \frac{1}{1 - x - y + xy^2}.$$

□

Lecture 16 — Wednesday, October 21, 2015

Solution 3 (more trickery). Observe that

$$\mathscr{S}\{0, 1\} = (\mathscr{S} \setminus \{\varepsilon\}) \cup \mathscr{S}\{011\},$$

and both pieces are unambiguous. Therefore,

$$\Phi_{\mathscr{S}}(x, y)(x + y) = \Phi_{\mathscr{S}}(x, y) - 1 + \Phi_{\mathscr{S}}(x, y)xy^2,$$

so

$$\Phi_{\mathscr{S}}(x, y) = \frac{1}{1 - x - y + xy^2}.$$

This method does not directly work for excluding 000, but it can be adapted by considering the set of strings that do not have 000 as a substring and end with 0, and the set of strings that do not have 000 as a substring and end with 1. □

Solution 4 (substitution). For a fourth solution, we introduce **substitution**. Let $\mathscr{S}$ and A be sets of strings on an alphabet X, and let $j \in X$. Consider the set

$$\mathscr{S}[j \to A] = \text{the set of strings obtained from } \sigma \in \mathscr{S} \text{ by replacing each occurrence of } j \text{ in } \sigma \text{ with an element of } A.$$

For example, if $\mathscr{S} = \{0, 010, 001\}$ and $A = \{10, 000\}$, then

$$\mathscr{S}[0 \to A] = \{10, 000, 10110, 101000, 000110, 0001000, 10101, 100001, 000101, 0000001\}.$$

Formally,

$$\mathscr{S}[j \to A] = \bigcup_{\substack{\sigma \in \mathscr{S} \\ \sigma = \sigma_1 \dots \sigma_n}} \Big\{ \sigma_1 \dots \sigma_{a_1 - 1} A \sigma_{a_1 + 1} \dots$$

$$\sigma_{a_2-1}A\ldots\Big\}$$

where $a_1, a_2, \ldots$ are the positions of the js in σ. This is unambiguous if union and concatenations are unambiguous.

Theorem 16.1 Let $X = \{0, 1, 2, \ldots, m\}$, let $A, \mathscr{S} \subseteq X^*$, and let $j \in X$. Define weight functions $w_0, \ldots, w_m : X^* \to \mathbb{N}$, where $w_i(\sigma)$ is the number of is in σ. The multivariate generating series for $\mathscr{S}$ is

$$\Phi_{\mathscr{S}}(x_0, \ldots, x_m) = \sum_{\sigma \in \mathscr{S}} x_0^{w_0(\sigma)} \cdots x_m^{w_m(\sigma)}.$$

Define $\Phi_A(x_0, \ldots, x_m)$ and $\Phi_{\mathscr{S}[j\to A]}(x_0, \ldots, x_m)$ similarly. If $\mathscr{S}[j \to A]$ is unambiguous, then

$$\Phi_{\mathscr{S}[j\to A]}(x_0, \ldots, x_m) = \Phi_{\mathscr{S}}(x_0, \ldots, x_{j-1}, \Phi_A(x_0, \ldots, x_m), x_{j+1}, \ldots, x_m).$$

Proof. We have

$$\Phi_{\mathscr{S}}(x_0, \ldots, x_m) = \sum_{\sigma \in \mathscr{S}} \prod_{i=1}^{\text{length}(\sigma)} x_{\sigma_i}.$$

Also,

$$\Phi_{\mathscr{S}[j\to A]}(x_0, \ldots, x_m) = \sum_{\sigma \in \mathscr{S}} \prod_{i=1}^{\text{length}(\sigma)} \begin{cases} x_{\sigma_i}, & \text{if } \sigma_i \neq j, \\ \Phi_A(x_0, \ldots, x_m), & \text{if } \sigma_i = j, \end{cases}$$

which is the desired right-hand side. □

Now, let Q be the set of 012-strings that do not have 011 as a substring. We claim that

$$Q[2 \to \{2, 011\}] = \{0, 1, 2\}^*.$$

Also, $\Phi_{\mathscr{S}}(x, y) = \Phi_Q(x, y, 0)$, because

$$\begin{aligned} \Phi_Q(x, y, 0) &= \sum_{\sigma \in Q} x^{w_0(\sigma)} y^{w_1(\sigma)} 0^{w_2(\sigma)} \\ &= \sum_{\sigma \in Q,\ w_2(\sigma)=0} x^{w_0(\sigma)} y^{w_1(\sigma)} \cdot 1 \\ &= \sum_{\sigma \in \mathscr{S}} x^{w_0(\sigma)} y^{w_1(\sigma)} \end{aligned}$$

$$= \Phi_{\mathscr{S}}(x, y).$$

By Theorem 16.1,

$$\Phi_Q(x, y, z + xy^2) = \Phi_{\{0,1,2\}^*}(x, y, z) = \frac{1}{1 - x - y - z}.$$

Setting $z = -xy^2$ gives

$$\Phi_{\mathscr{S}}(x, y) = \Phi_Q(x, y, 0) = \frac{1}{1 - (x + y - xy^2)}.$$

□

Solution 5 (blocked strings). A **blocked string** is a string with certain (possibly overlapping) substrings marked. For example,

$$1\boxed{011}0\boxed{100}0111\boxed{011}10.$$

Consider

$$\begin{aligned} R &= \{0, 1, \boxed{011}\}^* \\ &= \text{all 01-strings in which some occurrences of 011 are marked.} \end{aligned}$$

Let $Q \subseteq \left\{0, 1, \boxed{011}\right\}^*$ be the subset in which all occurrences of 011 are marked. Then $Q\left[\boxed{011} \to \{011, \boxed{011}\}\right] = R$, in the intended sense. Define a weight function $c(\sigma)$ to be the number of marked blocks in σ. With respect to w_0, w_1, c, we get

$$\Phi_Q(x, y, 1 + z) = \Phi_R(x, y, z) = \frac{1}{1 - (x + y + zy^2x)}.$$

Since $\Phi_{\mathscr{S}}(x, y) = \Phi_Q(x, y, 0)$, taking $z = -1$ gives

$$\Phi_{\mathscr{S}}(x, y) = \frac{1}{1 - (x + y - xy^2)}.$$

□

Lecture 17 — Friday, October 23, 2015

Solution 6 (unambiguous regular expression). There is an algorithm for turning an ambiguous regular expression into an unambiguous one. For example, $\{0,1\}^*\{011\}\{0,1\}^*$ is totally ambiguous. The algorithm is hideously impractical, and should never be used. However, it shows that all problems involving excluded substrings can be solved, at least in principle. □

Solution 7 (classification by ending). Consider subsets of $\mathscr{S}$ classified by how their strings end, and then consider which symbols can be appended while staying in $\mathscr{S}$.

Let

$$A = \{\text{strings ending in } 01\}, \qquad B = \{\text{strings ending in } 0\}, \qquad \text{and} \qquad C = \{1\}^*.$$

Then $\mathscr{S} = A \cup B \cup C$. Appending 0 to a string in A produces a string in B. Appending 0 to a string in B keeps the string in B, while appending 1 to a string in B produces a string in A. Appending 1 to a string in C keeps the string in C, while appending 0 to a string in C produces a string in B. Thus

$$A = B\{1\}, \qquad B = A\{0\} \cup B\{0\} \cup C\{0\}, \qquad \text{and} \qquad C = \{1\}^*.$$

This system gives

$$\begin{aligned} \Phi_A(x,y) &= y\Phi_B(x,y), \\ \Phi_B(x,y) &= x\Phi_A(x,y) + x\Phi_B(x,y) + x\Phi_C(x,y), \\ \text{and} \quad \Phi_C(x,y) &= \frac{1}{1-y}. \end{aligned}$$

Solving for $\Phi_A(x,y)$ and $\Phi_B(x,y)$ gives

$$\Phi_A(x,y) = \frac{xy}{(1-y)(1-x-xy)} \qquad \text{and} \qquad \Phi_B(x,y) = \frac{x}{(1-y)(1-x-xy)}.$$

Therefore,

$$\Phi_{\mathscr{S}}(x,y) = \Phi_A(x,y) + \Phi_B(x,y) + \Phi_C(x,y) = \frac{1}{1-x-y+xy^2}.$$

□

Example 17.1 Let $\mathcal{D}_k \subseteq \{0,1,2,\ldots,k\}^*$ be the set of strings in which every block has length 1. Let $w_i(\sigma)$ be the number of is in σ, for $i = 0,\ldots,k$. We want to count these strings with respect to the weights $w_0,\ldots,w_k$.

Solution 1 (block decomposition). We have the following block decomposition:

$$\mathcal{D}_k = \mathcal{D}_{k-1}(\{k\}(\mathcal{D}_{k-1} \setminus \{\varepsilon\}))^*\{\varepsilon, k\}.$$

This recursive decomposition gives

$$\Phi_{\mathcal{D}_k}(x_0,\ldots,x_k) = \Phi_{\mathcal{D}_{k-1}}(x_0,\ldots,x_{k-1})\frac{1}{1 - x_k(\Phi_{\mathcal{D}_{k-1}}(x_0,\ldots,x_{k-1}) - 1)}(1 + x_k).$$

If we know $\Phi_{\mathcal{D}_{k-1}}(x_0,\ldots,x_{k-1})$, then we can compute $\Phi_{\mathcal{D}_k}(x_0,\ldots,x_k)$, and hence we can compute $\Phi_{\mathcal{D}_k}(x_0,\ldots,x_k)$ for all k by induction. However, this is a very messy expression, and it is not clear how to simplify it. □

Solution 2 (substitution). A cleaner approach uses substitution:

$$\mathcal{D}_k[0 \to \{0\}\{0\}^*][1 \to \{1\}\{1\}^*]\cdots[k \to \{k\}\{k\}^*] = \{0,1,2,\ldots,k\}^*.$$

By Theorem 16.1,

$$\begin{aligned}\Phi_{\mathcal{D}_k}\left(\frac{x_0}{1-x_0}, \frac{x_1}{1-x_1}, \ldots, \frac{x_k}{1-x_k}\right) &= \Phi_{\{0,1,\ldots,k\}^*}(x_0, x_1, \ldots, x_k) \\ &= \frac{1}{1 - (x_0 + x_1 + \cdots + x_k)}.\end{aligned}$$

Let

$$y_i = \frac{x_i}{1 - x_i}.$$

Then $x_i = y_i/(1 + y_i)$, and hence

$$\Phi_{\mathcal{D}_k}(y_0,\ldots,y_k) = \frac{1}{1 - (y_0/(1+y_0) + \cdots + y_k/(1+y_k))}.$$

Does this give us what we want? We can rewrite the denominator as

$$1 - \sum_{i=0}^{k} \frac{y_i}{1+y_i} = \sum_{i=0}^{k} \frac{1}{1+y_i} - k.$$

Therefore,

$$\Phi_{\mathcal{D}_k}(y_0, \ldots, y_k) = \frac{1}{\frac{1}{1+y_0} + \cdots + \frac{1}{1+y_k} - k}.$$

It can be checked that this agrees with the result from the block decomposition. □

Stirling Numbers

Definition 17.2 Let $n, k \in \mathbb{N}$. A **partition** of $[n]$ into k subsets is a set $\{X_1, \ldots, X_k\}$ such that $[n] = X_1 \cup \cdots \cup X_k$ and $X_i \neq \emptyset$ for all i, and $X_i \cap X_j = \emptyset$ for $i \neq j$. The **subset Stirling number** (or **Stirling number of the second kind**) $\left\{ {n \atop k} \right\}$ is the number of partitions of $[n]$ into k subsets.

Example 17.3 The partitions of $[4]$ into 2 subsets are $\{\{1,2,3\},\{4\}\}$, $\{\{1,2,4\},\{3\}\}$, $\{\{1,3,4\},\{2\}\}$, $\{\{2,3,4\},\{1\}\}$, $\{\{1,2\},\{3,4\}\}$, $\{\{1,3\},\{2,4\}\}$, and $\{\{1,4\},\{2,3\}\}$.

Therefore, $\left\{ {4 \atop 2} \right\} = 7$.

Lemma 17.4 The subset Stirling numbers satisfy the following identities:

1. $$\left\{ {0 \atop 0} \right\} = 1, \qquad \left\{ {n \atop 1} \right\} = 1 \quad \text{for } n \geq 1.$$

2. $$\left\{ {n \atop 0} \right\} = \left\{ {0 \atop k} \right\} = 0$$

   for $n, k > 0$.

3.
$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k \left\{ {n-1 \atop k} \right\} \quad \text{for } n, k \geq 1.$$

Proof. The identities (1) and (2) are clear. For (3), suppose that $S = \{X_1, \ldots, X_k\}$ is a partition of $[n]$ into k subsets. There are two possibilities:

1. $\{n\} \in S$. In this case, $S \setminus \{\{n\}\}$ is a partition of $[n-1]$ into $k-1$ subsets, of which there are $\left\{ {n-1 \atop k-1} \right\}$ possibilities.
2. $n \in X_i$ for some i, but $\{n\} \neq X_i$. In this case, $\{X_1, X_2, \ldots, X_i \setminus \{n\}, \ldots, X_k\}$ is a partition of $[n-1]$ into k subsets. Thus there are k possibilities for i, and $\left\{ {n-1 \atop k} \right\}$ possibilities for the partition.

□

Lecture 18 — Monday, October 26, 2015

We continue with the subset Stirling numbers by introducing their bivariate exponential generating function. Define
$$P(x, y) = \sum_{n \geq 0} \sum_{k \geq 0} \left\{ {n \atop k} \right\} \frac{x^n}{n!} y^k.$$
Then $P(x, 0) = P(0, y) = 1$. Using the recurrence for subset Stirling numbers, we also get
$$\begin{aligned} \frac{\partial}{\partial x} P(x, y) &= \sum_{n \geq 0} \sum_{k \geq 0} \left\{ {n+1 \atop k} \right\} \frac{x^n}{n!} y^k \\ &= \sum_{n \geq 0} \sum_{k \geq 0} \left(\left\{ {n \atop k-1} \right\} + k \left\{ {n \atop k} \right\} \right) \frac{x^n}{n!} y^k \\ &= y P(x, y) + y \frac{\partial}{\partial y} P(x, y). \end{aligned}$$

Here the term with lower index $k-1=-1$ is interpreted as 0 when $k=0$. Thus P satisfies the partial differential equation

$$\frac{\partial}{\partial x}P(x,y) = y\left(P(x,y) + \frac{\partial}{\partial y}P(x,y)\right), \tag{18.1}$$

subject to the initial conditions $P(x,0) = P(0,y) = 1$.

Theorem 18.1

$$\sum_{n\geq 0}\sum_{k\geq 0}\left\{{n \atop k}\right\}\frac{x^n}{n!}y^k = e^{y(e^x-1)}.$$

Proof. Put $F(x,y) = e^{y(e^x-1)}$. Then $F(x,0) = F(0,y) = 1$, and

$$\frac{\partial}{\partial x}F = ye^xF = y\left(F + (e^x-1)F\right) = y\left(F + \frac{\partial}{\partial y}F\right),$$

so F satisfies (18.1). The equation determines the coefficient of x^{n+1} from the coefficients of x^n, while $P(0,y)=1$ supplies the initial coefficient. Thus the formal power-series solution is unique, and $P = F$. □

The recurrence relation for the subset Stirling numbers can be rearranged as

$$\left\{{n-1 \atop k-1}\right\} = \left\{{n \atop k}\right\} - k\left\{{n-1 \atop k}\right\}.$$

For the negative indices needed below, we make the extension precise. Set

$$\left\{{0 \atop 0}\right\} = 1, \qquad \left\{{0 \atop -n}\right\} = 0 \quad (n \geq 1), \qquad \left\{{-k \atop -n}\right\} = 0 \quad (k > n),$$

and, for $1 \leq k \leq n$, define recursively

$$\left\{{-k \atop -n}\right\} = \left\{{-k+1 \atop -n+1}\right\} + (n-1)\left\{{-k \atop -n+1}\right\}.$$

This is the negative-index form of the rearranged recurrence above. For example,

$$\begin{aligned}\left\{{-2 \atop -4}\right\} &= \left\{{-1 \atop -3}\right\} + 3\left\{{-2 \atop -3}\right\} \\ &= \left\{{0 \atop -2}\right\} + 2\left\{{-1 \atop -2}\right\} + 3\left(\left\{{-1 \atop -2}\right\} + 2\left\{{-2 \atop -2}\right\}\right)\end{aligned}$$

$$= 2\left(\left\{ {0 \atop -1} \right\} + \left\{ {-1 \atop -1} \right\}\right) + 3\left(\left\{ {0 \atop -1} \right\} + \left\{ {-1 \atop -1} \right\} + 2\left(\left\{ {-1 \atop -1} \right\} + \left\{ {-2 \atop -1} \right\}\right)\right)$$
$$= 2 \cdot 1 + 3 \cdot 1 + 6 \cdot 1 = 11.$$

In particular, the recursion gives $\left\{ {-n \atop -n} \right\} = 1$ for every $n \in \mathbb{N}$.

Definition 18.2 Let $\begin{bmatrix} n \\ k \end{bmatrix}$ be defined as the number of permutations of $[n]$ with exactly k cycles. This is called the **cycle Stirling number**.

Theorem 18.3

$$\begin{bmatrix} n \\ k \end{bmatrix} = \left\{ {-k \atop -n} \right\}.$$

Proof. Let

$$c(n,k) \coloneqq \begin{bmatrix} n \\ k \end{bmatrix} \qquad \text{and} \qquad t(n,k) \coloneqq \left\{ {-k \atop -n} \right\}$$

for $n, k \in \mathbb{N}$.

First we find a recurrence for $c(n,k)$. Consider a permutation of $[n]$ with exactly k cycles, and look at where n appears.

If n is a 1-cycle (n), then deleting (n) leaves a permutation of $[n-1]$ with $k-1$ cycles, giving $c(n-1,k-1)$ possibilities. If n is not a 1-cycle, start with a permutation of $[n-1]$ with k cycles and insert n immediately after one of the elements $1, 2, \ldots, n-1$ in cycle notation. This gives $(n-1)c(n-1,k)$ possibilities.

Combining the two possibilities gives

$$c(n,k) = c(n-1,k-1) + (n-1)c(n-1,k).$$

Next we compare this with $t(n,k)$. From

$$\left\{ {a-1 \atop b-1} \right\} = \left\{ {a \atop b} \right\} - b\left\{ {a-1 \atop b} \right\},$$

substitute $a = -k+1$ and $b = -n+1$. We obtain

$$\left\{ {-k \atop -n} \right\} = \left\{ {-k+1 \atop -n+1} \right\} + (n-1)\left\{ {-k \atop -n+1} \right\},$$

which is exactly

$$t(n,k) = t(n-1,k-1) + (n-1)t(n-1,k).$$

Thus c and t satisfy the same recurrence.

For initial values, $c(0,0) = 1$ and $c(n,n) = 1$, and by the extension above we also have $t(0,0) = \left\{ {0 \atop 0} \right\} = 1$ and $t(n,n) = \left\{ {-n \atop -n} \right\} = 1$. Also $c(n,0) = t(n,0) = 0$ for $n \geq 1$. Hence the two triangular arrays are determined by the same recurrence with the same boundary values, so

$$c(n,k) = t(n,k)$$

for all $n, k \in \mathbb{N}$. Therefore,

$$\begin{bmatrix} n \\ k \end{bmatrix} = \left\{ {-k \atop -n} \right\}.$$

□

3. Graph Theory and Planarity

Graph Theory

Definition 18.4 A **graph** G consists of a finite set of **vertices** $V(G)$ and another finite set of **edges** $E(G)$, whose elements are all unordered pairs of distinct elements of $V(G)$.

Remark 18.5 This will be the definition we use. Some authors allow infinite graphs, multiple edges, loops, or ordered edges; our convention excludes these possibilities. In particular, our graphs are **simple graphs**.

Example 18.6 Define G by

$$V(G) = \{1,2,3,4,5\},$$
$$\text{and} \quad E(G) = \{\{1,2\},\{1,3\},\{2,4\},\{3,4\},\{4,5\},\{2,5\}\}.$$

The **drawing** of G is not unique, but the structure of G is uniquely determined by $V(G)$ and $E(G)$. One possible drawing appears in Figure 5.

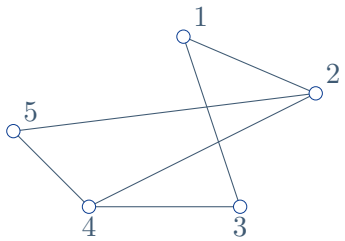


FIGURE 5

Note that simple graphs can also be thought of as symmetric, irreflexive relations on a finite set. For example, the graph G above corresponds to the relation

$$\{(x,y),(y,x) \mid \{x,y\} \in E(G)\} \subseteq V(G) \times V(G).$$

Curly brackets are a tremendous strain on our wrists, so we will usually write the set of edges as, for example, $\{12, 13, 24, 34, 45, 25\} = E(G)$.

Lecture 19 — Wednesday, October 28, 2015

Definition 19.1 If $x, y \in V(G)$ and $e = xy \in E(G)$, then x is **adjacent** to y, the vertices x and y are **neighbours**, the vertices x and y are **incident** with e, and the edge e **joins** x and y. The **degree** of a vertex is the number of edges to which it is incident. We denote the degree of $v \in V(G)$ by $\deg(v)$.

A map $f : V(G) \to V(H)$ is called an **isomorphism** if it satisfies the following two conditions:

1. The map f is a bijection.
2. The map f preserves adjacencies: for all $x, y \in V(G)$, we have $\{x, y\} \in E(G)$ if and only if $\{f(x), f(y)\} \in E(H)$.

If there exists an isomorphism between two graphs G and H, then we say that G and H are **isomorphic**. Isomorphic graphs share all structural properties, so when we are only interested in those properties, and not the labels, we sometimes draw the **isomorphism type** of the graph

with vertices unlabeled.

Example 19.2 Figure 6 shows the labelled graphs G and H used in this example.

FIGURE 6

The graphs G and H are completely different graphs, because $V(G) \neq V(H)$ and $E(G) \neq E(H)$. However, the map $f : V(G) \to V(H)$ defined by $f(1) = a$, $f(2) = b$, $f(3) = c$, and $f(4) = d$ is an isomorphism, so G and H are isomorphic. Their common isomorphism type is shown in Figure 7.

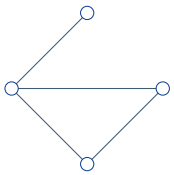


FIGURE 7

Theorem 19.3 (Handshake theorem) The sum of the degrees in a graph is twice the number of edges:

$$\sum_{v \in V(G)} \deg(v) = 2 \cdot |E(G)|.$$

Proof. Count incident pairs (v, e), where v is a vertex and e is an edge incident with v. Counting by vertices gives $\sum_{v \in V(G)} \deg(v)$. Counting by edges gives $2|E(G)|$, since each edge is incident with two vertices. The two counts are equal. □

Definition 19.4 A graph in which all vertices have degree k is called a ***k*-regular graph**.

Example 19.5 Figure 8 shows a 2-regular cycle.

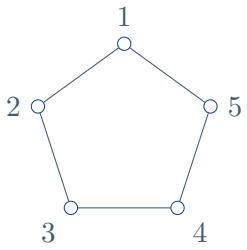


FIGURE 8

Example 19.6 Let Q_n be the graph such that

$$V(Q_n) = \{\text{01-strings of length } n\},$$

and $\sigma, \sigma' \in V(Q_n)$ are adjacent if σ and σ' differ in exactly one position.

For example, Q_2 is shown in Figure 9.

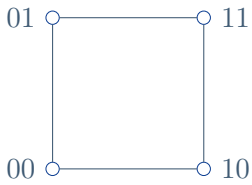


FIGURE 9

The graph Q_1 is shown in Figure 10.

FIGURE 10

The graph Q_3 is shown in Figure 11.

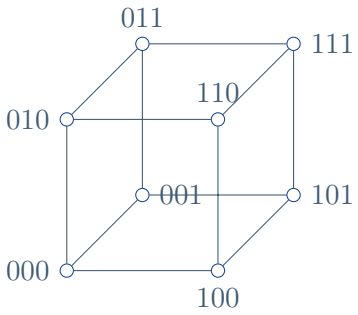


FIGURE 11

The graph Q_n is called the ***n*-cube**. It has $|V(Q_n)| = 2^n$ vertices and $|E(Q_n)| = n2^{n-1}$ edges.

Proof of the second formula. By Theorem 19.3,

$$\sum_{v \in V(Q_n)} \deg(v) = 2|E(Q_n)|.$$

Since every vertex of Q_n has degree n and $|V(Q_n)| = 2^n$, this gives $n2^n = 2|E(Q_n)|$, and hence $|E(Q_n)| = n2^{n-1}$.

Example 19.7 Is there a 3-regular graph with an odd number of vertices?

Solution. Absolutely not. If there were such a graph G, then by Theorem 19.3,

$$3|V(G)| = \sum_{v \in V(G)} \deg(v) = 2|E(G)|,$$

so $|V(G)|$ is even. □

Definition 19.8 For $0 \leq m \leq n$, the **Kneser graph** $KG(n, m)$ has the m-element subsets of $[n]$ as its vertices, with two vertices adjacent precisely when the corresponding subsets are disjoint. The **Petersen graph** is $KG(5, 2)$, drawn in Figure 12.

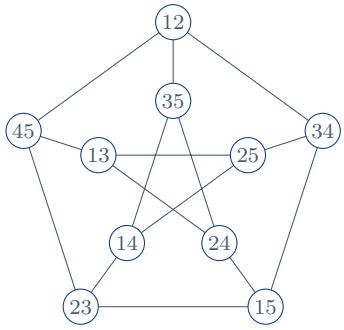
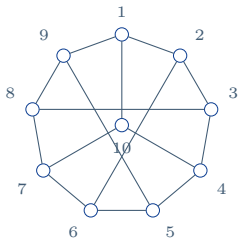


FIGURE 12

To prove that two graphs are isomorphic, exhibit an isomorphism. To prove that two graphs are not isomorphic, find a structural property that distinguishes them. This can be subtle when the graph has many automorphisms, as the Petersen graph does.

Exercise 19.9 The two candidates are displayed in Figure 13. Show that the Petersen graph is isomorphic to the graph on the left,



but not

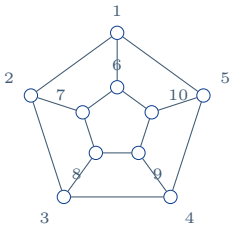


FIGURE 13

Definition 19.10 A **complete graph** K_n is a graph with $|V(K_n)| = n$ and $E(K_n) = \{xy \mid x, y \in V(K_n), x \neq y\}$.

A graph G is **bipartite** if $V(G) = A \cup B$, $A \cap B = \emptyset$, and every edge joins a vertex in A to a vertex in B. Usually both A and B are nonempty. The ordered pair (A, B) is called a **bipartition** of G.

The **complete bipartite graph** $K_{m,n}$ has a bipartition (A, B) where $|A| = m$, $|B| = n$, and every vertex in A is joined to every vertex in B.

Definition 19.11 The **complement** of a graph G is the graph $\overline{G}$ where $V(\overline{G}) = V(G)$ and $E(\overline{G}) = \{xy : x, y \in V(G), x \neq y, xy \notin E(G)\}$.

Example 19.12 Figure 14 shows the complete bipartite graph $K_{3,3}$.

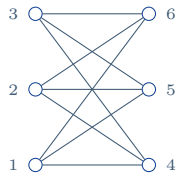


FIGURE 14

Connectivity and Bridges

Lecture 20 — Friday, October 30, 2015

Definition 20.1 A **subgraph** of G is a graph H such that $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. H is said to be **spanning** if $V(H) = V(G)$. H is said to be **induced** if $E(H) = \{xy \in E(G) \mid x \in V(H) \text{ and } y \in V(H)\}$.

Example 20.2 Start with the graph in Figure 15.

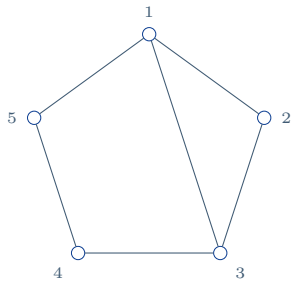


FIGURE 15

Figure 16 shows a subgraph that is neither spanning nor induced.

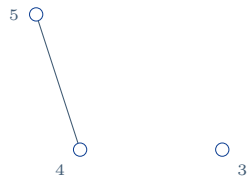


FIGURE 16

Figure 17 shows a subgraph that is induced but not spanning.

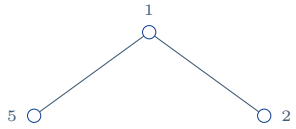


FIGURE 17

Definition 20.3 A **walk** on a graph G is an alternating sequence of vertices and edges $v_0, e_1, v_1, e_2, v_2, \ldots, e_n, v_n$, where $v_0, \ldots, v_n \in V(G)$, $e_1, \ldots, e_n \in E(G)$, and $e_i = \{v_{i-1}, v_i\}$. We say that the walk **starts** at v_0 and **ends** at v_n.

Since we are not allowing multiple edges, we sometimes just write the vertices, as in $v_0 v_1 \ldots v_n$. We do not write down only the edges, since this loses information such as the empty walk. The **length** of the walk $v_0 \ldots v_n$ is n, the number of edges counted with multiplicity. The **reverse walk** is $v_n \ldots v_0$.

Definition 20.4 A **trail** is a walk with no repeated edges, and a **path** is a walk with no repeated vertices. A **closed walk** is a walk $v_0, e_1, v_1, \ldots, e_n, v_n$ with $v_0 = v_n$. A **circuit** is a closed trail. A **cycle** is a closed walk with $v_i \neq v_j$ for $i \neq j$, except for $v_0 = v_n$, with no repeated edges and at least one edge.

In a simple graph, we require a cycle to have length at least 3 (i.e., at least three vertices and three edges), since a cycle of length 1 would be a loop, and a cycle of length 2 would be a pair of parallel edges.

Example 20.5 Figure 18 shows the graph used in this example.

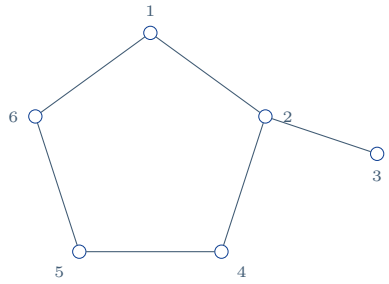
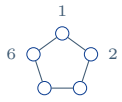


FIGURE 18

In this graph, 12324564 is a walk, 6123 is a path, and 124561 is a cycle.

We also identify cycles and paths with the subgraphs using vertices and edges. In the graph in Example 20.5,  is also called a cycle.

Theorem 20.6 Given a graph G, define a relation P_G on $V(G)$ as follows: uP_Gv if and only if there exists a path from u to v. Then P_G is an equivalence relation on $V(G)$.

Proof. To prove reflexivity, let $v \in V(G)$. There is a path of length 0 from v to v, and therefore vP_Gv.

To prove symmetry, suppose that uP_Gv. The reverse path is a path from v to u, and hence vP_Gu.

To prove transitivity, suppose that uP_Gv and vP_Gw. Let $u = u_0, u_1, \ldots, u_n = v$ be a path from u to v, and let $v = v_0, v_1, \ldots, v_m = w$ be a path from v to w. Then $u_0, u_1, \ldots, u_n, v_1, \ldots, v_m$ is a walk from u to w; we omit v_0 because $v_0 = u_n$. Consider a shortest walk $u = x_0, \ldots, x_k = w$ from u to w. We claim that this walk is a path. If not, then $x_i = x_j$ for some $i < j$, and $x_0, x_1, \ldots, x_i, x_{j+1}, \ldots, x_k$ is a shorter walk from u to w, a contradiction. Hence uP_Gw. □

Definition 20.7 A graph G is **connected** if P_G has exactly one equivalence class.

Definition 20.8 A **component** is the subgraph of G induced by one equivalence class of P_G. Hence a graph is connected if and only if it has one component. In particular, the empty graph is *not* connected, because it has 0 components rather than 1 component.

Equivalently, a component is a maximal connected subgraph.

Example 20.9 Let G be the graph in Figure 19.

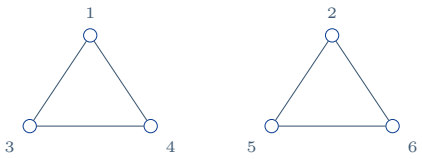


FIGURE 19

Figure 20 shows a subgraph that is not a component of G.

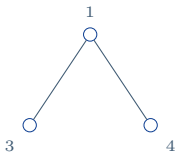


FIGURE 20

Figure 21 shows a component of G.

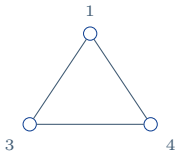


FIGURE 21

Note that the proof of Theorem 20.6 shows that if there is a walk from u to v, then there is a path from u to v. Indeed, if we have a walk $u = u_0, \ldots, u_n = v$, then $u_0 P_G u_1, u_1 P_G u_2, \ldots, u_{n-1} P_G u_n$. By transitivity, $u P_G v$.

Lecture 21 — Monday, November 02, 2015

Example 21.1 Show that Q_n is connected.

Solution. Let $v \in V(Q_n)$. There is a path from v to $000\ldots0$ obtained by changing 1s to 0s, one at a time. Hence every vertex lies in the same component as $000\ldots0$, so Q_n is connected. □

Definition 21.2 Let $X \subseteq V(G)$. The **cut** on X is the subset $\{xy \in E(G) : x \in X, y \notin X\} \subseteq E(G)$.

To show a graph G is connected, the following lemma may come in handy.

Lemma 21.3 Assume that G is not the empty graph. Then G is connected if and only if, for every proper nonempty subset $X \subsetneq V(G)$, the cut on X is nonempty.

Proof. *Forward implication.* Suppose G is connected. Let X be a proper nonempty subset of $V(G)$. Choose $x \in X$ and $y \in V(G) \setminus X$. A path from x to y must leave X, so some edge on that path has one endpoint in X and the other outside X. Hence the cut on X is nonempty.

Reverse implication. Suppose G is not connected, and let H be a component of G. Then $X = V(H)$ is a proper nonempty subset of $V(G)$ whose cut is empty. □

Definition 21.4 Let $e \in E(G)$ be an edge of G. Then $G - e$ is the spanning subgraph of G with $E(G - e) = E(G) \setminus \{e\}$.

Example 21.5 If G is the graph in Figure 22,

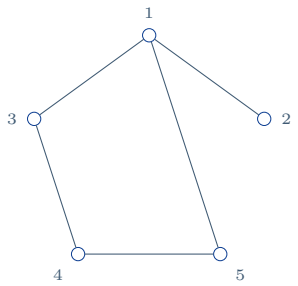


FIGURE 22

then $G - 15$ is the subgraph in Figure 23.

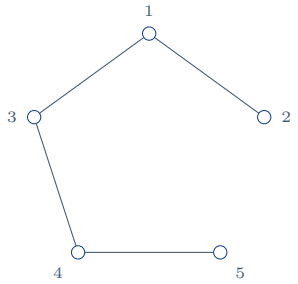


FIGURE 23

Definition 21.6 If G is a connected graph, we say $e \in E(G)$ is a **bridge** (or a **cut-edge**) if $G - e$ is not connected. In general, we say $e \in E(G)$ is a bridge if it is a bridge in its connected component. A typical bridge is highlighted in Figure 24.

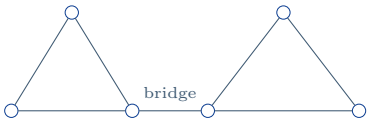


FIGURE 24

Theorem 21.7 If G is a connected graph and $xy = e \in E(G)$ is a bridge, then $G - e$ has exactly two components, one of which contains x, the other of which contains y.

Proof. Let $v \in V(G)$ and consider a path in G from v to x. If e does not appear in this path,

then the same path lies in $G - e$, so v is in the component of x in $G - e$. Otherwise, the path has the form $[v, \ldots, y], e, x$. Deleting the last step gives a path from v to y in $G - e$, so v is in the component of y in $G - e$. Thus every vertex of $G - e$ lies in the component of x or in the component of y. Since e is a bridge, $G - e$ has at least two components. Hence $G - e$ has exactly two components, with x and y in different components. □

Exercise 21.8 Prove that a 4-regular graph has no bridges.

Solution. Suppose instead that e is a bridge in one component of G. Deleting e separates that component into two vertex sets; let X be the one containing one endpoint of e. Summing degrees in the original graph over X gives $4|X|$. Every edge internal to X contributes 2 to this sum, while the bridge e contributes 1, so the same sum is odd. This contradicts the fact that $4|X|$ is even. Hence G has no bridges. □

Theorem 21.9 Let G be a graph, and let $e \in E(G)$. Then e is a bridge if and only if there is no cycle containing e.

Proof. *Forward implication.* Suppose that $e = xy$ lies on a cycle. Then in $G - e$ there is still an x–y path, namely the rest of the cycle. Thus removing e does not disconnect the graph, so e is not a bridge.

Reverse implication. Suppose that e is not on any cycle. If e were not a bridge, then $G - e$ would still contain an x–y path. Adding e to that path would create a cycle containing e, a contradiction. Hence e is a bridge. □

Theorem 21.10 Let G be a graph. The following are equivalent:

1. There exist vertices $u, v \in V(G)$ such that there are two different paths from u to v.
2. There exists a cycle in G.

Proof. (2) *implies* (1). If G contains a cycle, then any two distinct vertices on that cycle are joined by two different paths obtained by going around the cycle in the two possible directions.

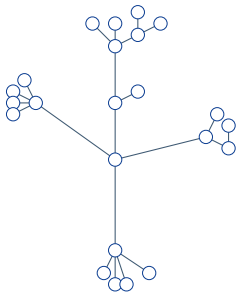
(1) *implies* (2). Let P_1 and P_2 be distinct u–v paths. Follow them from u until the first vertex a at which they take different next edges. Since both paths end at v, they meet again; let b be their first common vertex after this divergence. The a–b segment of P_1 and the a–b segment of P_2 have no internal vertex in common. Traversing one segment from a to b and the other in reverse therefore gives a cycle. □

Example 21.11 In Exercise 19.9, the pentagonal prism graph on the right contains a cycle of length 4, whereas the Petersen graph does not. This proves that the two graphs are not isomorphic.

Trees and Spanning Trees

Definition 21.12 A **tree** is a connected graph with no cycles. A **forest** is a graph in which every component is a tree.

Example 21.13 Figure 25 shows two trees.



and

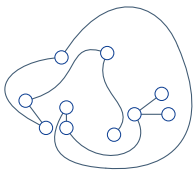


FIGURE 25

By contrast, Figure 26 shows a graph that is not a tree.

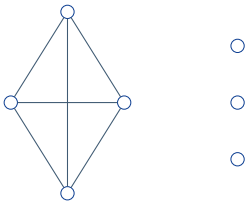


FIGURE 26

Lecture 22 — Wednesday, November 04, 2015

Theorem 22.1 Let T be a connected graph with p vertices and q edges. The following are equivalent:

1. T is a tree.
2. Every edge of T is a bridge.
3. There is exactly one path joining any pair of vertices.
4. $q = p - 1$.

Proof. (1) *is equivalent to* (2). By Theorem 21.9, an edge is not a bridge precisely when it is contained in a cycle. Hence T has no cycles if and only if every edge of T is a bridge.

(1) *is equivalent to* (3). If there were two distinct paths between the same pair of vertices, then Theorem 21.10 would give a cycle. Conversely, any cycle gives two distinct paths between two vertices on the cycle.

(1) *implies* (4). We prove this by strong induction on p. If $p = 1$, then T has no edges, so $q = 0 = 1 - 1$.

Assume that $p > 1$ and that the result holds for trees with fewer than p vertices. Since T is connected, there exists an edge $e \in E(T)$. By (2), the edge e is a bridge, so $T - e$ has two components, say T_1 and T_2. Deleting an edge cannot create a cycle, so T_1 and T_2 are trees. Let $|V(T_1)| = p_1$ and $|V(T_2)| = p_2$. By the inductive hypothesis, $|E(T_1)| = p_1 - 1$ and $|E(T_2)| = p_2 - 1$. Therefore

$$|E(T)| = |E(T_1)| + |E(T_2)| + 1 = (p_1 - 1) + (p_2 - 1) + 1 = p - 1.$$

(4) *implies* (2). Suppose to the contrary that $q = p - 1$ and that T has an edge e that is not a bridge. Then $T - e$ is connected and has fewer than $p - 1$ edges. Continue removing edges that are not bridges until every remaining edge is a bridge. The resulting connected graph is a tree with fewer than $p - 1$ edges, contradicting what we have just proved. Hence every edge of T is a bridge. □

Exercise 22.2 Let F be a forest with p vertices and q edges and c components. Prove that $q = p - c$.

Definition 22.3 A **spanning tree** of a graph G is a spanning subgraph that is a tree.

Theorem 22.4 G is connected if and only if G has a spanning tree.

Proof. *Forward implication.* If G has a spanning tree, then G is connected because the spanning tree contains all vertices and is connected.

Reverse implication. Assume that G is connected. If G has no cycle, then G itself is a spanning tree. Otherwise, remove one edge from a cycle; connectivity is preserved because the remaining edges of the cycle still give a path between the endpoints of the removed edge. Repeat until no cycle remains. Since G has finitely many edges, the process terminates and yields a connected spanning subgraph with no cycles, namely a spanning tree. □

Theorem 22.5 Let T be a tree with $p \geq 2$ vertices. Let n_i be the number of vertices of degree i. Then the following hold:

1. $n_0 = 0$.
2. $$\sum_{i \geq 1} n_i = p \tag{22.1}$$
3. $$\sum_{i \geq 1} i n_i = 2(p-1) \tag{22.2}$$

Proof. For (1), if $n_0 > 0$, then there is an isolated vertex, contradicting the fact that T is connected.

For (2), we are counting the total number of vertices.

For (3), we are counting the total degree of all vertices, which is twice the number of edges by Theorem 19.3. Since T has $p - 1$ edges, the total degree is $2(p-1)$. □

Taking twice (22.1) minus (22.2) gives

$$\sum_{i\geq 1}(2-i)n_i = 2,$$

and hence

$$n_1 = 2 + \sum_{i\geq 3}(i-2)n_i.$$

Notice that n_2 does not appear in this equation. This is related to **edge subdivision**, which we return to later.

Definition 22.6 A vertex of degree 1 in a tree is called a **leaf**.

Corollary 22.7 Every tree with at least two vertices has at least two leaves.

Example 22.8 How many trees are there? There are 16 trees with vertex set $[4]$, falling into the two isomorphism types shown in Figure 27.

FIGURE 27

It is not a coincidence that the number of trees with vertex set $[4]$ is $4^{4-2} = 16$, as we will see.

Lecture 23 — Friday, November 06, 2015

Definition 23.1 A **rooted tree** is a pair (T, r) where T is a tree and $r \in V(T)$. Here, r is called a **root**.

We can mark the root in a drawing, or we can draw a rooted tree by drawing arrows on the edges of the tree, with all arrows pointing towards the root. The two representations in Figure 28 describe the same rooted tree.

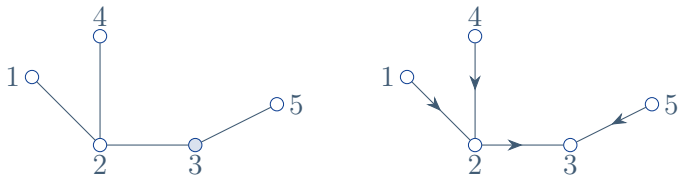


Figure 28

In the first drawing, the root is the shaded vertex. In the second drawing, we can identify the root by looking for the vertex with no arrows leaving it.

The orientation in Figure 29 has no root. We can detect this by checking whether any vertex has two arrows leaving it.

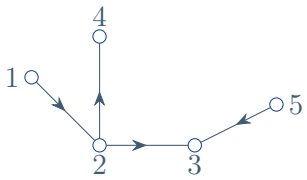


Figure 29

Theorem 23.2 (Cayley's formula) For $p \geq 1$, there are p^{p-2} trees with vertex set $[p]$.

Proof. The case $p = 1$ is immediate, so assume $p \geq 2$. We use Prüfer codes. Given a tree T on $[p]$, repeatedly delete its smallest leaf and record that leaf's unique neighbour. Stop when two vertices remain. This produces a sequence

$$(a_1, \ldots, a_{p-2}) \in [p]^{p-2}.$$

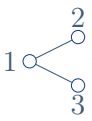
We reconstruct T from the sequence. At each stage, let v be the smallest available label that does not appear in the remaining sequence. Join v to the first remaining entry, delete v from the available labels, and delete that first entry from the sequence. When the sequence is empty, join the two remaining labels.

To justify the key claim, observe that at every stage a current vertex occurs in the remaining code exactly one fewer time than its current degree: it is recorded once for each incident edge removed before the vertex itself is deleted. Consequently, the labels absent from the remaining code are

precisely the current leaves. The chosen v is therefore exactly the smallest leaf that would be deleted at that stage of the encoding. Hence decoding reverses encoding, so trees on $[p]$ are in bijection with the p^{p-2} sequences in $[p]^{p-2}$. □

Adjacency Matrices and Graph Search

Definition 23.3 Let G be a simple graph with $V(G) = [p]$. The **adjacency matrix** of G is the $p \times p$ matrix $\mathbf{A} = \mathbf{A}(G)$ given by $\mathbf{A}_{u,v} = \mathbb{1}_{\{uv \in E(G)\}}$. The assumption $V(G) = [p]$ is only a convenient labelling convention.

Example 23.4 For the graph G: 

, the adjacency matrix is $\mathbf{A}(G) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

$\operatorname{diag}(\mathbf{A}) = 0$ always and $\mathbf{A}$ is a real symmetric matrix. Hence $\mathbf{A}$ is diagonalizable over $\mathbb{R}$.

Theorem 23.5 Let G be a k-regular graph with adjacency matrix $\mathbf{A}$. Then k is an eigenvalue of $\mathbf{A}$, with multiplicity equal to the number of components in G. Here algebraic and geometric multiplicities are equal.

Proof. We will show the following: if $\mathbf{A}\boldsymbol{x} = k\boldsymbol{x}$ where $\boldsymbol{x} = (x_1, \ldots, x_p)^\top$, then $\boldsymbol{x}_u = \boldsymbol{x}_v$ for all $uv \in E(G)$. It follows that

$$\dim\{\boldsymbol{x} \in \mathbb{R}^p \mid \mathbf{A}\boldsymbol{x} = k\boldsymbol{x}\} = \text{the number of components of } G,$$

as required. In fact, we will prove the converse as well.

Reverse implication. Write $\mathbf{A}\boldsymbol{x} = \boldsymbol{y} = (y_1, \ldots, y_p)^\top$. Then

$$\boldsymbol{y}_u = \sum_{v=1}^{p} \mathbf{A}_{u,v}\boldsymbol{x}_v = \sum_{v : uv \in E(G)} \boldsymbol{x}_v.$$

If $\boldsymbol{x}_u = \boldsymbol{x}_v$ for every edge $uv \in E(G)$, then

$$\boldsymbol{y}_u = \sum_{v : uv \in E(G)} \boldsymbol{x}_u = \boldsymbol{x}_u \cdot \deg(u) = k\boldsymbol{x}_u,$$

and so $\mathbf{A}\boldsymbol{x} = \boldsymbol{y} = k\boldsymbol{x}$.

Forward implication. Suppose $\mathbf{A}\boldsymbol{x} = k\boldsymbol{x}$, and fix a component H of G. Choose $u \in V(H)$ for which $\boldsymbol{x}_u$ is maximal on H. Then

$$k\boldsymbol{x}_u = \sum_{v:uv\in E(G)} \boldsymbol{x}_v \leq k\boldsymbol{x}_u.$$

Equality forces every neighbour v of u to satisfy $\boldsymbol{x}_v = \boldsymbol{x}_u$. Applying the same argument successively along paths from u shows that $\boldsymbol{x}$ is constant on all of H. Since this holds for every component, $\boldsymbol{x}_u = \boldsymbol{x}_v$ for every edge uv. □

Lecture 24 — Monday, November 09, 2015

Example 24.1 We use the Petersen graph in Figure 30 to compute a concrete adjacency spectrum.

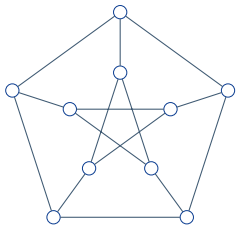


FIGURE 30

Let P be the Petersen graph, and let $\mathbf{A} = \mathbf{A}(P)$ be its adjacency matrix. We determine the eigenvalues of $\mathbf{A}$ and their multiplicities. By Theorem 23.5, 3 is an eigenvalue, with eigenvector $\mathbf{1}$. What else?

We claim that $\mathbf{A}^2 + \mathbf{A} = 2\mathbf{I} + \mathbf{J}$, where $\mathbf{J}$ is the all-ones matrix $\mathbf{J} = \mathbf{1}\mathbf{1}^\top$.

To verify the claim, first note that

$$(\mathbf{A}^2)_{uv} = \sum_{w=1}^{10} \mathbf{A}_{uw}\mathbf{A}_{wv}.$$

Note that $\mathbf{A}_{uw}\mathbf{A}_{wv} = 1$ if uwv is a walk and is 0 otherwise. Thus $(\mathbf{A}^2)_{uv}$ is the number of walks of length 2 from u to v, while $\mathbf{A}_{uv}$ is the number of walks of length 1 from u to v. More generally, an induction shows that $(\mathbf{A}^k)_{uv}$ is the number of walks of length k from u to v. Therefore, $(\mathbf{A}^2 + \mathbf{A})_{uv}$ is the number of walks of length 1 or 2 from u to v.

Now

$$(2\mathbf{I} + \mathbf{J})_{uv} = \begin{cases} 3 & \text{if } u = v \\ 1 & \text{if } u \neq v \end{cases}.$$

The claim says that if $u = v$, then there are three walks of length 1 or 2 from u to v, and if $u \neq v$, then there is one walk of length 1 or 2 from u to v.

This follows directly from the Kneser description $P = KG(5, 2)$. Each vertex has three disjoint 2-subsets as neighbours, giving three length-2 closed walks. If two distinct vertices are adjacent, their 2-subsets are disjoint and no 2-subset is disjoint from both, so the edge between them is their unique walk of length 1 or 2. If they are nonadjacent, their subsets meet in one element; the complement of their three-element union is the unique 2-subset disjoint from both, and hence determines their unique length-2 walk. This proves $\mathbf{A}^2 + \mathbf{A} = 2\mathbf{I} + \mathbf{J}$.

We can now use this identity to determine the remaining eigenvalues. Let $\lambda \neq 3$ be an eigenvalue of $\mathbf{A}$, and let $\boldsymbol{x}$ be a nonzero eigenvector such that $\mathbf{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. Since $\mathbf{A}$ is symmetric, $\mathbf{1}^\top\mathbf{A} = 3\mathbf{1}^\top$. Therefore $\mathbf{1}^\top\mathbf{A}\boldsymbol{x} = 3\mathbf{1}^\top\boldsymbol{x}$, while also $\mathbf{1}^\top\mathbf{A}\boldsymbol{x} = \lambda\mathbf{1}^\top\boldsymbol{x}$. Since $\lambda \neq 3$, we must have $\mathbf{1}^\top\boldsymbol{x} = 0$. Hence $\mathbf{J}\boldsymbol{x} = \mathbf{1}\mathbf{1}^\top\boldsymbol{x} = \mathbf{0}$. Now,

$$\mathbf{A}^2\boldsymbol{x} + \mathbf{A}\boldsymbol{x} = 2\mathbf{I}\boldsymbol{x} + \mathbf{J}\boldsymbol{x},$$

so $\lambda^2\boldsymbol{x} + \lambda\boldsymbol{x} = 2\boldsymbol{x}$. Since $\boldsymbol{x} \neq 0$, we have $\lambda^2 + \lambda - 2 = 0$, and therefore $\lambda = 1$ or $\lambda = -2$.

Thus the eigenvalues of $\mathbf{A}$ are in $\{3, 1, -2\}$. To determine the multiplicities, use

$$\sum_{\lambda \in \text{spec}(\mathbf{A})} \lambda = \text{tr}(\mathbf{A}) = 0.$$

Since $\mathbf{A}$ has ten eigenvalues counted with multiplicity and 3 has multiplicity 1, the only possibility is

$$3, 1, 1, 1, 1, 1, -2, -2, -2, -2.$$

Definition 24.2 Let G be a graph. A **Hamiltonian cycle** is a subgraph of G which is a spanning subgraph and also a cycle.

Theorem 24.3 The Petersen graph P has no Hamiltonian cycle.

Proof. Suppose to the contrary that P has a Hamiltonian cycle H. Let $\mathbf{C}$ be the adjacency matrix of H. If we delete the edges in H from P, we are left with the 1-regular subgraph M shown in Figure 31. Let $\mathbf{B} = \mathbf{A}(M)$. Then $\mathbf{A} = \mathbf{B} + \mathbf{C}$. The graph M is a matching with five

FIGURE 31

edges, so, after reordering the vertices, $\mathbf{B}$ is block diagonal with five blocks of the form $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Hence the eigenvalues of $\mathbf{B}$ are $1, 1, 1, 1, 1, -1, -1, -1, -1, -1$. Recall from Example 24.1 that the eigenvalues of $\mathbf{A}$ are $3, 1, 1, 1, 1, 1, -2, -2, -2, -2$.

Let $V = \{\boldsymbol{x} \in \mathbb{R}^{10} : \mathbf{1}^\top \boldsymbol{x} = 0\}$, with $\dim(V) = 9$. Let $E = \{\boldsymbol{x} \in \mathbb{R}^{10} : \mathbf{A}\boldsymbol{x} = \boldsymbol{x}\}$ and $\tilde{E} = \{\boldsymbol{x} \in \mathbb{R}^{10} : \mathbf{B}\boldsymbol{x} = -\boldsymbol{x}\}$. We showed that $E \subseteq V$, and a similar argument shows $\tilde{E} \subseteq V$. Since $\dim(E) = 5 = \dim(\tilde{E})$, these two 5-dimensional subspaces lie inside the 9-dimensional space V, so there exists a nonzero vector $\boldsymbol{y} \in E \cap \tilde{E}$. Now

$$\mathbf{C}\boldsymbol{y} = \mathbf{A}\boldsymbol{y} - \mathbf{B}\boldsymbol{y} = \boldsymbol{y} - (-\boldsymbol{y}) = 2\boldsymbol{y}.$$

Thus $\mathbf{C}\boldsymbol{y} = 2\boldsymbol{y}$ and $\mathbf{C}\mathbf{1} = 2\mathbf{1}$. Since $\boldsymbol{y} \in V$, we have $\boldsymbol{y} \perp \mathbf{1}$, so $\boldsymbol{y}$ and $\mathbf{1}$ are linearly independent. Hence 2 is an eigenvalue of $\mathbf{C}$ with multiplicity at least 2. By Theorem 23.5, this means that H has at least two components, contradicting the fact that H is a cycle. □

Theorem 24.4 It is not possible to partition the edges of K_{10} into three subgraphs, all isomorphic to the Petersen graph P. That is, there do not exist edge-disjoint subgraphs P_1, P_2, P_3 such that

$$E(K_{10}) = E(P_1) \,\dot\cup\, E(P_2) \,\dot\cup\, E(P_3) \qquad \text{and} \qquad P_1 \cong P_2 \cong P_3 \cong P.$$

Proof. Suppose to the contrary that such a decomposition exists, and let $\mathbf{A}_i = \mathbf{A}(P_i)$ for $i = 1, 2, 3$. Since the edge sets partition $E(K_{10})$, we have

$$\mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3 = \mathbf{J} - \mathbf{I},$$

where $\mathbf{J}$ is the all-ones matrix.

Let

$$V = \{\boldsymbol{x} \in \mathbb{R}^{10} : \mathbf{1}^\top \boldsymbol{x} = 0\}.$$

By Example 24.1, each $\mathbf{A}_i$ has eigenvalue 1 with multiplicity 5, and the corresponding eigenspace is contained in V. Let

$$E_i = \{\boldsymbol{x} \in V : \mathbf{A}_i \boldsymbol{x} = \boldsymbol{x}\}.$$

Then $\dim(E_1) = \dim(E_2) = 5$, while $\dim(V) = 9$. Hence

$$\dim(E_1 \cap E_2) \geq \dim(E_1) + \dim(E_2) - \dim(V) = 1.$$

Choose a nonzero vector $\boldsymbol{x} \in E_1 \cap E_2$. Since $\boldsymbol{x} \in V$, we have $(\mathbf{J} - \mathbf{I})\boldsymbol{x} = -\boldsymbol{x}$. Therefore

$$\mathbf{A}_3 \boldsymbol{x} = (\mathbf{J} - \mathbf{I})\boldsymbol{x} - \mathbf{A}_1 \boldsymbol{x} - \mathbf{A}_2 \boldsymbol{x} = -\boldsymbol{x} - \boldsymbol{x} - \boldsymbol{x} = -3\boldsymbol{x}.$$

This says that -3 is an eigenvalue of $\mathbf{A}_3$, contradicting the spectrum of the Petersen graph computed in Example 24.1. □

Lecture 25 — Wednesday, November 11, 2015

Recall from Theorem 24.3 that the Petersen graph has no Hamiltonian cycle. Its spectral proof uses the same dimension-intersection mechanism as Theorem 24.4; we will now turn from spectral arguments to graph search.

Definition 25.1 A **breadth first search tree** (**BFST**) is a rooted tree produced by the **BFST algorithm**. An ambient graph on which to run the algorithm is shown in Figure 32.

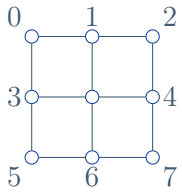


FIGURE 32

Algorithm 1: BFST Algorithm

Input: a graph G and a chosen root vertex r.
Output: a rooted breadth first search tree T.

1. Start with a queue containing only the root r.
2. Start the output tree with the root alone.
3. Take the first vertex in the queue and inspect its neighbors that have not yet appeared in the tree.
4. Add those new neighbors as children of the active vertex.
5. Pop the queue, append the new neighbors to the queue, and repeat until the queue is empty.

Definition 25.2 The vertices currently in the queue are called **unexhausted**. The vertex at the head of the queue is **active**. The **parent** of a vertex x is the active vertex such that x is joined in the tree (note: the root has no parent). The **level** of x is defined by $\text{level}(\text{root}) = 0$ and $\text{level}(x) = \text{level}(\text{parent}(x)) + 1$.

Theorem 25.3 A breadth first search tree is a tree. Moreover, it is spanning if and only if the ambient graph is connected.

Proof. By induction, at each stage of the breadth first search algorithm, T is connected and $|E(T)| = |V(T)| - 1$, so T is a tree.

For the spanning assertion, first suppose that T is spanning. Since T is connected and contains all vertices of G, the graph G is connected.

Conversely, suppose that G is connected and that T is a breadth first search tree with root r. Given $v \in V(G)$, let $v = v_0, v_1, \ldots, v_k = r$ be a path from v to r. If $v \notin V(T)$, then there is an index i such that $v_i \notin V(T)$ but $v_{i+1} \in V(T)$. When v_{i+1} was active in the breadth first search algorithm, all of its undiscovered neighbors, including v_i, were added to T, a contradiction. Hence T is spanning. □

Lemma 25.4 Suppose $x \in V(G)$ is active. If $\text{level}(x) = k$, then all vertices in the queue have level k or $k + 1$.

Proof. In breadth first search, vertices are processed in queue order. When a level-k vertex is active, every earlier level-k vertex has already been processed, and any newly discovered neighbor is assigned level $k+1$ and appended to the queue. No vertex of level less than k can remain in the queue, and no vertex of level greater than $k+1$ can be present before the level-k vertices are processed. Hence all queued vertices have level k or $k+1$. □

Theorem 25.5 Let G be a connected graph and let T be a breadth first search tree. If $e = xy \in E(G)$, then $|\text{level}(x) - \text{level}(y)| \leq 1$.

Proof. Without loss of generality, assume that x is active before y. Let $k = \text{level}(x)$. If y is not in the queue when x is active, then y is not in $V(T)$ at that time, so x becomes the parent of y. On the other hand, if y *is* in the queue when x is active, then $\text{level}(y) = k$ or $\text{level}(y) = k+1$ by Lemma 25.4. □

Definition 25.6 Let G be a connected graph, and let $x, y \in V(G)$. The **distance** between x and y is the length of the shortest path from x to y.

Theorem 25.7 Let G be connected and let T be a breadth first search tree rooted at x. Then the unique path from x to y in T is a shortest path; in particular, $\text{dist}(x, y) = \text{level}(y)$.

Proof. The path $y, \text{parent}(y), \text{parent}(\text{parent}(y)), \ldots, x$ is a path from y to x of length $\text{level}(y)$. Suppose that $y = v_0, v_1, \ldots, v_k = x$ is any other path from y to x. Then

$$\begin{aligned}
\text{level}(y) &= |\text{level}(y) - \text{level}(x)| && \text{since } \text{level}(x) = 0 \\
&= \left| \sum_{i=1}^{k} \big(\text{level}(v_{i-1}) - \text{level}(v_i)\big) \right| \\
&\leq \sum_{i=1}^{k} |\text{level}(v_{i-1}) - \text{level}(v_i)| && \text{by the triangle inequality} \\
&\leq \sum_{i=1}^{k} 1 && \text{since } T \text{ is a breadth first search tree} \\
&= k.
\end{aligned}$$

□

Lecture 26 — Friday, November 13, 2015

Definition 26.1 The **girth** of a graph is the length of the shortest cycle.

Theorem 26.2 Let G be a connected graph containing a cycle. For each $r \in V(G)$, construct a breadth first search tree T_r. Let

$$m_r = \min\{\text{level}_{T_r}(x) + \text{level}_{T_r}(y) + 1 : xy \in E(G) \setminus E(T_r)\}.$$

Then

$$\text{girth}(G) = \min\{m_r : r \in V(G)\}.$$

Proof. Let $r \in V(G)$ and $xy \in E(G) \setminus E(T_r)$. First we show that

$$\text{level}_{T_r}(x) + \text{level}_{T_r}(y) + 1 \geq \text{girth}(G).$$

Hence it will follow that $\min\{m_r\} \geq \text{girth}(G)$. Consider the walk

$$x, \text{pr}(x), \text{pr}(\text{pr}(x)), \ldots, r, \ldots, \text{pr}(\text{pr}(y)), \text{pr}(y), y, x.$$

Let H be the subgraph of G defined by the vertices and edges in this walk. The number of edges in H satisfies

$$|E(H)| \leq \text{length of walk} = \text{level}(x) + \text{level}(y) + 1$$

and H contains two different x–r paths: the tree path from x to r, and the path $x, y, \text{pr}(y), \ldots, r$. These paths are different because $xy \notin E(T_r)$. Therefore H contains a cycle whose length is at most $|E(H)|$. Thus

$$\text{girth}(G) \leq |E(H)| \leq \text{level}_{T_r}(x) + \text{level}_{T_r}(y) + 1.$$

Since this holds for every r and every $xy \in E(G) \setminus E(T_r)$, we get

$$\text{girth}(G) \leq \min_r m_r.$$

For the reverse inequality, let $C = v_0 v_1 \cdots v_{g-1} v_0$ be a shortest cycle, where $g = \text{girth}(G)$, and choose $r = v_0$. In T_r, not all edges of C can be tree edges (a tree has no cycle), so pick $v_i v_{i+1} \in E(C) \setminus E(T_r)$. Then

$$m_r \leq \text{level}_{T_r}(v_i) + \text{level}_{T_r}(v_{i+1}) + 1.$$

Also $\text{level}_{T_r}(v_j) = \text{dist}_G(r, v_j) \leq \text{dist}_C(r, v_j)$. The two distances in C from r to the adjacent vertices v_i and v_{i+1} sum to at most $g - 1$, so

$$\text{level}_{T_r}(v_i) + \text{level}_{T_r}(v_{i+1}) + 1 \leq g.$$

Therefore $\min_r m_r \leq g = \text{girth}(G)$. Combining both inequalities gives

$$\text{girth}(G) = \min\{m_r : r \in V(G)\}.$$

□

Here is one last application of breadth first search trees:

Theorem 26.3 Let G be a connected graph, and let T be a breadth first search tree of G. The following are equivalent:

1. G is bipartite.
2. G has no cycles of odd length.
3. For every $xy \in E(G)$, $\text{level}(x) \neq \text{level}(y)$.

Proof. (1) *implies* (2). Let (A, B) be a bipartition.

Suppose $v_0, v_1, v_2, \ldots, v_n = v_0$ is a cycle. Consecutive vertices alternate between A and B. Without loss of generality, suppose $v_0 \in A$. Since $v_n = v_0 \in A$, the number of edges in the cycle is even.

(2) *implies* (3). Suppose to the contrary that there is an edge $xy \in E(G)$ such that $\text{level}(x) = \text{level}(y)$. Following the two tree paths toward the root until their first common vertex, and then adding xy, gives a closed walk of odd length. Every odd closed walk contains an odd cycle: a shortest odd closed walk cannot repeat an internal vertex, since splitting at a repetition would produce a shorter odd closed walk. This contradicts (2).

(3) *implies* (1). Let $A = \{x \in V(G) : \text{level}(x) \text{ is even}\}$ and $B = \{y \in V(G) : \text{level}(y) \text{ is odd}\}$. By Theorem 25.5 and (3), every edge of G joins a vertex in A to a vertex in B, since the levels of adjacent vertices differ by exactly 1. Then (A, B) is a bipartition. □

Planar Graphs

Definition 26.4 A **drawing** of a graph in the plane is a **planar embedding** if no edges cross. A graph is **planar** if a planar embedding exists.

Example 26.5 The planar embedding of the 3-cube Q_3 in Figure 33 proves that Q_3 is planar.

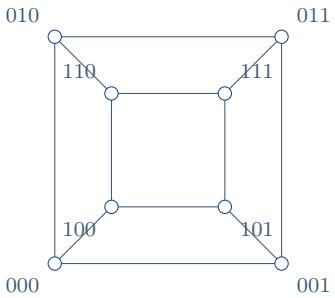


Figure 33

In contrast, Q_4, K_5, and $K_{3,3}$ are not planar. Try to draw drawings of them without crossings, and fail miserably.

The question is how to prove that a graph cannot be drawn without crossings.

Definition 26.6 A planar embedding divides the plane into regions called **faces**. Two faces are called **adjacent** if they are incident with a common edge. A **boundary** of a face is the subgraph consisting of all vertices and edges incident with the face.

Definition 26.7 The **degree** of a face is the number of boundary edges that are not bridges plus twice the number of bridges in the boundary.

Euler's Formula and Edge Bounds

Lecture 27 — Monday, November 16, 2015

We now explain why bridges must be treated separately when counting face degrees, using the example in Figure 34.

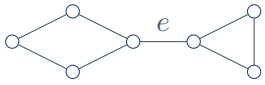


FIGURE 34

There are two possibilities.

1. If the trace ends on the other side of e, then we have traced a closed path γ in the plane. One end of e lies inside γ, the other lies outside, and no other edges are crossed. Hence $P - e$ has one end of e inside γ and one end outside γ, so e is a bridge.
2. If the trace returns to the same side of e, then we have traced a cycle in P containing e, so e is not a bridge.

In (1), the trace moves from one side of e to the other inside a face, so e is incident with only one face. In (2), the trace runs around the entire boundary of a face and does not reach the other side of e, so e is incident with two faces.

We have the following dichotomy:

Theorem 27.1 An edge e is a bridge if and only if e is incident with only one face. An edge e is not a bridge if and only if it is incident with two faces.

A complete proof of Theorem 27.1 uses the Jordan curve theorem. It is too hard.

Figure 35 shows fundamentally different planar embeddings of the same graph. The face degrees

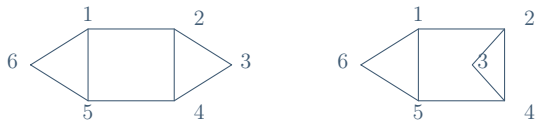


FIGURE 35

in the first embedding are $6, 3, 4, 3$, while the face degrees in the second embedding are $5, 3, 5, 3$. In both cases, the sum is the same.

Theorem 27.2 (Handshake theorem for faces (or "faceshake theorem")) For a planar embedding with q edges and faces $f_1, f_2, \ldots, f_s$, we have

$$\sum_{i=1}^{s} \deg(f_i) = 2q.$$

Proof. Count pairs (f, e) where f is a face and e is an edge incident with f, but count bridges twice. Let

$$A = \{(f, e) : f \text{ is a face and } e \text{ is not a bridge and is incident with } f\}$$

and

$$B = \{(f, e) : f \text{ is a face and } e \text{ is a bridge incident with } f\}.$$

Then

$$2q = \underbrace{|A| + 2|B|}_{\text{counting by edges}} = \underbrace{\sum_{i=1}^{s} \deg(f_i)}_{\text{counting by faces}}.$$

□

Now for the most fundamental theorem in topology [citation needed]:

Theorem 27.3 (Euler's formula) For a planar embedding with p vertices, q edges, s faces, and c components, we have that

$$p - q + s = c + 1.$$

Proof. We prove the formula by induction on q.

If $q = 0$, then the embedding has no edges, so $s = 1$ and $p = c$. Hence $p - q + s = c + 1$.

Now suppose $q > 0$ and that the result holds for planar embeddings with $q - 1$ edges. Let P be a planar embedding with p vertices, q edges, s faces, and c components. Choose an edge $e \in E(P)$ and consider $P - e$.

If e is a bridge, then e is incident with one face, so $P - e$ has p vertices, $q - 1$ edges, s faces, and

$c+1$ components. By the inductive hypothesis,

$$p-(q-1)+s=(c+1)+1,$$

and therefore $p-q+s=c+1$.

If e is not a bridge, then e is incident with two faces, and those two faces merge when we delete e. Thus $P-e$ has p vertices, $q-1$ edges, $s-1$ faces, and c components. By the inductive hypothesis,

$$p-(q-1)+(s-1)=c+1,$$

and therefore $p-q+s=c+1$. □

Lecture 28 — Wednesday, November 18, 2015

Theorem 28.1 Let P be a planar embedding that contains a cycle. Then the following hold:

1. The boundary of every face contains a cycle.
2. The degree of every face is at least girth(P).

Proof. For (1), let $f \in F(P)$. In a forest, we have $q=p-c$. P contains a cycle, so it is not a forest, and so $q \geq p-c+1$. By Euler's formula, $s=c+1-p+q \geq 2$. Therefore f cannot be the whole plane, so f has a boundary, say Q. Then f is also a face of Q, and because it is not the whole plane, Q has at least two faces. Hence Q contains a cycle, and that cycle is contained in the boundary of f.

For (2), we have just shown that the boundary of f contains a cycle, and every cycle has length at least girth(P), so the degree of f is at least girth(P). □

Example 28.2 Show that K_5 is not planar.

Solution. Suppose that P is a planar embedding of K_5. Since K_5 has girth 3, Theorem 28.1

implies that every face of P has degree at least 3. By the handshake theorem for faces, $2q = 2 \cdot 10 = 20 \geq 3s$. By Euler's formula, $s = 2 - p + q = 2 - 5 + 10 = 7$. Hence $20 \geq 3s = 21$, a contradiction. □

Theorem 28.3 Let G be a connected simple planar graph with $p \geq 3$ vertices and q edges. Then

$$q \leq 3p - 6.$$

Proof. First suppose that G has no cycles. Then G is a forest with $p \geq 3$ vertices, so $q = p - 1 < 3p - 6$.

Now suppose that G has a cycle. Let P be a planar embedding of G. By Theorem 28.1, we have

$$2q = \sum_{f \in F(P)} \deg(f) \geq 3s,$$

so $s \leq 2q/3$. Since G is connected, Euler's formula gives

$$p - q + s = 2,$$

so

$$q \leq p + s - 2 \leq p + \frac{2q}{3} - 2,$$

and rearranging yields the claim. □

Consider connected simple planar embeddings for which every vertex has the same degree $d \geq 2$ and every face has the same degree d^*. Such an embedding contains a cycle, so $d^* \geq 3$. By the handshake theorem for vertices and the handshake theorem for faces, we have $pd = 2q = sd^*$. By Euler's formula, $p - q + s = 2$, since these graphs are connected. Eliminating p and s, we obtain

$$\frac{2}{d} - 1 + \frac{2}{d^*} = \frac{2}{q},$$

and some tedious rearranging gives

$$\frac{d(d^* - 2)}{2d^*} = \frac{p - 2}{p} < 1,$$

and in particular,

$$(d - 2)(d^* - 2) < 4.$$

There are six possibilities for the pair (d, d^*):

1. $d = 2$ and $d^* \geq 3$. This corresponds to a plain old cycle, shown in Figure 36.

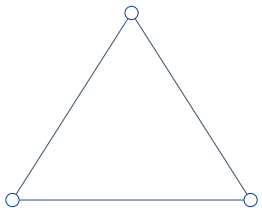


FIGURE 36

2. $d = 3$ and $d^* = 3$. This corresponds to a **tetrahedron**, shown in Figure 37.

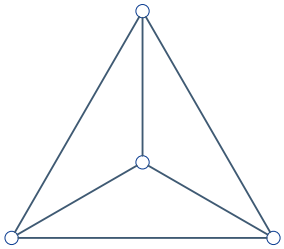


FIGURE 37

3. $d = 3$ and $d^* = 4$. This corresponds to a **cube**, shown in Figure 38.

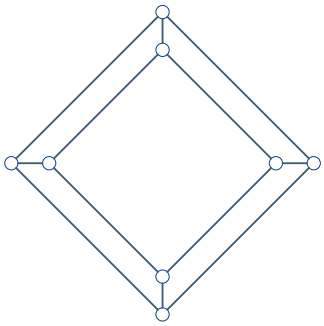


FIGURE 38

4. $d = 4$ and $d^* = 3$. This corresponds to an **octahedron**, shown in Figure 39.

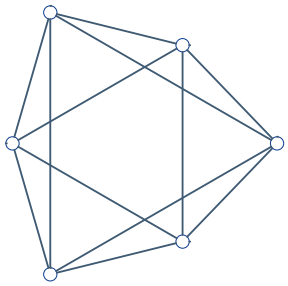


FIGURE 39

5. $d = 3$ and $d^* = 5$. This corresponds to a **dodecahedron**, shown in Figure 40.

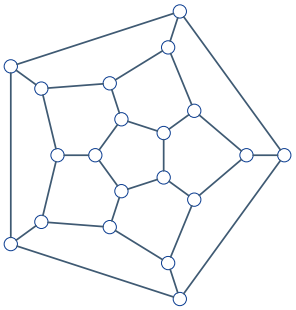


FIGURE 40

6. $d = 5$ and $d^* = 3$. This corresponds to an **icosahedron**, shown in Figure 41.

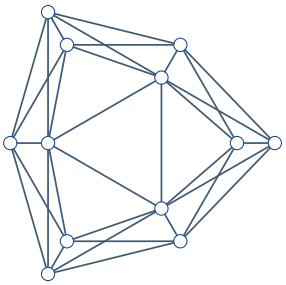


FIGURE 41

The last five correspond to the **five Platonic solids**.

Theorem 28.4 Let G be a connected simple planar graph with p vertices, q edges, and girth k. Then

$$q \leq \frac{k(p-2)}{k-2}.$$

Proof. Let P be a plane embedding of G with s faces. Every face boundary contains a cycle, and every cycle has length at least k, so every face has degree at least k. Hence

$$ks \leq \sum_{f \in F(P)} \deg(f) = 2q.$$

Since G is connected, Theorem 27.3 gives $s = 2 - p + q$. Substituting,

$$k(2 - p + q) \leq 2q,$$

and rearranging yields the claim. □

Note that the converse of Theorem 28.4 is false. For example, attach one leaf to $K_{3,3}$. The resulting graph is still nonplanar, since it contains $K_{3,3}$ as a subgraph, but it has $p = 7$, $q = 10$, and girth 4, so it satisfies the bound with equality.

Corollary 28.5 $K_{3,3}$ is not planar.

Proof. The graph $K_{3,3}$ has $p = 6$, $q = 9$, and girth 4. If it were planar, then Theorem 28.4 would give

$$9 \leq \frac{4(6-2)}{4-2} = 8,$$

a contradiction. □

Theorem 28.6 Any subgraph of a planar graph is planar.

Proof. Start with a planar embedding of the larger graph and delete the unwanted vertices and edges. No crossings are created. □

Subdivisions and Minors

Lecture 29 — Friday, November 20, 2015

Definition 29.1 An **edge subdivision** of a graph G is obtained by replacing an edge uv with a path from u to v whose internal vertices are new.

Subdividing edges does not change the essential topological shape of a graph, as illustrated in Figure 42.

FIGURE 42

A graph is planar if and only if any edge subdivision of it is planar. Subdividing an edge merely inserts degree-2 vertices along that edge.

Theorem 29.2 (Kuratowski's theorem) A graph is nonplanar if and only if it has a subgraph that is an edge subdivision of K_5 or $K_{3,3}$.

The proof is too much work.

Remark 29.3 In practice, Theorem 29.2 is often used by finding a subgraph that is "really" a copy of K_5 or $K_{3,3}$ after suppressing degree-2 vertices.

Example 29.4 Show that the Petersen graph is nonplanar.

Solution. By Theorem 29.2, it is enough to find in the Petersen graph a subgraph that is an edge subdivision of $K_{3,3}$. The correspondence is displayed in Figure 43.

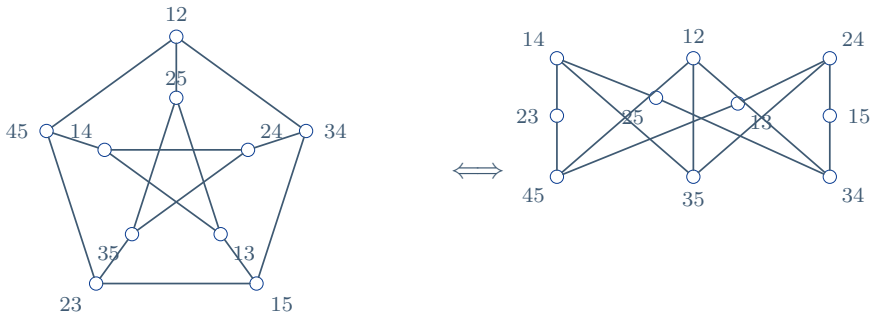


FIGURE 43

The graph on the right is obtained from $K_{3,3}$ by subdividing edges, so it is an edge subdivision of $K_{3,3}$. Hence the Petersen graph contains a subgraph that is an edge subdivision of $K_{3,3}$. Therefore, by Theorem 29.2, the Petersen graph is nonplanar. □

Definition 29.5 Let $e = xy$ be an edge of a graph G. The **contraction** G/e is obtained by identifying x and y to a single new vertex and deleting any loops that arise. Thus

$$V(G/e) = (V(G) \setminus \{x, y\}) \cup \{v_e\},$$

where v_e is the new vertex resulting from the contraction of e, and

$$\begin{aligned} E(G/e) = \; & \{ab \in E(G) : \{a, b\} \cap \{x, y\} = \emptyset\} \\ & \cup \{av_e : ax \in E(G) \text{ or } ay \in E(G) \text{ with } a \notin \{x, y\}\}. \end{aligned}$$

Example 29.6 The contraction in Figure 44 shows how an edge contraction merges its endpoints.

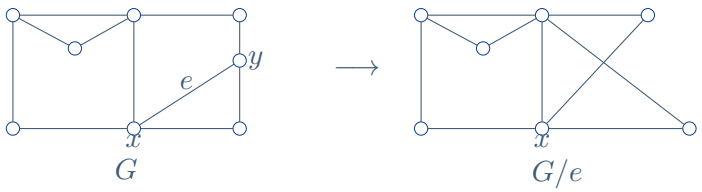
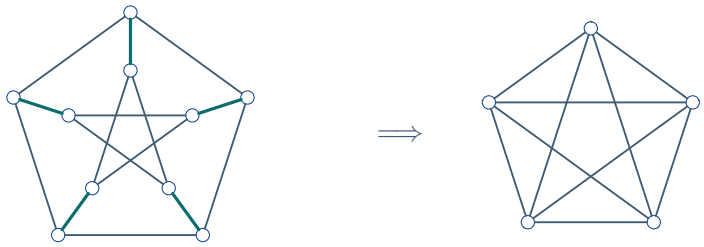


FIGURE 44

Definition 29.7 A graph H is a **minor** of a graph G if H can be obtained from G by deleting vertices, deleting edges, and contracting edges.

Minors are often easier to spot than subdivisions. Contracting a suitable collection of edges can expose a hidden K_5 or $K_{3,3}$.

Example 29.8 The Petersen graph is nonplanar. Figure 45 shows how contracting the "spokes" produces a K_5 minor.



P K_5

FIGURE 45

Theorem 29.9 (Wagner's theorem) A graph is nonplanar if and only if it has K_5 or $K_{3,3}$ as a minor.

Lecture 30 — Monday, November 23, 2015

4. Graph Coloring, Matchings, and Polyhedral Methods

Graph Colorings

Definition 30.1 A ***k*-coloring** of a graph G is an assignment of at most k colors to the vertices of G such that adjacent vertices receive different colors. Formally speaking, it is a function $\varphi : V(G) \to [k]$ such that $\varphi(x) \neq \varphi(y)$ whenever $xy \in E(G)$.

Definition 30.2 A graph is ***k*-colorable** if it has a k-coloring.

There is no known polynomial-time algorithm to determine whether a graph is k-colorable for $k \geq 3$. In fact, the decision problem of k-colorability is NP-complete for $k \geq 3$; so even a polynomial-time algorithm for 3-colorability would imply that P = NP. Conversely, if P = NP, then there is a polynomial-time algorithm to determine 3-colorability.

Example 30.3 The proper colouring in Figure 46 shows that the Petersen graph is 3-colorable.

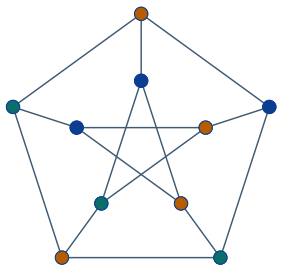


FIGURE 46

Theorem 30.4 A graph is 2-colorable if and only if it is bipartite.

Proof. *Forward implication.* If G is 2-colorable, then the two color classes form a bipartition.

Reverse implication. If G is bipartite with bipartition (A, B), color the vertices of A with one color and those of B with the other. □

Theorem 30.5 Every nonempty simple planar graph has a vertex of degree at most 5.

Proof. Suppose for a contradiction that there exists some planar graph G with $\deg(v) \geq 6$ for every vertex v. Write $p = |V(G)|$ and $q = |E(G)|$. Every vertex then has six distinct neighbours, so $p \geq 7$. Thus Theorem 28.3 gives

$$q \leq 3p - 6.$$

On the other hand, the handshake theorem for vertices gives

$$2q = \sum_{v \in V(G)} \deg(v) \geq 6p.$$

Combining these inequalities gives $6p \leq 2q \leq 6p - 12$, which is known to be a contradiction. □

How many colors does it take to color a planar graph? Let us start off with a crude upper bound.

Theorem 30.6 (Six color theorem) Every simple planar graph is 6-colorable.

Proof. We proceed by induction on $p = |V(G)|$. The cases $p \leq 1$ are trivial. Now, suppose that $p \geq 2$ and that the claim holds for all planar graphs with $p - 1$ vertices. By Theorem 30.5, we can choose a vertex v of degree at most 5. By the induction hypothesis, $G - v$ has a 6-coloring. At most five colors appear on the neighbours of v, so one color remains available for v. Thus G is 6-colorable. □

But we can do better than six colors...

Theorem 30.7 (Five color theorem) Every simple planar graph is 5-colorable.

Proof. We proceed by induction on $p = |V(G)|$. The cases $p \leq 1$ are immediate. Assume $p \geq 2$, and assume that every simple planar graph with fewer than p vertices is 5-colorable.

Let $v \in V(G)$ be a vertex of degree at most 5, which exists by Theorem 30.5. If $\deg(v) \leq 4$, then by induction $G - v$ has a proper 5-coloring, and one of the five colors is not used on $N(v)$, so we can color v with that color.

It remains to treat the case $\deg(v) = 5$. Let

$$N(v) = \{a, b, u_1, u_2, u_3\}.$$

Two vertices in $N(v)$ must be nonadjacent. Indeed, if every pair in $N(v)$ were adjacent, then

the induced subgraph $G[N(v)]$ would be K_5, which is impossible in a planar graph. Choose nonadjacent vertices $a, b \in N(v)$.

Let G' be obtained from G by contracting the two edges va and vb, and let v' be the new vertex obtained by identifying a, v, b. The contraction is illustrated in Figure 47.

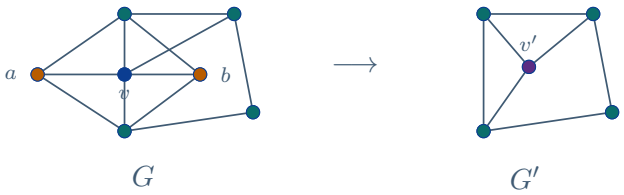


FIGURE 47

The graph G' is planar and has fewer vertices than G, so by induction G' has a proper 5-coloring, say φ'. We now construct a proper 5-coloring φ of $G - v$ as follows:

$$\varphi(x) = \varphi'(x) \quad \text{for } x \in V(G) \setminus \{v, a, b\},$$

and

$$\varphi(a) = \varphi(b) = \varphi'(v').$$

This is proper on $G - v$ because a and b are nonadjacent.

Now consider the neighbors of v in G. The vertices a and b have the same color, and the other three neighbors contribute at most three additional colors. Therefore at most four colors appear on $N(v)$, so one of the five colors remains available for v. Assign that color to v. This extends φ to a proper 5-coloring of G. □

Can we do even better? Yes, we can. But the proof is somewhat harder.

Theorem 30.8 (Four color theorem) Every simple planar graph is 4-colorable.

The four color theorem was first conjectured in 1852 by Francis Guthrie, and was finally proved in 1976 by Kenneth Appel and Wolfgang Haken using a computer-assisted proof, which was controversial at the time. Neil Robertson, Daniel P. Sanders, Paul Seymour, and Robin Thomas (1997) gave a simpler proof and an efficient algorithm for finding a 4-coloring of a planar graph.

The idea of the proof is similar to that of the five color theorem, but it looks for configurations that are more complicated than a single vertex of degree at most 5. The proof is by contradiction, and

the minimal counterexample is shown to contain one of a finite list of unavoidable configurations using "discharging" arguments. Each of these configurations is then shown to be reducible, meaning that it cannot appear in a minimal counterexample.

Lecture 31 — Wednesday, November 25, 2015

Chromatic Polynomials

Theorem 31.1 Let G be a graph with p vertices. There exists a polynomial $\chi_G(t)$ of degree at most p with integer coefficients such that, for every positive integer k, the number of k-colorings of G is $\chi_G(k)$.

Proof. We proceed by induction on the number of edges of G.

If G has no edges, then every function $V(G) \to [k]$ is a proper k-coloring, so the number of k-colorings is k^p. Hence we may take

$$\chi_G(t) = t^p.$$

Now assume that G has at least one edge, and assume the claim is already true for graphs with fewer edges. Let $e = xy \in E(G)$.

Every proper k-coloring of G is also a proper k-coloring of $G - e$. Conversely, a proper k-coloring of $G - e$ is a proper k-coloring of G if and only if x and y receive different colors.

Therefore, among the proper k-colorings of $G - e$, the ones that are not colorings of G are exactly those in which x and y have the same color. These are in bijection with proper k-colorings of G/e: if x and y have the same color, identify them to color the contracted vertex, and conversely pull a coloring of G/e back to $G - e$ by giving x and y that common color.

Hence

$$|\{\text{proper } k\text{-colorings of } G\}| = |\{\text{proper } k\text{-colorings of } G - e\}| - |\{\text{proper } k\text{-colorings of } G/e\}|.$$

By the induction hypothesis, there are polynomials $\chi_{G-e}(t)$ and $\chi_{G/e}(t)$ (with integer coefficients

and degree at most p) that count these two quantities. We therefore define

$$\chi_G(t) = \chi_{G-e}(t) - \chi_{G/e}(t).$$

This polynomial has integer coefficients and degree at most p, and for each positive integer k it equals the number of proper k-colorings of G. It is unique because two polynomials that agree at every positive integer are equal. □

Definition 31.2 The polynomial $\chi_G(t)$ is called the **chromatic polynomial** of G.

Example 31.3 For the complete graph K_p, there are k ways to color the first vertex, $k-1$ ways to color the second vertex, $k-2$ ways to color the third vertex, and so on, giving

$$\chi_{K_p}(k) = k(k-1)(k-2)\cdots(k-p+1) = k^{\underline{p}}.$$

Remark 31.4 The deletion–contraction recurrence is theoretically very useful, but computing $\chi_G(t)$ by repeated contractions is usually exponential-time.

Lecture 32 — Friday, November 27, 2015

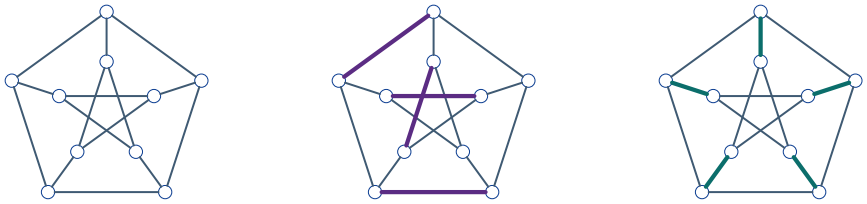
Matchings

Definition 32.1 A **matching** in a graph G is a subset $M \subseteq E(G)$ such that every vertex of G is incident with at most one edge of M.

The empty set is a matching, so every graph has a matching. We are more interested in large matchings.

Definition 32.2 A matching is **maximum** if it has as many edges as possible, and **maximal** if it is not properly contained in a larger matching.

Example 32.3 Figure 48 shows several matchings in the Petersen graph.



$M = \emptyset$ | matching of size 4 (maximal) | matching of size 5 (maximum)

FIGURE 48

Definition 32.4 A vertex is **saturated** by a matching M if it is incident with an edge of M.

Example 32.5 In the Petersen graph, use the same maximal matching of size 4 as in Example 32.3. The saturated vertices are exactly those incident with one of those four purple edges, as displayed in Figure 49.

M is maximal with $|M| = 4$

FIGURE 49

Definition 32.6 A **perfect matching** of G is a matching of G with $|V(G)|/2$ edges (i.e. a matching that saturates every vertex of G).

Note that a perfect matching is maximum, but a maximum matching need not be perfect. Some graphs have no perfect matching at all, while others have many perfect matchings.

Example 32.7 The six perfect matchings of the Petersen graph are shown in Figure 50.

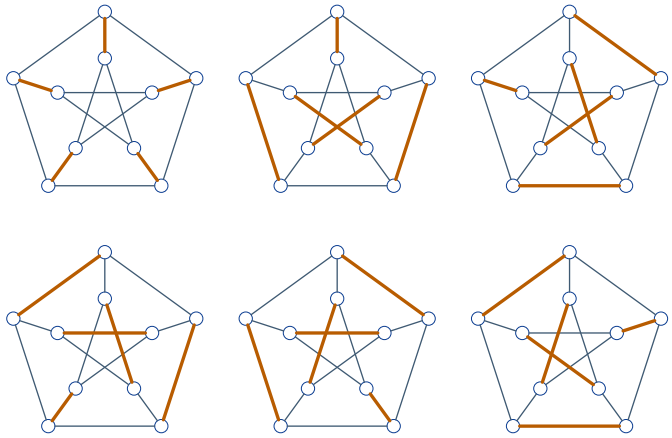


FIGURE 50

Definition 32.8 Let M be a matching in a graph G. An **alternating path** with respect to M is a path whose edges alternate between belonging to M and not belonging to M. An **augmenting path** with respect to M is an alternating path whose two endpoints are both unsaturated.

Example 32.9 Let $M = \{u_2u_3,\, u_4u_5\}$ in the path graph on vertices $u_1, \ldots, u_6$. Figure 51 distinguishes an alternating path from an augmenting path for this matching.

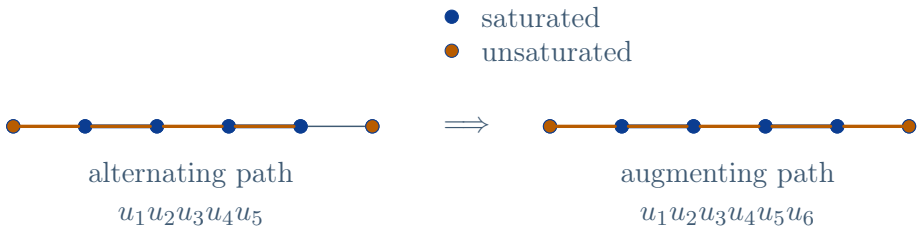


FIGURE 51

Theorem 32.10 A matching M is not maximum if and only if there exists an augmenting path with respect to M.

Proof. *Forward implication.* If P is an augmenting path, replace the edges of M on P by the edges of $P \setminus M$. This increases the size of the matching by 1, so M is not maximum.

Reverse implication. Suppose M is not maximum, and choose a larger matching N. Consider the graph with edge set $M \triangle N = (M \setminus N) \cup (N \setminus M)$. Every vertex has degree at most 2, so each component is a path or a cycle. Cycles and even paths contain equally many edges of M and N, while the larger size of N forces some component to be a path with one more edge of N than of M. That path is an augmenting path for M. □

If P is an augmenting path for M, we replace M by

$$M \leftarrow M \triangle E(P),$$

where

$$A \triangle B = (A \setminus B) \cup (B \setminus A).$$

This produces a strictly larger matching, as illustrated in Figure 52.

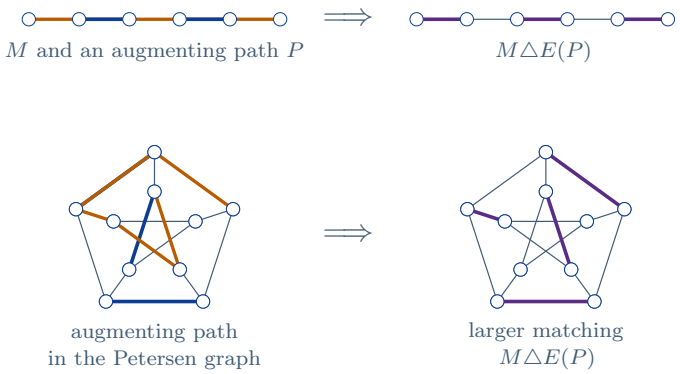


FIGURE 52

Finding Augmenting Paths

In the remainder of this section, we focus on bipartite graphs because they're easier to work with, and because they have many applications.

Example 32.11 Suppose we have a bipartite graph whose top vertices represent co-op students/job applicants/etc. and whose bottom vertices represent jobs/positions/etc. We draw an edge when a student is qualified for a job.

Then matching as many students to jobs as possible is exactly the problem of finding a maximum matching. The bipartite model is shown in Figure 53.

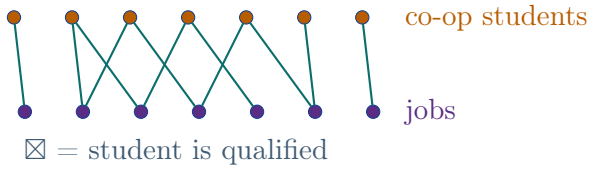


FIGURE 53

Definition 32.12 A **cover** (or **vertex cover**) of a graph G is a set $C \subseteq V(G)$ such that every edge of G is incident with at least one vertex of C.

The set $V(G)$ is always a cover, so every graph has a cover. We are interested in minimum covers.

Example 32.13 In the Petersen graph, the whole vertex set is a cover, but much smaller covers also exist, as Figure 54 illustrates.

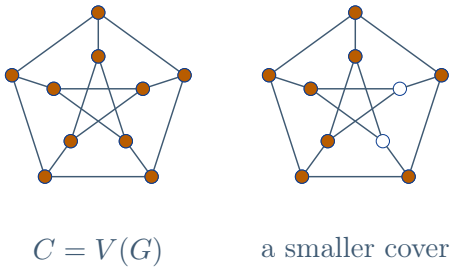


FIGURE 54

Finding a minimum cover is NP-hard in general.

Theorem 32.14 If M is a matching and C is a cover in the same graph, then

$$|M| \leq |C|.$$

Proof. Every edge of M must be incident with a vertex of C. Since distinct edges of a matching are disjoint, different edges of M require different vertices of C. Therefore $|M| \leq |C|$. □

Corollary 32.15 If M is a matching and C is a cover in the same graph, and if $|M| = |C|$, then M is maximum and C is minimum.

Proof. Let M' be any matching and C' be any cover. Since every matching has size at most every cover, we have

$$|M'| \leq |C| = |M| \leq |C'|.$$

Hence no matching is larger than M, and no cover is smaller than C. □

Lecture 33 — Monday, November 30, 2015

Theorem 33.1 (Kőnig's theorem) Let G be a bipartite graph. Then there exists a matching M and a cover C of the same size.

Equivalently, in any bipartite graph, the maximum size of a matching equals the minimum size of a cover.

Proof. Start with a matching M. We will search for an augmenting path and, in the process, build a cover. Either an augmenting path exists, in which case M is not maximum and we should enlarge it, or we end with a cover C satisfying $|C| = |M|$.

Let $G = (A, B)$ be bipartite. Any augmenting path has one end in A and the other in B, because it has odd length. Without loss of generality, we search from unsaturated vertices of A.

Define

$$X_0 = \{\text{unsaturated vertices of } A\},$$

let $X \subseteq A$ be the set of reachable vertices of A, let $Y \subseteq B$ be the set of reachable vertices of B, and let

$$Y_0 = \{\text{unsaturated vertices of } B \text{ that lie in } Y\}.$$

Here a vertex is called reachable if there is an alternating path from some vertex of X_0 to that vertex. Every vertex of X_0 is reachable by a path of length 0, so $X_0 \subseteq X$. Also, an augmenting path exists if and only if $Y_0 \neq \emptyset$. The partition used in the proof is summarized in Figure 55.

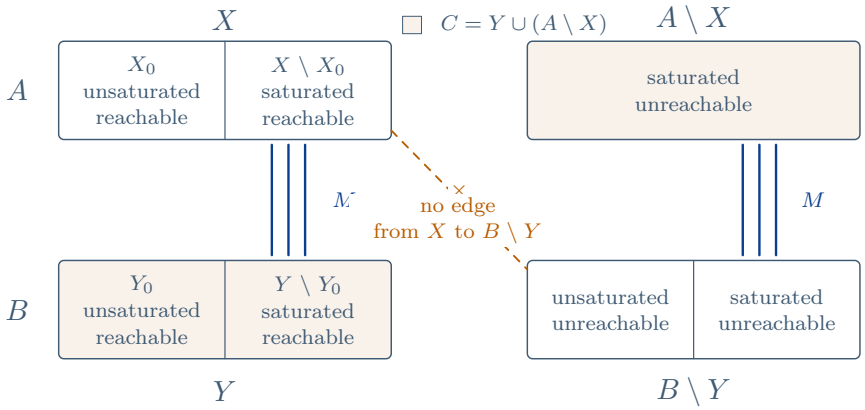


FIGURE 55

Lemma 1. If $u \in X$ and $uv \in E(G)$, then $v \in Y$.

Proof. If $uv \in M$, then uv must be the last edge of the alternating path to u, because that path starts in A with an edge not in M and therefore reaches vertices of A after an even number of steps. Since M is a matching, the final matching edge incident with u is unique, so it must be uv. Thus the path already contains v, and hence $v \in Y$.

If $uv \notin M$, then either the alternating path to u already contains v, in which case $v \in Y$, or appending uv to that path produces an alternating path to v. Again $v \in Y$. This proves the lemma.

Lemma 2. If $v \in Y$ and $uv \in M$, then $u \in X$.

Proof. The proof is similar. Let $P(v)$ be an alternating path from X_0 to v. Either u already lies on $P(v)$, in which case $u \in X$, or appending the matching edge vu to $P(v)$ gives an alternating path to u. Thus $u \in X$.

We now build a cover.

Lemma 3. Let

$$C = Y \cup (A \setminus X).$$

Then C is a cover. Moreover,

$$|C| = |M| + |Y_0|.$$

Proof. By Lemma 1, every edge of G has at least one endpoint in $A \setminus X$ or in Y, so C is a cover.

By Lemma 2, every edge of M is incident with exactly one vertex in $A \setminus X$ or exactly one vertex in $Y \setminus Y_0$, and never with both. Therefore

$$|M| = |A \setminus X| + |Y \setminus Y_0| = |C| - |Y_0|,$$

which is equivalent to the claimed identity.

Finally, suppose M is a maximum matching. Then no augmenting path exists, so $Y_0 = \emptyset$. Lemma 3 therefore gives a cover C with

$$|C| = |M|.$$

By the previous corollary, M is maximum and C is minimum. This proves Kőnig's theorem. □

Example 33.2 Kőnig's theorem is special to bipartite graphs. In nonbipartite graphs, the maximum matching size can be strictly smaller than the minimum cover size; Figure 56 contrasts the two settings.

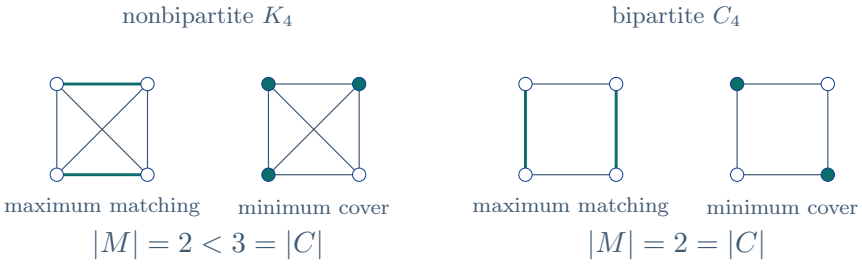


FIGURE 56

Lecture 34 — Wednesday, December 02, 2015

Bipartite Matching Algorithm

To turn the proof of Kőnig's theorem into an algorithm for maximum matchings and minimum covers, we need a systematic way to compute the sets X and Y. The plan is to modify the BFST algorithm as follows:

1. Every vertex of X_0 is a root, rather than starting with a single root.

2. If v is the active vertex and $v \in A$, so that v is at an even level, then we only follow edges that are not in M.
3. If v is the active vertex and $v \in B$, so that v is at an odd level, then we only follow edges that are in M.
4. At the end, X is the set of vertices at even levels, Y is the set of vertices at odd levels, X_0 is the set of vertices at level 0, and Y_0 is the set of vertices at odd levels with no children.
5. If $Y_0 = \emptyset$, then we have a maximum matching and $C = Y \cup (A \setminus X)$ is a minimum cover. If $Y_0 \neq \emptyset$, then a path from a vertex of Y_0 to its root is an augmenting path.

Example 34.1 Start with the bipartite graph in Figure 57, where $A = \{1, 2, 3, 4, 5\}$ is the top row, $B = \{6, 7, 8, 9, 10\}$ is the bottom row, and the initial matching is

$$M = \{27, 38, 5\,10\}.$$

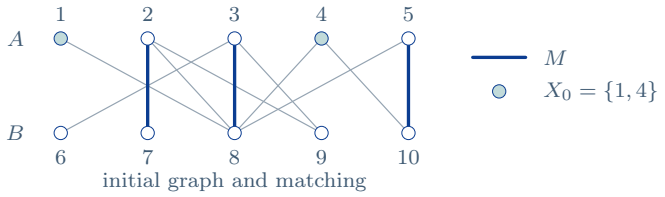


FIGURE 57

We have $X_0 = \{1, 4\}$. Running the algorithm gives the first search forest in Figure 58 and produces the augmenting path $1, 8, 3, 6$.

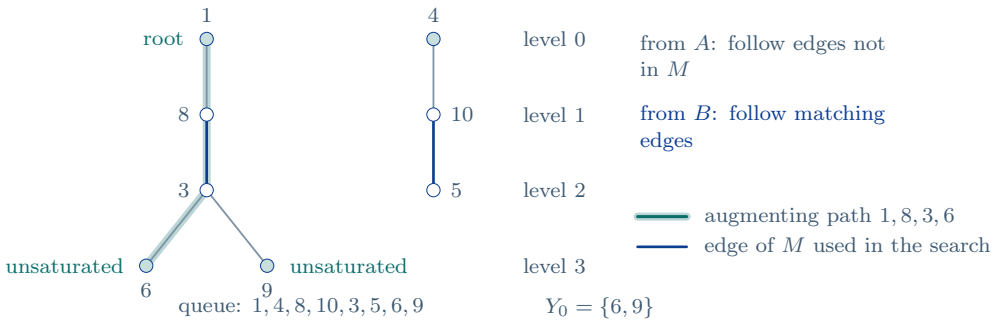


FIGURE 58

After augmenting, we obtain

$$M' = \{18, 27, 36, 5\,10\}.$$

Figure 59 compares the toggled path with the new matching.

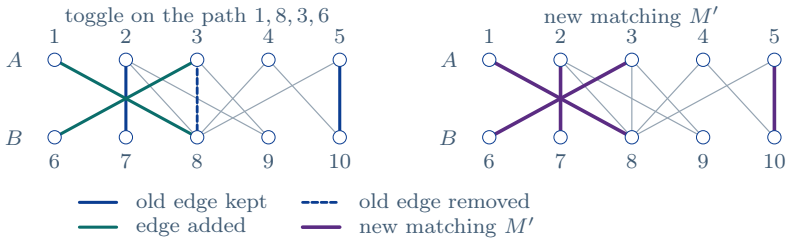


FIGURE 59

Running the algorithm again gives $X_0 = \{4\}$, $X = \{1, 4, 5\}$, $Y = \{8, 10\}$, and $Y_0 = \emptyset$. Hence M' is maximum and

$$C = Y \cup (A \setminus X) = \{8, 10, 2, 3\}$$

is a minimum cover. The final search forest and cover are shown in Figure 60.

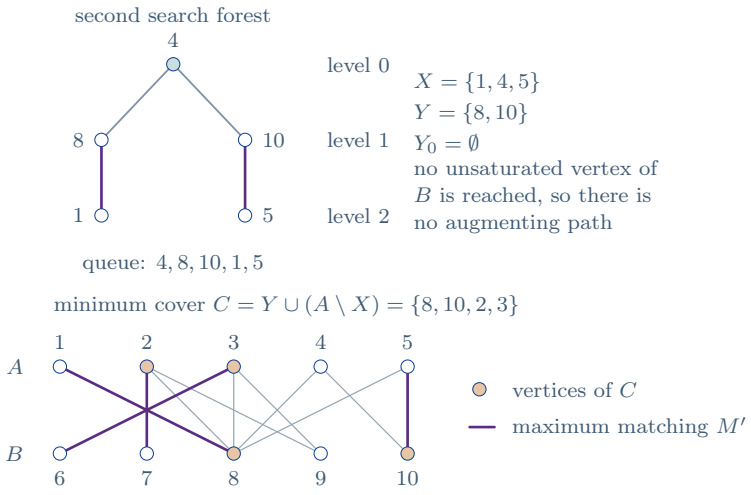


FIGURE 60

Theorem 34.2 (Hall's marriage theorem) Let G be a bipartite graph with bipartition (A, B). For $D \subseteq A$, let $N(D) = \cup_{v \in D} N(v)$, where $N(v)$ is the set of neighbors of v in B. Then there exists a matching of size $|A|$ if and only if

$$|N(D)| \geq |D|$$

for every subset $D \subseteq A$.

Proof. *Forward implication.* Suppose M is a matching of size $|A|$. Then every vertex of A is saturated. Let $D \subseteq A$, and let M' be the set of edges of M incident with vertices of D. Let N' be the set of vertices of B incident with edges of M'. Since M is a matching, we have

$$|D| = |M'| = |N'|.$$

Because every vertex of N' is adjacent to a vertex of D, we have $N' \subseteq N(D)$, so $|D| \leq |N(D)|$.

Reverse implication. Suppose there is no matching of size $|A|$. Let M be a maximum matching. Then $|M| < |A|$, so some vertex of A is unsaturated. Construct the sets X_0, X, Y, and Y_0 as in Kőnig's theorem. Since M is maximum, there is no augmenting path, hence $Y_0 = \emptyset$. The resulting partition is shown in Figure 61.

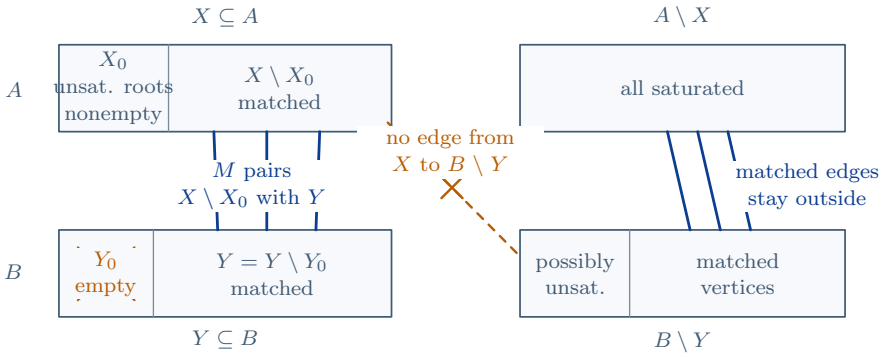


FIGURE 61

Since $|M| < |A|$, we have $X_0 \neq \emptyset$, so $X \neq \emptyset$. Also,

$$|X| = |X_0| + |X \setminus X_0| > |X \setminus X_0| = |Y \setminus Y_0| = |Y|,$$

where the middle equality comes from the matching between $X \setminus X_0$ and Y.

Moreover, no edge can go from X to $B \setminus Y$; otherwise a vertex of $B \setminus Y$ would be reachable. Hence $N(X) \subseteq Y$, so

$$|N(X)| \leq |Y| < |X|.$$

Therefore $D = X$ violates Hall's condition. □

If a matching of size $|A|$ exists, then it is automatically a maximum matching. Also, if $|B| < |A|$, then taking $D = A$ shows immediately that no such matching can exist.

Example 34.3 A deck of 52 cards is dealt as evenly as possible into k piles, where $1 \leq k \leq 13$. Then it is always possible to choose one card from each pile so that no two chosen cards have the same rank.

Proof. We use Hall's theorem. Let G be the bipartite graph with bipartition (A, B), where A is the set of piles and $B = \{\mathrm{A}, 2, \ldots, \mathrm{K}\}$ is the set of ranks. Put an edge pr between a pile $p \in A$ and a rank $r \in B$ if pile p contains a card of rank r. A schematic version of this bipartite graph appears in Figure 62.

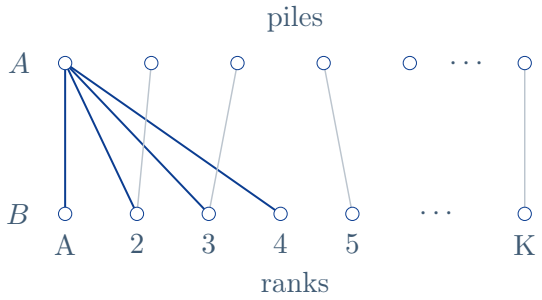


FIGURE 62

Let $D \subseteq A$. Consider all cards in the piles from D. Since the cards are dealt as evenly as possible and $k \leq 13$, each pile has at least 4 cards, so these piles contain at least $4|D|$ cards altogether. There are only four cards of each rank in the deck, so at least $|D|$ ranks are represented among these cards. In other words, $|D| \leq |N(D)|$.

Thus Hall's condition holds, so by Hall's theorem, G has a matching of size $|A|$. Such a matching chooses one rank from each pile, and no two chosen cards have the same rank. □

Theorem 34.4 If $k \geq 1$, then every k-regular bipartite graph has a perfect matching.

Proof. Let G be a k-regular bipartite graph with bipartition (A, B), and let $D \subseteq A$. Since G is k-regular, the number of edges incident with a vertex in D is $k|D|$. Every such edge has its other endpoint in $N(D)$, while the total number of edges incident with vertices in $N(D)$ is $k|N(D)|$. Therefore

$$k|D| \leq k|N(D)|,$$

so $|D| \leq |N(D)|$ for every $D \subseteq A$. Hall's theorem gives a matching of size $|A|$.

The same argument, with the roles of A and B reversed, gives a matching of size $|B|$. Also, counting all edges gives $k|A| = |E(G)| = k|B|$, so $|A| = |B|$. Thus these matchings are perfect. □

Lecture 35 — Friday, December 04, 2015

The Birkhoff–von Neumann Theorem

Definition 35.1 Let $S \subseteq \mathbb{R}^m$, say $S = \{s_1, \ldots, s_k\}$. Then the **convex hull** of S is

$$\operatorname{conv}(S) = \left\{ a_1 s_1 + \cdots + a_k s_k : a_i \geq 0 \text{ for all } i,\ \sum_{i=1}^{k} a_i = 1 \right\}.$$

Several representative convex hulls are collected in Figure 63.

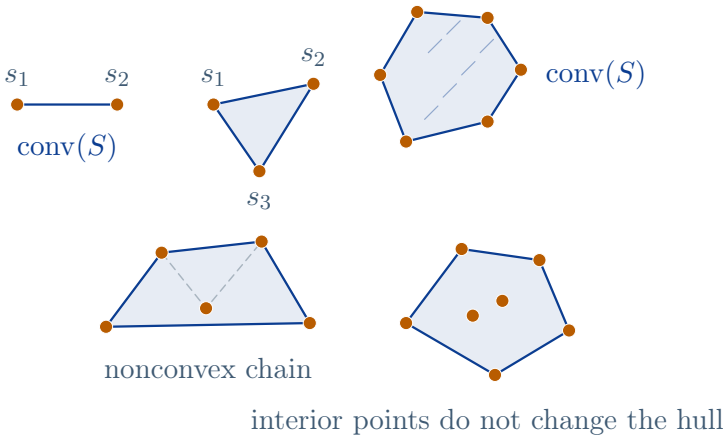


FIGURE 63

Definition 35.2 A **doubly stochastic matrix A** is an $n \times n$ matrix such that the following hold:

1. Every entry of $\mathbf{A}$ is nonnegative.
2. $\mathbf{A1} = \mathbf{1}$, where $\mathbf{1}$ is the all-ones vector.
3. $\mathbf{A}^\top \mathbf{1} = \mathbf{1}$.

We write DSM_n for the set of all $n \times n$ doubly stochastic matrices.

Definition 35.3 A **permutation matrix P** is an $n \times n$ matrix such that the following hold:

1. Every entry of $\mathbf{P}$ is either 0 or 1.
2. $\mathbf{P1} = \mathbf{1}$.
3. $\mathbf{P}^\top \mathbf{1} = \mathbf{1}$.

We write PM_n for the set of all permutation matrices.

Theorem 35.4 (Birkhoff–von Neumann theorem)

$$\mathrm{DSM}_n = \mathrm{conv}(\mathrm{PM}_n).$$

Proof. The inclusion $\mathrm{conv}(\mathrm{PM}_n) \subseteq \mathrm{DSM}_n$ is obvious, since every permutation matrix is doubly stochastic, and DSM_n is convex.

For the reverse inclusion, let $\mathbf{A} = (A_{ij}) \in \mathrm{DSM}_n$. Form a bipartite graph G with bipartition

$$A_0 = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n\} \qquad \text{and} \qquad B_0 = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\},$$

where $\boldsymbol{r}_i$ represents the ith row and $\boldsymbol{c}_j$ represents the jth column. We join $\boldsymbol{r}_i$ to $\boldsymbol{c}_j$ if and only if $A_{ij} \neq 0$.

We claim that G has a perfect matching. Let $D \subseteq A_0$. Then

$$\sum_{\boldsymbol{r}_i \in D} \sum_{j=1}^{n} A_{ij} = |D|,$$

since every row sum is 1. Every nonzero term on the left occurs in a column indexed by a vertex of $N(D)$, so

$$|D| = \sum_{\boldsymbol{r}_i \in D} \sum_{j=1}^{n} A_{ij} \leq \sum_{\boldsymbol{c}_j \in N(D)} \sum_{i=1}^{n} A_{ij} = |N(D)|,$$

since every column sum is also 1. Hall's theorem therefore gives a perfect matching in G.

Let $\mathbf{P} \in \mathrm{PM}_n$ be the permutation matrix corresponding to this perfect matching, and let

$$q = \min\{A_{ij} : \boldsymbol{r}_i\boldsymbol{c}_j \text{ is an edge of the matching}\}.$$

We now argue by induction on the number of nonzero entries of $\mathbf{A}$. The minimum possible number is n, in which case the row and column sums force $\mathbf{A}$ to be a permutation matrix. For the inductive step, use the perfect matching above. We have $0 < q \leq 1$. If $q = 1$, every matched entry is 1, so again $\mathbf{A} = \mathbf{P}$. Otherwise $0 < q < 1$, and the matrix

$$\mathbf{A}' = \frac{\mathbf{A} - q\mathbf{P}}{1 - q}$$

is a doubly stochastic matrix with fewer nonzero entries than $\mathbf{A}$, so $\mathbf{A}' \in \mathrm{conv}(\mathrm{PM}_n)$ by induction,

and

$$\mathbf{A} = q\mathbf{P} + (1-q)\mathbf{A}'.$$

Hence $\mathbf{A} \in \mathrm{conv}(\mathrm{PM}_n)$, as required. □

Appendix: Lecture and Tutorial Index