Counting Trees

UConn Math Club

Jeremy Teitelbaum

February 28, 2018

UCONN

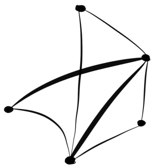
Basic Definitions

Definition

A graph G is a set of vertices V, together with a set of edges E, each of which connects together two distinct vertices. We sometimes use the word "node" instead of vertex.

Remark

This definition disallows loops and multiple edges between two vertices, and so these objects are sometimes called simple graphs.

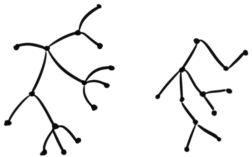


Definition

A graph is *finite* if it has finitely many vertices. A *path* between vertices a and b in a graph G is a determined by finitely many distinct vertices $\{v_i\}_{i=0}^n$ so that $v_0 = a$, $v_n = b$, and (v_i, v_{i+1}) is an edge for each $i = 0, \ldots, n-1$.

Definition

A graph is *connected* if there is (at least) one path between any two vertices. A graph is a *forest* if there is at most one path between any two vertices. A graph is a *tree* if there is exactly one path between any two vertices.



A forest with two trees

Rooted Trees

Definition

A tree is *rooted* if it has a distinguished node, called the root. In a rooted tree, the *children* of a node are the nodes adjacent to, but one step farther away from the root. The parent of a node is the unique node that is one step closer to the root.

Euler Characteristic

If G is a finite graph, its Euler Characteristic $\chi_G = V - E$ where V is the number of vertices of G and E is its number of edges.

Theorem

If G is a connected finite graph, then $\chi_G \leq 1$ with equality if and only if G is a tree.

Euler Characteristic 2

Theorem

G a tree implies $\chi_G = 1$

Proof.

Suppose G is a tree. If G has 1 vertex (and zero edges) then $\chi_G = 1$. Suppose G has n vertices. Choose a vertex with only one edge leaving it – this must exist since G is finite. Delete this vertex and edge to get a tree with $n-1$ vertices and, by induction, euler characteristic 1. Since G is obtained from this smaller tree by adding one edge and one vertex, $\chi_G = 1$. Thus every tree has euler characteristic 1. □

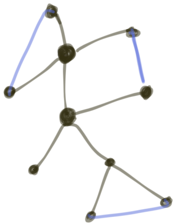
Euler Characteristic 3

Theorem

G not a tree implies $\chi_G < 1$.

Proof.

Suppose G is a finite connected graph that is not a tree. Choose a maximal subtree T of G. Every vertex of G belongs to T since otherwise we could extend T by adding an edge leading to a vertex v' adjacent to T. Thus G is obtained from T by adding some number, say $k > 0$, edges. Thus $\chi_G = 1 - k < 1$. □



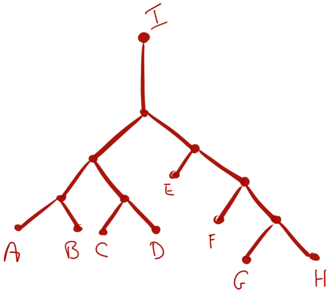
Remark

Note that $1 - \chi_G$ *is the minimal number of edges that must be removed from* G *to make it a tree.*

Phylogenetic Trees

Definition

A phylogenetic tree is a tree where every node has either one or three adjacent vertices. The nodes with one adjacent vertex, called *leaves*, are labelled.



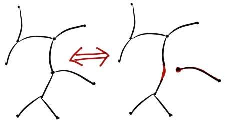
Phlogenetic Trees: Formulae

V is the number of *vertices*, I is the number of *internal nodes*, L is the number of *leaves*, and E is the number of *edges*.

1. $V = I + L$
2. $2E = 3I + L$ because every internal node has 3 edges, every leaf has 1; but every edge is counted twice.
3. $V - E = 1$ because this is a connected tree.
4. $L - I = 2$; $V = 2L - 2$; $E = 2L - 3$.

Counting Phylogenetic Trees

Let $T(n)$ be the number of phylogenetic trees with n leaves. Given a phylogenetic tree, you can delete a leaf and the attached edge; this leaves an internal vertex with two edges.



Delete that vertex and join the edges. Now you have a phylogenetic tree with $n-1$ leaves.
To go in the other direction, pick an edge, split it by adding a vertex in the middle, and then add a leaf.

Counting Phylogenetic Trees

We have a bijection between phylogenetic trees with $n-1$ leaves and a chosen edge, and phylogenetic trees with n leaves. Since there are $(2n-5)$ edges in a tree with $n-1$ leaves we have $T(3)=1$ and $T(n)=(2n-5)T(n-1)$.

1. $T(4)=3$
2. $T(5)=15$
3. $T(n)$ is the product of the first $n-2$ odd numbers. $T(7)=(1)(3)(5)(7)(9)$.

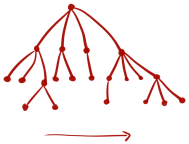
Ordered Trees

Definition

A tree T is *ordered* (or, equivalently, *planar*), if

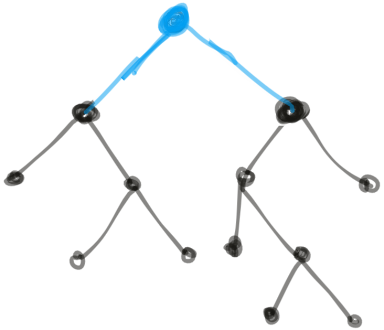
1. it is rooted;
2. the descendants of a node are totally ordered.

One way to think of this ordering is to think of the descendants of a node as having a left-to-right ordering corresponding to the way they would be drawn in a picture of the tree.



Binary Trees

A binary tree is a particularly common example of an ordered tree.
A (rooted) binary tree is an ordered tree in which every node has zero or two children.



Binary Trees 2

Binary trees have a recursive organization, since a binary tree on n vertices is constructed by taking two binary trees, one on x vertices and one on y vertices with $x + y = n - 1$, and then adding a vertex and joining it to the roots of the original two trees.

Lemma

The number $b(n)$ of binary trees on n vertices satisfies $b(0) = 0$, $b(1) = 1$ and the recurrence relation

$$b(n) = \sum_{x+y=n-1} b(x)b(y)$$

Binary Trees 3

Let $F(x) = \sum_{n=0}^{\infty} b(n)x^n$ be the "generating function" for the number of binary trees. Notice that

$$F(x)^2 = \sum_{i+j=n} b(i)b(j)x^n.$$

Lemma

The generating function $F(x)$ satisfies the quadratic equation

$$xF_2(x)^2 - F_2(x) + x = 0$$

Binary Trees 4

From the lemma above we find:

$$F_2(x) = \frac{1-\sqrt{1-4x^2}}{2x} = \sum_{i=0}^{\infty} (-1)^i \binom{1/2}{i} 2^{2i-1} x^{2i-1}.$$

Binary Trees 5

The *Catalan Numbers* C_n are defined by

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

(see sequence A000108 in the **OEIS**). By fiddling around with binomial coefficients, we may rewrite the generating function $F_2(x)$ as

$$F_2(x) = \sum_{i=1}^{\infty} C_{i-1}x^{2i-1}.$$

Lemma

The number of binary trees on $2n-1$ vertices is C_{n-1}.

Catalan Numbers

The Catalan numbers arise in a huge number of combinatorial situations and entry A000108 of the OEIS is "probably the longest entry" in the encyclopedia. A few other interpretations:

- The number of ways to insert n pairs of parentheses in a word of $n + 1$ letters.
- Iterate $f(x) = x^2 + x$ infinitely often to get convergence to the Catalan series $F(x)$.

m-ary Trees

Definition

For an integer $m \geq 1$, an m-ary tree is a (rooted) tree where every node has zero or m (ordered) children.

The same sort of generating function calculation says that if $b_m(n)$ is the number of m-ary trees with n vertices, and $F_m(x) = \sum b_m(n)x^n$, then

$$xF_m(x)^m - F_m(x) + x = 0.$$

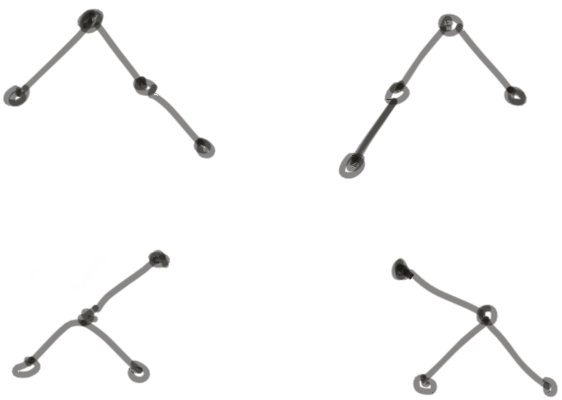
Alternatively we can write this

$$F_m(x) = x(1 + F_m(x)^m).$$

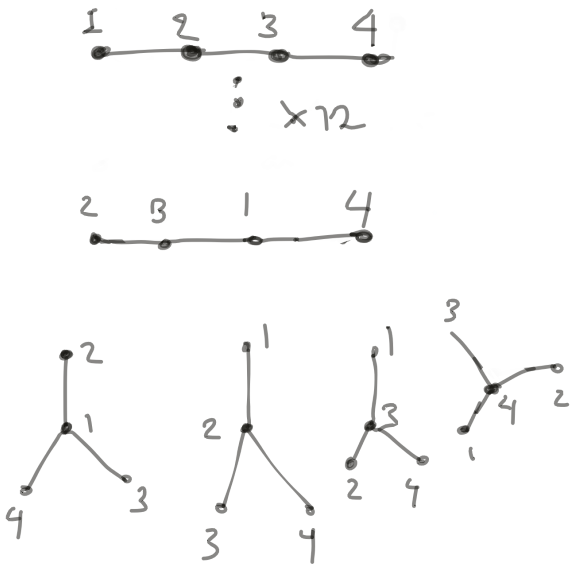
Three more counting problems

1. Ordered trees. The children of a node are ordered left to right, but there can be any number of them.
2. Labelled trees. The n nodes are labelled $1 \ldots, n$ and two trees are counted as the same only if the identification between them associates nodes with the same label.
3. General trees. The trees are treated as general graphs with no ordering or labelling.

Ordered trees: 4 nodes



Labelled trees: 4 nodes



General trees: 4 nodes

Prüfer sequences

Prüfer sequences are a concrete realization of Cayley's theorem. Let T be a labelled (unrooted) tree on n vertices labelled with

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

Definition

The Prüfer sequence for T is the sequence $a_0, \dots, a_{n-2}$ constructed inductively as follows. Find the leaf of T with the smallest label and let a_0 be the label of the unique adjacent node; delete this leaf with its attached edge and repeat the process. Continue until only two nodes connected by an edge remain.

Prüfer sequences 2

Theorem

The Prüfer sequence gives a bijection between the sequences of integers between 0 and $n-1$ of length $n-2$ and the labelled trees on n nodes.