Path Length and External Path Length in Random Trees

Jacob Lundblad
Department of Mathematics, Uppsala University, Box 480, 751 06 Uppsala, Sweden

Stephan Wagner
Institute of Discrete Mathematics, TU Graz, Steyrergasse 30, 8010 Graz, Austria
Department of Mathematics, Uppsala University, Box 480, 751 06 Uppsala, Sweden

Abstract

We consider two closely related concepts in rooted trees: the path length is the sum of all distances from the root to the vertices of the tree, while the external path length is the sum of all distances from the root to the leaves of the tree. Upon dividing by the number of vertices and leaves respectively, we obtain the average distance to the root. For two important classes of random trees, we show that the average distance of the root to a random leaf is almost the same as the average distance to a random vertex. For Bienaymé–Galton–Watson trees, the difference is bounded in probability. For three varieties of increasing trees (recursive trees, d-ary increasing trees, generalised plane-oriented recursive trees), on the other hand, the difference converges to a constant in probability.

2012 ACM Subject Classification Mathematics of computing → Generating functions; Mathematics of computing → Enumeration; Theory of computation → Randomness, geometry and discrete structures

Keywords and phrases Path length, external path length, Bienaymé–Galton–Watson trees, increasing trees

Digital Object Identifier 10.4230/LIPIcs.AofA.2026.2

Supplementary Material *Software (SageMath computations)*: `https://github.com/JacobLundblad/ExternalPathLengthAndTerminalWienerIndex`

Funding *Jacob Lundblad*: Supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.
Stephan Wagner: Supported by the Swedish research council (VR), grant 2022-04030.

1 Introduction

The *path length*, defined as the sum of the depths (distances to the root), is a fundamental quantity associated with a rooted tree. Different notions of path length exist depending on the range of summation: the *internal path length* is the sum of the depths of all internal vertices (non-leaves), while the *external path length* is the sum of the depths of all external vertices (leaves). When the sum is taken over all vertices, one obtains the *(total) path length* (also known as *total height*) of the tree.

The path length in its different versions can be used as a measure for the performance of algorithms based on trees. As a result, its asymptotic distribution has been studied for various types of random trees.

In this extended abstract, we will be particularly interested in how the different notions of path length differ. If we divide the internal/external/total path length of a tree by the internal/external/total number of vertices, we obtain the respective average distance to the root. It is intuitive to assume that the average distance from a leaf to the root should be greater than the average distance from an arbitrary vertex to the root (though this is not deterministically the case: consider for example a tree with n leaves and a path of length n attached to a common root as a counterexample). However, the question remains how much these averages differ.

We study this question for two different types of random trees – conditioned Bienaymé–Galton–Watson trees (critical, aperiodic, with finite variance) and linear preferential attachment trees (which encompass recursive trees, d-ary increasing trees and generalised plane-oriented recursive trees). In both cases, we will show that the difference between the average depth of a leaf and the average depth of an arbitrary vertex is only of constant order. This implies in particular that the different notions of path length lead to the same limiting distribution.

In particular, for both types of trees, we obtain the asymptotic behaviour of the first and second moments of the difference between the average distance from the root to a random vertex and the average distance from the root to a random leaf. In both cases, these are asymptotically constant. For linear preferential attachment trees, the second moment is asymptotically equal to the square of the first, so that the variance goes to zero and we obtain convergence in probability to a constant. For Bienaymé–Galton–Watson trees, we do not get quite as strong results, but the difference is shown to be bounded in probability.

For all these families, the path length and related depth parameters have previously been studied individually. For conditioned Bienaymé–Galton–Watson trees, it is well known that the expectation of the path length is of asymptotic order $n^{3/2}$. After a proper rescaling, it converges weakly to the *Airy distribution*, which can be seen as the area under a Brownian excursion. This was first shown by Takács (in special cases, [27, 28]) and Aldous (for more general Bienaymé–Galton–Watson trees, see [1, Corollary 9]) in the 1990s, see also [6, Theorem 4.9].

On the other hand, for trees of logarithmic height (such as linear preferential attachment trees), the path length is typically of order $n \log n$. The expected value in the case of recursive trees has the nice closed form $n(H_n - 1)$, where H_n denotes the n-th harmonic number. This goes back to Moon in 1974 [19].

The path length of recursive trees was further investigated by Szymański [26], Dobrow and Smythe [5], Mahmoud [14] and Dobrow and Fill [4]. Other types of linear preferential attachment trees have been considered as well: plane-oriented recursive trees [15], binary search trees [22, 23], and (weighted) d-ary increasing trees [20]. See also [24] for a general treatment of linear preferential attachment trees.

Linear preferential attachment trees can also be seen as varieties of increasing trees. Bergeron, Flajolet and Salvy [2] showed that the path length grows in the order of $n \log n$ for rather general classes of increasing trees (that include those of interest in this paper as special cases). Let us finally also mention that the distribution of the path length has also been studied in the very general setting of split trees [3].

By contrast, the external path length has received much less attention (see [12, 13] for two instances), but by comparing it to the ordinary path length we can now transfer results from one to the other.

The paper is structured as follows: we start with a formal definition of the tree models and the quantities of interest and state the main results in the following section. An outline of the proofs, which are based on the analysis of suitable generating functions, is given in Section 3, with a more detailed discussion of conditioned Bienaymé–Galton–Watson trees and linear preferential attachment trees in Sections 4 and 5 respectively. A brief section on possible extensions and generalisations concludes the paper. The proofs of some technical lemmas are given in the appendix.

2 Definitions and main results

We study two closely related depth parameters of rooted trees, the path length and the external path length. The *path length* (also known as total height) of a tree T is defined as the sum of all distances between vertices in the tree and the root r, i.e.,

$$P(T) = \sum_{v \in V(T)} d_T(r, v).$$

The *external path length* of the tree (terminal path length, total leaf height) is defined analogously, but with the summation restricted to leaves (vertices without children), i.e.,

$$P_E(T) = \sum_{l \in L(T)} d_T(r, l),$$

where $L(T)$ is the set of all leaves of the tree T. Normalizing these parameters by the number of vertices and the number of leaves yields the average depth of a uniformly chosen vertex and a uniformly chosen leaf, respectively.

A **simply generated family of trees**, as introduced by Meir and Moon [17], is described by a weight sequence $\phi_0, \phi_1, \ldots$, which for each tree T associates a weight

$$\omega(T) = \prod_{v \in V(T)} \phi_{\deg_+(v)} = \prod_{i \geq 0} \phi_i^{c_i(T)}, \tag{1}$$

where $\deg_+(v)$ denotes the outdegree of a vertex v and $c_i(T)$ is the number of vertices in T with outdegree i. The generating function of these weights, $G(x) = \sum_T \omega(T) x^{|T|}$, then satisfies the equation

$$G(x) = x\Phi(G(x)), \tag{2}$$

where $\Phi(t) = \sum_{i \geq 0} \phi_i t^i$. Whenever the radius of convergence of Φ is positive, there exists a probability distribution ξ such that the distribution on a family of simply generated trees of size n with probabilities proportional to the weights is the same as the distribution of a random **conditioned Bienaymé–Galton–Watson tree** with offspring distribution ξ. This type of random tree is constructed by a branching process that starts with a root vertex and then attaches a number of children to it according to ξ. This is then repeated independently for all successors. Conditioning this process on resulting in a finite tree with n vertices yields a conditioned Bienaymé–Galton–Watson tree.

Setting $\phi_i = \mathbb{P}(\xi = i)$ turns $\omega(T)$ into the probability that the (unconditioned) process yields the tree T. Translating from a general weight sequence to an equivalent probability distribution can be achieved by a transformation known as *tilting*, where ϕ_i is replaced by $\phi_i \cdot ab^i$ for suitable constants a and b – see for example Drmota's book [6] or Janson's survey [11] for details.

In this paper, we will make the standard assumption that $\mathbb{E}\,\xi = 1$, i.e., that the process is *critical* (see again [6] or [11]), and also assume that $\operatorname{Var}\xi = \sigma^2 < \infty$. For technical convenience, we only consider the aperiodic case, where $\operatorname{span}(\xi) = \gcd\{i : \mathbb{P}(\xi = i) > 0\} = 1$. However, it would not be difficult to extend all results to the case where $\operatorname{span}(\xi) > 1$.

Under these conditions ($\mathbb{E}\,\xi = 1$, $\operatorname{Var}\xi = \sigma^2 < \infty$ and $\operatorname{span}(\xi) = 1$), the generating function $G(x)$ has a dominant square root singularity at 1 and is otherwise analytic in what is known as a Δ-domain. This is well known in the case where ξ has exponential moments, but perhaps less well known under the weaker assumption that the variance is finite. The proof of the following result, which will be crucial for our approach, is due to Janson [10, Lemma A2].

▶ **Lemma 1.** *Consider conditioned Bienaymé–Galton–Watson trees with an offspring distribution ξ that satisfies $\mathbb{E}\,\xi = 1$, $Var\xi = \sigma^2 < \infty$ and $\operatorname{span}(\xi) = 1$, and let $G(x)$ be the associated generating function. Then $G(x)$ is analytic in a domain of the form*

$$\{x \in \mathbb{C} : |x| < 1+\epsilon, x \neq 1, |\arg(x-1)| > \delta\}, \tag{3}$$

where $\epsilon > 0$ and $0 < \delta < \frac{\pi}{2}$, and it also satisfies

$$G(x) = 1 - \frac{\sqrt{2}}{\sigma}\sqrt{1-x} + o(|x-1|^{1/2})$$

as $x \to 1$ within this domain.

Let us now state the main result of this paper for Bienaymé–Galton–Watson trees.

▶ **Theorem 2.** *Let T denote a conditioned Bienaymé–Galton–Watson tree of size $|T| = n$ with offspring distribution ξ satisfying $\mathbb{E}\,\xi = 1$, $Var\xi = \sigma^2 < \infty$ and $\operatorname{span}(\xi) = 1$. Set $\phi_0 = \mathbb{P}(\xi = 0)$. Then, as $n \to \infty$,*

$$\mathbb{E}\left(\frac{P_E(T)}{|L(T)|} - \frac{P(T)}{|T|}\right) \to \frac{1}{\sigma^2}, \tag{4}$$

$$\mathbb{E}\left(\left(\frac{P_E(T)}{|L(T)|} - \frac{P(T)}{|T|}\right)^2\right) \to \frac{2\phi_0 + (1-\phi_0)\sigma^2}{3\phi_0\sigma^2}, \tag{5}$$

and

$$\frac{P_E(T)}{|L(T)|} - \frac{P(T)}{|T|} = O_p(1),$$

where $O_p(1)$ denotes boundedness in probability.

▶ Remark 3. In principle, it would be possible to determine the asymptotic behaviour of higher moments as well. However, we were not able to identify a general formula for all moments that would potentially characterise the limiting distribution of the difference $\frac{P_E(T)}{|L(T)|} - \frac{P(T)}{|T|}$. This is left as an open problem.

As a side result to the fact that the average distances to the root, i.e., $\frac{P(T)}{|T|}$ and $\frac{P_E(T)}{|L(T)|}$, are close to each other, we also obtain the limiting distribution of $P_E(T)$. As mentioned in the introduction, $\frac{P(T)}{n^{3/2}}$ converges weakly to a multiple of the Airy distribution. We now find that this is also the case for $\frac{P_E(T)}{n^{3/2}}$.

▶ **Corollary 4.** *Let T denote a conditioned Bienaymé–Galton–Watson tree of size $|T| = n$ with offspring distribution ξ satisfying the same conditions as in Theorem 2. Then there exists a constant η such that*

$$\frac{P_E(T)}{\eta n^{3/2}}$$

converges weakly to the Airy distribution as $n \to \infty$.

▶ Remark 5. It was shown by Marckert and Mokkadem [16] that the height process of the leaves has the same limit as the height process associated with all vertices in the tree. In fact, this holds in greater generality for vertices with given outdegree. As such, it should not come as a surprise that the first order asymptotics of $\frac{P_E(T)}{|L(T)|}$ and $\frac{P(T)}{|T|}$ agree. However, it does not seem obvious that the difference is even bounded in probability.

Random linear preferential attachment trees are trees constructed using a growth process, starting with a single root, where the probability of attachment to a vertex depends linearly on its number of children. For a real number α, a random (linear) preferential attachment tree $\mathcal{T}_{\alpha,n}$ is obtained from $\mathcal{T}_{\alpha,n-1}$ by attaching the n-th vertex to $v \in V(\mathcal{T}_{\alpha,n-1})$ with probability

$$\frac{\alpha \deg_+(v) + 1}{\sum_{u \in V(\mathcal{T}_{\alpha,n-1})}(\alpha \deg_+(u) + 1)} = \frac{\alpha \deg_+(v) + 1}{\alpha(n-1) + n},$$

where $\deg_+(v)$ is again the number of children of v. To avoid negative probability of choosing a vertex v, this is restricted to $\alpha \in \{\dots, -\frac{1}{3}, -\frac{1}{2}\} \cup [0, \infty)$. This also excludes the case $\alpha = -1$, which simply produces a rooted path. In order to perform our analysis, we need to consider three separate cases: $\alpha = 0$, $\alpha \in \{\dots, -\frac{1}{3}, -\frac{1}{2}\}$, and $\alpha > 0$.

- $\alpha = 0$: When $\alpha = 0$, the choice between existing vertices is uniform and we get **random recursive trees**. They can be defined either using the growth process described above, or as labelled trees with vertices labelled $1, \dots, n$ in an increasing fashion, chosen uniformly at random. These trees have been studied since at least 1967 (see Tapia and Myers [29]). There is a bijection between trees of size n and permutations of $n-1$ elements, and thus the number of recursive trees of size n is $(n-1)!$.
- $\alpha \in \{\dots, -\frac{1}{3}, -\frac{1}{2}\}$: When $\alpha = -\frac{1}{d}$ for a positive integer d, we get **random d-ary increasing trees**. From a more combinatorial viewpoint, these can be seen as trees where each vertex has d places where a child can be attached (thus up to d children), and vertices are again labelled $1, \dots, n$ in an increasing fashion.
- $\alpha > 0$: these are preferential attachment trees in the sense that vertices with many children have a greater probability of attracting new vertices. These trees are also known as **random generalised plane-oriented recursive trees** (GPORTs), as they extend the special case of **plane oriented recursive trees** (introduced by Szymański [25]), or PORTs, which corresponds to $\alpha = 1$.

All three cases can also be obtained from a combinatorial construction similar to that of simply generated families of trees: on the set of all rooted ordered trees with vertices labelled $1, \dots, n$ in an increasing fashion (i.e., the labels increase as one moves away from the root, see also [2]), one defines a weight function as in (1), with ϕ_i as follows:

- $\phi_i = \frac{1}{i!}$ to obtain recursive trees,
- $\phi_i = \binom{d}{i}$ to obtain d-ary increasing trees,
- $\phi_i = \binom{i-1+\frac{1}{\alpha}}{i}$ to obtain GPORTs.

Random trees in each of these cases are obtained by choosing trees of a given size with probabilities proportional to the weights. As in the case of simply generated families, one could multiply ϕ_i by ab^i for arbitrary positive constants a and b to obtain a probabilistically equivalent choice of weights. While the choices of weights might look rather different, we remark that they are always equal to $\frac{1}{i!}\prod_{j=0}^{i-1}(1 + \alpha j)$ times a factor of the form b^i.

It was shown by Panholzer and Prodinger [21] that these are the only choices of weights for which the probabilistic model is equivalent to a suitable growth process.

Aiming towards a similar analysis as in the simply generated case, we consider exponential generating functions of the form

$$\sum_T \omega(T) \frac{x^{|T|}}{|T|!},$$

where the sum is over all rooted ordered increasing trees. We use G_{r} for the exponential generating function associated with recursive trees, G_d for d-ary increasing trees and G_α for GPORTs. Each of these satisfies a first order differential equation (see [6]). Specifically,

$$G_{\mathrm{r}}'(x) = e^{G_{\mathrm{r}}(x)} \tag{6}$$

for recursive trees, with the explicit solution

$$G_{\mathrm{r}}(x) = \log\left(\frac{1}{1-x}\right).$$

Likewise,

$$G_d'(x) = (1 + G_d(x))^d \tag{7}$$

with the explicit solution

$$G_d = (1-(d-1)x)^{\frac{1}{1-d}} - 1.$$

Finally,

$$G_\alpha'(x) = (1 - G_\alpha(x))^{-\frac{1}{\alpha}} \tag{8}$$

with the explicit solution

$$G_\alpha = 1 - \left(1 - \left(1 + \frac{1}{\alpha}\right)x\right)^{\frac{\alpha}{1+\alpha}}.$$

These differential equations and the slightly different shapes of their solutions are a main reason for differentiating between the three cases.

Let us now state the main result for linear preferential attachment trees, which parallels Theorem 2.

▶ **Theorem 6.** *Let T denote a random linear preferential attachment tree $\mathcal{T}_{\alpha,n}$ of size $|T| = n$, with $\alpha \in \{\dots, -\frac{1}{3}, -\frac{1}{2}\} \cup [0, \infty)$. Then, as $n \to \infty$,*

$$\mathbb{E}\left(\frac{P_E(T)}{|L(T)|} - \frac{P(T)}{|T|}\right) \to \frac{1}{(1+\alpha)(2+\alpha)},$$
$$\mathbb{E}\left(\left(\frac{P_E(T)}{|L(T)|} - \frac{P(T)}{|T|}\right)^2\right) \to \frac{1}{(1+\alpha)^2(2+\alpha)^2},$$

and

$$\frac{P_E(T)}{|L(T)|} - \frac{P(T)}{|T|} \xrightarrow{P} \frac{1}{(1+\alpha)(2+\alpha)},$$

where $\xrightarrow{P}$ denotes convergence in probability.

An analogue of Corollary 4 follows as well. For random linear preferential attachment trees, $\frac{P(T)-\mathbb{E}(P(T))}{n}$ converges in distribution to a limiting random variable X_α (whose distribution depends on α) as $n \to \infty$ (in fact, even almost sure convergence holds, see [24]).

▶ **Corollary 7.** *Let T denote a random linear preferential attachment tree $\mathcal{T}_{\alpha,n}$ of size $|T| = n$, with $\alpha \in \{\dots, -\frac{1}{3}, -\frac{1}{2}\} \cup [0, \infty)$. Then, as $n \to \infty$,*

$$\frac{P_E(T) - \mathbb{E}(P_E(T))}{n} \xrightarrow{d} \frac{1+\alpha}{2+\alpha} X_\alpha.$$

3 Proof outline

The proofs of Theorems 2 and 6 are based on the analysis of multivariate generating functions encoding the relevant statistics $|T|$, $|L(T)|$, $P(T)$ and $P_E(T)$.

The functional and differential equations satisfied by the various generating functions reflect the recursive structure of trees: a tree can be decomposed into a root and smaller trees (branches) that are attached to it.

In order to study the path length and external path length we will therefore consider the following recursions that also make use of this branch structure. If T has more than one vertex and $T_1, T_2, \ldots, T_m$ are the branches of T (the rooted trees that remain as components when the root is removed), then we have the recursions

$$L(T) = \sum_{i=1}^{m} L(T_i), \tag{9}$$

$$P(T) = \sum_{i=1}^{m} \left(P(T_i) + |T_i|\right) \tag{10}$$

and

$$P_E(T) = \sum_{i=1}^{m} \left(P_E(T_i) + |L(T_i)|\right). \tag{11}$$

In the Bienaymé–Galton–Watson case, this leads to an implicit functional equation extending (2) to the multivariate setting. For the different varieties of increasing trees, we work with an exponential multivariate generating function and obtain partial differential equations that extend the ODEs (6), (7) and (8). Repeated differentiation then allows us to determine generating functions for the first and second moments of the random variable

$$|T|P_E(T) - |L(T)|P(T).$$

Using singularity analysis and the known behaviour of the univariate generating functions for each tree family, we are able to determine the asymptotic behaviour of these moments. The calculations related to the generating functions are carried out in SageMath [30]; the complete code used for this project is available in a public repository.[1] Using concentration results for the number of leaves $|L(T)|$ allows for these moment asymptotics to be used to determine the asymptotic behaviour of the first and second moments of the quotient

$$\frac{|T|P_E(T) - |L(T)|P(T)}{|T||L(T)|}.$$

To handle this ratio, we rely on the technical lemma stated below, the proof of which is given in the appendix.

▶ **Lemma 8.** *Let X_n and Y_n be random variables. Assume that $\frac{\mathbb{E}(X_n)}{\mu_n}$ and $\frac{\mathbb{E}(X_n)}{\mathbb{E}(|X_n|)}$ are bounded below by positive constants, that $|X_n| \le h(n)$ for some function h and that $|Y_n| \ge 1$. If Y_n satisfies*

$$\mathbb{P}\left(|Y_n - \mu_n| > g(n)\right) \le f(n)$$

[1] All code used for symbolic computations is available at: `https://github.com/JacobLundblad/ExternalPathLengthAndTerminalWienerIndex`.

for some functions f and g such that $f(n) = o(1)$, $h(n) = o(1/f(n))$ and $g(n) = o(\mu_n)$, then

$$\mathbb{E}\left(\frac{X_n}{Y_n}\right) \sim \frac{\mathbb{E}(X_n)}{\mu_n} \quad \text{as } n \to \infty.$$

▶ Remark 9. Note that we do not require $\mathbb{E}(Y_n) = \mu_n$.

For some calculations we will also use the following lemma.

▶ **Lemma 10.** *Let Y_n be a non-negative random variable satisfying*

$$\mathbb{P}(|Y_n - \mu_n| > g(n)) \leq f(n)$$

for some functions μ_n, f and g and assume that $Y_n \leq h(n)$ and $0 \leq \mu_n \leq h(n)$ for some function h. Then

$$\mathbb{P}\left(|Y_n^2 - \mu_n^2| > 2g(n)h(n)\right) \leq f(n).$$

4 Bienaymé–Galton–Watson trees

In this section, we discuss details of the proof in the case of conditioned Bienaymé–Galton–Watson trees. The main ingredient for the proof of Theorem 2 is the following.

▶ **Lemma 11.** *Let T be as in Theorem 2. Then we have*

$$\mathbb{E}\left(|T|P_E(T) - |L(T)|P(T)\right) \sim \frac{\phi_0}{\sigma^2}n^2$$

and

$$\mathbb{E}\left(\left(|T|P_E(T) - |L(T)|P(T)\right)^2\right) \sim \frac{\phi_0(2\phi_0 + (1-\phi_0)\sigma^2)}{3\sigma^4}n^4$$

as $n \to \infty$.

Proof. Aiming to extend (2) to include all quantities of interest, namely $|T|$, $|L(T)|$, $P(T)$ and $P_E(T)$, we consider the generating function

$$F(x, y, z, u) = \sum_T \omega(T)x^{|T|}y^{|L(T)|}z^{P_E(T)}u^{P(T)}.$$

This allows us to use the recursions (9), (10) and (11) for $|L(T)|$, $P(T)$ and $P_E(T)$ to get the following functional equation.

$$F(x, y, z, u) = x\Phi\left(F(xu, yz, z, u)\right) + \phi_0 x(y - 1). \tag{12}$$

Note here that the final term $\phi_0 x(y-1)$ accounts for the fact that the single vertex tree has one leaf.

Let us now introduce differential operators δ_x, δ_y, δ_z and δ_u as follows: for a function $A(x, y, z, u)$, we define

$$(\delta_x A)(x, y, z, u) = \left(x \cdot \frac{\partial}{\partial x}A\right)(x, y, z, u).$$

The operators δ_y, δ_z and δ_u are defined analogously. Note that these operators commute. For a polynomial $p = \sum_{i=0}^k c_i x^{x_i} y^{y_i} z^{z_i} u^{u_i}$, where $c_i \in \mathbb{R}$ and $x_i, y_i, z_i, u_i \in \mathbb{N}$ for all i, we define a new operator Δ_p, where p acts as an indicator of which differential operators to apply. We define

$$(\Delta_p A)(x,y,z,u) = \left(\sum_{i=0}^{k} c_i \delta_x^{x_i}\delta_y^{y_i}\delta_z^{z_i}\delta_u^{u_i} A\right)(x,y,z,u).$$

Now, we can express the generating functions for the expressions in our lemma by applying Δ_p to F for a suitable p. In particular, for $p_1 = xz - yu$ and $p_2 = (xz-yu)^2$,

$$(\Delta_{p_1}F)(x,1,1,1) = \sum_T \omega(T)\big(|T|P_E(T) - |L(T)|P(T)\big)x^{|T|}$$

and

$$(\Delta_{p_2}F)(x,1,1,1) = \sum_T \omega(T)\big(|T|P_E(T) - |L(T)|P(T)\big)^2 x^{|T|}.$$

This also motivates the definition of Δ_p, as we note that writing $\Delta_{p_2}F$ in terms of the various δ-operators would be somewhat cumbersome. To then get the expectations in the lemma we need to divide the coefficient of x^n in these generating functions by the coefficient of x^n in

$$G(x) := F(x,1,1,1) = \sum_T \omega(T)x^{|T|}.$$

As previously stated, $G(x) = x\Phi(G(x))$ and thus

$$\Phi\big(F(xu,yz,z,u)\big)\Big|_{y=z=u=1} = \Phi(G(x)) = \frac{G(x)}{x}.$$

Applying Δ_{p_i} to the right hand side of (12) allows us to express all generating functions in terms of G and its derivatives. Consider Example 12 for a more detailed calculation.

For $(\Delta_{p_1}F)(x,1,1,1)$, it is possible to perform the calculations by hand, but for the rest we use SageMath [30] for the somewhat lengthy symbolic computations. The resulting functions are explicit expressions in terms of G and its derivatives, but for readability we only write out one of them here.

$$(\Delta_{p_1}F)(x,1,1,1) = \frac{\phi_0 x^4 G'(x)^3}{G(x)^3} - \frac{\phi_0 x^3 G'(x)^2}{G(x)^2}.$$

Now we can use the analytic information on the generating function $G(x)$ that Lemma 1 provides: analyticity in the Δ-domain (3) carries over from G to its derivatives. Moreover, the equation $G(x) = x\Phi(G(x))$ ensures that $G(x)$ does not have any zeros other than $x=0$ in its domain of analyticity. Thus we know now that $(\Delta_{p_1}F)(x,1,1,1)$ is analytic in the same Δ-domain.

Now we recall that, close to $x=1$,

$$G(x) = 1 - \frac{\sqrt{2}}{\sigma}\sqrt{1-x} + o(|x-1|^{1/2}).$$

Such an asymptotic expansion carries over to derivatives of G, see [8, Section VI.10.1] for details. So taking derivatives and plugging into the expressions for $(\Delta_{p_1}F)(x,1,1,1)$ and $(\Delta_{p_2}F)(x,1,1,1)$ respectively, we get

$$(\Delta_{p_1}F)(x,1,1,1) = \frac{\sqrt{2}\phi_0}{4\,\sigma^3}(1-x)^{-\frac{3}{2}} + o\Big(|x-1|^{-\frac{3}{2}}\Big),$$
$$(\Delta_{p_2}F)(x,1,1,1) = \frac{5\sqrt{2}}{8\sigma^3}\left(-\frac{\phi_0^2}{2} + \frac{\phi_0}{2} + \frac{\phi_0^2}{\sigma^2}\right)(1-x)^{-\frac{7}{2}} + o\left(|x-1|^{-\frac{7}{2}}\right).$$

We can now use the Flajolet–Odlyzko singularity analysis [7] (see also the books [8] and [18]). This yields asymptotic formulas for the n-th coefficients of the generating functions, which we then divide by the n-th coefficient of G (whose asymptotics are obtained in the same way) to get the statement of the lemma. ◀

▶ **Example 12.** We illustrate the general procedure with a simple example. Suppose that we are interested in the expected number of leaves in a tree. To this end, we differentiate both sides of (12) with respect to y and evaluate both sides at $y = u = z = 1$ to get

$$F_y(x) = x\Phi'(G(x)) \cdot F_y(x) + x\phi_0,$$

where $F_y(x) := \left(\frac{\partial}{\partial y} F(x, y, z, u)\right)\big|_{y=u=z=1}$. Here, it was also used that

$$\left(\frac{\partial}{\partial y} F(xu, yz, z, u)\right)\Big|_{y=u=z=1} = \left(\frac{\partial}{\partial y} F(x, y, z, u)\right)\Big|_{y=u=z=1} = F_y(x).$$

Solving this for $F_y(x)$, we get

$$F_y(x) = \frac{x\phi_0}{1 - x\Phi'(G(x))}.$$

To continue, we return to (2), which we differentiate to get

$$1 - x\Phi'(G(x)) = \frac{G(x)}{xG'(x)}.$$

This can then be plugged into the above expression to end up with

$$F_y(x) = \frac{\phi_0 x^2 G'(x)}{G(x)}.$$

From Lemma 1, we know that

$$G(x) = 1 - \frac{\sqrt{2}}{\sigma}\sqrt{1-x} + o(|x-1|^{1/2}),$$

and by differentiation

$$G'(x) = \frac{\sqrt{2}}{2\sigma}(1-x)^{-1/2} + o(|1-x|^{-1/2}).$$

Plugging these in gives us

$$\begin{aligned} F_y(x) &= \phi_0 x^2 \frac{\frac{\sqrt{2}}{2\sigma}(1-x)^{-1/2} + o(|1-x|^{-1/2})}{1 - \frac{\sqrt{2}}{\sigma}\sqrt{1-x} + o(|x-1|^{1/2})} \\ &= \frac{\phi_0\sqrt{2}}{2\sigma}(1-x)^{-1/2} + o(|1-x|^{-1/2}). \end{aligned}$$

Now singularity analysis yields

$$[x^n]F_y(x) \sim \frac{\phi_0\sqrt{2}}{2\sigma}\frac{n^{-1/2}}{\Gamma(1/2)} = \frac{\phi_0\sqrt{2}}{2\sigma\sqrt{\pi}}n^{-1/2}$$

and

$$[x^n]G(x) \sim -\frac{\sqrt{2}}{\sigma}\frac{n^{-3/2}}{\Gamma(-1/2)} = \frac{\sqrt{2}}{2\sigma\sqrt{\pi}}n^{-3/2}.$$

So we can finally conclude that

$$\mathbb{E}(|L(T)|) = \frac{[x^n]F_y(x)}{[x^n]G(x)} \sim \phi_0 n.$$

In addition, we need a concentration result for the number of leaves, so that we can apply Lemma 8 and Lemma 10 to the quotient of the random variables $|T|P_E(T) - |L(T)|P(T)$ and $|L(T)|$. Its proof is given in the appendix.

▶ **Lemma 13.** *Let T denote a conditioned Bienaymé–Galton–Watson tree of size $|T| = n$ with offspring distribution ξ satisfying $\mathbb{E}\,\xi = 1$, $Var\xi = \sigma^2 < \infty$ and $\operatorname{span}(\xi) = 1$. The number of leaves $L(T)$ satisfies the concentration inequality*

$$\mathbb{P}(||L(T)| - \phi_0 n| > x) \leq \kappa\sqrt{n}e^{-\lambda x^2/n}$$

for certain positive constants κ and λ that only depend on ξ.

We now have the required ingredients to prove Theorem 2.

Proof of Theorem 2. We use Lemma 13 with $x = n^{3/4}$, which gives us

$$\mathbb{P}(||L(T)| - \phi_0 n| > n^{3/4}) \leq \kappa\sqrt{n}e^{-\lambda\sqrt{n}}.$$

Thus, we can apply Lemma 8, combined with the first part of Lemma 11, to $X_n = |T|P_E(T) - |L(T)|P(T)$ and $Y_n = |T||L(T)|$ with $f(n) = \kappa\sqrt{n}e^{-\lambda\sqrt{n}}$, $g(n) = n^{7/4}$, $h(n) = n^3$ and $\mu_n = \phi_0 n^2$. Boundedness of $\frac{\mathbb{E}(X_n)}{\mu_n}$ from below is clear, and boundedness of $\frac{\mathbb{E}(X_n)}{\mathbb{E}(|X_n|)}$ from below follows from the Cauchy–Schwarz inequality and Lemma 11. This readily proves the first statement (4) of the theorem. For the remaining parts of the theorem, we use Lemma 10 with the bounds $|L(T)| \leq n$ and $\phi_0 n \leq n$ to get

$$\mathbb{P}(||L(T)|^2 - \phi_0^2 n^2| > 2n^{7/4}) \leq \kappa\sqrt{n}e^{-\lambda\sqrt{n}}.$$

This allows us to apply Lemma 8 with appropriately chosen X_n and Y_n together with the second part of Lemma 11 to finish the proof of (5) in Theorem 2. Equations (4) and (5) show that the variance of the difference tends to a constant, and Chebyshev's inequality yields boundedness in probability. ◀

Proof of Corollary 4. Since

$$\frac{P_E(T)}{|L(T)|} - \frac{P(T)}{|T|}$$

is bounded in probability,

$$n^{-1/2}\Big(\frac{P_E(T)}{|L(T)|} - \frac{P(T)}{|T|}\Big) = \frac{P_E(T)}{\sqrt{n}|L(T)|} - \frac{P(T)}{n^{3/2}}$$

goes to 0 in probability. It follows from Slutsky's theorem that $\frac{P_E(T)}{\sqrt{n}|L(T)|}$ converges weakly to the same distribution as $\frac{P(T)}{n^{3/2}}$, which we already know to be a multiple of the Airy distribution. Since $\frac{|L(T)|}{n}$ converges in probability to ϕ_0 (a consequence of Lemma 13), a second application of Slutsky's theorem proves the statement. ◀

5 Increasing trees

We proceed along the same lines as in the previous section. The main lemma needed to prove Theorem 6 is

▶ **Lemma 14.** *Let T be as in Theorem 6. Then we have*

$$\mathbb{E}\left(|T|P_E(T) - |L(T)|P(T)\right) \sim \frac{1}{(\alpha+2)^2}n^2$$

and

$$\mathbb{E}\left(\left(|T|P_E(T) - |L(T)|P(T)\right)^2\right) \sim \frac{1}{(\alpha+2)^4}n^4$$

as $n \to \infty$.

Proof. Similarly to the Bienaymé–Galton–Watson case, we aim to extend (6), (7) and (8) to include all of the variables of interest. In general, we construct the exponential generating function

$$F(x,y,z,u) = \sum_T \omega(T)\frac{1}{|T|!}x^{|T|}y^{|L(T)|}z^{P_E(T)}u^{P(T)},$$

where the sum is over all rooted ordered increasing trees. For all three families, we can use the same recursions as before to get a differential equation

$$\frac{\partial}{\partial x}F(x,y,z,u) = \varphi(F(xu,yz,z,u)) + y - 1,$$

where the correction term $y-1$ accounts for the case of the single vertex tree and $\varphi(\cdot)$ is a function depending on the family. In particular, we get

$$\begin{aligned}
\frac{\partial}{\partial x}F_{\mathrm{r}}(x,y,z,u) &= \exp\left(F_{\mathrm{r}}(xu,yz,z,u)\right) + y - 1, \qquad (13)\\
\frac{\partial}{\partial x}F_d(x,y,z,u) &= \left(1 + F_d(xu,yz,z,u)\right)^d + y - 1,\\
\frac{\partial}{\partial x}F_\alpha(x,y,z,u) &= \left(1 - F_\alpha(xu,yz,z,u)\right)^{-\alpha} + y - 1.
\end{aligned}$$

The subscript of F indicates the type of tree, with r indicating recursive trees, d indicating d-ary trees and α indicating GPORTs. Using the same notation as previously, we get that

$$\begin{aligned}
(\Delta_{p_1}F)(x,1,1,1) &= \sum_T \omega(T)\big(|T|P_E(T) - |L(T)|P(T)\big)\frac{x^{|T|}}{|T|!},\\
(\Delta_{p_2}F)(x,1,1,1) &= \sum_T \omega(T)\big(|T|P_E(T) - |L(T)|P(T)\big)^2\frac{x^{|T|}}{|T|!}.
\end{aligned}$$

Dividing the n-th coefficient in the respective generating functions by the corresponding n-th coefficient in G yields the expected values of interest. Note here that the factor $\frac{1}{|T|!}$ is automatically cancelled. To illustrate the calculations, Example 15 considers the expected number of leaves in a random recursive tree. Finding other derivatives of F is done in a similar manner, which always involves solving an ODE of the form

$$f'(x) = a(x)f(x) + b(x),$$

where $a(x)$ and $b(x)$ are functions dependent on lower order derivatives of F resulting from the iterative calculation of inner derivatives and applications of the product rule. In fact, we always have $a(x) = \varphi'(G(x))$. Such an ODE has the general solution

$$f(x) = Ce^{A(x)} + e^{A(x)}\int b(x)e^{-A(x)}dx,$$

where $A(x)$ is any antiderivative of $a(x)$. Carrying out these calculations[2] in SageMath [30], we get explicit functions in all these cases, but the expressions are quite large and also require some extra case distinctions. Using error terms to increase readability, we have

$$(\Delta_{p_1} F_{\mathrm{r}})(x,1,1,1) = \frac{1}{4}(1-x)^{-2} + O(\log(1-x)),$$
$$(\Delta_{p_2} F_{\mathrm{r}})(x,1,1,1) = \frac{3}{8}(1-x)^{-4} + O(\log(1-x)(1-x)^{-3}),$$

for recursive trees. Similarly,

$$(\Delta_{p_1} F_d)(x,1,1,1) = \frac{(\beta-1)^2(\beta)_2}{(\beta-2)^2} t^{\beta-2} + O(t^{\beta-1}\log t),$$
$$(\Delta_{p_2} F_d)(x,1,1,1) = \frac{(\beta-1)^4(\beta)_4}{(\beta-2)^4} t^{\beta-4} + O(t^{\beta-3}\log t),$$

for d-ary trees with $\beta = \frac{1}{1-d}$ and $t = 1-(d-1)x$, where $(\cdot)_k$ denotes the falling factorial. In the final case of GPORTs, we get, with $\beta = \frac{\alpha}{1+\alpha}$ and $t = 1-(1+\frac{1}{\alpha})x$,

$$(\Delta_{p_1} F_\alpha)(x,1,1,1) = -\frac{(\beta-1)^2(\beta)_2}{(\beta-2)^2} t^{\beta-2} + O(t^{\beta-1}\log t),$$
$$(\Delta_{p_2} F_\alpha)(x,1,1,1) = -\frac{(\beta-1)^4(\beta)_4}{(\beta-2)^4} t^{\beta-4} + O(t^{\beta-3}\log t).$$

Consider the latter pair for example. These generating functions have coefficients

$$[x^n](\Delta_{p_1} F_\alpha)(x,1,1,1) \sim -\frac{(\beta-1)^2(\beta)_2}{(\beta-2)^2}(1+1/\alpha)^n \frac{n^{-\beta+1}}{\Gamma(-\beta+2)},$$
$$[x^n](\Delta_{p_2} F_\alpha)(x,1,1,1) \sim -\frac{(\beta-1)^4(\beta)_4}{(\beta-2)^4}(1+1/\alpha)^n \frac{n^{-\beta+3}}{\Gamma(-\beta+4)}.$$

Dividing by

$$[x^n]G_\alpha(x) \sim -(1+1/\alpha)^n \frac{n^{-\beta-1}}{\Gamma(-\beta)}$$

proves the lemma, using the fact that $\Gamma(-\beta+k) = (-1)^k(\beta)_k\Gamma(-\beta)$ for any positive integer k. At the end, everything can be expressed in terms of α as $\beta = \frac{\alpha}{1+\alpha}$. Similar calculations also apply to recursive trees (with $\alpha = \beta = 0$) and d-ary increasing trees (with $\alpha = -\frac{1}{d}$ and $\beta = \frac{1}{1-d}$). ◀

▶ **Example 15.** Again, we illustrate the general procedure with a simple example. Suppose we aim to calculate the expected number of leaves in a random recursive tree. To do this, we differentiate both sides of (13) with respect to y and evaluate both sides at $y = z = u = 1$ to get

$$\frac{\partial}{\partial x} f(x) = 1 + f(x) \cdot \exp(G_{\mathrm{r}}(x)),$$

[2] Once again, all code is available at the following repository: `https://github.com/JacobLundblad/ExternalPathLengthAndTerminalWienerIndex`.

where $f(x) := \left(\frac{\partial}{\partial y}F_{\mathrm{r}}(x,y,z,u)\right)\big|_{y=z=u=1}$ and we use that $\left(\frac{\partial}{\partial y}F_{\mathrm{r}}(xu,yz,z,u)\right)\big|_{y=z=u=1} = f(x)$. Plugging in $G_{\mathrm{r}}(x) = \log\left(\frac{1}{1-x}\right)$, we get the first order linear ODE

$$f'(x) = 1 + f(x)\cdot(1-x)^{-1},$$

which is solved by

$$f(x) = C(1-x)^{-1} + (1-x)^{-1}\left(x - \frac{x^2}{2}\right).$$

To find the constant C, we consider first that $f(0) = C$. Since the constant coefficient of F_{r}, and thus of f, is 0, we have $C = 0$ and thus

$$\begin{aligned} f(x) &= (1-x)^{-1}\left(x - \frac{x^2}{2}\right) \\ &= \frac{1}{2}(1-x)^{-1} - \frac{1}{2}(1-x). \end{aligned}$$

The asymptotic behaviour of the coefficients of f is thus determined by the term $\frac{1}{2}(1-x)^{-1}$, which is the dominating term around the singularity of the function. Using singularity analysis, we get that

$$[x^n]f(x) \sim \frac{1}{2}\frac{n^0}{\Gamma(1)} = \frac{1}{2}.$$

In fact, this holds with equality for $n > 1$, which however is not needed for our purposes. To get the expected number of leaves, we then have to multiply by n (since $[x^n]G_{\mathrm{r}}(x) = \frac{1}{n}$), which yields the result

$$\mathbb{E}(|L(T)|) \sim \frac{n}{2}.$$

Once again, we also need a concentration inequality for the number of leaves. Specifically, we have the following lemma, whose proof is given in the appendix. In the special case of recursive trees, such a concentration result was proven by Zhang in [31].

▶ **Lemma 16.** *Let T be a random linear preferential attachment tree $\mathcal{T}_{\alpha,n}$ of size $|T| = n$ with $\alpha \in \{\dots, -\frac{1}{3}, -\frac{1}{2}\} \cup [0,\infty)$. The number of leaves $L(T)$ satisfies the concentration inequality*

$$\mathbb{P}\Big(\Big|L(T) - \frac{1+\alpha}{2+\alpha}n\Big| > x\Big) \le \kappa e^{-\lambda x^2/n}$$

for certain positive constants κ and λ that only depend on α.

Now the proof of Theorem 6 follows exactly the same lines as the proof of Theorem 2, with Lemma 14 and Lemma 16 playing the roles of Lemma 11 and Lemma 13, respectively.

Proof of Corollary 7. We can argue in a similar fashion as in the proof of Corollary 4: by Lemma 14,

$$\frac{nP_E(T) - |L(T)|P(T)}{n^2} \xrightarrow{P} \frac{1}{(\alpha+2)^2}$$

and

$$\mathbb{E}\Big(\frac{nP_E(T) - |L(T)|P(T)}{n^2}\Big) \to \frac{1}{(\alpha+2)^2}$$

as $n \to \infty$. Consequently, the difference of the two, which can be written as

$$\frac{P_E(T) - \mathbb{E}(P_E(T))}{n} - \frac{|L(T)|P(T) - \mathbb{E}(|L(T)|P(T))}{n^2},$$

converges in probability to 0. So by Slutsky's theorem, it suffices to show that the random variable $\frac{|L(T)|P(T) - \mathbb{E}(|L(T)|P(T))}{n^2}$ has the desired limiting distribution. We split it into three parts:

$$\begin{aligned} \frac{|L(T)|P(T) - \mathbb{E}(|L(T)|P(T))}{n^2} = &\frac{|L(T)|(P(T) - \mathbb{E}(P(T)))}{n^2} \\ &+ \frac{(|L(T)| - \mathbb{E}(|L(T)|))\mathbb{E}(P(T))}{n^2} \\ &+ \frac{\mathbb{E}(|L(T)|)\mathbb{E}(P(T)) - \mathbb{E}(|L(T)|P(T))}{n^2}. \end{aligned} \tag{14}$$

Now, we know that $\frac{P(T) - \mathbb{E}(P(T))}{n}$ has limiting distribution X_α, and $\frac{|L(T)|}{n}$ goes to $\frac{1+\alpha}{2+\alpha}$ in probability by the concentration property in Lemma 16. So it only remains to show that the second and third term go to 0 in probability, so that another application of Slutsky's theorem yields the statement:

- Observe that $\frac{|L(T)| - \mathbb{E}(|L(T)|)}{n^{1/2+\epsilon}} \xrightarrow{P} 0$ for every $\epsilon > 0$ by Lemma 16. Since we know that $\mathbb{E}(P(T)) = O(n \log n)$, the second term on the right hand side of (14) goes to 0 in probability.
- Finally, one can show by a direct computation or by another application of Lemma 16 that $\mathbb{E}(|L(T)|)\mathbb{E}(P(T)) - \mathbb{E}(|L(T)|P(T)) = o(n^2)$, which shows that the last term in (14) also goes to 0.

Putting everything together, we obtain the statement of the corollary. ◀

6 Further extensions

The methods used in this paper can be adapted to other tree functionals, provided that they have a similar recursive decomposition that allows for the extension of the univariate generating function and the corresponding functional or differential equation. In particular, one can obtain similar results for the Wiener index (the sum of all distances between pairs of vertices) and the terminal Wiener index (the sum of all distances between pairs of leaves) with a similar analysis, albeit with more variables and more complicated expressions.

One may also expect similar phenomena for other families of random trees beyond those considered here. One such example is the family of unlabelled trees, which also has a generating function satisfying a functional equation and should therefore be amenable to the same type of multivariate analysis.

References

1 David Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991. doi:10.1017/CBO9780511662980.003.

2 François Bergeron, Philippe Flajolet, and Bruno Salvy. Varieties of increasing trees. In *CAAP '92 (Rennes, 1992)*, volume 581 of *Lecture Notes in Comput. Sci.*, pages 24–48. Springer, Berlin, 1992. doi:10.1007/3-540-55251-0_2.

3 Nicolas Broutin and Cecilia Holmgren. The total path length of split trees. *Ann. Appl. Probab.*, 22(5):1745–1777, 2012. doi:10.1214/11-AAP812.

4 Robert P. Dobrow and James Allen Fill. Total path length for random recursive trees. *Combin. Probab. Comput.*, 8(4):317–333, 1999. doi:10.1017/S0963548399003855.

5 Robert P. Dobrow and Robert T. Smythe. Poisson approximations for functionals of random trees. *Random Structures Algorithms*, 9(1-2):79–92, 1996. doi:10.1002/(sici)1098-2418(199608/09)9:1/2<79::aid-rsa5>3.0.co;2-8.

6 Michael Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability. doi:10.1007/978-3-211-75357-6.

7 Philippe Flajolet and Andrew Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3(2):216–240, 1990. doi:10.1137/0403019.

8 Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009. doi:10.1017/CBO9780511801655.

9 B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1968.

10 Svante Janson. Random cutting and records in deterministic and random trees. *Random Structures Algorithms*, 29(2):139–179, 2006. doi:10.1002/rsa.20086.

11 Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012. doi:10.1214/11-PS188.

12 M. Javanian and M. G. Vahidi-Asl. External path length of random m-oriented recursive. *J. Natur. Sci. Math.*, 50(1-2):11–18, 2010.

13 Mehri Javanian. On the external path length of random recursive k-ary trees. *Ital. J. Pure Appl. Math.*, 31:21–26, 2013.

14 Hosam M. Mahmoud. Limiting distributions for path lengths in recursive trees. *Probab. Engrg. Inform. Sci.*, 5(1):53–59, 1991. doi:10.1017/S0269964800001881.

15 Hosam M. Mahmoud. Distances in random plane-oriented recursive trees. *J. Comput. Appl. Math.*, 41(1-2):237–245, 1992. doi:10.1016/0377-0427(92)90252-S.

16 Jean-François Marckert and Abdelkader Mokkadem. The depth first processes of Galton-Watson trees converge to the same Brownian excursion. *Ann. Probab.*, 31(3):1655–1678, 2003. doi:10.1214/aop/1055425793.

17 A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian J. Math.*, 30(5):997–1015, 1978. doi:10.4153/CJM-1978-085-0.

18 Stephen Melczer. *An invitation to analytic combinatorics—from one to several variables*. Springer, Cham, 2021. doi:10.1007/978-3-030-67080-1.

19 J. W. Moon. The distance between nodes in recursive trees. In *Combinatorics (Proc. British Combinatorial Conf., Univ. Coll. Wales, Aberystwyth, 1973)*, volume No. 13 of *London Math. Soc. Lecture Note Ser.*, pages 125–132. Cambridge Univ. Press, London-New York, 1974.

20 Götz Olaf Munsonius and Ludger Rüschendorf. Limit theorems for depths and distances in weighted random b-ary recursive trees. *J. Appl. Probab.*, 48(4):1060–1080, 2011. doi:10.1017/s0021900200008627.

21 Alois Panholzer and Helmut Prodinger. Level of nodes in increasing trees revisited. *Random Structures & Algorithms*, 31(2):203–226, 2007. doi:10.1002/rsa.20161.

22 Mireille Régnier. A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.*, 23(3):335–343, 1989. doi:10.1051/ita/1989230303351.

23 Uwe Rösler. A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.*, 25(1):85–100, 1991. doi:10.1051/ita/1991250100851.

24 Henning Sulzbach. On martingale tail sums for the path length in random trees. *Random Structures Algorithms*, 50(3):493–508, 2017. doi:10.1002/rsa.20674.

25 Jerzy Szymański. On a nonuniform random recursive tree. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 297–306. North-Holland, Amsterdam, 1987.

26 Jerzy Szymański. On the complexity of algorithms on recursive trees. *Theoret. Comput. Sci.*, 74(3):355–361, 1990. doi:10.1016/0304-3975(90)90084-U.

27 Lajos Takács. On the total heights of random rooted trees. *J. Appl. Probab.*, 29(3):543–556, 1992. doi:10.1017/s0021900200043370.

28 Lajos Takács. The asymptotic distribution of the total heights of random rooted trees. *Acta Sci. Math. (Szeged)*, 57(1-4):613–625, 1993.

29 M. A. Tapia and B. R. Myers. Generation of concave node-weighted trees. *IEEE Trans. Circuit Theory*, CT-14:229–230, 1967.

30 The Sage Developers. *SageMath, the Sage Mathematics Software System*, 2022. DOI 10.5281/zenodo.6259615. URL: https://www.sagemath.org.

31 Yazhe Zhang. On the number of leaves in a random recursive tree. *Braz. J. Probab. Stat.*, 29(4):897–908, 2015. doi:10.1214/14-BJPS252.

A Further proof details

In this appendix, we provide the proofs of some technical lemmas that do not fit the main body of the paper.

Proof of Lemma 8. Consider the event $A = \{|Y_n - \mu_n| > g(n)\}$, and split the expectation as follows:

$$\mathbb{E}\left(\frac{X_n}{Y_n}\right) = \mathbb{E}\left(\frac{X_n}{Y_n} \,\Big|\, A\right)\mathbb{P}(A) + \mathbb{E}\left(\frac{X_n}{Y_n} \,\Big|\, \bar{A}\right)\mathbb{P}(\bar{A}). \tag{15}$$

Starting with the first term, we can use the bounds for X_n and Y_n to obtain

$$\mathbb{E}\left(\Big|\frac{X_n}{Y_n}\Big| \,\Big|\, A\right) \le \mathbb{E}\left(|X_n| | A\right) \le h(n).$$

Thus, we have

$$\left|\mathbb{E}\left(\frac{X_n}{Y_n} \,\Big|\, A\right)\mathbb{P}(A)\right| \le \mathbb{E}\left(\Big|\frac{X_n}{Y_n}\Big| \,\Big|\, A\right) \cdot \mathbb{P}(A) \le h(n) \cdot f(n) = o(1).$$

For the second term, we note that the expectation of X_n can be written as

$$\mathbb{E}(X_n) = \mathbb{E}(X_n|A)\mathbb{P}(A) + \mathbb{E}(X_n|\bar{A})\mathbb{P}(\bar{A}),$$

and so

$$\mathbb{E}(X_n|\bar{A}) = \frac{\mathbb{E}(X_n) - \mathbb{E}(X_n|A)\mathbb{P}(A)}{\mathbb{P}(\bar{A})}.$$

Using the bounds on X_n and $\mathbb{P}(A)$, we get that

$$\mathbb{E}(X_n|\bar{A}) = \frac{\mathbb{E}(X_n)}{\mathbb{P}(\bar{A})} + O\left(\frac{h(n) \cdot f(n)}{1 - f(n)}\right) \sim \mathbb{E}(X_n),$$

since $\mathbb{P}(\bar{A}) \xrightarrow{n\to\infty} 1$ by the assumption that $f(n) = o(1)$. Likewise, we have

$$\mathbb{E}\big(|X_n| \,\big|\, \bar{A}\big) \sim \mathbb{E}(|X_n|).$$

Under the condition $\bar{A}$, we have $Y_n = \mu_n + O(g(n))$. Thus, it follows that

$$\begin{aligned}
\mathbb{E}\left(\frac{X_n}{Y_n} \,\Big|\, \bar{A}\right) &= \mathbb{E}\left(\frac{X_n}{\mu_n + O(g(n))} \,\Big|\, \bar{A}\right) \\
&= \mathbb{E}\left(\frac{X_n}{\mu_n} + O\Big(\frac{|X_n| g(n)}{\mu_n^2}\Big) \,\Big|\, \bar{A}\right) \\
&= \frac{\mathbb{E}(X_n \,|\, \bar{A})}{\mu_n} + O\Big(\frac{\mathbb{E}\big(|X_n| \,\big|\, \bar{A}\big) g(n)}{\mu_n^2}\Big) \\
&\sim \frac{\mathbb{E}(X_n)}{\mu_n}.
\end{aligned}$$

Combining (15) with our estimates for $\mathbb{E}\left(\frac{X_n}{Y_n}\,\middle|\,A\right)$ and $\mathbb{E}\left(\frac{X_n}{Y_n}\,\middle|\,\bar{A}\right)$ finally gives us

$$\mathbb{E}\left(\frac{X_n}{Y_n}\right) \sim \frac{\mathbb{E}(X_n)}{\mu_n},$$

which proves the lemma. ◀

Proof of Lemma 10. We have

$$\begin{aligned} f(n) &> \mathbb{P}(|Y_n - \mu_n| > g(n)) \\ &= \mathbb{P}(2h(n)|Y_n - \mu_n| > 2g(n)h(n)) \\ &\geq \mathbb{P}\left((Y_n + \mu_n)|Y_n - \mu_n| > 2g(n)h(n)\right) \\ &= \mathbb{P}\left(|Y_n^2 - \mu_n^2| > 2g(n)h(n)\right). \end{aligned}$$ ◀

Proof of Lemma 13. We take the viewpoint of the Bienaymé–Galton–Watson process. Let ξ_i be the offspring of the i-th vertex, so that $\xi_1, \xi_2, \ldots$ are independent random variables with the same distribution ξ. Conditioning on the event that $S_n = \sum_{i=1}^n \xi_i = n-1$ while $S_k \geq k$ for $k < n$, we get a conditioned Bienaymé–Galton–Watson tree of size n. Note that the leaves are precisely the vertices for which $\xi_i = 0$. By the cycle lemma, we can equivalently condition on $S_n = \sum_{i=1}^n \xi_i = n-1$ only, since we are only interested in the number of ξ_i that are equal to 0.

Define $\hat{\xi}_i$ as ξ_i conditioned on being non-zero, and set $\hat{S}_{n-l} = \sum_{i=1}^{n-l} \hat{\xi}_i$. Then we have

$$\begin{aligned} \mathbb{P}(|L(T)| = l) &= \mathbb{P}(\text{exactly } l \text{ of the } \xi_i \text{ are } 0 \,|\, S_n = n-1) \\ &= \frac{\mathbb{P}(S_n = n-1, \text{exactly } l \text{ of the } \xi_i \text{ are } 0)}{\mathbb{P}(S_n = n-1)} \\ &= \binom{n}{l} \frac{\mathbb{P}(S_{n-l} = n-1, \xi_i \neq 0 \text{ for } i \leq n-l,\, \xi_i = 0 \text{ for } i > n-l)}{\mathbb{P}(S_n = n-1)} \\ &= \binom{n}{l} \phi_0^l \frac{\mathbb{P}(S_{n-l} = n-1, \xi_i \neq 0 \text{ for } i \leq n-l)}{\mathbb{P}(S_n = n-1)} \\ &= \binom{n}{l} \phi_0^l (1-\phi_0)^{n-l} \frac{\mathbb{P}(S_{n-l} = n-1 \,|\, \xi_i \neq 0 \text{ for } i \leq n-l)}{\mathbb{P}(S_n = n-1)} \\ &= \binom{n}{l} \phi_0^l (1-\phi_0)^{n-l} \frac{\mathbb{P}(\hat{S}_{n-l} = n-1)}{\mathbb{P}(S_n = n-1)} \\ &\leq \mathbb{P}(\mathrm{Bi}(n, \phi_0) = l) \cdot c\sqrt{n} \end{aligned}$$

for some constant c, where $\mathrm{Bi}(n, \phi)$ denotes the binomial distribution with parameters n and ϕ. The last inequality bounds the numerator by 1 and uses the classical local central limit theorem for discrete distributions, see e.g. [9, Chapter 9]. Thus we get

$$\begin{aligned} \mathbb{P}(||L(T)| - \phi_0 n| > x) &= \sum_{l=1}^{n-1} \mathbb{P}(|L(T)| = l) I_{|l - \phi_0 n| > x} \\ &\leq c\sqrt{n} \sum_{l=1}^{n-1} \mathbb{P}(\mathrm{Bi}(n, \phi_0) = l) I_{|l - \phi_0 n| > x} \\ &= c\sqrt{n}\, \mathbb{P}(|\mathrm{Bi}(n, \phi_0) - \phi_0 n| > x). \end{aligned}$$

Combining this with the Chernoff bound

$$\mathbb{P}(|\mathrm{Bi}(n, \phi_0) - \phi_0 n| \geq \delta \phi_0 n) \leq 2e^{-\phi_0 n \frac{\delta^2}{3}},$$

which is valid for any $\delta \in (0,1)$, yields the desired concentration inequality, proving the lemma. ◀

Proof of Lemma 16. Let us consider the growth process that attaches a new leaf in each step as described in Section 2. Suppose that L_k is the number of leaves after k steps. Then we have (deterministically) $L_1 = L_2 = 1$, and the random variable $L(T)$ we are interested in is simply L_n. After k steps, there are L_k leaves and $k - L_k$ internal vertices. By definition of the process, the probability to attach to a fixed leaf in the $(k+1)$-th step is $\frac{1}{\alpha(k-1)+k}$. Thus, the $(k+1)$-th vertex is attached to a leaf with probability $\frac{L_k}{\alpha(k-1)+k}$, and to an internal vertex with probability $1 - \frac{L_k}{\alpha(k-1)+k}$. In the former case, the number of leaves stays the same, in the latter case it increases by 1. Hence we have

$$\mathbb{E}\big(L_{k+1} \,\big|\, L_k\big) = L_k + 1 - \frac{L_k}{\alpha(k-1)+k} = \frac{(1+\alpha)(k-1)}{\alpha(k-1)+k} L_k + 1 = \frac{k-1}{k - \frac{\alpha}{1+\alpha}} L_k + 1.$$

We multiply by $\frac{\Gamma(k+1/(1+\alpha))}{\Gamma(k)} = \frac{\Gamma(k+1-\alpha/(1+\alpha))}{\Gamma(k)}$ to find

$$\mathbb{E}\Big(\frac{\Gamma(k+1/(1+\alpha))}{\Gamma(k)} L_{k+1} \,\Big|\, L_k\Big) = \frac{\Gamma(k-\alpha/(1+\alpha))}{\Gamma(k-1)} L_k + \frac{\Gamma(k+1/(1+\alpha))}{\Gamma(k)}.$$

Thus, if we set

$$X_k = \frac{\Gamma(k-\alpha/(1+\alpha))}{\Gamma(k-1)} L_k - \sum_{j=1}^{k-1} \frac{\Gamma(j+1/(1+\alpha))}{\Gamma(j)}$$

for $k \geq 2$, then X_k is a martingale with $X_2 = 0$. Note also that

$$X_{k+1} - X_k = \begin{cases} \frac{\Gamma(k-\alpha/(1+\alpha))L_k}{(1+\alpha)\Gamma(k)} & L_{k+1} = L_k + 1, \\ \frac{\Gamma(k-\alpha/(1+\alpha))(L_k-(1+\alpha)k+\alpha)}{(1+\alpha)\Gamma(k)} & L_{k+1} = L_k, \end{cases}$$

thus (as $L_k \leq k$) we have $|X_{k+1} - X_k| = O(k^{1/(1+\alpha)})$ by Stirling's formula. In other words, there exists a constant C such that $|X_{k+1} - X_k| \leq Ck^{1/(1+\alpha)}$ for all $k \geq 2$. By the Azuma–Hoeffding inequality, we have, with

$$d_n^2 = \sum_{j=2}^{n-1} C^2 j^{2/(1+\alpha)} = O\big(n^{1+2/(1+\alpha)}\big),$$

that

$$\mathbb{P}(|X_n| > t) \leq 2e^{-t^2/(2d_n^2)}$$

for all $t \geq 0$. Let us write $a_n = \frac{\Gamma(n-\alpha/(1+\alpha))}{\Gamma(n-1)}$ and $b_n = \sum_{j=1}^{n-1} \frac{\Gamma(j+1/(1+\alpha))}{\Gamma(j)}$, so that $X_n = a_n L_n - b_n$. Note that (again by Stirling's formula)

$$a_n = \frac{\Gamma(n-\alpha/(1+\alpha))}{\Gamma(n-1)} = n^{1/(1+\alpha)}(1 + O(n^{-1}))$$

and thus

$$b_n = \sum_{j=1}^{n-1} \frac{\Gamma(j+1/(1+\alpha))}{\Gamma(j)} = \frac{1+\alpha}{2+\alpha} n^{(2+\alpha)/(1+\alpha)} + O\big(n^{1/(1+\alpha)}\big) = \frac{1+\alpha}{2+\alpha} a_n (n + O(1)).$$

Putting everything together, we obtain the estimate

$$\begin{aligned}
\mathbb{P}\Big(\Big|L_n - \frac{1+\alpha}{2+\alpha}n\Big| > x\Big) &= \mathbb{P}\Big(\Big|a_n L_n - \frac{1+\alpha}{2+\alpha}na_n\Big| > a_n x\Big) \\
&\le \mathbb{P}\Big(\Big|a_n L_n - b_n\Big| > a_n x - \Big|b_n - \frac{1+\alpha}{2+\alpha}na_n\Big|\Big) \\
&= \mathbb{P}\Big(|X_n| > a_n x - \Big|b_n - \frac{1+\alpha}{2+\alpha}na_n\Big|\Big) \\
&\le 2\exp\Big(-\frac{(a_n x - |b_n - \frac{1+\alpha}{2+\alpha}na_n|)^2}{2d_n^2}\Big) \\
&= 2\exp\Big(-\frac{a_n^2(x - O(1))^2}{2d_n^2}\Big),
\end{aligned}$$

which is valid at least for $x > \Big|\frac{b_n}{a_n} - \frac{1+\alpha}{2+\alpha}n\Big| = O(1)$. Since $\frac{d_n^2}{a_n^2} = O(n)$, it follows that there exist constants $\kappa, \lambda > 0$ such that

$$\mathbb{P}\Big(\Big|L_n - \frac{1+\alpha}{2+\alpha}n\Big| > x\Big) \le \kappa e^{-\lambda x^2/n}$$

holds for all n and all $x > 0$, which proves the lemma. ◀