Balanced vertices in rooted trees

Miklós Bóna

Department of Mathematics
University of Florida
Gainesville FL 32611-8105

bona@ufl.edu

May 20, 2017

Rooted Trees

Various parameters of a many kinds of rooted trees are fairly well understood *if they originate at the root*.

Rooted Trees

Various parameters of a many kinds of rooted trees are fairly well understood *if they originate at the root*.

For instance, the average number of vertices at distance d from the root can be computed, the average root degree can be computed, and so on.

Rooted Trees

Various parameters of a many kinds of rooted trees are fairly well understood *if they originate at the root*.

For instance, the average number of vertices at distance d from the root can be computed, the average root degree can be computed, and so on.

However, much less is known when we start counting *from the bottom up*, that is, from the leaves.

k-protected vertices

Why should we care?

k-protected vertices

Why should we care?

Vertices of a network that are close to a leaf may be vulnerable to attacks.

k-protected vertices

Why should we care?

Vertices of a network that are close to a leaf may be vulnerable to attacks.

Alternatively, people corresponding to vertices far from leaves may represent people who were not active lately.

Decreasing binary trees

While we studied several tree varieties, most of our results are about *decreasing binary trees*, also called *binary search trees*. In the first part of the talk, we will discuss results about these trees.

Decreasing binary trees

While we studied several tree varieties, most of our results are about *decreasing binary trees*, also called *binary search trees*. In the first part of the talk, we will discuss results about these trees.

These are plane trees in which every vertex has at most two children, and each child is a left or right child of its parent, even if it is an only child. The vertices are bijectively labeled by the numbers $1, 2, \cdots, n$, and the label of each vertex is smaller than that of its parent.

Decreasing binary trees

While we studied several tree varieties, most of our results are about *decreasing binary trees*, also called *binary search trees*. In the first part of the talk, we will discuss results about these trees.

These are plane trees in which every vertex has at most two children, and each child is a left or right child of its parent, even if it is an only child. The vertices are bijectively labeled by the numbers $1, 2, \cdots, n$, and the label of each vertex is smaller than that of its parent.

These treees are in natural bijection with permutations of length n, so their number is $n!$.

A decreasing binary tree

In the tree $T(\pi)$ of the permutation π, the root will have label n, the entries on the left of n will go in the left subtree, and the entries on the right of n will go in the right subtree. These subtrees will be defined recursively by the same rule.

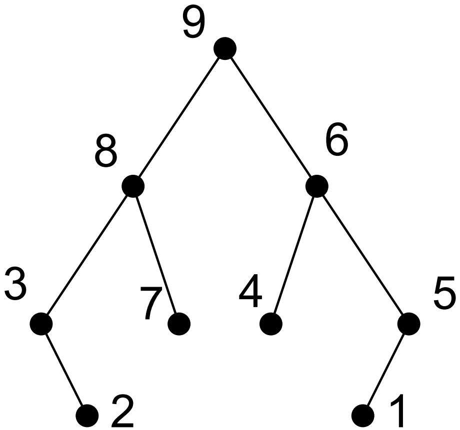

Figure: The tree $T(\pi)$ for $\pi = 328794615$.

Rank

Let the rank of a vertex v in a tree be the distance (number of edges) of the shortest path from v to any leaf that is a descendant of v.

Rank

Let the rank of a vertex v in a tree be the distance (number of edges) of the shortest path from v to any leaf that is a descendant of v.

So leaves have rank 0, neighbors of leaves have rank 1, and so on.

The ratio of vertices of rank k

In earlier work, we proved that if $F_{n,k}$ is the number of all vertices of rank k in all trees of size n, then

$$\lim_{n\to\infty} \frac{F_{n,k}}{n \cdot n!} = c_k,$$

for a positive rational number c_k, and we computed the numbers c_k for $k \leq 5$.

The ratio of vertices of rank k

In earlier work, we proved that if $F_{n,k}$ is the number of all vertices of rank k in all trees of size n, then

$$\lim_{n\to\infty} \frac{F_{n,k}}{n \cdot n!} = c_k,$$

for a positive rational number c_k, and we computed the numbers c_k for $k \leq 5$.

It follows from those numbers c_k that for large n, about 99.75 percent of vertices are of rank five or less.

Balanced vertices

A vertex v is called *balanced* if *all* descending paths from v to a leaf have the same length.

Balanced vertices

A vertex v is called *balanced* if *all* descending paths from v to a leaf have the same length. In Figure 8, all vertices are balanced, except for 6, 8, and 9.

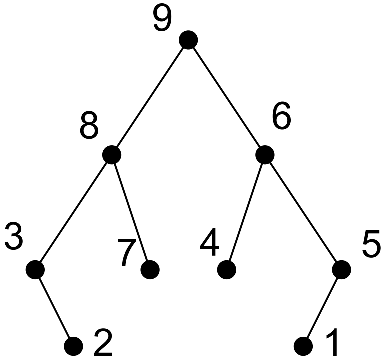

Figure: The tree $T(\pi)$ for $\pi = 328794615$.

Small ranks

All leaves are balanced.

Small ranks

All leaves are balanced.

A vertex of $T(\pi)$ is a leaf iff the corresponding entry of π is smaller than both of its neighbors, and that happens one third of the time. So $C_0 = 1/3$.

Small ranks

All leaves are balanced.

A vertex of $T(\pi)$ is a leaf iff the corresponding entry of π is smaller than both of its neighbors, and that happens one third of the time. So $C_0 = 1/3$.

A vertex of rank 1 is balanced iff all of its children are leaves.

Small ranks

All leaves are balanced.

A vertex of $T(\pi)$ is a leaf iff the corresponding entry of π is smaller than both of its neighbors, and that happens one third of the time. So $C_0 = 1/3$.

A vertex of rank 1 is balanced iff all of its children are leaves.

It follows from elementary considerations (like those for leaves) concerning the neighbors and the second neighbors of an entry of π that one fifth of all vertices are like this, so $C_1 = 1/5$.

General Rank

For larger values of k, this type of argument will not work, since the parent of a vertex of rank k does not have to have rank $k+1$. It can have any rank between 1 and $k+1$.

General Rank

For larger values of k, this type of argument will not work, since the parent of a vertex of rank k does not have to have rank $k+1$. It can have any rank between 1 and $k+1$.

So an analytic approach is needed.
Let $A_k(x)$ be the exponential generating function for the number of balanced vertices of rank k in all $n!$ trees of size n.

General Rank

For larger values of k, this type of argument will not work, since the parent of a vertex of rank k does not have to have rank $k+1$. It can have any rank between 1 and $k+1$.

So an analytic approach is needed.
Let $A_k(x)$ be the exponential generating function for the number of balanced vertices of rank k in all $n!$ trees of size n.

Let $B_k(x)$ be the exponential generating function for the number of trees of size n in which *the root* is balanced, and is of rank k.

Differential equations

Then by the Exponential formula, we have

Lemma

For $k \geq 1$, the linear differential equation

$$A_k'(x) = \frac{2}{1-x} \cdot A_k(x) + B_k'(x)$$

holds, with initial condition $A_k(0) = 0$.

Differential equations

Then by the Exponential formula, we have

Lemma

For $k \geq 1$, the linear differential equation

$$A'_k(x) = \frac{2}{1-x} \cdot A_k(x) + B'_k(x)$$

holds, with initial condition $A_k(0) = 0$.

Crucially, and this is different from the problem of counting all vertices of a given rank, $B_k(x)$ is a *polynomial*, since a tree whose root is balanced and of rank k can have at most $2^{k+1} - 1$ vertices.

The form of A_k

Solving the linear differential equation for $A_k(x)$, we get that

$$A_k(x) = \frac{\int (1-x)^2 B_k'(x)\ dx}{(1-x)^2}.$$

The form of A_k

Solving the linear differential equation for $A_k(x)$, we get that

$$A_k(x) = \frac{\int (1-x)^2 B_k'(x)\ dx}{(1-x)^2}.$$

So $A_k(x)$ is a rational function with denominator $(1-x)^2$, and so the C_k exist, and are computable from $A_k(x)$.

First few values

- $C_0 = 1/3$
- $C_1 = 1/5$
- $C_2 = 52/567$
- $C_3 = 7175243/222660900.$

First few values

- $C_0 = 1/3$
- $C_1 = 1/5$
- $C_2 = 52/567$
- $C_3 = 7175243/222660900$.

This shows that for large n, about 65.7 percent of all vertices of decreasing binary trees are balanced and of rank at most three.

First few values

- $C_0 = 1/3$
- $C_1 = 1/5$
- $C_2 = 52/567$
- $C_3 = 7175243/222660900.$

This shows that for large n, about 65.7 percent of all vertices of decreasing binary trees are balanced and of rank at most three.

More computation shows that for n sufficiently large, about 66.62 percent of all vertices are balanced and of rank at most four, and about 66.84 percent are balanced and of rank at most five.

Monotonicity

Let P_n be the probability that vertex chosen uniformly from the set of all vertices of all decreasing binary trees on $[n]$ is balanced. Our goal is to prove the following.

Monotonicity

Let P_n be the probability that vertex chosen uniformly from the set of all vertices of all decreasing binary trees on $[n]$ is balanced. Our goal is to prove the following.

Theorem

The sequence $P_1, P_2, \cdots$ is weakly decreasing.

Fixed rank

Let $p_{n,k}$ be the probability that the *root* of a randomly selected tree on n vertices is balanced, and is of rank k. Set $p_{0,i} = 1$ for all i.

Fixed rank

Let $p_{n,k}$ be the probability that the *root* of a randomly selected tree on n vertices is balanced, and is of rank k. Set $p_{0,i} = 1$ for all i.

Lemma

For all $n \geq 1$ and all fixed $k \leq n$, the inequality $p_{n+1,k} \leq p_{n,k}$ holds.

Induction on n. True for all k if $n \geq 3$, since then $p_{n,k} = 1$. Now let us assume that the statement is true for n and prove it for $n+1$.

Induction on n. True for all k if $n \geq 3$, since then $p_{n,k} = 1$. Now let us assume that the statement is true for n and prove it for $n+1$.

Let π be a permutation of length $n+1$. The probability that the largest entry of π is in position $i+1$ for any $i \in [0, n]$ is $1/(n+1)$. The root of $T(\pi)$ is balanced of rank k if and only if all its children are balanced of rank $k-1$,

Induction on n. True for all k if $n \geq 3$, since then $p_{n,k} = 1$. Now let us assume that the statement is true for n and prove it for $n+1$.

Let π be a permutation of length $n+1$. The probability that the largest entry of π is in position $i+1$ for any $i \in [0,n]$ is $1/(n+1)$. The root of $T(\pi)$ is balanced of rank k if and only if all its children are balanced of rank $k-1$,

so

$$p_{n+1,k} = \frac{\sum_{i=0}^{n} p_{i,k-1} p_{n-i,k-1}}{n+1}. \tag{1}$$

pause Replacing $n+1$ by n, we get the analogous formula

Induction on n. True for all k if $n \geq 3$, since then $p_{n,k} = 1$. Now let us assume that the statement is true for n and prove it for $n+1$.

Let π be a permutation of length $n+1$. The probability that the largest entry of π is in position $i+1$ for any $i \in [0, n]$ is $1/(n+1)$. The root of $T(\pi)$ is balanced of rank k if and only if all its children are balanced of rank $k-1$,

so

$$p_{n+1,k} = \frac{\sum_{i=0}^{n} p_{i,k-1} p_{n-i,k-1}}{n+1}. \tag{1}$$

pause Replacing $n+1$ by n, we get the analogous formula

$$p_{n,k} = \frac{\sum_{i=0}^{n-1} p_{i,k-1} p_{n-1-i,k-1}}{n}. \tag{2}$$

Trick

Compare

$$p_{0,2}p_{6,2} + p_{1,2}p_{5,2} + p_{2,2}p_{4,2} + p_{3,2}p_{3,2} + p_{4,2}p_{2,2} + p_{5,2}p_{1,2} + p_{6,2}p_{0,2}$$

Trick

Compare

$$p_{0,2}p_{6,2} + p_{1,2}p_{5,2} + p_{2,2}p_{4,2} + p_{3,2}p_{3,2} + p_{4,2}p_{2,2} + p_{5,2}p_{1,2} + p_{6,2}p_{0,2}$$

and

$$p_{0,2}p_{5,2} + p_{1,2}p_{4,2} + p_{2,2}p_{3,2} + p_{3,2}p_{2,2} + p_{4,2}p_{1,2} + p_{5,2}p_{0,2}.$$

Trick

Compare

$$p_{0,2}p_{6,2} + p_{1,2}p_{5,2} + p_{2,2}p_{4,2} + p_{3,2}p_{3,2} + p_{4,2}p_{2,2} + p_{5,2}p_{1,2} + p_{6,2}p_{0,2}$$

and

$$p_{0,2}p_{5,2} + p_{1,2}p_{4,2} + p_{2,2}p_{3,2} + p_{3,2}p_{2,2} + p_{4,2}p_{1,2} + p_{5,2}p_{0,2}.$$

Say the jth summand of the top sum is minimal. Then compare the ith summand on the top with the ith summand at the bottom if $i < j$, and the ith summand at the top with the $(i-1)$st summand at the bottom if $i > j$.

This shows that the top sum is at most $7/6$ (or, in the general case, $(n+1)/n$) times the bottom sum, proving the lemma.

Corollary

Let p_n be the probability that the root of a decreasing binary tree on $[n]$ is balanced. Then $p_n \geq p_{n+1}$.

Proof: It follows from our definitions that

Corollary

Let p_n be the probability that the root of a decreasing binary tree on $[n]$ is balanced. Then $p_n \geq p_{n+1}$.

Proof: It follows from our definitions that

$$p_n = \sum_{k=1}^{n-1} p_{n,k}$$

and

Corollary

Let p_n be the probability that the root of a decreasing binary tree on $[n]$ is balanced. Then $p_n \geq p_{n+1}$.

Proof: It follows from our definitions that

$$p_n = \sum_{k=1}^{n-1} p_{n,k}$$

and

$$p_{n+1} = \sum_{k=1}^{n} p_{n+1,k}.$$

As the lemma shows that $p_{n+1,k} \leq p_{n,k}$ for $k \leq n$, the only issue that we must consider is that the sum yielding p_{n+1} has one more summand than that yielding p_n.

As the lemma shows that $p_{n+1,k} \leq p_{n,k}$ for $k \leq n$, the only issue that we must consider is that the sum yielding p_{n+1} has one more summand than that yielding p_n.

However, this is not a problem, since for all $n \geq 2$, we have $p_{n,n-1} = 2^{n-1}/n!$, while $p_{n+1,n-1} = 2^{n-1}/(n+1)!$ and $p_{n+1,n} = 2^n/(n+1)!$, so

As the lemma shows that $p_{n+1,k} \leq p_{n,k}$ for $k \leq n$, the only issue that we must consider is that the sum yielding p_{n+1} has one more summand than that yielding p_n.

However, this is not a problem, since for all $n \geq 2$, we have $p_{n,n-1} = 2^{n-1}/n!$, while $p_{n+1,n-1} = 2^{n-1}/(n+1)!$ and $p_{n+1,n} = 2^n/(n+1)!$, so

$$p_{n,n-1} = \frac{2^{n-1}}{n!} \geq \frac{3 \cdot 2^{n-1}}{n+1} = p_{n+1,n-1} + p_{n+1,n}.$$

This inequality, and applying the lemma for all $k \leq n-2$, proves our claim.

Finishing the proof of monotonicity

Induction on n. In order to prove that $P_n \geq P_{n+1}$, note that a random vertex of a tree of size n has $1/n$ probability to be the root, and it has, for each $i \in [n-1]$, exactly $1/n$ probability to be a vertex in a subtree of size i which is the left subtree or right subtree of the root. Therefore, the inequality $P_n \geq P_{n+1}$ is equivalent to the inequality

Finishing the proof of monotonicity

Induction on n. In order to prove that $P_n \geq P_{n+1}$, note that a random vertex of a tree of size n has $1/n$ probability to be the root, and it has, for each $i \in [n-1]$, exactly $1/n$ probability to be a vertex in a subtree of size i which is the left subtree or right subtree of the root. Therefore, the inequality $P_n \geq P_{n+1}$ is equivalent to the inequality

$$P_n = \frac{p_n + \sum_{i=1}^{n-1} P_i}{n} \geq \frac{p_{n+1} + \sum_{i=1}^{n} P_i}{n+1} = P_{n+1}. \qquad (3)$$

The inequality in (3) is true, since the first equality in (3) shows that P_n is obtained as the average of the n values in the set $S = \{p_n, P_1, P_2, \cdots, P_{n-1}\}$.

The inequality in (3) is true, since the first equality in (3) shows that P_n is obtained as the average of the n values in the set $S = \{p_n, P_1, P_2, \cdots, P_{n-1}\}$.

That average does not change if we extend S by adding P_n (the average of the values in S) to it. Then, if we replace p_n by p_{n+1}, the average of the new set $S' = \{p_{n+1}, P_1, P_2, \cdots, P_n\}$ is at most as large as the average of S, (since $p_{n+1} \leq p_n$ by Corollary 4), while the average of S' is P_{n+1} by the second equality in (3).

Limiting probability

As the sequence of the P_n is monotone decreasing, its limit exists.

Limiting probability

As the sequence of the P_n is monotone decreasing, its limit exists.

The limit is in the interval $[0.6684, 0.66965]$.

Limiting probability

As the sequence of the P_n is monotone decreasing, its limit exists.

The limit is in the interval $[0.6684, 0.66965]$.

This is obtained by computing C_k for $k \leq 5$, then saying that there are very few vertices (balanced or not) of rank more than five.

What can be said about other trees?

There are numerous other varieties of labeled rooted trees for which we could ask the same question.

What can be said about other trees?

There are numerous other varieties of labeled rooted trees for which we could ask the same question.

These include plane trees in which every vertex can have at most k children, or any number of children, and non-plane trees with similar conditions.

What can be said about other trees?

There are numerous other varieties of labeled rooted trees for which we could ask the same question.

These include plane trees in which every vertex can have at most k children, or any number of children, and non-plane trees with similar conditions.

When counting all vertices of a given rank, one fact that makes life harder is that the analogous versions of $A_k(x)$ will not be elementary functions if $k \geq 2$.

An example: non-plane 1-2 trees

In such trees, each vertex has a label smaller than its parent, each vertex has at most two children, but left or right does not matter.

An example: non-plane 1-2 trees

In such trees, each vertex has a label smaller than its parent, each vertex has at most two children, but left or right does not matter.

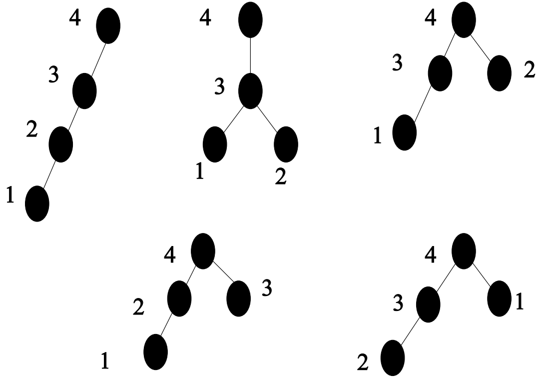

Figure: The five rooted non-plane 1-2 trees on vertex set [4].

Euler numbers

It is well known that the number of such trees on $[n]$ is the *Euler number* E_n, which counts, among other things, alternating permutations of length n.

Euler numbers

It is well known that the number of such trees on $[n]$ is the *Euler number* E_n, which counts, among other things, alternating permutations of length n.

It is also well known that

$$\sum_{n\geq 0} E_n \frac{x^n}{n!} = \tan x + \sec x.$$

All vertices

When counting all vertices of a given rank, we can set up a sequence of differential equations like before. Then, using some complex analysis, we can compute that

All vertices

When counting all vertices of a given rank, we can set up a sequence of differential equations like before. Then, using some complex analysis, we can compute that

$$c_0 = 1 - \frac{2}{\pi} \approx 0.36338,$$

and

$$c_1 = 2 - \frac{\pi^2}{24} - \frac{4}{\pi} \approx 0.31553.$$

All vertices

When counting all vertices of a given rank, we can set up a sequence of differential equations like before. Then, using some complex analysis, we can compute that

$$c_0 = 1 - \frac{2}{\pi} \approx 0.36338,$$

and

$$c_1 = 2 - \frac{\pi^2}{24} - \frac{4}{\pi} \approx 0.31553.$$

We cannot get any further, since we cannot solve the relevant linear differential equations. This is because functions like $x \tan x$ do not have an elementary antiderivative.

Results for balanced vertices

Theorem

Let $H_k(x)$ be the exponential generating function for the number of balanced vertices of rank k in all non-plane 1-2 trees of size n. Then

Results for balanced vertices

Theorem

Let $H_k(x)$ be the exponential generating function for the number of balanced vertices of rank k in all non-plane 1-2 trees of size n. Then

$$H_k(x) = \frac{\int b_k'(x)(1-\sin x)\ dx}{1-\sin x},$$

where $b_k(x)$ is the exponential generating function of such trees in which the root is balanced and is of rank k.

Results for balanced vertices

Theorem

Let $H_k(x)$ be the exponential generating function for the number of balanced vertices of rank k in all non-plane 1-2 trees of size n. Then

$$H_k(x) = \frac{\int b_k'(x)(1 - \sin x)\ dx}{1 - \sin x},$$

where $b_k(x)$ is the exponential generating function of such trees in which the root is balanced and is of rank k.

Note that $b_k'(x)$ is is a *polynomial*. Therefore, integral in the numerator is an elementary function since the integral of $x^n \sin x$ is an elementary function for all positive integers n.

Numerical results

For $k = 0$, we get nothing new, since vertices of rank 0 are leaves, and they are all balanced.

Numerical results

For $k = 0$, we get nothing new, since vertices of rank 0 are leaves, and they are all balanced.

For $k = 1$, we have

$$H_1(x) = \frac{6x\cos(x) - 6\cos(x) + 3x^2\cos(x) - 6x\sin(x) - 6\sin(x) + P(x)}{6(1 - \sin(x))},$$

where $P(x) = x^3 + 6 + 3x^2$. This yields that for large n, about

$$\frac{\pi}{4} + \frac{\pi^2}{24} - 1 \approx 0.1966$$

of all vertices are balanced and of rank 1.