Vertical Paths in Simple Varieties of Trees

Keith Copenhaver

University of Florida

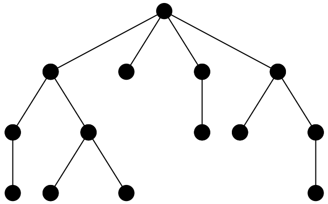
Simple Varieties of Trees

A graph $G = (V, E)$ consists of a set of vertices V and a set of edges E.

Definition

A rooted tree is a connected graph without cycles with one vertex distinguished as the root. A class of trees is called *simple variety* if the generating function for the number of trees on n vertices satisfies an equation of the form

$$T(x) = x\phi(T(x)).$$



Motivation

Things modelled by rooted trees:

- Computer network access
- Social networks
- Distribution networks

Rooted trees are also used as data structures.

A Few Simple Varieties of Trees

$$T(x) = x\phi(T(x)).$$

A Few Simple Varieties of Trees

UF UNIVERSITY of FLORIDA

$$T(x) = x\phi(T(x)).$$

- General trees: unlabeled, plane, any number of children.
  $T(x) = x\dfrac{1}{1 - T(x)}.$

A Few Simple Varieties of Trees

$$T(x) = x\phi(T(x)).$$

- General trees: unlabeled, plane, any number of children.
  $T(x) = x\dfrac{1}{1 - T(x)}$.
- Motzkin trees: unlabeled, plane, 0, 1, or 2 children.
  $T(x) = x(1 + T(x) + T(x)^2)$

A Few Simple Varieties of Trees

$$T(x) = x\phi(T(x)).$$

- General trees: unlabeled, plane, any number of children. $T(x) = x\dfrac{1}{1 - T(x)}$.
- Motzkin trees: unlabeled, plane, 0, 1, or 2 children. $T(x) = x(1 + T(x) + T(x)^2)$
- Binary trees: unlabeled, plane, children are designated left or right, 0, 1, or 2 children. $T(x) = x(1 + 2T(x) + T(x)^2)$.

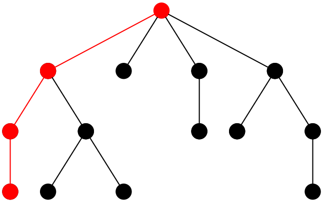
A Few Simple Varieties of Trees

$$T(x) = x\phi(T(x)).$$

- General trees: unlabeled, plane, any number of children. $T(x) = x\dfrac{1}{1-T(x)}$.
- Motzkin trees: unlabeled, plane, 0, 1, or 2 children. $T(x) = x(1 + T(x) + T(x)^2)$
- Binary trees: unlabeled, plane, children are designated left or right, 0, 1, or 2 children. $T(x) = x(1 + 2T(x) + T(x)^2)$.
- Cayley trees: labeled, nonplane, any number of children. $T(x) = xe^{T(x)}$.

Some Traits of Trees/Vertices

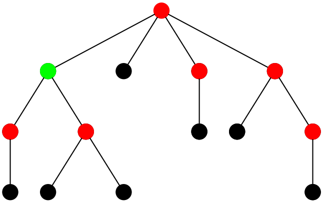
- The *height* of a tree is the distance from the root to the furthest vertex plus one.



Some Traits of Trees/Vertices

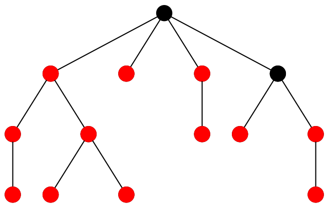
UF UNIVERSITY of FLORIDA

- The *height* of a tree is the distance from the root to the furthest vertex plus one.
- The *rank/protection number* of a vertex is the shortest distance from that vertex to its closest descendant leaf.



Some Traits of Trees/Vertices

- The *height* of a tree is the distance from the root to the furthest vertex plus one.
- The *rank/protection number* of a vertex is the shortest distance from that vertex to its closest descendant leaf.
- A tree or vertex is *balanced* if its height is one greater than its rank.



Previous Results

Theorem (Flajolet and Odlyzko, 1981)

The expected height of any simple variety of tree is asymptotically $k\sqrt{\pi n}$ for some $k > 0$.

Previous Results

UF UNIVERSITY of FLORIDA

Theorem (Flajolet and Odlyzko, 1981)

The expected height of any simple variety of tree is asymptotically $k\sqrt{\pi n}$ for some $k > 0$.

Theorem (Flajolet and Sedgewick, 2009)

The sum of the lengths of paths where the root is one endpoint is of order $n^{3/2}$.

Previous Results

Theorem (C, 2016)

The expected rank of the root of a uniformly chosen general tree approaches

$$\sum_{k=1}^{\infty} \frac{9}{4^{1-k} + 4 + 4^k} \approx 1.62297.$$

The expected rank of a uniformly chosen vertex in a uniformly chosen general tree approaches

$$\sum_{k=1}^{\infty} \frac{3}{4^k + 2} \approx 0.727649.$$

Counting Leaves

Proposition

Let $T(x)$ be the generating function for the number of trees on n vertices in some simple variety of trees, and let $L(x)$ be the generating function for the number of trees with a marked leaf in the same family. Then

$$L(x) = \frac{x^2 T'(x)}{T(x)}.$$

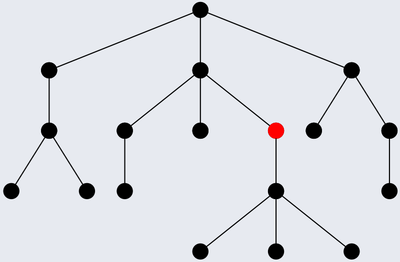
It suffices to show that there is a bijection between trees with a marked vertex and pairs of trees and trees with a marked leaf with one vertex removed. This would show that

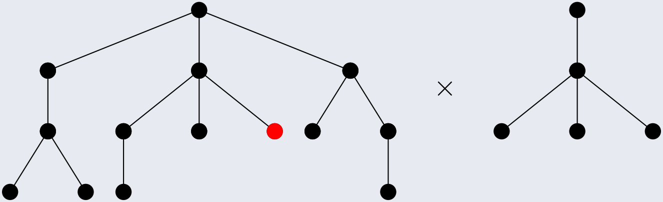
$$xT'(x) = V(x) = \frac{T(x)L(x)}{x}.$$

Counting Leaves

UF UNIVERSITY of FLORIDA

Proof.





□

Counting Paths

- Paths from the root: $V(x) - T(x)$

Counting Paths

- Paths from the root: $V(x) - T(x)$
- Paths from the root to a leaf: $L(x) - x$

Counting Paths

- Paths from the root: $V(x) - T(x)$
- Paths from the root to a leaf: $L(x) - x$
- Any vertical path: $\frac{L(x)}{x}(V(x) - T(x))$

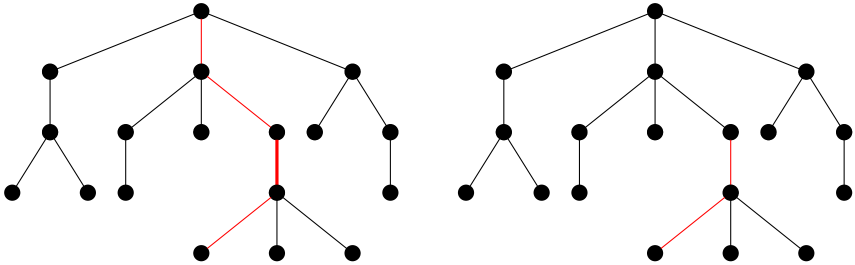
Counting Paths

UF UNIVERSITY of FLORIDA

- Paths from the root: $V(x) - T(x)$
- Paths from the root to a leaf: $L(x) - x$
- Any vertical path: $\frac{L(x)}{x}(V(x) - T(x))$
- Vertical paths that end in a leaf: $\frac{L(x)}{x}(L(x) - x)$

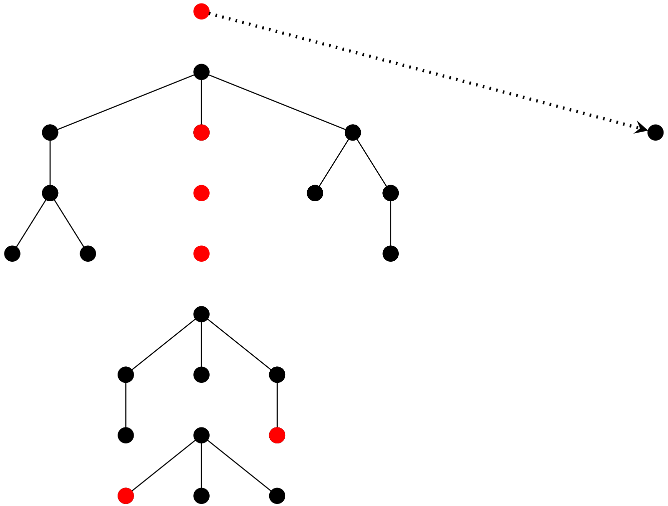
Counting Paths

- Paths from the root: $V(x) - T(x)$
- Paths from the root to a leaf: $L(x) - x$
- Any vertical path: $\frac{L(x)}{x}(V(x) - T(x))$
- Vertical paths that end in a leaf: $\frac{L(x)}{x}(L(x) - x)$
- Edges in paths from the root: $\frac{L(x)}{x}(V(x) - T(x))$



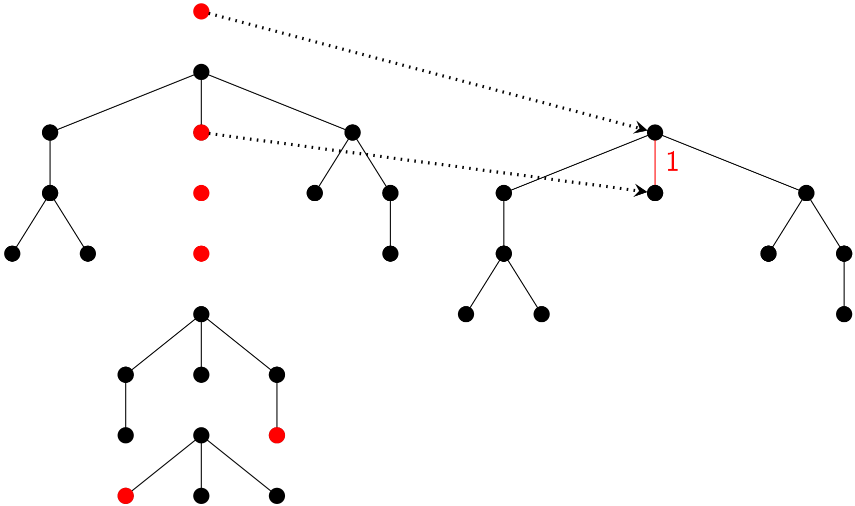
Higher Moments

Consider the generating function $\left(\frac{L(x)}{x}\right)^k (V(x) - T(x))$.



Higher Moments

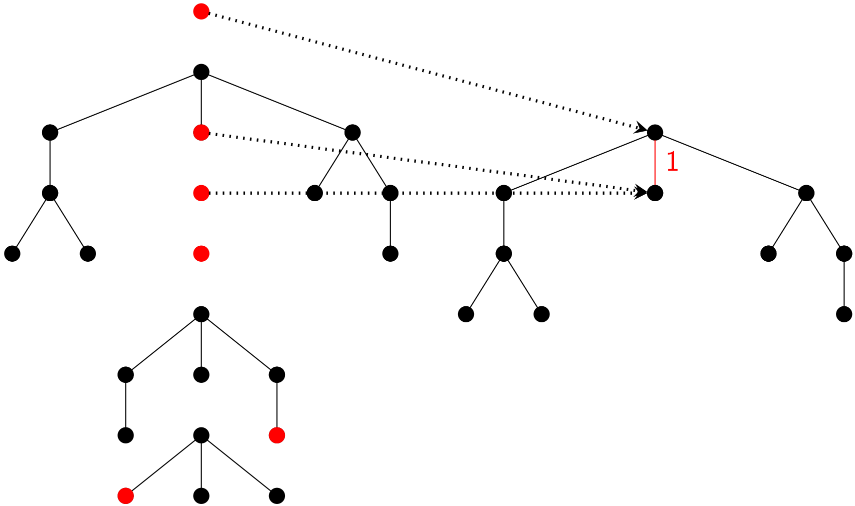
Consider the generating function $\left(\frac{L(x)}{x}\right)^k (V(x) - T(x))$.



Higher Moments

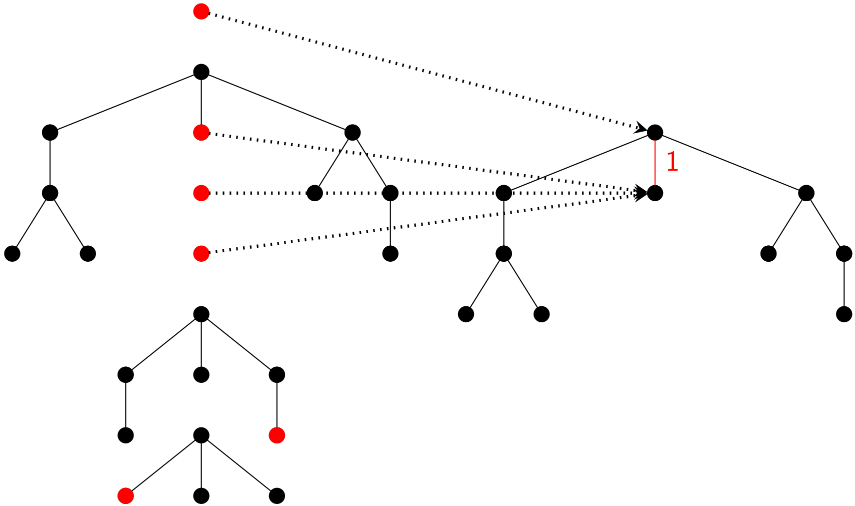
UF UNIVERSITY of FLORIDA

Consider the generating function $\left(\frac{L(x)}{x}\right)^k (V(x) - T(x))$.



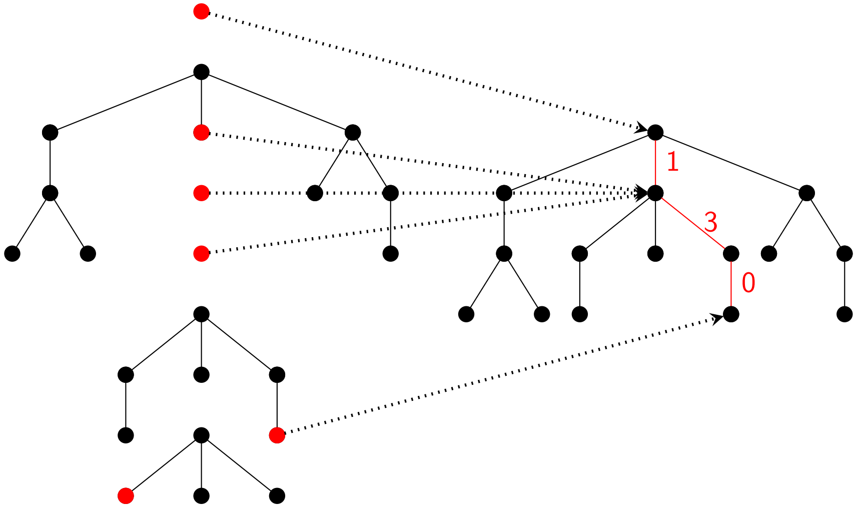
Higher Moments

Consider the generating function $\left(\frac{L(x)}{x}\right)^k (V(x) - T(x))$.



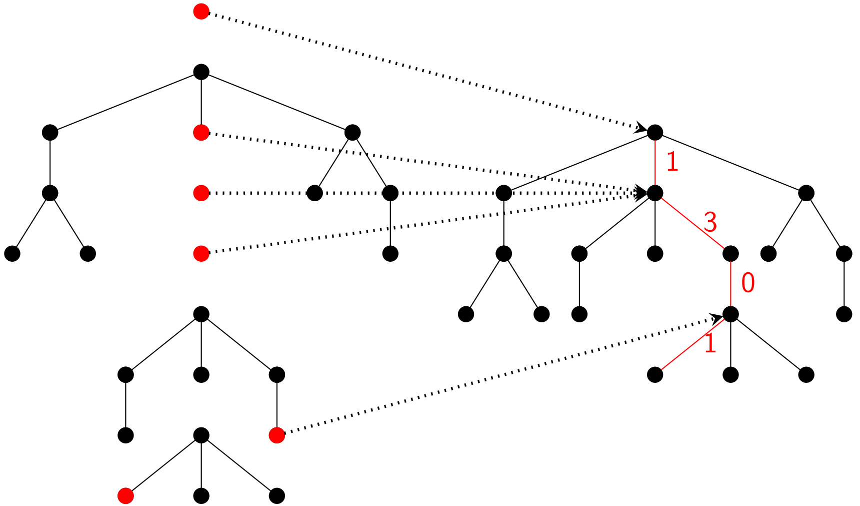
Higher Moments

Consider the generating function $\left(\frac{L(x)}{x}\right)^k (V(x) - T(x))$.



Higher Moments

Consider the generating function $\left(\frac{L(x)}{x}\right)^k (V(x) - T(x))$.



Higher Moments

Proposition

The generating function $\left(\frac{L(x)}{x}\right)^k (V(x) - T(x))$ *counts the number of paths in all trees of some variety on* n *vertices, with each path weighted by the number of* k *element multi-sets of the edges.*

Proposition

The generating function for the number of paths from the root in all trees of some variety on n *vertices, with each path weighted by the length of the path to the* k*th power is a polynomial of degree* k *of the form*

$$\left(k!\left(\frac{L(x)}{x}\right)^k - \frac{k!(k-1)}{2}\left(\frac{L(x)}{x}\right)^{k-1} + Q_k\left(\frac{L(x)}{x}\right)\right)(V(x) - T(x)),$$

where $Q_k(x)$ *is a polynomial in* $\frac{L(x)}{x}$ *of degree* $k-2$ *or less.*

Higher Moments

UF UNIVERSITY of FLORIDA

Proof.

Let the length of a given path be ℓ. Then in the generating function $\left(\frac{L(x)}{x}\right)^k (V(x) - T(x))$ has weight $\binom{\ell+k-1}{k} = \frac{1}{k!}\left(\ell^k - \frac{k(k-1)}{2}\ell^{k-1} + \dots\right)$, this gives the leading coefficient. Proceeding by induction, we can remove the lower order terms by subtracting lower order polynomials.

Higher Moments

UF UNIVERSITY of FLORIDA

Proof.

Let the length of a given path be ℓ. Then in the generating function $\left(\frac{L(x)}{x}\right)^k (V(x) - T(x))$ has weight $\binom{\ell+k-1}{k} = \frac{1}{k!}\left(\ell^k - \frac{k(k-1)}{2}\ell^{k-1} + \dots\right)$, this gives the leading coefficient. Proceeding by induction, we can remove the lower order terms by subtracting lower order polynomials. □

Asymptotics

UF UNIVERSITY of FLORIDA

In any simple variety of tree,

$$T(x) = a_0 - a_1\sqrt{1-\frac{x}{\rho}} + a_2\left(1-\frac{x}{\rho}\right) + O\left(\left(1-\frac{x}{\rho}\right)^{3/2}\right),$$

with $a_0, a_1 > 0$.

$$V(x) = \frac{a_1}{2\sqrt{1-\frac{x}{\rho}}} - a_2 + O\left(\sqrt{1-\frac{x}{\rho}}\right),$$

$$\frac{L(x)}{x} = \frac{a_1}{2a_0\sqrt{1-\frac{x}{\rho}}} + \frac{\left(a_1^2 - 2a_0a_2\right)}{2a_0^2} + O\left(\sqrt{1-\frac{x}{\rho}}\right).$$

Asymptotics

Let $X(n)$ be the r.v. whose value is the length of a uniformly randomly selected path from the root of a tree on n vertices to any vertex. Then we can compute

$$\mathbb{E}[X(n)^k] = n^{k/2}\left(\left(\frac{a_1}{a_0}\right)^k \Gamma\left(\frac{k}{2}+1\right) - \frac{a_1^{k-1}\left((k+1)a_0^2 + 2a_2(k+1)a_0 - a_1^2 k\right) k(k-1)\Gamma\left(\frac{k-1}{2}\right)}{4a_0^{k+1}\sqrt{n}}\right) + O\left(n^{k/2-1}\right),$$

as well as corresponding expectations for three other variations of vertical paths.

Asymptotics

UF UNIVERSITY of FLORIDA

In the family of Cayley trees $a_0 = 1, a_1 = \sqrt{2}, a_2 = \frac{2}{3}$. The expected length of a path from

- the root to any vertex: $\sqrt{\frac{\pi n}{2}} - \frac{4}{3}$.
- the root to any leaf: $\sqrt{\frac{\pi n}{2}} - \frac{1}{3}$.
- any vertex to any of its descendants: $\sqrt{\frac{2n}{\pi}} + \frac{8}{3\pi} - 1$.
- any vertex to any of its descendant leaves: $\sqrt{\frac{2n}{\pi}} + \frac{2}{3\pi}$.

Asymptotics

In the family of Cayley trees $a_0 = 1, a_1 = \sqrt{2}, a_2 = \frac{2}{3}$. The expected length of a path from

- the root to any vertex: $\sqrt{\frac{\pi n}{2}} - \frac{4}{3}$.
- the root to any leaf: $\sqrt{\frac{\pi n}{2}} - \frac{1}{3}$.
- any vertex to any of its descendants: $\sqrt{\frac{2n}{\pi} + \frac{8}{3\pi}} - 1$.
- any vertex to any of its descendant leaves: $\sqrt{\frac{2n}{\pi} + \frac{2}{3\pi}}$.

These trends are universal, if done with general terms, fixing a type of top point, leaves are a constant further away ($\frac{a_1^2}{2a_0^2}$ with roots, $\frac{a_1^2(\pi-2)}{2a_0^2(\pi)}$ with vertices) plus an error term, fixing a type of bottom point, roots are further by a factor of $\frac{\pi}{2}$ plus an error term (which depends on the type of tree).

Related Problems

- Expected length of paths from the root to a leaf in various families
- Bijective proofs of correspondence between all paths and vertical paths