RANDOM PLANE INCREASING TREES: ASYMPTOTIC ENUMERATION OF VERTICES BY DISTANCE FROM LEAVES

MIKLÓS BÓNA AND BORIS PITTEL

We dedicate this paper to the precious memories of Miklós Bóna, the father of the first author, and of Irina Pittel, the wife of the second author.

Abstract. We prove that for any fixed $k$, the probability that a random vertex of a random increasing plane tree is of rank $k$, that is, the probability that a random vertex is at distance $k$ from the leaves, converges to a constant $c_k$ as the size $n$ of the tree goes to infinity. We prove that $1 - \sum_{j \leq k} c_k < \frac{k^{k+1}}{(2k+1)!}$, so that the tail of the limiting rank distribution is super-exponentially narrow. We prove that the latter property holds uniformly for all finite $n$ as well. More generally, we prove that the ranks of a finite uniformly random set of vertices are asymptotically independent, each with distribution $\{c_k\}$. We compute the exact value of $c_k$ for $0 \leq k \leq 3$, demonstrating that the limiting expected fraction of vertices with rank $\leq 3$ is 0.9997. . . We show that with probability $1 - n^{-0.99c}$ the highest rank of a vertex in the tree is sandwiched between $(1 - \varepsilon) \log n / \log \log n$ and $(1.5 + \varepsilon) \log n / \log \log n$, and that this rank is asymptotic to $\log n / \log \log n$ with probability $1 - o(1)$.

1. Introduction

1.1. Definitions and Background. The analysis of various parameters of random rooted trees is a classic subject if those parameters relate to a near-root part of a tree or to global tree structure. See [11] for a comprehensive treatment of these results. A more recent area of interest is the random tree fringe, that is, the part close to the leaves. A list of articles on this subject can be found in [6].

A particularly interesting problem is the enumeration of vertices of a given rank. If $v$ is a vertex of a tree $T$, then let the rank of $v$ be the number of edges in the shortest path from $v$ to a leaf of $T$ that is a descendant of $v$. So leaves are of rank 0, parents of leaves are of rank 1, and so on. Let us
fix a variety of trees, such as binary search trees, or non-plane binary trees, or increasing trees. Let \( n \) be a fixed positive integer, and let us compute the probability that a randomly selected vertex of a randomly selected tree of size \( n \) is of rank \( k \). What can we say about this probability as \( n \) goes to infinity?

This question for various models of random trees has been the subject of vigorous research in the last few years. Perhaps the first was a random binary search tree process \( \{T_n\} \): conditioned on a binary tree \( T_n \) with \( n \) vertices, \( T_{n+1} \) is obtained by selecting uniformly at random (u.a.r.) one of \( n \) available positions for a new leaf and putting the next vertex into it. It was proved in [5] that for any fixed \( k \), the probability in question converges to a rational number \( c_k \), and \( c_k \) was computed for \( k \leq 3 \). Then in [6], the present authors computed \( c_4 \) and \( c_5 \), and proved lower and upper bounds for \( c_k \) in general. They proved that as \( k \) grows, \( c_k \) exponentially decays. They also showed that the ranks of the uniformly random, fixed size, sample of vertices are asymptotically independent. For increasing non-plane trees where the next vertex is attached to a previous vertex selected u.a.r. from the existing vertices, known also as recursive trees, ranks were extensively studied by Holmgren and Janson in [16]. They also studied ternary trees from this aspect [15]. In [7], the first author and István Mező proved the existence of the limits \( c_k \) for non-plane 1-2 trees and plane 1-2-trees.

A natural tree variety that is missing from the above list is plane increasing trees. These are rooted plane trees whose vertices are bijectively labeled with the elements of \([n]\) so that each vertex has a larger label than its parent, and the children of every parent form an ordered set. This tree variety is the subject of our paper. As these are the only kind of trees that will appear in this work, we will often simply call these rooted plane increasing trees just trees. These trees have been studied under various names, such as plane-oriented trees [17, 21], PORTs (see page 14 of [9]), or scale free trees [1].

Given a tree with \( n-1 \) vertices of total degrees \( d_1, \ldots, d_{n-1} \), there are \( d_{\text{root}} + 1 + \sum_{j \in [2, n-1]} d_j = 2n-3 \) different ways to attach vertex \( n \) to a vertex in \([n-1]\). Indeed, there are \( d_{\text{root}} + 1 \) ways to attach \( n \) to the root, and \( d_j \) ways to attach \( n \) to any other non-root vertex \( j \in [2, n-1] \). So the total number of these trees is \((2n-3)!! := 1 \cdot 3 \cdots \cdot (2n-3)\). Our tree \( T_n \) is chosen u.a.r. from the set of all \((2n-3)!!\) trees. We can view \( T_1, T_2, \ldots \) as snapshots of the tree growing (Markov) process: conditioned on \( T_{n-1} \), vertex \( n \) becomes a child of a vertex \( j \in [n-1] \) taking up one of \( d_j \) possible positions in the ordered set of the other children of \( j \). So, unlike in recursive trees, a host vertex for the vertex \( n \) is chosen from existing vertices with probability proportional to total degree of that potential host. The second author [25] proved that
the height, that is, the length of the longest path from the root to a leaf of this tree, scaled by \( \log n \) converges, almost surely, to \( 1.795560738 \), which is the root of the equation \( 2y/e = \exp(1/2y) \). Later, Drmota \[8\] extended this result to a broad variety of increasing trees introduced by Bergeron, Flajolet, and Salvy \[2\].

In a different direction, \( \{T_n\}_{n \geq 1} \) is a well-known case of the random tree (graph, if cycles allowed) process with “preferential attachment”. See Barabási and Albert \[1\], Móri \[22\], \[23\], Bollobás and Riordan \[4\], Bollobás \[3\], van der Hofstad \[14\], Frieze and Karoński \[10\], Pittel \[26\].

1.2. Our Main Results. As we mentioned above, in this paper, the tree variety we study is that of increasing plane trees. In Section \[3\] we prove that for this tree variety, the numbers \( c_k \) exist for all \( k \), and that \( c_k \) is equal to the value of a certain integral. We provide the explicit value of this integral for \( k = 0 \) and \( k = 1 \), and its approximate value (to eleven decimals) for \( k = 2 \) and \( k = 3 \). We also prove a non-obvious result that \( \sum_{k \geq 0} c_k = 1 \), so the numbers \( c_k \) form a probability distribution.

In order to see that this fact is not obvious and has to be proved, consider the trivial variety of trees in which every non-leaf vertex has exactly one child. In that case, the ratio of vertices of rank \( k \) in a tree of size \( n \) is \( c_k(n) = 1/n \) if \( k < n \), so \( c_k = 0 \) for any fixed \( k \). That is, for this tree variety, \( \sum_{k} c_k = 0 < 1 \).

For an arbitrary tree variety, only the following holds. As \( \sum_{k=0}^{n} c_k(n) = 1 \), it follows that for each fixed bound \( B \), we have \( \sum_{k \leq B} c_k(n) \leq 1 \), and taking limits as \( n \) goes to infinity, \( \sum_{k=0}^{B} c_k \leq 1 \). Now taking limits as \( B \) goes to infinity, we get \( \sum_{k \geq 0} c_k \leq 1 \). So the sum of the \( c_k \) is never more than 1, i.e. \( \sum_{k \geq 0} c_k \in [0, 1] \). To show for a particular increasing tree that \( \sum_{k \geq 0} c_k = 1 \), one has to prove tightness condition, namely that \( \sum_{k=K}^{n} c_k(n) \) can be made arbitrarily small for all \( n \) by selecting a constant \( K \) that is sufficiently large. Less formally, one has to show that, as a tree in question grows indefinitely, with high probability no substantial portion of the vertices will have higher and higher ranks, exceeding any fixed \( k \). That’s what we will prove for the tree growing process studied in this paper.

In fact, we will prove not only that \( \sum_{k} c_k = 1 \), but also the stronger statement that \( 1 - \sum_{j \leq k} c_j \leq \frac{3^{k+1}}{(2k+1)!} \). So the tail of the limiting rank distribution is super-exponentially narrow. The latter property holds uniformly for all finite \( n \) as well. These results together imply that as \( n \) goes to infinity, the limiting probability that a random vertex of a random tree has rank four or more is approximately 0.0002843360. So about 99.97 percent of all vertices have rank three or less.
Using our results on the expected number of vertices of a given rank, we are able to prove lower and upper bounds on the highest rank $R_n$ of a vertex in a tree of size $n$ selected uniformly at random. First, in Theorem 3.3, we show that, for each $\epsilon > 0$, the inequality $R_n \leq (1.5 + \epsilon) \log n / \log \log n$ holds with probability $1 - n^{-\epsilon-o(1)}$. (Bollobás and Riordan [3] proved that with probability $1-o(1)$, the diameter of the preferential attachment tree is asymptotic to $\log n / \log \log n$.) Later, in Theorem 3.11, focusing on vertices whose descendant tree is a path we prove that if $0 < \epsilon < 1$, then with probability $1 - n^{-\epsilon-o(1)}$, there exist $\Theta(n^\epsilon)$ vertices with rank $k = (1-\epsilon) \log n / \log \log n$. (This result obviously implies Bollobás-Riordan’s lower bound for the diameter.) In other words, with probability $1-o(1)$, the highest rank is asymptotic to $\log n / \log \log n$.

In Section 3.2 we turn to the counting of $r$-tuples of vertices with given ranks. We are able to prove that the ranks of a finite, uniformly random, set of vertices are asymptotically independent. In other words, if $R_n(1), R_n(2), \cdots, R_n(r)$ denote the ranks of $r$ vertices, chosen uniformly at random, with order and without replacement, from the set of $n$ vertices of a random tree, then

$$P \left( \bigcap_{i=1}^{r} (R_n(i) = k_i) \right) = (1 + O(n^{-1})) \prod_{j=1}^{r} c_{k_j}.$$ 

2. Counting vertices by ranks via generating functions

2.1. Individual vertices and tuples of vertices. The following result is well-known [21], and we have proved it in a purely combinatorial way in the introduction. Nevertheless, here we include another proof, to show the method we will use later in more complicated circumstances.

**Proposition 2.1.** Let $N \geq 2$. Then the number $t_n$ of all rooted increasing plane trees on vertex set $[n]$ is given by the formula $t(n) = (2n-3)!!$.

**Proof.** Note that $t(1) = 1$ and for $n \geq 2$ we have the recurrence

$$t(n) = \sum_{s \geq 1} \sum_{j_1, \ldots, j_s = n-1} \frac{(n-1)!}{j_1! \cdots j_s!} \prod_{i=1}^{s} t(j_i), \quad (t(0) := 0),$$

where $s$ is the number of the children of the root. Multiplying both sides by $z^{n-1}/(n-1)!$, summing over $n \geq 2$ and adding $z t(1)$, we obtain that the power series $T(z) := \sum_{n \geq 1} \frac{t(n)}{n!} z^n$ satisfies the equation

$$T'(z) = \sum_{j \geq 0} T^j(z) = (1 - T(z))^{-1}, \quad T(0) = 0,$$
if \(|T(z)| < 1\). The solution of this differential equation is \(T(z) = 1 - \sqrt{1 - 2z}\), where we use the main branch of the square root, and \(|T(z)| < 1\) if \(|z| < 1/2\). It follows that

\[
\frac{t(n)}{n!} = -[z^n](1 - 2z)^{1/2} = \frac{(2n - 3)!!}{n!} \rightarrow t(n) = (2n - 3)!!.
\]

\[\square\]

**Note.** We could have used the symbolic method (Flajolet and Sedgewick [11]) and obtain (2.2) directly, but we decided to present an explicit, admittedly pedestrian, argument where we start with tree-based recurrence for the tree counts in question and then derive recurrences for generating functions. This short proof will serve as a self-contained template for our considerably more technical proofs coming next, especially when we will turn to counting tuples of vertices of given ranks.

**Notation.** Suppose \(|z| < 1/2\). Let \(A_k(z) = \sum_{n \geq 1} a_k(n)z^n/n!\) be the (exponential) generating function for the number of all vertices of rank \(k\) in all rooted plane trees on vertex set \([n]\). Let \(B_k(z) = \sum_{n \geq 1} b_k(n)z^n/n!\) be the generating function for the number of trees in which the root is of rank \(k\), so we have \(B_0(z) = z\). We will also need \(B_{\geq k}(z)\), the generating function for the number of trees in which the root is of rank \(\geq k\). So we have \(B_{\geq 0}(z) = T(z)\), and \(B_{\geq 1}(z)\) is the generating function of numbers of trees with at least two vertices, i.e. \(B_{\geq 1}(z) = T(z) - z\). We will use the inequality \(B_{\geq k+1}(z) \leq B_{\geq k}(z)\) to express the fact that for all \(n\), the coefficient \([z^n]B_{\geq k}(z)\) decreases as \(k\) increases.

**Proposition 2.2.** For all \(k\), the generating functions \(A_k(z)\) and \(B_k(z)\) satisfy

\[
\begin{align*}
A_k'(z) &= \frac{A_k(z)}{(1 - T(z))^2} + B_k'(z), \quad B_k(z) = B_{\geq k}(z) - B_{\geq k+1}(z), \quad k \geq 0, \\
B_{\geq k}'(z) &= \frac{B_{\geq k-1}(z)}{1 - B_{\geq k-1}(z)}, \quad k \geq 1, \quad B_{\geq 0}(z) = T(z).
\end{align*}
\]

**Proof.** Let \(X_k(n)\) stand for the total number of rank \(k\) vertices in the uniformly random tree, so that \(\mathbb{E}[X_k(n)] = \frac{a_k(n)}{t(n)}\). Let \(p_k(n)\) (resp. \(p_{\geq k}(n)\)) stand for probability that the root 1 is of rank \(k\) (resp. rank \(\geq k\)). So, the formulas \(b_k(n) = t(n)p_k(n)\) and \(b_{\geq k}(n) = t(n)p_{\geq k}(n)\) provide the total number of trees on \([n]\) with root of rank \(k\) and of rank \(\geq k\), respectively. For \(k \geq 1\), the root is of rank at least \(k\) if and only if the root has children and each of them is of rank at least \(k - 1\). Therefore, for \(n \geq 2\),

\[
p_{\geq k}(n) = \frac{(n-1)!}{t(n)} \sum_{s \geq 1} \sum_{j_1 + \cdots + j_s = n-1} \prod_{r=1}^{s} \frac{t(j_r)p_{\geq k-1}(j_r)}{j_r!}, \quad j_1, \ldots, j_s \geq 1.
\]
Indeed, by Proposition 2.2 with \( j \) subtrees of sizes \( s \) children of the root. The probability that the root has \( k \) plus the total sum of all vertices of rank \( k \). Multiplying both sides by \( t \), \( B \) The series \( B'_{\geq k}(z) \) above converges for \( |z| < 1/2 \) and \( k \geq 1 \) because

\[
|B_{\geq k-1}(z)| \leq B_{\geq k-1}(1) \leq B_{\geq 0}(1) = T(1/2) = 1.
\]

Let us turn to \( A_k(z) \). Now, in the random tree \( X_k(n) \) equals \( I(\text{root is of rank } k) \) plus the total sum of all vertices of rank \( k \) in the subtrees rooted at the children of the root. The probability that the root has \( s \) children with ordered subtrees of sizes \( j_1, \ldots, j_s \) is \( n! \prod_{r=1}^{s} \frac{t(j_r)}{j_r!} \). So, by linearity of expectation, for \( n \geq 2 \), the following holds.

\[
\mathbb{E}[X_k(n)] = p_k(n)
\]

\[
+ \frac{(n-1)!}{t(n)!} \sum_{s \geq 1} \sum_{j_1 + \ldots + j_s = n-1} \prod_{r=1}^{s} \frac{t(j_r)}{j_r!} \left( \sum_{j_r \geq 1} \mathbb{E}[X_k(j_r)] \right).
\]

Multiplying both sides by \( \frac{t(n)z^{n-1}}{(n-1)!} \), and summing over \( n \geq 2 \), we obtain

\[
\frac{d}{dz} \sum_{n \geq 1} \frac{t(n)\mathbb{E}[X_k(n)]}{n!} z^n = \frac{d}{dz} \sum_{n \geq 1} \frac{t(n)p_k(n)}{n!} z^n
\]

\[
+ \sum_{s \geq 1} \sum_{j_1 \geq 1} \sum_{j_r \geq 1} \frac{t(j_r)\mathbb{E}[X_k(j_r)]}{j_r!} z^{j_r} \sum_{j_r \geq 1, j_r' \neq j_r} \prod_{r' \neq r} \frac{t(j_{r'})}{j_{r'}}! z^{j_{r'}},
\]

or equivalently

\[
A'_k(z) = B'_k(z) + A_k(z) \sum_{s \geq 1} sT^{s-1}(z) = B'_k(z) + \frac{A_k(z)}{(1-T(z))^2}.
\]

\[\square\]

**Example 2.3.** Let \( n \geq 2 \). Then the number of all leaves in all trees of size \( n \) is \((2n-1)!!/3\). Indeed, by Proposition 2.2 with \( k = 0 \) and \( B_0(z) = z \), we have

\[
A_0(z) = \frac{2z - 1}{3} + \frac{1}{3\sqrt{1-2z}} \implies A_0(n) = \frac{1}{3} \frac{(2n-1)!!}{n!}.
\]
2.1.1. Counting ordered tuples of vertices by ranks. Given an ordered \( r \)-tuple \( \mathbf{k} = (k_1, \ldots, k_r) \) of (non-negative) integers, let \( a_k(n) \) denote the total number of occurrences of ordered \( r \)-tuples of distinct vertices \( \mathbf{v} = (v_1, \ldots, v_r) \) with ranks \( \mathbf{k} \) in all trees on \([n]\). (Proposition 2.2 deals with a special case \( r = 1 \).) Given \( k_1 \) and an \((r - 1)\)-long ordered tuple \( \mathbf{k}' = (k_2, \ldots, k_r) \), let \( b_k(n) \), \( (\mathbf{k} := (k_1, \mathbf{k}')) \), denote the total number of occurrences of \( r \)-long tuples \((\text{root}, \mathbf{v}')\) such that the root has rank \( k_1 \) and \( r - 1 \) distinct vertices \( \mathbf{v}' \) have ranks \( \mathbf{k}' \). Let \( A_k(z) \) and \( B_k(z) \) denote the respective generating functions.

An ordered partition of a set or \( r \)-tuple is a partition of that set or \( r \)-tuple into nonempty blocks in which the set of blocks is linearly ordered. There is no new ordering on each individual block; the individual blocks inherit the original ordering.

**Proposition 2.4.** Let us assume that \( r > 1 \) and \( \mathbf{k} \) is an ordered \( r \)-tuple. Let \((\mathbf{k}^{(1)}, \ldots, \mathbf{k}^{(r)})\) denote an ordered partition of \( \mathbf{k} \) into \( \tau \leq r \) non-empty sub-tuples, each sub-tuple inheriting its ordering from \( \mathbf{k} \). Then

\[
A_k^{(\tau)}(z) = \sum_{\tau \leq r} (1 - T(z))^{-\tau - 1} \left( \sum_{\mathbf{k}^{(1)}, \ldots, \mathbf{k}^{(r)} \neq \emptyset} \prod_{u=1}^{\tau} A_{k^{(u)}}(z) \right)
+ \sum_{u \in [r]} B_{k^{(u)}}(z) \left( \prod_{v \in [r] \setminus \{u\}} A_{k^{(v)}}(z) \right).
\]

Further, with \( \mathbf{k} = (k_1, \mathbf{k}') \), let \((\mathbf{k}^{(1)}, \ldots, \mathbf{k}^{(\tau)})\) denote a generic ordered partition of \( \mathbf{k}' \) into \( \tau \) non-empty ordered sub-tuples. If \( \mathbf{k}' \neq \emptyset \), then

\[
B_{k^{(u)}}(z) \leq \sum_{\tau < r} (1 - T(z))^{-\tau - 1} \left( \sum_{\mathbf{k}^{(1)}, \ldots, \mathbf{k}^{(r)} \neq \emptyset} \prod_{u=1}^{\tau} A_{k^{(u)}}(z) \right).
\]

This inequality means that, for each \( \nu \geq 0 \), \([z^n]B_{k^{(u)}}(z)\) is at most the coefficient of \( z^n \) in the generating function on the RHS.

**Example 2.5.** Let \( \mathbf{k} = (k_1, k_2, k_3) = (3, 7, 7) \). Then \( a_k(n) \) denotes the total number of occurrences of ordered triples \( \mathbf{v} = \{v_1, v_2, v_3\} \) of vertices in all trees on \([n]\) so that \( v_1 \) is of rank 3, and \( v_2 \) and \( v_3 \) are both of rank 7. Furthermore, \( b_k(n) \) denotes the number of occurrences of all ordered triples \((\text{root}, v_2, v_3)\) in all such trees such that the root has rank 3, and \( v_2 \) and \( v_3 \) are both of rank 7.

Then \( \mathbf{k} \) has 13 ordered partitions into non-empty sub-tuples, each inheriting its ordering from \( \mathbf{k} \): \((\{k_1, k_2, k_3\}, \{k_1, k_2\}, \{k_3\}), (\{k_1, k_2\}, \{k_3\}, \{k_1, k_2\}), (\{k_1, k_3\}, \{k_2\}, \{k_1, k_3\}), (\{k_1\}, \{k_2, k_3\}, \{k_1\}), (\{k_1\}, \{k_2\}, \{k_3\}, \{k_2\}), (\{k_1\}, \{k_3\}, \{k_1\}), (\{k_2\}, \{k_3\}, \{k_1\}), (\{k_2\}, \{k_3\}, \{k_1\}), (\{k_3\}, \{k_1\}, \{k_2\})\), \((\{k_3\}, \{k_2\})\), \((\{k_3\}, \{k_1\}, \{k_2\})\).
tuples of ranks $k_1, \ldots, k_r$. Therefore, the first equality in Proposition 2.4 reduces to

\[ A'_{(3,7,7)}(z) = \frac{A_{(3,7,7)}(z)}{(1-T(z))^2} + \frac{2A_{(3,7,7)}(z)+2A_{(3,7,7)}(z)+A_{(3,7,7)}(z)}{(1-T(z))^3} + B'_{(3,7,7)}(z) + 2B'_{(7,7,3)}(z). \]

Further, $B_{\geq(3,7,7)}(z)$ is the generating function for the numbers of ordered triples $(v_1, v_2, v_3)$ so that the root has rank at least three, and $v_2, v_3$ are both of rank 7. So, the inequality in Proposition 2.4 reduces to

\[ B'_{(3,7,7)}(z) \leq \frac{A_{(7,7)}(z)}{(1-T(z))^2} + \frac{2A_{(3,7,7)}(z)}{(1-T(z))^3}. \]

Likewise

\[ B'_{(7,7,3)}(z) \leq \frac{A_{(3,7,7)}(z)}{(1-T(z))^2} + \frac{2A_{(3,7,7)}(z)}{(1-T(z))^3}. \]

Proof. Let $X_k(n)$ stand for the total number of $r$-long tuples of distinct vertices of ranks $k = (k_1, \ldots, k_r)$ in the random tree, so that $E[X_k(n)] = \frac{a_k(n)}{t(n)}$. Let $Y_k(n)$ (resp., $Y_{\geq k_1,k'}$) stand for the number of $r$-long tuples of distinct vertices $(1, v_2, \ldots, v_r) =: (1, v')$ such that the root 1 is of rank $k_1$ (of rank $\geq k_1$ resp.) and vertices forming $v'$ are of ranks $k' = (k_2, \ldots, k_r)$. So, $E[Y_k(n)] = \frac{b_k(n)}{t(n)}$ and $E[Y_{\geq k_1,k'}(n)] = \frac{b_{\geq k_1,k'}(n)}{t(n)}$.

Let us write a recurrence for $E[X_k(n)]$. The $r$-long tuples $v$ that contain the root 1 contribute $\sum_{u \in \{r\}} E[Y_{k_u,(k_t)_{t \in \{r\}\backslash\{u\}}}]$ to this expectation. (The $u$-th term corresponds to the case when $k_u$, i.e. the $u$-th component of $k$, is the root’s rank.) Consider the tuples $v$ that do not contain the root 1. Let $s \geq 1$ be a generic number of the root’s ordered children, and let $[r]$ be a disjoint, union of $s$ subsets, some of which may be empty. This ordered partition of $[r]$ determines the partition of the ordered tuple $k$ into $s$ sub-tuples $k^{(1)}, \ldots, k^{(s)}$ of total length $r$, where each sub-tuple inheriting the order of $k$. Conditioned on $s$ and the sizes $j_1, \ldots, j_s$ of the subtrees rooted at $s$ children of the root 1, the subtrees are independent. Therefore, the conditional expected number of $r$-long tuples of rank $k$, composed of $s$ sub-tuples of ranks $k^{(1)}, \ldots, k^{(s)}$, is $\prod_{u=1}^{s} \frac{E[X_{k^{(u)}}(j_u)]}{E[X_{k^{(u)}}(j_u)]}$. Here $E[X_{k^{(u)}}(j_u)] := 1$ if $k^{(u)} = \emptyset$. Therefore, for $n \geq 2$, we have

\[ E[X_k(n)] = \sum_{u \in [r]} E[Y_{k_u,(k_t)_{t \in [r]\backslash\{u\}}}(n)] \]

\[ + \frac{(n-1)!}{(n-1)!} \sum_{s \geq 1} \sum_{k^{(1)}, \ldots, k^{(s)}} \sum_{j_1 + \ldots + j_s = n-1} \prod_{r=1}^{s} \frac{t(j_r)}{t(r)} \prod_{u=1}^{s} E[X_{k^{(u)}}(j_u)]. \]

Using $t(n)E[X_k(n)] = a_k(n)$ with $a_0(n) := t(n)$ and $t(n)E[Y_k(n)] = b_k(n)$, then multiplying the resulting equation by $\frac{z^{n-1}}{(n-1)!}$, and summing over $n \geq 2$,
we obtain

\[ A'_k(z) = \sum_{u \in [r]} B'_{k_0,(k_1)_{i \in [r] \setminus \{u\}}}(z) + \sum_{s \geq 1} \sum_{k^{(1)}, \ldots, k^{(s)}} \prod_{u=1}^{s} A_k^{(u)}(z), \]

\[ = \sum_{u \in [r]} B'_{k_0,(k_1)_{i \in [r] \setminus \{u\}}}(z) + \sum_{\tau \geq 1} (1 - T(z))^{-\tau - 1} \sum_{k^{(1)}, \ldots, k^{(\tau)} \neq \emptyset} \prod_{u=1}^{\tau} A_k^{(u)}(z). \]

Explanation: The roots of \( \tau \) subtrees that contain non-empty sub-tuples of vertices of ranks \( k^{(1)}, \ldots, k^{(\tau)} \) form \( \tau + 1 \) consecutive intervals. Therefore, \((1 - T(z))^{-\tau - 1}\) is the EGF of finite forests of ordered trees whose roots are hosted by these intervals.

Let \( k = (k_1, \ldots, k_r) =: (k_1, k') \). To handle \( B_k(z) \), introduce \( B_{\geq k_1, k'}(z) \), the generating function of the numbers of occurrences of \( v = (\text{root}, v_2, \ldots, v_r) =: (\text{root}, v') \) such that the root is of rank \( \geq k_1 \), and \( v' \) is of rank \( k' \). Clearly \( B_k(z) \leq B_{\geq k_1, k'}(z) \). Analogously to the case of \( A_k(z) \), we obtain

\[ B'_{\geq k_1, k'}(z) \leq \sum_{\tau \geq 1} \frac{1}{(1 - T(z))^{\tau + 1}} \left( \sum_{\cup_{u=1}^{\tau} k^{(u)} = k', \ k^{(u)} \neq \emptyset} \prod_{u=1}^{\tau} A_k^{(u)}(z) \right). \]

Here \( \tau \) is the number of children of the root 1 whose subtrees split among themselves the ordered \((r - 1)\)-tuple \( k' \) according to a generic ordered partition of \( k' \) into \( \tau \) non-empty sub-tuples \( k^{(1)}, \ldots, k^{(\tau)} \), each ordered in compliance with the ordering of \( k' \). Note that (2.6) is an inequality, because we neglected the condition that the roots of all the subtrees, rooted at the children of the root 1, should have ranks \( \geq k_1 - 1 \) as well. \( \square \)

3. Asymptotics for the uniformly random tree \( T_n \)

We start this section by a series of lemmas that will allow us to prove an upper bound for the expected number of vertices of rank at least \( k \) in the uniformly random tree in Theorem 3.3. We then turn to the expected number of vertices of rank exactly \( k \), and prove an asymptotic formula for that number in Theorem 3.4. That theorem will show that this number is asymptotically equal to \( c_k n \), for a positive constant \( c_k \). We compute the values of \( c_k \) for \( k \leq 3 \), and then, in Theorem 3.5, we prove an upper bound for \( 1 - \sum_{j \leq k} c_j \), which proves that the numbers \( c_j \) form a probability distribution.

3.1. Counting individual vertices by rank in \( T_n \). Recall that \( A_k(z) \) is the (exponential) generating function of all occurrences of vertices of rank \( k \) in all the trees on \([n], n \geq 1\), and \( B_k(z) \) is the generating function of number of trees with root of rank \( k \). We start with a simple, but instrumental lemma.
Lemma 3.1. For \( z \in \mathbb{C} \) and \( |z| < 1/2 \), we have
\[
A_k(z) = \frac{I_k(z)}{\sqrt{1-2z}}, \quad I_k(z) := \int_0^z \sqrt{1-2\xi} B'_k(\xi) \, d\xi,
\]
where the integral is taken over any path from 0 to \( z \) within the disc \( |\xi| < 1/2 \).

Proof. According to Proposition 2.2, we have
\[
A_k(z) = \frac{A_k(z)}{(1-T(z))^2} + B'_k(z), \quad |z| < 1/2.
\]
Since \( T'(z) = (1-T(z))^{-1} \), we have \( (1-T(z))^{-2} = T''(z)/T'(z) \), whence
\[
A_k(z) = \frac{T''(z)}{T'(z)} A_k(z) + B'_k(z) \implies \left( \frac{A_k(z)}{T'(z)} \right)' = \frac{B'_k(z)}{T'(z)} = (1-T(z))B'_k(z),
\]
where \( 1-T(z) = \sqrt{1-2z} \). Our claim is now proved by integration. \( \Box \)

Let us introduce \( A_{\geq k}(z) = \sum_{j \geq k} A_j(z) \), the exponential generating function for the number of all vertices of rank \( \geq k \) in all rooted trees on vertex set \([n]\). By Lemma 3.1., we obtain
\[
(3.1) \quad A_{\geq k}(z) = \frac{I_{\geq k}(z)}{\sqrt{1-2z}}, \quad I_{\geq k}(z) := \int_0^z \sqrt{1-2\xi} B'_{\geq k}(\xi) \, d\xi.
\]

Lemma 3.2. For \( z \in \mathbb{C} \) and \( |z| < 1/2 \) and \( k > 2 \), we have \( |B'_{\geq k}(z)| \leq 1/(k-2)! \). Consequently, \( |I_{\geq k}(z)| \leq 1/(k-2)! \).

Proof. Let \( z \in \mathbb{R} \) and \( z \in [0,1/2) \). By Proposition 2.2, we have: for \( \ell \geq 2 \),
\[
B'_{\geq \ell}(z) = \frac{B_{\geq \ell-1}(z)}{1-B_{\geq \ell-1}(z)} \leq \frac{B_{\geq \ell-1}(z)}{1-B_{\geq 1}(z)} = \frac{B_{\geq \ell-1}(z)}{1+z-T(z)} \leq 2B_{\geq \ell-1}(z).
\]
In the last step, we used the fact that if \( z \in [0,1/2) \), then \( z - T(z) \geq -1/2 \).

Integrating both sides of the resulting inequality \( B'_{\geq \ell}(z) \leq 2B_{\geq \ell-1}(z) \), by induction we obtain
\[
B_{\geq \ell}(z) \leq \frac{2^{\ell-2}z^{\ell-1}}{(\ell-1)!}.
\]
Hence, for \( k > 2 \), we have
\[
B'_{\geq k}(z) \leq 2B_{\geq k-1}(z) \leq \frac{(2z)^{k-2}}{(k-2)!} \leq \frac{1}{(k-2)!}.
\]
So, for \( z \in \mathbb{C} \) and \( |z| < 1/2 \), we have
\[
|B'_{\geq k}(z)| \leq B'_{\geq k}(|z|) \leq \frac{1}{(k-2)!}, \quad |\sqrt{1-2z}| \leq \sqrt{2}.
\]
It follows that \( |I_{\geq k}(z)| \leq 2^{-1/2}/(k-2)! \leq 1/(k-2)! \). \( \Box \)
Notation. To proceed, recall that $X_k(n)$ and $X_{\geq k}(n)$ stand for the total number of vertices of rank $k$ and of rank $\geq k$, respectively. Let $R_n$ stand for the largest rank of a vertex in $T_n$.

**Theorem 3.3.** Uniformly for $k \leq n$, we have $\mathbb{E}[X_{\geq k}(n)] = O(n^{3/2}/(k-2)!)$.

Consequently, $\mathbb{P}(R_n \leq (1.5 + \varepsilon) \log n / \log \log n) = 1 - O(n^{-\varepsilon + o(1)})$.

Proof. By the Cauchy integral theorem, for a circular contour $C$ of radius $\rho < 1/2$ centered at zero, we have

$$\frac{a_{\geq k}(n)}{n!} = \frac{1}{2\pi i} \oint_{C} \frac{I_{\geq k}(z)}{z^{n+1} \sqrt{1 - 2z}} \, dz.$$ 

Let $C^*$ be the circle $|z| = 1/2$. Then letting $C$ tend to $C^*$, substituting $z = \frac{1}{2} e^{i\theta}$, and using Lemma 3.2 we have

$$\frac{a_{\geq k}(n)}{n!} \leq \frac{2^{n+2}}{2\pi(k-2)!} \int_0^{\pi/2} |1 - e^{i\theta}|^{-1/2} d\theta = \frac{2^{n+3/2}}{2\pi(k-2)!} \int_0^{\pi/2} \sin^{-1/2}(\theta/2) d\theta \leq \frac{2^{n+2}}{2\pi(k-2)!} \int_0^{\pi/2} \theta^{-1/2} d\theta \leq \frac{2^{n+1}}{(k-2)!}.$$ 

So using $2^{2n/(2n-1)} \leq 2n^{1/2}$, we get

$$\mathbb{E}[X_{\geq k}(n)] = \frac{a_{\geq k}(n)}{(2n-3)!!} \leq \frac{2^{n+1} n!}{(2n-3)!!(k-2)!} \leq \frac{4n}{(k-2)!} \cdot \frac{2^{2n}}{2n} \leq \frac{8n^{3/2}}{(k-2)!}.$$ 

Since $X_k(n) \leq n$, this bound is useless if $k$ is fixed, or grows very slowly as $n$ grows. However, setting $k_n = \lceil (1.5 + \varepsilon) \log n / \log \log n \rceil$, $\varepsilon > 0$ being arbitrary, and using $m! \geq (m/e)^m$, we conclude that

$$\mathbb{P}(R_n \geq k_n) \leq \mathbb{E}[X_{\geq k_n}(n)] = O(n^{1.5}(k_n/e)^{-k_n}) = O(n^{-\varepsilon'}), \quad \forall \varepsilon' \in (0, \varepsilon).$$

Let us turn now to $\mathbb{E}[X_k(n)]$.

**Theorem 3.4.** For every fixed $k$, we have

$$\mathbb{E}[X_k(n)] = c_k n (1 + O(n^{-1})), \quad c_k := 2 \int_0^{1/2} \sqrt{1 - 2\xi} B_k' (\xi) \, d\xi.$$ 

Proof. By Lemma 3.1 for $|z| < 1/2$,

$$\mathbb{E}[X_{n,k}] = \frac{a_k(n)}{(2n-3)!!} = \frac{n!}{(2n-3)!!} [z^n] A_k(z),$$

$$A_k(z) = \frac{I_k(z)}{\sqrt{1 - 2z}}, \quad I_k(z) = \int_0^z \sqrt{1 - 2\xi} B_k' (\xi) \, d\xi.$$ 

Since $B_0(z) = z$, the integral $I_0(z)$ is $(1 - (1 - 2z)^{3/2})/3$, so we view $I_0(z)$ as being analytically extended to $\mathbb{C} \setminus \{z \in \mathbb{R} : z \in [1/2, \infty)\}$. 


Further, recall that $B_{k}(z) = B_{\geq k}(z) - B_{\geq k+1}(z)$, and

\begin{equation}
B_{\geq j+1}(z) = \int_{0}^{z} \frac{B_{\geq j}(\xi)}{1 - B_{\geq j}(\xi)} d\xi, \quad |z| < 1/2,
\end{equation}

and the integral is taken along any path from 0 to $z$ in the circle $|z| < 1/2$. (Recall that $|B_{j}(z)| < 1$ for $|z| < 1/2$.) Simply by the definition of $B_{\geq 1}(z)$, we have

$B_{\geq 1}(z) = B_{\geq 0}(z) - z = \sum_{\nu \geq 2} \frac{t(\nu)}{\nu!} z^{\nu} = T(z) - z = 1 - \sqrt{1 - 2z} - z,$

with $\sqrt{\xi} = |\xi|^{1/2} e^{i \text{Arg}(\xi)/2}$, $\text{Arg}(\xi) \in (-\pi, \pi)$. The last expression defines an analytic extension of $B_{\geq 1}(z)$ to $\mathbb{C} \setminus [1/2, \infty)$. For $|z| < 1/2$, the series formula implies that $|B_{\geq 1}(z)| \leq B_{\geq 1}(1/2) \leq T(1/2) - 1/2 = 1/2$. Since $B_{\geq j+1}(z) \leq B_{\geq j}(z)$ for those $z$’s, we obtain that $|B_{\geq j}(z)| \leq 1/2$ for $j > 1$ and $|z| < 1/2$.

So, by induction on $j$, it follows from (3.4) that for $|z| < 1/2$ and $j \geq 1$ the following holds.

(a) $|B_{j}(z)| \leq 1/2$, whence the integral in (3.4) is well-defined;
(b) $|B'_{\geq j+1}(z)| \leq 1$.

We use (3.4) to assert, by induction on $j$, the existence of the analytic extension of each $B_{\geq j}(z)$ to the domain $D_{j} := \{z \in \mathbb{C} : |z| < \rho_{j}\} \setminus [1/2, \rho_{j})$, where $\rho_{1} \geq \rho_{2} \geq \cdots > 1/2$, such that—keeping $B_{j}(z)$ as notation for this extension—keeping $B_{\geq 1}(z)$ as notation for this extension—$|B_{\geq j}(z)| \leq 3/4$ for $z \in D_{j}$, and (3.4) holds for $z \in D_{j+1}$. We use $D_{j}$ for fixed $j$ only. So, setting $\rho > 1/2$ equal to the minimum of $\rho_{j}$ over the finitely many $j$’s in question, we may and will focus on the analytic continuations $B_{\geq j}(z)$ to the same $D := \{z \in \mathbb{C} : |z| < \rho\} \setminus [1/2, \rho)$.

Consequently, each $B_{k}(z)$ analytically extends to $D$. Keeping $B_{k}(z)$ and $I_{k}(z)$ to denote the extensions of $B_{k}(z)$ and the corresponding $I_{k}(z)$ given by (3.3), we obtain the following. Uniformly for $z \in D$, we have

\begin{equation}
I_{k}(z) = \int_{0}^{z} \sqrt{1 - 2\xi} B'_{k}(\xi) d\xi = \gamma_{k} + \sum_{j \geq 0} \lambda_{k,j} (1 - 2z)^{j+2},
\end{equation}

\begin{align*}
\gamma_{k} &:= \int_{0}^{1/2} \sqrt{1 - 2\xi} B'_{k}(\xi) d\xi;
\end{align*}

here the integral is taken over any path which is within $D$. Recall that $B_{1}(z) = z$, so in particular, $B'_{1}(1/2)$ does exist. A bounded factor implicit in the big-O notation here, and in what follows, may depend on $k$.

Let $C_{m}$ be a contour that consists of the counter-clockwise circular arc $z = pe^{i\theta}$, $\theta \in (0, 2\pi)$, and a detour part formed by two opposite-directed line segments, one from $z = pe^{i(2-\ell/m)\pi}$ to $z = 1/2 e^{i(2-\ell/m)\pi}$ and another from $z = 1/2 e^{i(-\ell/m)}$ back to $z = pe^{i(\ell/m)}$. 

Now consider the contour $C$ that consists of the counter-clockwise circular arc $z = \rho e^{i\theta}$, $\theta \in (0, 2\pi)$, and a detour part formed by two opposite-directed line segments, one from $z = \rho e^{i(2\pi-0)}$ to $z = 1/2e^{i(2\pi-0)}$ and another from $z = 1/2e^{i(+0)}$ back to $z = \rho e^{i(+0)}$. Then $C$ is obviously a limit of the contours $C_m$, which are covered by the Cauchy theorem. See Figure ?? for an illustration. (Readers can consult the Transfer theorem in [11] for more on this technique.)

Applying this theorem and going to the limit, we have, for $\alpha < 1/2$,

$$\left[z^n\right] \frac{I_k(z)}{\sqrt{1-2z}} = \frac{1}{2\pi i} \oint_{z \in C : |z| = \alpha} \frac{I_k(z)}{z^{n+1} \sqrt{1-2z}} dz$$

$$= \gamma_k[z^n](1-2z)^{-1/2} + \lambda_{k,0}[z^n](1-2z)^{1/2} + \frac{1}{2\pi i} \oint_{z \in C} \frac{O(|1-2z|)}{z^{n+1}} dz.$$

Here

$$\gamma_k[z^n](1-2z)^{-1/2} + \lambda_{k,0}[z^n](1-2z)^{1/2} = \gamma_k \frac{(2n-1)!!}{n!} + \lambda_{k,0} \frac{(2n-3)!!}{n!}$$

$$= \gamma_k \frac{(2n-1)!!}{n!} (1 + O(n^{-1})) = \Theta(n^{-1/2}2^n).$$
and the remainder integral is of order
\[
\int_{1/2}^{\rho} \frac{2\eta - 1}{\eta^n} d\eta + \frac{1}{\rho^n} \leq n^{-2} \rho^n + O(\rho^{-n}).
\]
We conclude that, for every fixed \(k\),
\[
[z^n] I_k(z) \frac{I_k(z)}{\sqrt{1 - 2z}} = \gamma_k \frac{(2n - 1)!!}{n!} \left(1 + O(n^{-1})\right).
\]
\[\mathbb{E}[X_k(n)] = \frac{n!}{(2n - 3)!!} [z^n] I_k(z) \frac{I_k(z)}{\sqrt{1 - 2z}} = (2n - 1)\gamma_k \left(1 + O(n^{-1})\right),\]
implying that \(\mathbb{E}[X_k(n)] = n\gamma_k \left(1 + O(n^{-1})\right), \ c_k := 2\gamma_k. \]

3.1.1. Computing \(c_k\) for \(k \leq 3\). Recall that \(c_k = \lim_{n \to \infty} n^{-1} \mathbb{E}[X_k(n)]\), where
\[
c_k = 2\gamma_k = 2 \int_0^{1/2} \sqrt{1 - 2\xi} B_k'\left(\xi\right) d\xi.
\]
Here \(B_0(z) = z\), and for \(k \geq 1\) we have the recurrence
\[
B_k(z) = B_{>k}(z) - B_{>k+1}(z), \quad B_{>k}' = \frac{B_{>k-1}(z)}{1 - B_{>k-1}(z)}, \quad |z| < 1/2.
\]
So,
\[
c_0 = 2 \int_0^{1/2} \sqrt{1 - 2\xi} d\xi = \frac{2}{3}.
\]
Next, we compute
\[
B_1' = B_{>1}' - B_{>2}' = \frac{B_{>0}(z)}{1 - B_{>0}(z)} - \frac{B_{>1}(z)}{1 - B_{>1}(z)},
\]
here \(B_{>0}(z) = T(z) = 1 - \sqrt{1 - 2z}\), and
\[
B_{>1}(z) = \int_0^z \frac{T(\xi)}{1 - T(\xi)} d\xi = \int_0^z \frac{1 - \sqrt{1 - 2\xi}}{\sqrt{1 - 2\xi}} d\xi
= 1 - z - \sqrt{1 - 2z}.
\]
Therefore
\[
B_1' = \frac{1 - \sqrt{1 - 2z}}{\sqrt{1 - 2z}} - \frac{1 - z - \sqrt{1 - 2z}}{z + \sqrt{1 - 2z}} = \frac{z}{\sqrt{1 - 2z} (z + \sqrt{1 - 2z})}.
\]
Consequently,
\[
c_1 = 2 \int_0^{1/2} \sqrt{1 - 2z} B_1'(z) dz = 2 \int_0^{1/2} \frac{z}{z + \sqrt{1 - 2z}} dz
= 5 + 4 \log 2 - 6 \sqrt{2} \arctanh(2^{-1/2}) \approx 0.2938858406.
\]
Next,

\[
B_2'(z) = B_{\geq 2}'(z) - B_{\geq 3}'(z) = \frac{1}{1-B_{\geq 1}(z)} - \frac{1}{1-B_{\geq 2}(z)}
\]

\[
= \frac{1}{z + \sqrt{1-2z}} - \frac{1}{1-B_{\geq 2}(z)},
\]

where

\[
B_{\geq 2}(z) = \int_0^z \frac{B_{\geq 1}(y)}{1-B_{\geq 1}(y)} dy
\]

\[
= -z + \int_0^z \frac{1}{y + \sqrt{1-2y}} dy
\]

\[
= -z + \log(z + \sqrt{1-2z}) - 2^{1/2} \arctanh(2^{-1/2}(\sqrt{1-2z} - 1)).
\]

So, using the two last formulas, we obtain

\[
c_2 = 2 \int_0^{1/2} \sqrt{1-2z} B_2'(z) dz
\]

\[
= 2 \int_0^{1/2} \sqrt{1-2z} \left( \frac{1}{z + \sqrt{1-2z}} - \frac{1}{1-B_{\geq 2}(z)} \right) dz
\]

\[
\approx 0.03589474655.
\]

Let us now turn to

\[
c_3 = 2 \int_0^{1/2} \sqrt{1-2z} B_3'(z) dz = 2 \int_0^{1/2} \sqrt{1-2z} (B_{\geq 3}'(z) - B_{\geq 4}'(z)) dz
\]

\[
= 2 \int_0^{1/2} \sqrt{1-2z} \left( (1-B_{\geq 2}(z))^{-1} - (1-B_{\geq 3}(z))^{-1} \right) dz.
\]

To compute \(c_3\), introduce

\[
C_3(z) = 2 \int_0^z \sqrt{1-2y} \left( (1-B_{\geq 2}(y))^{-1} - (1-B_{\geq 3}(y))^{-1} \right) dy.
\]

Then \(c_3 = c_3(1/2)\) and \(C_3(z), B_{\geq 3}(z)\) satisfy a system of two ODEs,

\[
\frac{dC_3}{dz} = 2\sqrt{1-2z} \left( (1-B_{\geq 2}(z))^{-1} - (1-B_{\geq 3}(z))^{-1} \right),
\]

\[
\frac{dB_{\geq 3}}{dz} = (1-B_{\geq 2}(z))^{-1} - 1,
\]

with \(C_3(0) = B_{\geq 3}(0) = 0\), and \(B_{\geq 2}(z)\) being defined in the computation of \(c_2\) above.

Sara Pollock, a colleague of the first author, devised an efficient integration scheme of this ODE system, based on a trapezoidal approximation of the integrals on each of intervals forming a partition of \([0, 1/2]\), and used Matlab to show that \(c_3 := C_3(1/2) = 0.0032684102\) if each interval has length \(h = 10^{-6}\). Selecting \(h = 10^{-7}\) did not change the approximate value of \(c_3\). We owe a debt of gratitude to Sara for her generous help.
The numbers \( \{c_k\} \) form a proper distribution. As we stated in the introduction, we will prove that \( \sum_{k \geq 0} c_k = 1 \), so the numbers \( c_k \) form a proper distribution. We saw in Section 1.2 that \( \sum_{k \geq 0} c_k \leq 1 \), so it suffices to show that
\[
\lim_{k \to \infty} \left( 1 - \sum_{j \leq k} c_j \right) = 0.
\]

We accomplish this by proving the following stronger statement.

**Theorem 3.5.** For all \( k \), the inequality \( 1 - \sum_{j \leq k} c_j \leq \frac{3^{k+1}}{(2k+1)!} \) holds.

**Note.** So the limiting expected fraction of vertices having rank at least 4 is \( 0.0002843360 \ldots \) exactly.

**Proof.** (i) Given fixed \( k \), let \( a_{\leq k}(n) \) and \( a_{> k}(n) \) stand for the total number of vertices of rank \( \leq k \) and \( > k \) in all trees on \( [n] \), and let \( X_{\leq k}(n) \), \( X_{> k}(n) \) stand for the number of vertices of rank \( \leq k \) and \( > k \) in the random tree. So \( t(n)E[X_{\leq k}(n)] = a_{\leq k}(n) \) and \( t(n)E[X_{> k}(n)] = a_{> k}(n) \). By (3.1),
\[
A_{\leq k}(z) = (1 - 2z)^{-1/2} \int_0^z \sqrt{1 - 2x} B'_{\leq k}(x) \, dx.
\]

The RHS of (3.10) is analytic within an open circle of a radius \( \rho > 1/2 \) minus a cut \([1/2, \rho)\). It follows that
\[
\frac{(2n-3)!!}{n!} E[X_{\leq k}(n)] = (1 + o(1)) \left[ z^n \right] (1 - 2z)^{-1/2} \int_0^{1/2} \sqrt{1 - 2x} B'_{\leq k}(x) \, dx
\]
\[
= (1 + o(1)) \frac{(2n-1)!!}{n!} \int_0^{1/2} \sqrt{1 - 2x} B'_{\leq k}(x) \, dx
\]
\[
= (1 + o(1)) \frac{(2n-1)!!}{n!} \int_0^{1/2} T'(x) B_{\leq k}(x) \, dx.
\]

Using
\[
E[X_{\leq k}(n)] = n - E[X_{> k}(n)], \quad B_{\leq k}(x) + B_{> k}(x) = T(x),
\]
\[
c_j = \lim_{n \to \infty} n^{-1} E[X_j(n)], \quad 2 \int_0^{1/2} T'(x) T(x) \, dx = 1,
\]
we obtain
\[
1 - \sum_{j \leq k} c_j = 2 \int_0^{1/2} \frac{B_{> k}(x)}{1 - T(x)} \, dx.
\]

Recall that, for \(|z| < 1/2\), the power series \( B_{> k}(z) \) is the EGF generating function of \( b_{> k}(n) \), where \( b_{> k}(n) \) is the total number of trees in which the root is of rank \( > k \). Clearly we need to find a good bound for \( b_{> k}(n) \), or equivalently for \( p_{> k}(n) = b_{> k}(n)/t(n) \), the probability that the root of the random tree has rank more than \( k \). To get such a bound we need a provably
well-performing randomized algorithm which is likely to find a competitively short path from the root to a leaf.

Here is the algorithm. If there are \( s \geq 2 \) children of the root, we keep a single tree, which we select with probability proportional to the total size of the deleted subtrees. For \( n \geq 1, k \geq -1 \), let \( \pi_{\geq k}(n) \) denote the probability that edge length of the resulting path strictly exceeds \( k \); obviously \( p_{\geq k}(n) \leq \pi_{\geq k}(n) \). Further, \( \pi_{\geq 1}(n) = 1 \), and for \( n > 1, k \geq 0 \),

\[
\pi_{k}(n) = \frac{(n-1)!}{t(n)} \pi_{\geq k-1}(n-1)
\]

\[
+ \frac{(n-1)!}{t(n)} \sum_{s \geq 2} \sum_{j_1 + \cdots + j_s = n-1} \sum_{r} \pi_{\geq k-1}(j_r) \prod_{r=1}^{s} \frac{t(j_r)}{j_r!},
\]

or equivalently

\[
\frac{t(n)\pi_{\geq k}(n)}{(n-2)!} = \frac{(n-1)\pi_{\geq k-1}(n)}{(n-2)!}\]

(3.12)

Introducing \( \mathbb{P}_{\geq k}(z) = \sum_{n \geq 1} \frac{t(n)\pi_{\geq k}(n)}{n!} z^n \), (so that \( \mathbb{P}_{\geq 1}(z) = T(z) \), and \( \mathbb{P}_{\geq 0}(z) = T(z) - z \)), we obtain a linear differential recurrence equation

\[
\frac{d^2\mathbb{P}_{\geq k}(z)}{dz^2} = \frac{d\mathbb{P}_{\geq k-1}(z)}{dz} + \mathbb{P}_{\geq k-1}(z) \cdot T'(z) \sum_{t \geq 0} T^t(z),
\]

or

\[
\frac{d^2\mathbb{P}_{\geq k}(z)}{dz^2} = \frac{d\mathbb{P}_{\geq k-1}(z)}{dz} + (1 - T(z))^{-2} \mathbb{P}_{\geq k-1}(z), \quad |z| < 1/2.
\]

Furthermore, by (3.11), for real-valued \( x \in [0, 1/2] \),

\[
1 - \sum_{j \leq k} c_j \leq 2 \int_{0}^{1/2} \frac{\mathbb{P}_{\geq k}(x)}{1 - T(x)} dx.
\]

Since \((1 - T(x))^{-1} = T'(x)\), we switch from \( x \) to \( \tau \in [0, 1] \) such that \( x = x(\tau) \) is inverse of \( T(x) \). So, introducing \( \Pi_{\geq k}(\tau) = \mathbb{P}_{\geq k}(x(\tau)) \), we obtain

(3.13)

\[
1 - \sum_{j \leq k} c_j \leq 2 \int_{0}^{1} \Pi_{\geq k}(\tau) d\tau.
\]

And, with a bit of work, the above differential equation becomes

(3.14) \( \Pi^{(2)}_{\geq k}(\tau) = \Pi^{(1)}_{\geq k-1}(\tau) (1 - \tau - \frac{1}{1 - \tau}) + \Pi_{\geq k-1}(\tau), \quad k \geq 0, \ \tau \in [0, 1], \)

subject to the initial conditions \( \Pi_{\geq k}(0) = \Pi^{(1)}_{\geq k}(0) = 0 \).
Thus we need to bound the RHS of (3.1.2). Introduce

\[(3.15) \quad I_{k,t} = \int_0^1 (1 - \tau)^t \Pi_{\geq k}^>(\tau) \, d\tau, \quad k \geq -1, \ t \geq 0;\]

so (3.13) becomes

\[1 - \sum_{j \leq k} c_j \leq 2I_{k,0}.\]

To bound \(I_{k,0}\), we use (3.15) to derive a recurrence inequality for all \(t \geq 0\) and \(k \geq -1\). First note that, for \(t \geq 0\),

\[I_{-1,t} = \int_0^1 (1 - \tau)^t \Pi_{\geq -1}^>(\tau) \, d\tau = \int_0^1 (1 - \tau)^t \, d\tau = \frac{1}{(t+2)^2}.\]

Let us turn to \(k \geq 0\). Integrating the RHS of (3.15) by parts twice, using the initial conditions, and (3.14), we have

\[I_{k,t} = \frac{1}{(t+2)^2} \int_0^1 (1 - \tau)^{t+2} \Pi_{\geq k}^>(\tau) \, d\tau\]

\[= \frac{1}{(t+2)^2} \left[ \int_0^1 [(1 - \tau)^{t+3} - (1 - \tau)^{t+1}] \Pi_{\geq k-1}^>(\tau) \, d\tau + I_{k-1,t+2} \right]\]

\[= \frac{1}{(t+2)^2} \left[ (t+4)I_{k-1,t+2} - (t+1)I_{k-1,t} \right]\]

\[\leq \frac{3}{(t+2)^2} \cdot I_{k-1,t+2}.\]

Here the third line results from yet another integration by parts, and the bottom line results from the observation that \(I_{k-1,s}\) decreases as \(s\) increases. It follows then that

\[I_{k,0} \leq \frac{3^k}{(2k)!} \cdot I_{0,2k} \leq \frac{3^k}{(2k+1)!}.\]

\[\square\]

3.1.3. How tight is \(\{n^{-1}\mathbb{E}[X_k(n)]\}_{k \geq 0}\)? It trivially follows from Theorem 3.5 that the sequence of probability distributions \(\{n^{-1}\mathbb{E}[X_k(n)]\}_{k \geq 0}\) is tight, meaning that \(\sup_n n^{-1}\mathbb{E}[X_{\geq k}(n)] \to 0\) as \(k \to \infty\). The question is how fast. The answer is super-exponentially fast.

**Theorem 3.6.** For every \(r \in (0,1)\), we have \(\sup_n n^{-1}\mathbb{E}[X_{\geq k}(n)] = O(r^k)\), with the coefficient implicit in the big-O notation dependent on \(r\) only.

**Proof.** (a) Let \(\rho > 1\) and \(\alpha \in (0,1)\). Introduce \(g_n(\rho) = \sum_{k \geq 0} \rho^k \pi_{\geq k}(n)\). We want to prove that \(g_n(\rho) \leq h_n(\rho) := An^\alpha\) if \(A\) is sufficiently large. Let
\( n > 1 \). The, by (3.12), we have:

\[
    g_n(\rho) = \frac{t(n-1)}{t(n)} (1 + \rho g_{n-1}(\rho)) \\
    + \frac{(n-1)!}{t(n)} \sum_{s \geq 2} \frac{s}{(n-1)(s-1)} \sum_{j=1}^{n-s} (1 + \rho g_j(\rho)) \frac{t(j)(n-1-j)}{j!} \\
    \times \sum_{j_1 + \cdots + j_{n-1} = n-1-j} \prod_{j_t}^{\scriptscriptstyle\pi^{-1}} t(j_t)!
\]

Since \( g_n(0) = \pi_0(n) = 1 \) for \( n > 1 \), and the \((s - 1)\)-fold sum equals \([z^{n-1-j}] T(z)^{s-1}\), the equation simplifies to

\[
    (3.16) \quad g_n(\rho) = 1 + \frac{\rho(n-1)}{t(n)} g_{n-1}(\rho) \\
    \quad + \frac{\rho(n-2)!}{t(n)} \sum_{s \geq 2} \frac{s}{s-1} \sum_{j=1}^{n-s} g_j(\rho) \frac{t(j)(n-1-j)}{j!} \cdot [z^{n-1-j}] T(z)^{s-1}.
\]

Changing the order of summation in the double sum, we transform it into

\[
    \sum_{j=1}^{n-2} g_j(\rho) \frac{t(j)(n-1-j)}{j!} \cdot [z^{n-1-j}] \sum_{s=2}^{n-j} \frac{s}{s-1} T(z)^{s-1} \\
    = \sum_{j=1}^{n-2} g_j(\rho) \frac{t(j)(n-1-j)}{j!} \cdot [z^{n-1-j}] \sum_{s=2}^{\infty} \frac{s}{s-1} T(z)^{s-1} \\
    = \sum_{j=1}^{n-2} g_j(\rho) \frac{t(j)}{j!} \cdot [z^{n-1-j}] \left( \frac{T(z)}{1-T(z)} + \log \frac{1}{1-T(z)} \right) \\
    = \sum_{j=1}^{n-2} g_j(\rho) \frac{t(j)}{j!} \cdot [z^{n-2-j}] \left( \frac{1}{(1-T(z))^2} + \frac{1}{1-T(z)} \right);
\]

for the last step we used \( \nu[z^\nu] F(z) = [z^{\nu-1}] F'(z) \) and \( T'(z) = \frac{1}{1-T(z)} \). So (3.16) becomes

\[
    (3.18) \quad g_n(\rho) = 1 + \frac{\rho(n-1)}{t(n)} g_{n-1}(\rho) \\
    \quad + \frac{\rho(n-2)!}{t(n)} \sum_{j=1}^{n-2} g_j(\rho) \frac{t(j)}{j!} \cdot [z^{n-1-j}] \left( \frac{1}{(1-T(z))^2} + \frac{1}{1-T(z)} \right).
\]

Our task is to prove that the RHS of (3.18) where \( g_j(\rho) \) is replaced with \( h_j(\rho) \) is at most \( h_n(\rho) \), provided that \( n \geq n(\rho) \) and \( A \geq A(\rho) \). We will use the identity \([z^\nu](1 - T(z))^{-\mu} = \frac{(\mu + 2\nu - 2)!}{\nu!} \), where the double factorial is the product of all positive integers at most \( \mu + 2\nu - 2 \), of the same parity as \( \mu \). Using the concavity of the function \( x^\alpha \), we have: \( \sum_j p_j j^\alpha \leq (\sum_j p_j j)^\alpha \) for \( p_j \geq 0 \), \( \sum_j p_j = 1 \).
To apply this inequality to
\[
\Sigma_n = \sum_{j=1}^{n-2} j^\alpha \frac{t(j)}{j^2} \cdot [z^{n-2-j}][(1 - T)^{-3} + (1 - T)^{-2}],
\]
introduce
\[
p_j := \frac{t(j)}{j^{2\alpha}} [z^{n-2-j}][(1 - T)^{-3} + (1 - T)^{-2}]
\sum_k \frac{j(t(k))}{k!} \cdot [z^{n-2-k}][(1 - T)^{-3} + (1 - T)^{-2}]
\]

The denominator equals
\[
[z^{n-2}] T(z)((1 - T)^{-3} + (1 - T)^{-2}) = [z^{n-2}] [(1 - T)^{-3} - (1 - T)^{-1}].
\]
Therefore, for \( n \geq 3 \),
\[
\Sigma_n \leq \left( [z^{n-2}] [(1 - T)^{-3} - (1 - T)^{-1}] \right)^{1-\alpha}
\times \left( \sum_{j=1}^{n-2} \frac{j(t(j))}{j!} \cdot [z^{n-2-j}][(1 - T)^{-3} + (1 - T)^{-2}] \right)^\alpha,
\]
Recalling \( T' = (1 - T)^{-1} \), we transform the second line sum into
\[
[z^{n-2}] \left( \sum_j \frac{j(t(j))}{j!} z^j \right) \cdot ((1 - T)^{-3} + (1 - T)^{-2})
= [z^{n-2}] z T'((1 - T)^{-3} + (1 - T)^{-2}) = [z^{n-3}] T'((1 - T)^{-3} + (1 - T)^{-2})
= [z^{n-3}] T'((1 - T)^{-4} + (1 - T)^{-3}).
\]

Collecting the pieces, we obtain
\[
\Sigma_n \leq \left( [z^{n-2}] [(1 - T)^{-3} - (1 - T)^{-1}] \right)^{1-\alpha}
\times \left( [z^{n-3}] [(1 - T)^{-4} + (1 - T)^{-3}] \right)^\alpha
= \frac{2(2n-5)!!}{(n-2)!} \left( \frac{(2n-4)!! + (2n-5)!!}{(n-3)!} \right)^\alpha.
\]
So the RHS of (3.18) with \( h_j(\rho) \) instead of \( g_j(\rho) \) is at most
\[
1 + A \frac{\rho}{2n-3} (n - 1)^\alpha + A \rho (n-2)! \frac{(2(2n-5)!!)}{(n-2)!} \left( \frac{(2n-4)!! + (2n-5)!!}{(n-3)!} \right)^\alpha
= 1 + A \frac{\rho}{2n-3} (n - 1)^\alpha + A \rho \frac{2}{2n-3} (n - 2) \left( \frac{(2n-4)!!}{(2n-3)!!} + \frac{1}{2n-3} \right)^\alpha
\leq 1 + A \rho (n-1)^\alpha + 2A \rho (n-1)^{1+2\alpha} \leq 1 + 3A \rho (n-1)^{1+2\alpha}.
\]
The last bound is indeed below \( An^\alpha \), if \( n > n(\rho) \) and \( A \geq A(\rho, \alpha) \), where
\[
n(\rho) := 1 + \max \{ 3, (3\rho) (1 - \alpha)^{-1} \}, \quad A(\rho, \alpha) := (n^\alpha (\rho) - 3\rho (2n-1)^{2\alpha}(\rho))^{-1}.
\]
Observe that, very crudely, \( g_n(\rho) \leq n_0 \rho^{n_0} \) if \( n \leq n_0 \). So, for those small values of \( n \) we have \( g_n(\rho) < An^\alpha \), if \( A \geq n_0 \rho^{n_0} \). By induction on \( n \geq n_0 \) we obtain that, for all \( n \), we have \( g_n(\rho) \leq A^* n^\alpha \) if \( A^* := \max\{A(\rho, \alpha), n_0 \rho^{n_0}\} \). Consequently, \( p_{\alpha k}(n) \leq \pi_{\alpha k}(n) \leq A^* n^\alpha \rho^{-k} \).

(b) Using \( h_n(\rho) \leq An^\alpha \), we will prove that \( \sup_{n \geq 1} n^{-1} \mathbb{E}[X_{>k}(n)] = O(r^k) \) for every \( r < 1 \). To begin, analogously to (2.5), we have: for all \( A \rho < 1 \),

\[
\mathbb{E}[X_{>k}(n)] = p_{\alpha k}(n) + \frac{(n-1)!}{t(n)} \sum_{k \geq 0} \sum_{s \geq 1} \prod_{r' = 1}^s \frac{t(j_r)}{j_r!} \left( \sum_{r = 1}^s \mathbb{E}[X_{>k}(j_r)] \right).
\]

By part (a), for all \( \rho \) and \( \alpha < 1 \), we have \( \sum_{k \geq 0} p_{\alpha k}(n) \rho^k \leq h_n(\rho) = An^\alpha \), with \( A = A(\rho, \alpha) \). Introduce \( H_n(\rho) = \sum_{k \geq 0} \rho^k \mathbb{E}[X_{>k}(n)] \); obviously \( H_n(\rho) = O(n) \) for \( \rho \leq 1 \). Let us show that \( H_n(\rho) = O(n) \) for every \( \rho > 1 \) as well. Of course, the coefficient hidden in the big-O notation will depend on \( \rho \). From the above equation, and part (a), it follows that

\[
H_n(\rho) \leq h_n(\rho) + \frac{(n-1)!}{t(n)} \sum_{k \geq 0} \rho^k \sum_{s \geq 1} \prod_{r' = 1}^s \frac{t(j_r)}{j_r!} \left( \sum_{r = 1}^s \mathbb{E}[X_{>k}(j_r)] \right)
\]

\[
= h_n(\rho) + \frac{(n-1)!}{t(n)} \sum_{j=1}^{n-1} \left( \sum_{k \geq 0} H_j(\rho) \frac{t(j)}{j!} [z^{n-1-j}] \sum_{s \geq 1} s T^{s-1}(z) \right)
\]

\[
= h_n(\rho) + \frac{(n-1)!}{t(n)} \sum_{j=1}^{n-1} H_j(\rho) \frac{t(j)}{j!} [z^{n-1-j}] (1 - T(z))^{-2}
\]

\[
= h_n(\rho) + \frac{2^{n-1}(n-1)!}{t(n)} \sum_{j=1}^{n-1} H_j(\rho) \frac{t(j)}{2j!},
\]

since \((1 - T(z))^{-2} = (1 - 2z)^{-1} = \sum_{u \geq 0} (2z)^u\). Clearly \( H_n(\rho) \leq H_n(\rho) \), where \( \mathcal{H}_1(\rho) = H_1(\rho) = 0 \) and

\[
\mathcal{H}_n(\rho) = h_n(\rho) + \frac{2^{n-1}(n-1)!}{t(n)} \sum_{j=1}^{n-1} H_j(\rho) \frac{t(j)}{2j!}, \quad n \geq 2.
\]

It follows that

\[
\frac{t(n+1)}{2^n n!} [\mathcal{H}_{n+1}(\rho) - h_{n+1}(\rho)] - \frac{t(n)}{2^{n-1}(n-1)!} [\mathcal{H}_n(\rho) - h_n(\rho)] = \mathcal{H}_n(\rho) \frac{t(n)}{2^n n!},
\]

or equivalently

\[
\frac{\mathcal{H}_{n+1}(\rho)}{2n+1} = \frac{\mathcal{H}_n}{2n-1} + D_n, \quad D_n := [h_{n+1}(\rho) - \frac{2n}{2n-1} h_n(\rho)] (2n + 1)^{-1}.
\]
Since \( h_n(\rho) = An^\alpha \), we have \( D_n = O(n^{\alpha-1}) \), and the recurrence equation implies that

\[
\mathcal{H}_{n+1}(\rho) = (2n + 1) \cdot \left( \frac{\mathcal{H}_2(\rho)}{\rho^2} + \sum_{j=2}^{n} D_n \right) = O(n + n^\alpha) = O(n).
\]

Therefore, uniformly for \( k \), and \( n \geq 1 \),

\[
n^{-1} \mathbb{E}[X_{>k}(n)] \leq n^{-1} \frac{\mathcal{H}_n(\rho)}{\rho^k} = O(\rho^{-k})
\]

for every \( \rho > 1 \). In other words, the constant hidden in \( O(\rho^{-k}) \) depends only on \( \rho \) and not on \( k \).

\[ \square \]

3.2. Counting \( r \)-tuples of vertices with given ranks in \( T_n \). We saw in Proposition 2.4 that

\[
A_k'(z) = \sum_{\tau \leq r} (1 - T)^{-\tau-1} \left( \sum_{k^{(1)}, \ldots, k^{(\tau)} \neq \emptyset} \prod_{u=1}^{\tau} A_{k^{(u)}}(z) \right)
+ \sum_{u \in [r]} B'_{k_u, (k_1), \ldots, (k_{|r|-1})(u)}(z),
\]

where \((k^{(1)}, \ldots, k^{(\tau)})\) denotes an ordered partition of an \( r \)-tuple \( k \) into \( \tau \) non-empty sub-tuples, and

\[
B'_k(z) \leq \sum_{\tau < r} (1 - T)^{-\tau-1} \left( \sum_{k^{(1)}, \ldots, k^{(\tau)} \neq \emptyset} \prod_{u=1}^{\tau} A_{k^{(u)}}(z) \right),
\]

where \( k = (k_1, k') \) is an \( r \)-tuple, and \((k^{(1)}, \ldots, k^{(\tau)})\) denotes an ordered partition of \( k' \) into \( \tau \) non-empty sub-tuples. Note that the right-hand side of (3.21) does not depend on \( k_1 \).

Lemma 3.7. Given an \( r \)-tuple \( k \), there exist positive constants \( \varepsilon_{k,\ell}^{(i)} \) such that we have \([z^\nu]A_k(z) \in \left[ [z^\nu]A_k^{(1)}(z), [z^\nu]A_k^{(2)}(z) \right]\) for each \( \nu \), \((A_k(z) \in [A_k^{(1)}(z), A_k^{(2)}(z)], \text{ in short})\) where

\[
A_k^{(i)}(z) = \frac{(2r - 3)!}{(1 - T)^{2r-1}} \prod_{j=1}^{r} \gamma_{k_j} + \sum_{\ell \geq 2} \varepsilon_{k,\ell}^{(i)} (1 - T)^{-2r+1+\ell}, \quad T = T(z).
\]

Proof. This claim certainly holds for \( r = 1 \); see (3.5). Suppose the claim holds for all \( r \leq r_0 \), for some \( r_0 \geq 1 \). Each \( k^{(u)} \) on the right-hand side of the inequality (3.21) is of cardinality \( \leq r - 1 \). So, applying the inductive hypothesis to each \( A_{k^{(u)}}(z) \) there, we bound the \( B' \)-sum in (3.20):

\[
\sum_{u \in [r]} B'_{k_u, (k_1), \ldots, (k_{|r|-1})(u)}(z) \in \left[ \sum_{\ell \leq 2r-1} \delta_{k,\ell}^{(1)} (1 - T)^{-\ell}, \sum_{\ell \leq 2r-1} \delta_{k,\ell}^{(2)} (1 - T)^{-\ell} \right].
\]
Let us turn to the $A$-sum in (3.20). For $\tau \in [2,r]$, each $|k^{(u)}| \leq r - 1$, where $|k^{(u)}|$ is just the number of components in $k^{(u)}$. So, by the inductive hypothesis again, the contribution of all ordered partitions of $[r]$ with $\tau \geq 2$ parts to this sum is at least

$$\frac{1}{(1 - T)^{r + 1}} \sum_{\cup_{u=1}^{r} k^{(u)} = k} \prod_{u=1}^{r} A_{k^{(u)}}(z)$$

$$= \frac{\prod_{j=1}^{r} \gamma_{kj}}{(1 - T)^{2r + 1}} \sum_{\cup_{u=1}^{r} k^{(u)} = k} \prod_{u=1}^{r} (2|k^{(u)}| - 3)!! \cdot \left(1 + \sum_{\ell \geq 2} \varepsilon_{k^{(u)}, \ell}(1 - T)^{\ell}\right)$$

$$= \frac{\prod_{j=1}^{r} \gamma_{kj}}{(1 - T)^{2r + 1}} \sum_{\cup_{u=1}^{r} k^{(u)} = k} \prod_{u=1}^{r} (2|k^{(u)}| - 3)!! + \sum_{\ell \geq 2} \frac{\Delta_{k, s, \ell}^{(1)}}{(1 - T)^{2r + 1 - \ell}}.$$

And we have the similar bound from above.

We need to sum these contributions over $\tau \in [2,r]$. The sums (those over $\ell \geq 2$) in the third line add up to the secondary sums

$$\sum_{\ell \geq 2} \frac{\Delta_{k, \ell}^{(i)}}{(1 - T)^{2r + 1 - \ell}}, \quad \Delta_{k, \ell}^{(i)} := \sum_{\tau = 2}^{r} \Delta_{k, \tau, \ell}^{(i)}, \quad i = 1, 2.$$

The factor $\frac{\prod_{j=1}^{r} \gamma_{kj}}{(1 - T)^{2r + 1}}$ aside, the first (primary) sum is the sum of the products $\prod_{u=1}^{r} (2|k^{(u)}| - 3)!!$ over all ordered partitions of the $r$ components of $k$, that is, ordered partitions of the set $[r]$, into $\tau \geq 2$ non-empty subsets. We know that the total number of trees with $\beta$ vertices is $(2\beta - 3)!!$. Therefore the sum of those products over all ordered partitions of $k$ is the total number of ordered trees with $(r + 1)$ vertices, (one root, some number of ordered trees, of total cardinality $r$, rooted at the root’s children) i.e. $2(r + 1) - 3)!! = (2r - 1)!!$. The contribution coming from the partition with a single part of size $r$ is the total number of trees with $r + 1$ vertices (one root, one subtree with $r$ vertices rooted at a single child of the root), which is $(2r - 3)!!$. Hence the sum of the products $\prod_{u=1}^{r} (2|k^{(u)}| - 3)!!$ over all ordered partitions $(k_1, \ldots, k_r)$ of $[r]$ into $\tau \geq 2$ parts is $(2r - 1)!! - (2r - 3)!! = 2(r - 1)(2r - 3)!!$. Consequently

$$\sum_{\cup_{u=1}^{r} k^{(u)} = k} \prod_{u=1}^{r} (2|k^{(u)}| - 3)!! = 2(r - 1)(2r - 3)!!.$$
So, according to (3.23), it remains to solve the differential equation
\[
dA_k(z) = A_k(z) + 2(r - 1)(2r - 3)!! \cdot \frac{\prod_{j=1}^r \gamma_{k_j}}{(1 - T)^{2r+1}} + \sum_{\ell \geq 2} \frac{\Delta_{k,\ell}}{(1 - T)^{2r+1-\ell}}.
\]

Using \( T''/T' = (1 - T)^{-2} \), we obtain
\[
d \left[ A_k(z)(1 - T) \right] = 2(r - 1)(2r - 3)!! \cdot \frac{\prod_{j=1}^r \gamma_{k_j}}{(1 - T)^{2r}} + \sum_{\ell \geq 2} \frac{\Delta_{k,\ell}}{(1 - T)^{2r-\ell}}.
\]

Recall that \( 1 - T = (1 - 2z)^{1/2} \). So, upon integration from 0 to \( z \) in the open disc \( |z| < 1/2 \), we have
\[
A_k(z) = \frac{(2r - 3)!! \prod_{j=1}^r \gamma_{k_j}}{(1 - 2z)^{r-1/2}} + \sum_{\ell \geq 2} \varepsilon_{k,\ell}(1 - 2z)^{-r+1/2+\ell/2},
\]
which is the equation (3.22). \( \square \)

**Theorem 3.8.** Given an \( r \)-tuple \( k = (k_1, \ldots, k_r) \), let \( X_k(n) \) denote the total number of \( r \)-tuples of distinct vertices \( (v_1, \ldots, v_r) \) in the random tree \( T_n \) with ranks \( k \). Then
\[
\mathbb{E}[X_k(n)] = (1 + O(n^{-1}))n^r \prod_{j=1}^r c_{k_j}.
\]

**Proof.** By the definition of \( A_k(z) \), we have
\[
\mathbb{E}[X_k(n)] = n! [z^n] A_k(z) \frac{(2n - 3)!!}{(2n - 3)!!}.
\]

To evaluate this coefficient, introduce \( C_k \), a contour enclosing \( z = 0 \), which is analogous to \( C \) in (3.5)-(3.6), with \( \rho := \rho_k \). By Lemma 3.7 and following
the derivation in those equations, we obtain

\[ z^n A_k(z) = (2r - 3)!! \left( \prod_{j=1}^r \gamma_{k_j} \right) [z^n](1 - 2z)^{-r+1/2} \]

\[ + \sum_{\ell=2}^{2r} \varepsilon_{k,\ell} [z^n](1 - 2z)^{-r+1/2+i/2} + \frac{1}{2\pi i} \oint_{z \in C_k} O\left( \frac{1 - 2z}{z^{n+1}} \right) dz \]

\[ = (2r - 3)!! \left( \prod_{j=1}^r \gamma_{k_j} \right) \frac{(2(n + r) - 3)!!}{n!(2r - 3)!!} \]

\[ + O\left( \frac{(2(n + r) - 5)!!}{n!} \sum_{\ell=2}^{2r} \varepsilon_{k,\ell} \right) + O\left( \int_{1/2}^{\rho_k} \frac{(2u - 1)}{u^{n+1}} du + \rho_k^{-n} \right) \]

\[ = (1 + O(n^{-1})) \left( \prod_{j=1}^r \gamma_{k_j} \right) \frac{(2(n + r) - 3)!!}{n!} + O(n^{-2}2^n). \]

We can drop the last error term, because the explicit term is of order $n^{r-3/2}2^n$ exactly. So, since $c_k = 2\gamma_k$, 

\[ \mathbb{E}[X_k(n)] = \frac{n! \cdot [z^n] A_k(z)}{(2n - 3)!!} = (1 + O(n^{-1})) \left( \prod_{j=1}^r \gamma_{k_j} \right) \cdot \prod_{u=1}^r (2(n + u) - 3) \]

\[ = (1 + O(n^{-1})) n^r \prod_{j=1}^r c_{k_j}. \]

\[ \square \]

**Corollary 3.9.** Let $R_n(1), \ldots, R_n(r)$ denote the ranks of $r$ vertices chosen uniformly at random, with order and without replacement, from $[n]$. Then, for every $(k_1, \ldots, k_r)$, we have

\[ \mathbb{P}(R_n(1) = k_1, \ldots, R_n(r) = k_r) = (1 + O(n^{-1})) \prod_{j=1}^r c_{k_j}. \]

In words, the ranks of a finite, uniformly random, set of vertices are asymptotically independent, each with distribution $\{c_k\}_{k \geq 0}$.

### 3.3. More on the largest rank $R_n$. In Section 3.1 we proved that with probability $\geq 1 - n^{-0.99\varepsilon}$, $R_n$ is at most $(1.5 + \varepsilon) \log n / \log \log n$. Here we will prove that with probability $\geq 1 - n^{-0.99\varepsilon}$, the inequality $R_n \geq (1 - \varepsilon) \log n / \log \log n$ also holds.

We focus on vertices $v$ for which the subtree rooted at $v$ is a path connecting $v$ to a leaf. For example, vertices whose only child is a leaf are vertices with this property. For brevity, we call such vertices path-type, or p-type. The rank of a p-type vertex $v$ is the edge-length of the path that connects
v to a unique leaf. Therefore $R_n \geq R_p^n$, where $R_p^n$ is the largest rank of a path-type vertex in $[n]$.

Recall that $A_k(z)$ is the exponential generating function for the number of all vertices of rank $k$ in all rooted plane trees on vertex set $[n]$, and $B_k(z)$ is the exponential generating function for the number of trees in which the root is of rank $k$. Furthermore, recall $A_{(k,k)}(z)$ is the exponential generating function for the number of ordered pairs of vertices of rank $k$ each in all trees on $[n]$, and $B_{(k,k)}(z)$ is the exponential generating function for the number of ordered pairs $(v, w)$ of rank $k$ each in all such trees so that $v$ is the root of the tree. We will need the following lemma.

**Lemma 3.10.** Let $A_k(z)$, $A_{(k,k)}(z)$, $B_k(z)$, and $B_{(k,k)}(z)$ be the counterparts of $A_k(z)$, $A_{(k,k)}(z)$, $B_k(z)$, and $B_{(k,k)}(z)$ for p-type vertices. Then $B_k(z) = z^k/k!$, while $B_{(k,k)}(z) = 0$, and consequently,

\begin{align}
A_k'(z) &= \frac{A_k(z)}{(1 - T(z))^2} + \frac{z^{k-1}}{(k-1)!}, \\
A_{(k,k)}'(z) &= \frac{A_{(k,k)}(z)}{(1 - T(z))^2} + \frac{2A_k^2(z)}{(1 - T(z))^3}.
\end{align}

The proof is a simpler version of the proofs of Proposition 2.2 and Proposition 2.4 and we omit it. Note that the part concerning $B_k(z)$ and $B_{(k,k)}(z)$ is straightforward. Indeed, there is only one increasing tree with a p-type root of rank $k$, namely the path $1 \to 2 \to \cdots \to k$. So $B_k(z)$, the exponential generating function of numbers of trees with a p-type root of rank $k$, is given by $B_k(z) = z^k/k!$. And there are no trees such that both the root and another vertex are p-type and are of the same rank $k$. So $B_{(k,k)}(z)$, the exponential generating function of numbers of trees with the root and another vertex both p-type and of the same rank $k$, is zero.

Given $k$, let $X_k(n)$ denote the total number of path-type vertices in the random tree $T_n$ of rank $k$.

**Theorem 3.11.** Given $\varepsilon \in (0, 1)$, define $k = [(1-\varepsilon)(\log n)/\log \log n]$. Then

\[ \frac{\mathbb{E}[X_k^2(n)]}{\mathbb{E}^2[X_k(n)]} = 1 + O(\varepsilon). \]

Consequently, with probability $\geq 1 - O(n^{-\varepsilon})$, there exist $\Theta(n^\varepsilon)$ vertices (in fact, p-type vertices) with rank $k(n)$. 
Before we can prove Theorem 3.11, we need to solve the system of ODEs that appear in Lemma 3.10. Arguing as in the proof of Lemma 3.1, we have

\[ A_k(z) = \frac{1}{\sqrt{1-2z}} \int_0^z \frac{d\xi}{\sqrt{1-2\xi}}. \]

(3.25)

\[ A_{k,k}(z) = \frac{1}{\sqrt{1-2z}} \int_0^z \frac{d\xi}{1-2\xi}. \]

This leads to the following Lemma that we will use to prove Theorem 3.11.

**Lemma 3.12.** For \( k < n \), we have

\[ [z^n] A_k(z) = \frac{(2n-1)!!}{n!(2k+1)!!}, \]

(3.26)

\[ [z^n] A_{k,k}(z) = (1 + O(n^{-1})) \frac{(2n+1)!!}{n![(2k+1)!!]^2}. \]

**Proof.** Let us start with \( A_k(z) \). We compute

\[
\sqrt{1-2\xi} \frac{\xi^{k-1}}{(k-1)!} = \frac{\sqrt{1-2\xi}}{2^{k-1}(k-1)!} (1 - (1 - 2\xi))^{k-1} = \frac{1}{2^{k-1}(k-1)!} \sum_{j<k} (-1)^j \binom{k-1}{j} (1 - 2\xi)^{j+1/2}.
\]

Therefore, integrating

\[
A_k(z) = \frac{(1-2z)^{-1/2}}{2^{k-1}(k-1)!} \sum_{j<k} \frac{(-1)^j}{2j+3} \binom{k-1}{j} [1 - (1-2z)^{j+3/2}] = \frac{(1-2z)^{-1/2}}{2^{k-1}(k-1)!} \sum_{j<k} \frac{(-1)^j}{2j+3} \binom{k-1}{j} - P_k(z),
\]

\[ P_k(z) := \frac{1}{2^{k-1}(k-1)!} \sum_{j<k} \frac{(-1)^j}{2j+3} \binom{k-1}{j} (1 - 2z)^{j+1}.
\]

Here \( P_k(z) \) is a polynomial of degree \( k \). Let us assume from now on that \( k < n \). Then

\[ [z^n] A_k(z) = \frac{1}{2^{k-1}(k-1)!} \sum_{j<k} \frac{(-1)^j}{2j+3} \binom{k-1}{j} \times [z^n](1 - 2z)^{-1/2}. \]

The familiar last factor is \((2n-1)!!/n!\), and (using Maple in the third line)

---

1. Note that \( \int_0^1 x^2(1-x^2)^{k-1} \, dx \) equals, by the change of variables \( u = x^2 \), the expression \( B(3/2, k)/2 \), where \( B \) is the Beta-function. Then using the formula for Beta in terms of the Gamma-function the formula follows.
\[
\sum_{j<k} \frac{(-1)^j}{2j+3} \binom{k-1}{j} = \sum_{j<k} (-1)^j \binom{k-1}{j} \int_0^1 x^{2j+2} \, dx \\
= \int_0^1 x^2 (1 - x^2)^{k-1} \, dx \\
= \int_0^1 (1 - x^2)^{k-1} \, dx - \int_0^1 (1 - x^2)^k \, dx \\
= \frac{2^{k-1} (k-1)!}{(2k+1)!!}.
\]

We conclude that
\[(3.27) \quad A_k(z) = \frac{(1 - 2z)^{-1/2}}{(2k+1)!!} - P_k(z), \quad [z^n]A_k(z) = \frac{(2n-1)!!}{n! (2k+1)!!}.
\]

It remains to plug the formula for \(A_k(z)\) into the bottom equation in \(3.25\).

It is easy to check that neglecting \(P_k(z)\) in this formula results in an asymptotic formula for \([z^n]A_{(k,k)}(z)\) within a factor \(1 + O(n^{-1})\) from the actual coefficient. With \(P_k(z)\) dropped, we have
\[
[z^n](1 - 2z)^{-1/2} \int_0^z \frac{2(1 - 2\xi)^{-2}}{(2k+1)!!^2} \, d\xi = \frac{1}{n! (2k+1)!!^2} \left[ \frac{1}{(1 - 2z)^{1/2}} - \frac{1}{(1 - 2z)^{1/2+1}} \right] \\
= \frac{1}{n! (2k+1)!!^2} \left[ (2n+1)!! - (2n - 1)!! \right] \\
= (1 + O(n^{-1})) \frac{(2n+1)!!}{n! (2k+1)!!^2}.
\]

Therefore
\[(3.28) \quad [z^n]A_{(k,k)}(z) = (1 + O(n^{-1})) \frac{(2n+1)!!}{n! (2k+1)!!^2}.
\]

□

Now we can prove Theorem 3.11

**Proof.** (of Theorem 3.11). Using \(3.27\), we have
\[
\mathbb{E}[X_k(n)] = \frac{n! [z^n]A_k(z)}{(2n-3)!!} = \frac{(2n-1)}{(2k+1)!!}.
\]

Applying Stirling formula, and setting \(k(n) := \left[ \frac{(1-\varepsilon)\log n}{\log \log n} \right] \), we obtain
\[
\mathbb{E}[X_{k(n)}(n)] = n^\varepsilon (1 + O(1/(\log \log n))) \to \infty.
\]

Furthermore
\[
\mathbb{E}[X_{k(n)}(n)] = \frac{n! [z^n]A_{k(n),k(n)}(z)}{(2n-3)!!} = (1 + O(n^{-1}))\mathbb{E}^2[X_{k(n)}(n)].
\]
So,
\[
\mathbb{E}[X_{k(n)}^2(n)] = \mathbb{E}[X_{k(n)}(n)] + \mathbb{E}[(X_{k(n)}(n))^2] = (1 + O(n^{-1}) + O(\mathbb{E}^{-1}[X_{k(n)}(n)]) \mathbb{E}^2[X_{k(n)}(n)]) = (1 + O(n^{-\epsilon}))\mathbb{E}^2[X_{k(n)}(n)].
\]
Consequently, by Chebyshev’s inequality, we have
\[
P\left[ \left| \frac{X_{k(n)}(n)}{\mathbb{E}[X_{k(n)}(n)]} - 1 \right| \geq \delta \right] = O(\delta^{-2}n^{-\epsilon}).
\]
\[\square\]

Combining Theorem 3.3 and Theorem 3.11, we arrive at the following result.

**Theorem 3.13.** For the largest rank \( R_n \) of a vertex, and \( \varepsilon \in (0, 1) \),
\[
P\left( \frac{R_n}{\log n} \in [1 - \varepsilon, 1.5 + \varepsilon] \right) = 1 - O\left( n^{-0.99\varepsilon} \right).
\]
Thus the largest vertex rank slowly but inexorably tends to infinity. Still the ranks of finitely many, uniformly random vertices approach finite random limits since \( \{c_j\} \) is proper. Can 1.5 be replaced with 1 without changing the RHS bound?

**Acknowledgment.** We are genuinely grateful to the two hard-working referees who helped us to significantly improve our paper.

**References**

[1] A-L. Barabási, R. Albert, *Emergence of scaling in random networks*, Science, 286 (1999), 509–512.

[2] F. Bergeron, P. Flajolet, and B. Salvy, *Varieties of increasing trees*, Lecture Notes in Comput. Sci. 581 (1992), 24–48.

[3] B. Bollobás, *Random graphs*, 2nd ed., Cambridge University Press, Cambridge, 2001.

[4] B. Bollobás and O. Riordan, *Mathematical results on scale-free random graphs*, in Ch 1. in Handbook of Graphs and Networks: from the Genome to the Internet, Wiley-VCH, Weinheim, (2003).

[5] M. Bóna, *k-protected vertices in binary search trees*, Adv. in Appl. Math 53 (2014), pp. 1–11.

[6] M. Bóna, B. Pittel, *On a random search tree: asymptotic enumeration of vertices by distance from leaves*, Adv. Appl. Prob. 49 (2017), 850–876.

[7] M. Bóna, I. Mezö, *Limiting probabilities for vertices of a given rank in 1-2 trees*, Electron. J. Combin. 26 (2019), no. 3, Paper No. 3.41, 19 pp.

[8] M. Drmota, *The height of increasing trees*, Annals of Combinatorics 12 (2009), 373–402.

[9] *Random Trees: An interplay between Combinatorics and Probability*, Springer Verlag, 2009.
Appendix. Originally we used the randomized algorithm for a suboptimal path from the root to a leaf for the search tree, \cite{B21}. This is a plane binary tree on vertex set \([n]\): recursively, vertex \(n+1\) joins the left subtree with probability proportional to its current size. We proved the existence of the limiting distribution \(\{c_j\}\) of \(R_n\), the rank of the random vertex in \([n]\), satisfying
\[
1 - \sum_{j=0}^{k-1} c_j \leq \frac{6k+7}{3} \left( \frac{1}{3} \right)^k.
\]
We also claimed that for every \(\rho \in (0, 3/2)\), the equality \(\lim E[\rho^{R_n}] = \sum_j \rho^j c_j\) holds, implying that \(R_n\) converges, in distribution, to the limit
rank \( R_\infty \) with all its moments. Looking at the proof anew, we have found an oversight. Here we present a corrected argument that proves a stronger result: 

\[ \lim_{n \to \infty} \mathbb{E}[\rho^{R_n}] = \sum_j \rho^j c_j \text{ for every } \rho \in [0, 3), \]  

so that 

\[ 1 - \sum_{j=0}^{k-1} \mathbb{P}(R_n = j) = O((1 + \epsilon)^k), \]

for every \( \epsilon > 0 \), almost matching the above tail bound for \( \{c_j\} \). The proof is based on the randomized algorithm. Let \( L_n \) and \( L_n \) denote, respectively, the rank of the root and the edge length of the path from the root to a leaf delivered by the algorithm. Then the inequality \( \mathbb{P}(L_n > k) \leq \mathbb{P}(L_n > k) \)

holds, meaning that \( h_n(\rho) := \mathbb{E}[\rho^{L_n}] \leq \mathbb{E}[\rho^{L_n}] \).

From the definition of plane binary trees and that of the algorithm, \( \pi_{n,k} := 1 \) for \( n \geq 1 \), and for \( k \geq 0 \) we have \( \pi_{1,k} = 0 \), while for \( n > 1 \) \( \pi_{n,k} \) satisfies a recurrence equation

\[ \pi_{n,k} = \frac{1}{n} \sum_{j=0}^{n-1} \left( \frac{n-j}{n-1} \pi_{j,k-1} + \frac{j}{n-1} \pi_{n-j,k-1} \right) \]

\[ = \frac{2}{n(n-1)} \sum_{j=0}^{n-1} (n-1-j) \pi_{j,k-1}; \quad \pi_{0,0} := 1. \]

So, for \( n > 1 \), \( g_n(\rho) := \sum_{k \geq -1} \rho^k \pi_{n,k} \) satisfies

\[ g_n(\rho) = \rho^{-1} + \sum_{k \geq 0} \rho^k \pi_{n,k} = \rho^{-1} + \frac{2}{n(n-1)} \sum_{j=0}^{n-1} (n-1-j)g_j(\rho). \]

Let us select \( \rho > 1 \) and \( \alpha \in (0, 1) \). Since \( g_j(\rho) < j \rho^j \), for every \( J > 0 \) there exists \( A := A(\alpha, \rho, J) > 0 \) such that \( g_j(\rho) \leq Aj^\alpha \) for \( j \leq J \). Now, for \( n > J \) we have

\[ \sum_{j=1}^{n-2} (n-1-j)j^\alpha \leq n \left[ \int_0^n y^\alpha \, dy - \int_0^n y^{\alpha+1} \, dy \right] = \frac{n^{\alpha+2}}{(\alpha+1)(\alpha+2)}. \]

So, the RHS of (3.29) with \( Aj^\alpha \) instead of \( g_j(\rho) \) is at most

\[ \rho^{-1} + A\frac{2\rho\alpha}{(\alpha+1)(\alpha+2)}(1 + O(J^{-1})), \]

which in turn is below \( An^\alpha \) for \( J \) large enough, provided that \( \rho < \frac{(\alpha+1)(\alpha+2)}{2} \) and

\[ A > \frac{\rho^{-1}}{1 - \frac{2\rho}{(\alpha+1)(\alpha+2)}(1 + O(J^{-1}))}. \]

Notice that \( \rho \) can be pushed arbitrarily close from below to \( \rho = 3 \) by making \( \alpha \) close enough to 1 from below. By induction on \( n \geq J \) we obtain that \( g_n(\rho) \leq An^\alpha \) for all \( n \).
Consequently, $h_n(\rho) = \mathbb{E}[\rho^{E_n}] = O(n^\alpha)$ for $\rho < \frac{(\alpha+1)(\alpha+2)}{2}$. Introduce $H_n(\rho) = \sum_{k<n} \rho^k \mathbb{E}[X_{n,k}]$, the expected value of $\sum_{v \in [n]} \rho^{R(v)}$, where $R(v)$ denotes the rank of a vertex $v$. Then

$$H_n(\rho) = h_n(\rho) + \frac{1}{n} \sum_{j=0}^{n-1} (H_j(\rho) + H_{n-1-j}(\rho)), \quad n > 1.$$

Here $\frac{h_n(\rho)}{n} = O(n^{\alpha-1})$; applying Lemma 2.1 from [6] we obtain: for each $\rho < \frac{(\alpha+1)(\alpha+2)}{2}$, there exists a finite $\lim_{n \to \infty} n^{-1} \sum_{v \in [n]} \rho^k \mathbb{E}[X_{n,k}]$. In other words, denoting $R_n$ the rank of the random vertex in the random search tree, we have: for each $\rho < 3$, $\lim_{n \to \infty} \sum_{k \geq 0} \rho^k \mathbb{P}(R_n = k) = \sum_{k \geq 0} \rho^k c_k$. 

Department of Mathematics, University of Florida, 358 Little Hall, PO Box 118105, Gainesville, FL, 32611 – 8105 (USA)
Email address: bona@ufl.edu

Department of Mathematics, The Ohio State University, 231 West 18th Avenue, Columbus, Ohio 43210 – 1175 (USA)
Email address: bgp@math.ohio-state.edu