RANDOM TREES HAVE HEIGHT $O(\sqrt{n})$

LOUIGI ADDARIO-BERRY AND SERTE DONDERWINKEL

Abstract. We obtain new non-asymptotic tail bounds for the height of uniformly random trees with a given degree sequence, simply generated trees and conditioned Bienaymé trees (the family trees of branching processes). In particular, we settle two conjectures of Janson [10], two conjectures of Addario-Berry, Devroye and Janson [2], and a conjecture of Addario-Berry [1]. Moreover, we define a partial ordering on degree sequences and show that it induces a stochastic ordering on the heights of uniformly random trees with given degree sequences. The latter result settles a conjecture of Addario-Berry, Donderwinkel, Maazoun and Martin [4], and can also be used to show that sub-binary random trees are stochastically the tallest trees with a given number of vertices and leaves.

Our proofs are based in part on a bijection introduced by Addario-Berry, Donderwinkel, Maazoun and Martin [4], which can be recast to provide a line-breaking construction of random trees with given vertex degrees.

1. Introduction

This paper proves optimal asymptotic and non-asymptotic tail bounds on the heights of random trees from several natural and well-studied classes, in the process proving conjectures and answering questions from [1, 2, 10]. We begin by stating our results for the *Bienaymé* trees, which constitute the most elementary and well-studied random graph model and the oldest stochastic model for studying population growth.

For $\mu = (\mu_k, k \geq 0)$ a probability distribution on $\mathbb{N} = \mathbb{Z} \cap [0, \infty)$, a Bienaymé tree with offspring distribution μ, denoted T_μ, is the family tree of a branching process with offspring distribution μ.[1] (An example appears in Figure 1.) We write $|\mathrm{T}_\mu|$ for the size (number of vertices) of T_μ. For $n \in \mathbb{N}$ such that $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$, we write $\mathrm{T}_{\mu,n}$ to denote a Bienaymé tree with offspring distribution μ conditioned to have size n. Bienaymé trees are random *plane* trees: rooted, unlabeled trees in which the set of children of each node is endowed with a (left-to-right) total order.

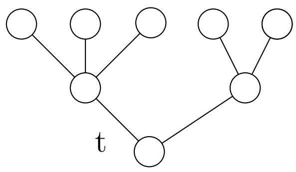


Figure 1. In the tree t, there are 5 vertices with no children, two vertices with two children and one vertex with 3 children, so for $\mu = (\mu_k, k \geq 0)$ a probability distribution on $\mathbb{N}$, we have that $\mathbf{P}(\mathrm{T}_\mu = \mathrm{t}) = \mu_0^5 \mu_2^2 \mu_3$.

For $p > 0$, write $|\mu|_p = (\sum_{k=0}^{\infty} k^p \mu_k)^{1/p}$.

Theorem 1. *Fix a probability distribution μ supported on $\mathbb{N}$ with $|\mu|_1 \leq 1$ and $|\mu|_2 = \infty$. Then, $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ as $n \to \infty$ along values n such that $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$, both in probability and in expectation.*

Date: January 27, 2022.

2010 *Mathematics Subject Classification.* 60C05,60J80,05C05.

Key words and phrases. Random trees, Bienaymé trees, Galton–Watson trees, simply generated trees, height.

[1]Bienaymé trees are often referred to as Galton–Watson trees, but we adopt the change in terminology suggested in [3].

Theorem 1 answers an open question from [2].

Theorem 2. *Fix a probability distribution $\mu = (\mu_k, k \geq 0)$ supported on $\mathbb{N}$ with $|\mu|_1 < 1$ and with $\sum_{c \geq 0} e^{tk}\mu_k = \infty$ for all $t > 0$. Then, $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ as $n \to \infty$ over all n such that $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$, both in probability and in expectation.*

Theorem 2 solves a slight variant of Problem 21.7 from [10]. (That problem is stated for simply generated trees, which are a slight generalization of Bienaymé trees. We in fact solve the problem from [10] in full; see Theorem 19, below.)

A collection $(X_i)_{i \in I}$ of random variables is *sub-Gaussian* if there exist constants $c, C > 0$ such that $\mathbf{P}(X_i \geq x) \leq C\exp(-cx^2)$ for all $i \in I$ and all $x > 0$. Our third theorem states that for any offspring distribution $\mu = (\mu_k, k \geq 0)$, the family of random variables $(n^{-1/2}\mathrm{ht}(T_{\mu,n}))$ has sub-Gaussian tails, with constants c, C that only depend on μ_0 and μ_1.

Theorem 3. *Fix $\epsilon \in [0, 1)$. Then, there exist constants $c = c(\epsilon)$ and $C = C(\epsilon)$ such that the following holds. For any probability distribution $\mu = (\mu_k, k \geq 0)$ on $\mathbb{N}$ with $\mu_0 + \mu_1 < 1 - \epsilon$ and $\mu_0/(\mu_0 + \mu_1) > \epsilon$, for all n sufficiently large and for which $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$, for all $x > 0$,*

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mu,n}) > xn^{1/2}\right) \leq C\exp(-cx^2).$$

This theorem has the following immediate corollary.

Corollary 4. *For any probability distribution $\mu = (\mu_k, k \geq 0)$ on $\mathbb{N}$ with $\mu_0 + \mu_1 < 1$, $\mathbf{E}[\mathrm{ht}(\mathrm{T}_{\mu,n})] = O(n^{1/2})$, and, more generally, for any fixed $r < \infty$, $\mathbf{E}[\mathrm{ht}(\mathrm{T}_{\mu,n})^r] = O(n^{r/2})$ as $n \to \infty$ over all n such that $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$.* □

Theorem 3 and Corollary 4 strengthen Theorem 1.2 and Corollary 1.3 in [2], as they are not restricted to critical offspring distributions with finite variance and the bounds only depend on μ via μ_0 and μ_1. This resolves conjectures stated in [1] and [2].

The requirement in Theorem 3 that n is sufficiently large (where "sufficiently large" depends on μ) is necessary. To see this, note that if μ has support $\{0, 1, N\}$ then for any $n < N + 1$, with probability one $\mathrm{T}_{\mu,n}$ is a path (so has height $n - 1$). The requirement in Theorem 3 that $1 - \mu_0 - \mu_1$ and $\mu_0/(\mu_0 + \mu_1)$ be bounded from below is also necessary. To see this, it suffices to consider probability distributions μ of the form

$$\mu_0 = q(1-p) \qquad \mu_1 = (1-q)(1-p) \qquad \mu_2 = p\,.$$

For any $x > 1$, it is possible to make $\liminf_{n\to\infty} \mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mu,n}) > xn^{1/2})$ arbitrarily close to one by either taking p fixed and q sufficiently small or q fixed and p sufficiently small; this fact is proved as Claim 23, below.

For random variables X, Y, we write $X \preceq_{\mathrm{st}} Y$ to mean that Y stochastically dominates X, which is to say that $\mathbf{P}(X \leq t) \geq \mathbf{P}(Y \leq t)$ for all $t \in \mathbb{R}$. We also write $X \prec_{\mathrm{st}} Y$, and say that Y strictly stochastically dominates X, if $X \preceq_{\mathrm{st}} Y$ and X and Y are not identically distributed. Our final result for Bienaymé trees states that, among all conditioned Bienaymé trees which almost surely have no vertices with exactly 1 child, the binary Bienaymé trees have the stochastically largest heights.

Theorem 5. *Fix any probability distribution $\mu = (\mu_k, k \geq 0)$ on $\mathbb{N}$ with $\mu_0 \in (0, 1)$ and $\mu_1 = 0$, and let ν be the probability distribution on $\mathbb{N}$ with $\nu(0) = \nu(2) = 1/2$. Then for all $n \geq 1$ such that $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$,*

$$\mathrm{ht}(\mathrm{T}_{\mu,n}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\nu,n}) \text{ if } n \text{ is odd}$$
$$\mathrm{ht}(\mathrm{T}_{\mu,n}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\nu,n+1}) \text{ if } n \text{ is even.}$$

The probability measure ν in the preceding theorem could be replaced by any nondegenerate probability measure ν' with $\nu'(0) + \nu'(2) = 1$, as for any such measure and any $m \in \mathbb{N}$, the tree

$\mathrm{T}_{\nu',2m+1}$ is uniformly distributed over rooted binary plane trees with m internal nodes and $m+1$ leaves. The parity requirement is due to the fact that binary trees always have an odd number of vertices. Also, the requirement that $\mu_1 = 0$ is necessary, since if $\mu_1 > 0$, then with positive probability $\mathrm{T}_{\mu,2m}$ (resp. $\mathrm{T}_{\mu,2m+1}$) is a path of length $2m-1$ (resp. $2m$), whereas $\mathrm{T}_{\nu,2m+1}$ has height at most $m+1$.

1.1. **Trees with fixed degree sequences.** The above results for Bienaymé trees are consequences of more refined results, presented in this subsection, on the heights of random trees with fixed degree sequences.

For a rooted tree t and a vertex v of t, the *degree* $d_{\mathrm{t}}(v)$ of v is the number of children of v in t. A *leaf* of t is a vertex of t with degree zero.

A *degree sequence* is a sequence of non-negative integers $\mathrm{d} = (d_1, \ldots, d_n)$ with $\sum_{i\in[n]} d_i = n-1$. We write $\mathcal{T}_{\mathrm{d}}$ for the set of finite rooted labeled trees t with vertex set labeled by $[n]$ and such that for each $i \in [n]$, the vertex with label i has degree d_i. Also write $\mathcal{T}(n)$ for the set of all finite rooted labeled trees with vertex set labeled by $[n]$. For $\mathrm{t} \in \mathcal{T}(n)$, we write $d_{\mathrm{t}}(i)$ for the degree of the vertex with label i and say that $(d_{\mathrm{t}}(1), \ldots, d_{\mathrm{t}}(n))$ is the degree sequence of t.

For $p > 0$ and a degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$ write $|\mathrm{d}|_p = (\sum_{i=1}^n d_i^p)^{1/p}$ and let $(\sigma_{\mathrm{d}})^2 = \frac{1}{n}\sum_{i=1}^n d_i(d_i - 1)$. Also, for $i \geq 0$ write $n_i(\mathrm{d}) = |\{j \in [n] : d_j = i\}|$ for the number of entries of d which equal i.

Given a finite set S, we write $X \in_u S$ to mean that X is a uniformly random element of S.

Theorem 6. *Fix any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$*, let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$*, and write* $\delta = (n - n_1(\mathrm{d}))/n$*. Then for all* $x > 0$*,*

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}) > xn^{1/2}\right) < 5\exp(-\delta x^2/2^{13}).$$

Theorem 7. *Fix any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$*, let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$*, and define* $(\sigma')^2 = \frac{n}{n-n_1(\mathrm{d})}\sigma_{\mathrm{d}}^2$*. Then for all* $x \geq 2^{14}$*,*

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}) > xn^{1/2}\frac{\log(\sigma'+1)}{\sigma_{\mathrm{d}}}\right) \leq 4\exp\left(-x\log(\sigma'+1)/2^{14}\right).$$

The following proposition is the tool which allows us to transfer results from the setting of random trees with given degree sequences to that of Bienaymé trees (and that of simply generated trees, which are introduced in Section 4, below).

Proposition 8. *Fix non-negative weights* $(w_i, i \geq 0)$*. Let* T *be a random plane tree of size* n *with distribution given by*

$$\mathbf{P}(\mathrm{T} = \mathrm{t}) \propto \prod_{v \in v(\mathrm{t})} w_{d_{\mathrm{t}}(v)}, \tag{1}$$

for plane trees t *of size* n*. Conditionally given* T*, let* $\hat{\mathrm{T}}$ *be the random tree in* $\mathcal{T}(n)$ *obtained as follows: label the vertices of* T *by a uniformly random permutation of* $[n]$*, then forget about the plane structure. For* $i \in [n]$ *let* D_i *be the degree of vertex* i *in* $\hat{\mathrm{T}}$ *of vertex* i*. Then for any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$*, conditionally given that* $(D_1, \ldots, D_n) = (d_1, \ldots, d_n)$*, we have* $\hat{\mathrm{T}} \in_u \mathcal{T}_{\mathrm{d}}$*.*

We provide the proof of Proposition 8 right away since it is short, but it is not necessary to read the proof in order to understand what follows.

Proof of Proposition 8. In this proof, by an *ordered labeled tree* we mean a finite tree $\mathbf{t}$ whose vertices are labeled by the integers $[n]$ (for some n) and such that the children of each node are endowed with a left-to-right order. From an ordered labeled tree, we may obtain a plane tree by ignoring the vertex labels, and we may obtain a labeled tree lying in $\mathcal{T}(n)$ by ignoring the orderings.

Let $\mathbf{T}$ be the random ordered labeled tree obtained from T by labeling the vertices of T by a uniformly random permutation of $[n]$.

Now fix any labeled tree $\hat{\mathrm{t}}$ with vertices labeled by $[n]$. For $j \geq 0$ write $n_j = \{i \in [n] : d_{\hat{\mathrm{t}}}(i) = j\}$. Then

$$\mathbf{P}(\hat{\mathrm{T}} = \hat{\mathrm{t}}) = \sum_{\mathbf{t}} \mathbf{P}(\mathbf{T} = \mathbf{t}),$$

where the sum is over ordered labeled trees $\mathbf{t}$ with vertices labeled by $[n]$ whose underlying (unordered) labeled tree is $\hat{\mathrm{T}}$. This sum has

$$\prod_{v \in \hat{\mathrm{t}}} d_{\hat{\mathrm{t}}}(v)! = \prod_{j \geq 0} (j!)^{n_j}$$

summands, corresponding to the number of ways to choose an ordering of the children of each vertex of $\hat{\mathrm{t}}$. Moreover, for any tree $\mathbf{t}$ included in the sum, writing t for the (unlabeled) plane tree underlying $\mathbf{t}$, we have

$$\mathbf{P}(\mathbf{T} = \mathbf{t}) = \frac{1}{n!} \mathbf{P}(\mathrm{T} = \mathrm{t}) \propto \prod_{v \in v(\mathrm{t})} w_{d_{\mathrm{t}}(v)} = \prod_{j \geq 0} w_j^{n_j}.$$

This value is the same for all summands, so it follows that

$$\mathbf{P}(\hat{\mathrm{T}} = \hat{\mathrm{t}}) \propto \prod_{j \geq 0} (j!)^{n_j} \prod_{j \geq 0} w_j^{n_j}.$$

The right-hand side depends on $\hat{\mathrm{t}}$ only through the number of vertices of each degree. In particular, for any degree sequence d, it is constant over trees in $\mathcal{T}_{\mathrm{d}}$. The result follows. □

Using this proposition, we can transfer results from the setting of random trees with given degree sequences to that of random plane trees, provided that we can obtain sufficiently precise control of the degree statistics of the random plane trees. (By "degree statistics" we mean the number of nodes of given degrees.) Once such control is established, Theorems 1, 2, and 3 follow straightforwardly from Theorems 6 and 7. Similarly, the stochastic inequality of Theorem 5 follows from a finer stochastic ordering on the heights of random trees with given degree sequences. We in fact obtain stochastic domination results for the heights of trees under two different partial orders on degree sequences We define these partial orders by specifying their corresponding covering relations[2].

Theorem 9. *Let $\prec_1$ be the partial order on degree sequences of length n defined by the following covering relation:* $\mathrm{d} = (d_1, \ldots, d_n)$ *covers* $\mathrm{d}' = (d'_1, \ldots, d'_n)$ *if there exist distinct $i, j \in [n]$ and a permutation $\nu : [n] \to [n]$ such that*

(1) $d_i \geq d_j$;
(2) $d'_{\nu(i)} = d_i + 1$;
(3) $d'_{\nu(j)} = d_j - 1$; *and*
(4) $d'_{\nu(k)} = d_k$ *for any* $k \in [n] \backslash \{i, j\}$.

Then with $\mathrm{T}_{\mathrm{d}} \in_u \mathcal{T}_{\mathrm{d}}$ *and* $\mathrm{T}_{\mathrm{d}'} \in_u \mathcal{T}_{\mathrm{d}'}$,

$$\mathrm{d}' \prec_1 \mathrm{d} \implies \mathrm{ht}(\mathrm{T}_{\mathrm{d}'}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}}).$$

Moreover, the stochastic domination of $\mathrm{ht}(\mathrm{T}_{\mathrm{d}})$ *over* $\mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$ *is strict if* $\mathrm{d}' \prec_1 \mathrm{d}$ *and* d *contains at least 3 non-leaf vertices.*

In words, to obtain d' from d in the definition of $\prec_1$, for $a \leq b$, we replace one vertex with a children and one vertex with b children by a vertex with $a-1$ children and one with $b+1$ children and potentially relabel the vertices; informally, the degrees in d' are more skewed than the degrees in d. Then, $\mathrm{d}' \prec_1 \mathrm{d}$ and $\mathrm{ht}(\mathrm{T}_{\mathrm{d}'}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}})$. So, informally, Theorem 9 states that more skewed

[2] For a partially ordered set $(\mathcal{P}, \prec)$, $y \in \mathcal{P}$ covers $x \in \mathcal{P}$ if $x \prec y$ and for all $z \in \mathcal{P}$, if $x \preceq z \preceq y$ then $z \in \{x, y\}$.

degrees yield shorter trees. This theorem has the following corollary, which resolves a conjecture from [4].

Corollary 10. *Let $\prec_2$ be the partial order on degree sequences of length n defined by the following covering relation:* $\mathrm{d} = (d_1, \ldots, d_n)$ *covers* $\mathrm{d}' = (d'_1, \ldots, d'_n)$ *if there is a permutation* $\nu : [n] \to [n]$ *and an* $i, j \in [n]$ *such that*

(1) $d_i, d_j \geq 1$*;*
(2) $d'_{\nu(i)} = d_i + d_j$*;*
(3) $d'_{\nu(j)} = 0$*; and*
(4) $d'_{\nu(k)} = d_k$ *for any* $k \in [n] \backslash \{i, j\}$*.*

Then with $\mathrm{T_d} \in_u \mathcal{T}_\mathrm{d}$ *and* $\mathrm{T_{d'}} \in_u \mathcal{T}_{\mathrm{d}'}$*,*

$$\mathrm{d}' \prec_2 \mathrm{d} \implies \mathrm{ht}(\mathrm{T_{d'}}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T_d}).$$

Corollary 10 follows immediately from Theorem 9 by observing that if $\mathrm{d}' \prec_2 \mathrm{d}$ then $\mathrm{d}' \prec_1 \mathrm{d}$. Arthur Blanc-Renaudie told us a proof of the stochastic domination presented in Corollary 10 (which he does not intend to write up), using a bijection presented in [8]. Our proof of Theorem 9 proceeds differently, but was inspired by Blanc-Renaudie's approach.

Before concluding the introduction, we state and prove two further corollaries of Theorem 9, and use the second of them to prove Theorem 5. The first states that height of any random tree with a fixed degree sequence is stochastically dominated by that of a random sub-binary tree with the same number of leaves and degree-one vertices. (We say d is sub-binary if $n_i(\mathrm{d}) = 0$ for all $i \geq 3$, and in this case also say that $\mathrm{T_d}$ is sub-binary.) The second essentially says that among all random trees with a given size and given number of vertices of degree 1, the sub-binary trees have the stochastically largest heights (with a very minor *caveat* that essentially addresses a potential parity issue).

Corollary 11. *Let* b *be a sub-binary degree sequence, let* d *be a degree sequence with* $n_0(\mathrm{d}) = n_0(\mathrm{b})$ *and* $n_1(\mathrm{d}) = n_1(\mathrm{b})$*, and let* $\mathrm{T_d} \in_u \mathcal{T}_\mathrm{d}$ *and* $\mathrm{T_b} \in_u \mathcal{T}_\mathrm{b}$*. Then* $\mathrm{ht}(\mathrm{T_d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T_b})$*.*

Proof. Note that necessarily $n_2(\mathrm{b}) = n_0(\mathrm{b}) - 1$, so b has length $2n_0(\mathrm{b}) + n_1(\mathrm{b}) - 1$ Write $\mathrm{d} = (d_1, \ldots, d_n)$. If $n = 2n_0(\mathrm{b}) + n_1(\mathrm{b}) - 1$ then d is a permutation of b and there is nothing to prove. So suppose that $n < 2n_0(\mathrm{b}) + n_1(\mathrm{b}) - 1$; in this case necessarily d contains at least one entry which is three or greater. We construct a degree sequence $\mathrm{d}' = (d'_1, \ldots, d'_{n+1})$ with $n_0(\mathrm{d}) = n_0(\mathrm{d}')$ and $n_1(\mathrm{d}) = n_1(\mathrm{d}')$ and $\mathrm{ht}(T_\mathrm{d}) \preceq \mathrm{ht}(T_{\mathrm{d}'})$. This proves the corollary (by induction on $2n_0(\mathrm{b}) + n_1(\mathrm{b}) - 1 - n$), as we can then transform d into a degree sequence of length sub-binary degree sequence $2n_0(\mathrm{b}) + n_1(\mathrm{b}) - 1$ without changing the number of zeros or of ones, and while stochastically increasing the height of the associated random tree.

To construct d', first let $\mathrm{d}^+ = (d^+_1, \ldots, d^+_{n+1}) = (d_1, \ldots, d_n, 1)$. The trees $\mathrm{T_d}$ and $\mathrm{T_{d^+}} \in_u \mathcal{T}_{\mathrm{d}^+}$ may be coupled as follows. If $n + 1$ is not the root of $\mathrm{T_{d^+}}$ then form $\mathrm{T_d}$ from $\mathrm{T_{d^+}}$ by replacing the two-edge path containing $n + 1$ in $\mathrm{T_{d^+}}$ by a single edge connecting its endpoints. If $n + 1$ *is* the root, then instead form $\mathrm{T_b}$ by deleting $n + 1$ and rerooting at its unique child. (The reverse operation is to add $n + 1$ as the parent of a uniformly random vertex of $\mathrm{T_d}$.) Under this coupling, the height of $\mathrm{T_{d^+}}$ is at least that of $\mathrm{T_d}$, so it follows that

$$\mathrm{ht}(\mathrm{T_d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T_{d^+}})\,.$$

Now choose $k \in [n]$ such that $d^+_k = d_k \geq 3$ and define $\mathrm{d}' = (d'_1, \ldots, d'_{n+1})$ by

$$d'_i = \begin{cases} d^+_i & \text{if } i \notin \{k, n+1\} \\ d^+_k - 1 & \text{if } i = k \\ 2 & \text{if } i = n+1\,. \end{cases}$$

Then $n_0(\mathrm{d}) = n_0(\mathrm{d}')$ and $n_1(\mathrm{d}) = n_1(\mathrm{d}')$. Moreover, $\mathrm{ht}(\mathrm{T}_{\mathrm{d}^+}) \preceq \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$ by Theorem 9, and so $\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$ as required. □

Corollary 12. *Let* $\mathrm{d} = (d_1, \ldots, d_n)$ *be any degree sequence and let* $\mathrm{T}_\mathrm{d} \in_u \mathcal{T}_\mathrm{d}$. *Let*

$$n^+ = \begin{cases} n & \text{if } n_1(\mathrm{d}) \geq 1 \text{ or if } n \text{ is odd} \\ n+1 & \text{if } n_1(\mathrm{d}) = 0 \text{ and } n \text{ is even.} \end{cases}$$

Then there is a sub-binary degree sequence $\mathrm{b} = (b_1, \ldots, b_{n^+})$ *with* $n_1(\mathrm{b}) \leq n_1(\mathrm{d})$, *such that with* $\mathrm{T}_\mathrm{b} \in_u \mathcal{T}_\mathrm{b}$, *then* $\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_\mathrm{b})$.

Proof. Suppose that d has at least one entry $d_k \geq 4$. Choose $j \in [n]$ with $d_j = 0$. Then the degree sequence $\mathrm{d}' = (d'_1, \ldots, d'_n)$ given by

$$d'_i = \begin{cases} d_i - 2 & \text{if } i = k \\ 2 & \text{if } i = j \\ d_i & \text{otherwise} \end{cases}$$

has the same length as d and satisfies $n_1(\mathrm{d}') = n_1(\mathrm{d})$. Moreover, by Theorem 9, $\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$. We have reduced the number of vertices of degree 4 while stochastically increasing the height, and without changing the number of vertices of degree 1. It follows that to prove the corollary we may restrict our attention to degree sequences corresponding to trees with maximum degree three.

Next suppose that d has $n_3(\mathrm{d}) \geq 2$ and choose distinct $k, \ell \in [n]$ with $d_k = d_\ell = 3$ and $j \in [n]$ with $d_j = 0$. Then the sequence $\mathrm{d}' = (d_1, \ldots, d'_n)$ given by

$$d'_i = \begin{cases} d_i - 1 & \text{if } i \in \{k, \ell\} \\ 2 & \text{if } i = j \\ d_i & \text{otherwise} \end{cases}$$

has the same length as d and satisfies $n_1(\mathrm{d}') = n_1(\mathrm{d})$ and has two fewer vertices of degree three, and two applications of Theorem 9 give that $\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$. This shows that we may restrict our attention to degree sequences corresponding to trees of maximum degree three and with at most one vertex of degree three.

Among such degree sequences, if $n_3(\mathrm{d}) = 0$ there is nothing to prove – the tree T_d is already sub-binary. If $n_3(\mathrm{d}) = 1$ and $n_1(\mathrm{d}) \geq 1$ then let k be such that $d_k = 3$ and choose $j \in [n]$ with $d_j = 1$. Then the sequence $\mathrm{d}' = (d'_1, \ldots, d'_n)$ given by

$$d'_i = \begin{cases} 2 & \text{if } i \in \{j, k\} \\ d_i & \text{otherwise} \end{cases}$$

is sub-binary and has $n_1(\mathrm{d}') < n_1(\mathrm{d})$, and $\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$.

Finally, if $n_3(\mathrm{d}) = 1$ and $n_1(\mathrm{d}) = 0$ then necessarily n is even. In this case, choose $k \in [n]$ such that $d_k = 3$, and define $\mathrm{d}' = (d'_1, \ldots, d'_{n+1})$ by

$$d'_i = \begin{cases} d_i & \text{if } i \notin \{k, n+1\} \\ d_k - 1 & \text{if } i = k \\ 2 & \text{if } i = n+1\,. \end{cases}$$

Then d′ is sub-binary. Moreover, the same argument as that in the final paragraph of the proof of Corollary 11 shows that with $\mathrm{T}_{\mathrm{d}'} \in_u \mathcal{T}_{\mathrm{d}'}$, then $\mathrm{ht}(\mathrm{T}_\mathrm{d}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathrm{d}'})$. This completes the proof. □

Proof of Theorem 5. For any $n \in \mathbb{N}$ with $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$ and any plane tree t of size n,

$$\mathbf{P}(\mathrm{T}_{\mu,n} = \mathrm{t}) = \frac{1}{\mathbf{P}(|\mathrm{T}_\mu| = n)} \prod_{v \in \mathrm{t}} \mu_{d_\mathrm{t}(v)}$$

and likewise if $\mathbf{P}(|\mathrm{T}_\nu| = n) > 0$ then $\mathbf{P}(\mathrm{T}_{\nu,n} = \mathrm{t}) = (\mathbf{P}(|\mathrm{T}_\mu| = n))^{-1} \prod_{v \in \mathrm{t}} \nu_{d_\mathrm{t}(v)}$. Thus, the laws of $\mathrm{T}_{\mu,n}$ and $\mathrm{T}_{\nu,n}$ both have the product structure required by Proposition 8.

Now fix n for which $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$ and let $n^+ = n + \mathbf{1}_{[n \text{ even}]}$, so that $\mathbf{P}(|\mathrm{T}_\nu| = n^+) > 0$. Label the vertices of $\mathrm{T}_{\mu,n}$ (resp. T_{ν,n^+}) by a uniformly random permutation of $[n]$ and write D (resp. B) for the resulting degree sequence. Then D is a degree sequence of length n with $n_1(\mathrm{D}) = 0$ and B is a binary degree sequence of length n^+. Corollary 12 yields that with $\mathrm{T}_\mathrm{B} \in_u \mathcal{T}_\mathrm{B}$ and $\mathrm{T}_\mathrm{D} \in_u \mathcal{T}_\mathrm{D}$, then

$$\mathrm{ht}(\mathrm{T}_\mathrm{D}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_\mathrm{B}) \,.$$

But Proposition 8 implies that $\mathrm{ht}(\mathrm{T}_{\mu,n}) \stackrel{\mathrm{d}}{=} \mathrm{ht}(\mathrm{T}_\mathrm{D})$ and $\mathrm{ht}(\mathrm{T}_{\nu,n^+}) \stackrel{\mathrm{d}}{=} \mathrm{ht}(\mathrm{T}_\mathrm{B})$; the result follows. □

The remainder of the paper is structured as follows. Section 2 presents the proofs of our results on random trees with fixed degree sequences, Theorems 6 and 7. Section 3 uses these results to prove our results on Bienaymé trees, Theorems 1, 2 and 3. This section also introduces simply generated trees, and presents our results on their heights. Finally, Section 4, which can be read independently of the rest of the paper, contains the proof of Theorem 9.

2. **Proofs of Theorems 6 and 7**

In this section we present the proofs of Theorems 6 and 7. (To streamline the presentation, a few of the longer proofs are postponed to Section 2.1.) Our approach is based on a bijection between rooted trees on $[n]$ and sequences in $[n]^{n-1}$ that was very recently introduced by Addario-Berry, Donderwinkel, Maazoun and Martin in [4]. We use the version of the bijection presented in [4, Section 3], specialized to trees with a fixed degree sequence.

To explain the bijection, it's convenient to focus on degree sequences with a particular form. Given a degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$, define another degree sequence $\mathrm{d}' = (d'_1, \ldots, d'_n)$ as follows. Let m be the number of non-zero entries of d; necessarily $1 \le m \le n-1$. List the non-zero entries of d in order of appearance as $d'_1, \ldots, d'_m$, and then set $d'_{m+1} = \ldots = d'_n = 0$. So, for example, if $\mathrm{d} = (1,0,3,0,0,2,0)$ then $\mathrm{d}' = (1,3,2,0,0,0,0)$. Say that the degree sequence d' is *compressed*. (So a degree sequence is compressed if all of its non-zero entries appear before all of its zero entries.)

There is a natural bijection between $\mathcal{T}_\mathrm{d}$ and $\mathcal{T}_{\mathrm{d}'}$: from a tree $\mathrm{t} \in \mathcal{T}_\mathrm{d}$, construct a tree $\mathrm{t}' \in \mathcal{T}_{\mathrm{d}'}$ by relabeling the non-leaf vertices of t as $1, \ldots, m$ and the leaves of t as $m+1, \ldots n$, in both cases in increasing order of their original labels. Using this bijection provides a coupling $(\mathrm{T}, \mathrm{T}')$ of uniformly random elements of $\mathcal{T}_\mathrm{d}$ and $\mathcal{T}_{\mathrm{d}'}$, respectively, such that T and T' have the same height. It follows that any tail bound for the height of a uniformly random tree in $\mathcal{T}_{\mathrm{d}'}$ applies *verbatim* to the height of a uniformly random tree in $\mathcal{T}_\mathrm{d}$.

Now, let $\mathrm{d} = (d_1, \ldots, d_n)$ be a compressed degree sequence, so there is $1 \le m \le n-1$ such that $d_i = 0$ if and only if $i > m$. Write $n_0 = n_0(\mathrm{d}) = n - m$ for the number of leaves in a tree with degree sequence d. Then define

$$\mathcal{S}_\mathrm{d} := \{(v_1, \ldots, v_{n-1}) : |\{k : v_k = i\}| = d_i \text{ for all } i \in [n]\} \,.$$

For example, if $\mathrm{d} = (1,3,2,0,0,0,0)$ then $\mathcal{S}_\mathrm{d}$ is the set of all permutations of the vector $(1,3,3,3,2,2)$, so has size $\binom{6}{1,3,2} = 60$.

The following bijection between $\mathcal{S}_\mathrm{d}$ and $\mathcal{T}_\mathrm{d}$ appears in [4, Section 3]. For $\mathrm{v} = (v_1, \ldots, v_{n-1}) \in \mathcal{S}_\mathrm{d}$, we say that $j \in \{2, \ldots, n-1\}$ is the location of a repeat of v if $v_j = v_i$ for some $i < j$.

Bijection t between $\mathcal{S}_{\mathrm{d}}$ and $\mathcal{T}_{\mathrm{d}}$.

- Let $j(0) = 1$, let $j(1) < j(2) < \cdots < j(n_0 - 1)$ be the locations of the repeats of the sequence v, and let $j(l) = m + n_0 = n$.
- For $i = 1, \ldots, n_0$, let P_i be the path $(v_{j(i-1)}, \ldots, v_{j(i)-1}, m + i)$.
- Let $\mathrm{t}(\mathrm{v}) \in \mathcal{T}_{\mathrm{d}}$ have root v_1 and edge set given by the union of the edge sets of the paths $P_1, P_2, \ldots, P_{n_0}$.

The inverse of the bijection works as follows. Fix a tree $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$. Let $S_0 = \{r(\mathrm{t})\}$ consist of the root of t. Recursively, for $1 \le i \le n_0$ let P_i be the path from S_{i-1} to $m + i$ in t, and let P_i^* be P_i excluding its final point $m + i$. Then let $\mathrm{v} = \mathrm{v}(\mathrm{t})$ be the concatenation of $P_1^*, \ldots, P_{n_0}^*$. For later use, we observe that this bijection implies that

$$|\mathcal{T}_{\mathrm{d}}| = \binom{n-1}{d_1, \ldots, d_n} = \frac{(n-1)!}{\prod_{i \in [n]} d_i!}. \tag{2}$$

This formula (which appears as Theorem 5.3.4 in [12]) holds for all degree sequences, not just compressed ones, by the observation from earlier in the section.

We now discuss how to use the bijection to bound the height of $\mathrm{t}(\mathrm{v})$. For this, we think of the bijection as constructing t from $\mathrm{v}(\mathrm{t}) = (v_1, \ldots, v_{n-1})$ by adding vertices one-at-a-time, in order of their first appearance in the concatenation of $P_1, \ldots, P_{n_0}$. Formally, define a permutation $(w(1), \ldots, w(n)) = (w_{\mathrm{t}}(1), \ldots, w_{\mathrm{t}}(n))$ of $[n]$ as follows. For $1 \le k \le n$, let $w(k) = v_k$ if k is not the location of a repeat, and let $w(k) = m + r$ if $j(r) = k$ (so either $r < n_0$ and k is the location of the r'th repeat in v, or $r = n_0$ and $k = n$). Then let t_k be the subtree of t with vertices $w(1), \ldots, w(k)$. An example appears in Figure 2.

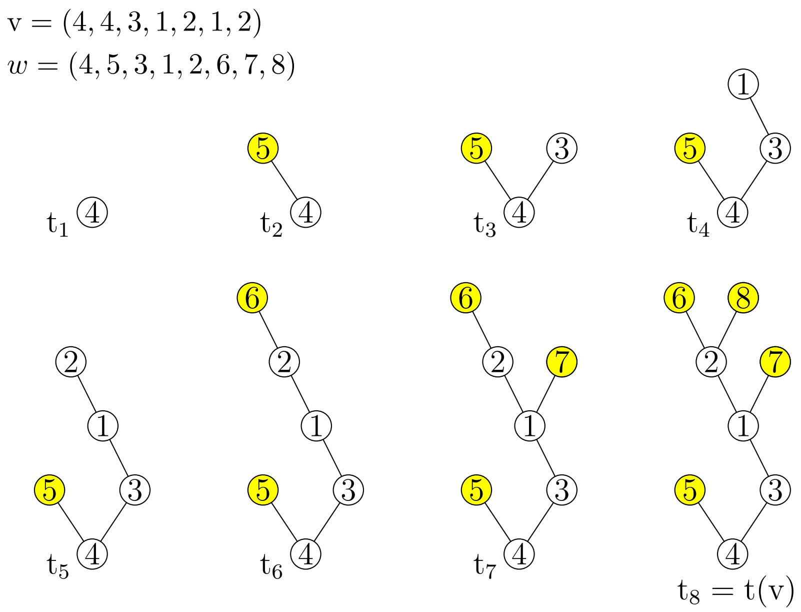


Figure 2. This figure illustrates the bijection and the sequential construction. In this example, we have $\pi(2) = 5$ since t_5 is the first tree in the sequence containing vertex 2. We also have $\rho(2) = 4$, since 3 is the minimal k such that $\sum_{1 \le j \le k} (d_{i(j)} - 1)$ is at least 2 and t_4 is the first tree in the sequence to contain vertices $\{i(1), i(2), i(3)\} = \{4, 3, 1\}$.

Let $(\pi(i), i \in [n]) = (\pi_{\mathrm{t}}(i), i \in [n])$ be the inverse permutation of w, so $\pi(i) = w^{-1}(i)$ is the step at which the vertex with label i is added when constructing t from v(t). List the non-leaf vertices $\{1, \ldots, m\}$ of t in the order they appear when constructing t as $(i(1), \ldots, i(m)) = (i_{\mathrm{t}}(1), \ldots, i_{\mathrm{t}}(m))$. In other words, $(i(1), \ldots, i(m))$ is the permutation of $\{1, \ldots, m\}$ such that $(\pi(i(1)), \ldots, \pi(i(m)))$ is increasing. Then for $0 \le x \le n_0 - 1$, let $k = k(x) = k_{\mathrm{t}}(x)$ be minimal so that

$$\sum_{j=1}^{k} (d_{i(j)} - 1) \ge x,$$

and let $\rho(x) = \rho_{\mathrm{t}}(x) = \pi(i(k(x)))$, so that $v_{\rho(x)} = i(k(x))$. Equivalently, $\rho(x)$ is the smallest integer ρ such that the subtree of t consisting of the nodes of t_ρ and all their children has strictly more than $\lceil x \rceil$ leaves; this is also the smallest integer ρ such that t_ρ contains vertices $i(1), \ldots, i(k(x))$. Note that $\rho(x)$ is non-decreasing, and since

$$\sum_{j=1}^{m} (d_{i(j)} - 1) = (\sum_{j=1}^{m} d_j) - m = n - 1 - m = n_0 - 1\,,$$

the tree $\mathrm{t}_\rho(l-1)$ has n_0 leaves. It follows that all vertices of t that do not belong to $\mathrm{t}_{\rho(n_0-1)}$ have degree either 0 or 1 in t. This in particular implies that if t has no vertices of degree exactly 1 then $\rho(n_0 - 1) = \pi(i(m))$ and $\mathrm{t}_{\rho(n_0-1)}$ contains all vertices of t except a subset of its leaves, so $\mathrm{ht}(\mathrm{t}) \le \mathrm{ht}(\mathrm{t}_{\rho(n_0-1)}) + 1$.

For any real sequence $0 \le y_0 < y_1 < \ldots < y_N = n_0 - 1$, writing $\mathrm{t} = \mathrm{t}(\mathrm{v})$ and $\mathrm{t}_k = \mathrm{t}_k(\mathrm{v})$, we may now bound $\mathrm{ht}(\mathrm{t})$ via the telescoping sum

$$\mathrm{ht}(\mathrm{t}) \le \mathrm{ht}(\mathrm{t}_{\rho(y_0)}) + \Big(\sum_{i=1}^{N} \big(\mathrm{ht}(\mathrm{t}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{t}_{\rho(y_{i-1})}) \big) \Big) + (\mathrm{ht}(\mathrm{t}) - \mathrm{ht}(\mathrm{t}_{\rho(n_0-1)})), \tag{3}$$

and the final term in the sum is at most 1 if t has no vertices of degree 1. Here is the key lemma which makes such a decomposition useful. Given $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$, let $\mathrm{v} = \mathrm{v}(\mathrm{t}) = (v_1, \ldots, v_{n-1})$.

Lemma 13. *Fix any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$*. Then for all* $0 \le x \le y$*, if* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ *then*

$$\mathbf{P}(\mathrm{dist}(v_{\rho(y)}, \mathrm{T}_{\rho(x)}) > b) \le \left(1 - \frac{x}{n-1}\right)^b$$

for all $b \ge 1$.

The proof of Lemma 13 appears in Section 2.1, below. This lemma is is not quite enough to control an increment of the form $\mathrm{ht}(\mathrm{T}_{\rho(y)}) - \mathrm{ht}(\mathrm{T}_{\rho(x)})$, since for that we need to control the distance from $v_{\rho_{\mathrm{t}}(z)}$ to $\mathrm{t}_{\rho(x)}$ for *all* $x \le z \le y$ – i.e. we need a "maximal inequality" version of Lemma 13. We achieve this in the following corollary. (Although in this section we defer the proofs of most of the supporting results, we prove the corollary immediately, as its proof is quite short.)

Corollary 14. *Fix any degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$ *with* $n_1(\mathrm{d}) = 0$*, and let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$*. Then for any integers* $0 \le x \le y$ *and any* $p \in (0,1)$ *and* $b > 0$*,*

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\rho(y)}) - \mathrm{ht}(\mathrm{T}_{\rho(x)}) > b + 1) \le \frac{y - x}{(1-p)b} \left(1 - \frac{x}{n-1}\right)^{\lfloor pb \rfloor}.$$

Proof. By relabeling the vertices, we may that assume without loss of generality that d is compressed. Write $m = \max(i : d_i \ne 0)$.

Since T has no vertices of degree 1, for all $k \in [m]$ we have we have $d_{i(k)} \ge 2$ and so $\sum_{j=1}^{k} (d_{i(j)} - 1) > \sum_{j=1}^{k-1} (d_{i(j)} - 1)$. It follows that the non-leaf vertices of $\mathrm{T}_{\rho(y)}$ lying outside of $\mathrm{T}_{\rho(x)}$ are a subset of $\{i(k(x+1)), \ldots, i(k(y))\} = \{v_{\rho(x+1)}, \ldots, v_{\rho(y)}\}$, and so

$$\mathrm{ht}(\mathrm{T}_{\rho(y)}) - \mathrm{ht}(\mathrm{T}_{\rho(x)}) \le 1 + \max\big(\mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)}) : z \in \{x+1, \ldots, y\}\big).$$

Now, if for a given $z \in \{x+1,\ldots,y\}$ we have

$$\mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)})) > b$$

then by considering the path in T from $v_{\rho(z)}$ to $\mathrm{T}_{\rho(x)}$, we see that there must be at least $(1-p)b$ distinct vertices $v \in \{v_{\rho(x+1)},\ldots,v_{\rho(y)}\}$ for which $\mathrm{dist}(v, \mathrm{T}_{\rho(x)}) > pb$. It follows that

$$\begin{aligned}
&\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\rho(y)}) - \mathrm{ht}(\mathrm{T}_{\rho(x)}) > b+1) \\
&\le \mathbf{P}\Big(\max\big(\mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)}) : z \in \{x+1,\ldots,y\}\big) > b\Big) \\
&\le \mathbf{P}\Big(\big|\{z \in \{x+1,\ldots,y\} : \mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)}) > pb\}\big| > (1-p)b\Big) \\
&\le \frac{1}{(1-p)b}\mathbf{E}\left[\big|\{z \in \{x+1,\ldots,y\} : \mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)}) > pb\}\big|\right] \\
&= \frac{1}{(1-p)b}\sum_{z=x+1}^{y} \mathbf{P}\big(\mathrm{dist}(v_{\rho(z)}, \mathrm{T}_{\rho(x)}) > pb\big)\,,
\end{aligned}$$

from which the corollary follows by the bound from Lemma 13. □

Combining the telescoping sum bound (3) with Corollary 14 will allow us to prove tail bounds for the heights of random trees $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ for degree sequences d with $n_1(\mathrm{d}) = 0$. The next lemma allows us to transfer tail bounds from this setting to that of general trees with fixed degree sequences.

Lemma 15. *Fix a degree sequence* $\mathrm{d} = (d_1,\ldots,d_n)$ *and write* $n_0 = n_0(\mathrm{d})$ *and* $n_1 = n_1(\mathrm{d})$*. Let* d' *be obtained from* d *by removing all entries which equal* 1*. Let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ *and* $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$*. Then for any* $h \ge 2$ *and* $y \ge 8hn/(n-n_1)$*,*

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}) > y) \le \mathbf{P}(\mathrm{ht}(\mathrm{T}') > \lceil h \rceil) + n_0 \exp(-y/4)\,.$$

Lemma 15, whose proof appears in Section 2.1, is the last tool we need to prove Theorem 6.

Proof of Theorem 6. Note that if $n \le 64$ then the result is trivially true, since $5e^{-\delta x^2/2^{13}} > 1$ for $x < 8$, and $\mathbf{P}(\mathrm{ht}(\mathrm{t}) > xn^{1/2}) = 0$ for $x \ge 8$. We thus hereafter assume that $n > 64$.

As described above, we shall bound $\mathbf{P}(\mathrm{ht}(\mathrm{T}) > xn^{1/2})$ by bounding the height via the telescoping sum (3), which for the random tree T becomes

$$\mathrm{ht}(\mathrm{T}) \le \mathrm{ht}(\mathrm{T}_{\rho(y_0)}) + \Big(\sum_{i=1}^{N}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})})\big)\Big) + (\mathrm{ht}(\mathrm{T}) - \mathrm{ht}(\mathrm{T}_{\rho(n_0-1)})), \tag{4}$$

for a suitably chosen sequence $0 \le y_0 < y_1 < \ldots < y_N = n_0 - 1$, and then controlling the contribution of each summand.

We first treat the case that d does not contain any entries equal to 1. In this case, the final term in (4) is at most 1, since when there are no vertices of degree one, all vertices of T not lying in $\mathrm{T}_{\rho(n_0-1)}$ are leaves. It follows that for any positive constants $b_0,\ldots,b_N$ such that $\sum_{i=0}^{N} b_i \le x$,

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}) > \lceil xn^{1/2}\rceil) \le \sum_{i=0}^{N}\mathbf{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\big)\,, \tag{5}$$

where we use the convention that $\mathrm{ht}(\mathrm{T}_{\rho(y_{-1})}) = 0$. We choose $b_0 = x/2$ and $b_i = x \cdot i/2^{i+2}$, so that $\sum_{i\ge0} b_i = x$ and $\sum_{i=0}^{N} b_i < x$ for all $N \in \mathbb{N}$.

Provided that $b_i n^{1/2} \geq 4$, Corollary 14 applied with $b = b_i n^{1/2} - 1$ and $p = (b-1)/2b$ yields

$$\begin{aligned}\mathbf{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\big) &\leq \frac{y_i - y_{i-1}}{(1-p)b}\left(1 - \frac{y_{i-1}}{n-1}\right)^{\lfloor pb \rfloor} \\ &= \frac{2(y_i - y_{i-1})}{b+1}\left(1 - \frac{y_{i-1}}{n-1}\right)^{(b-1)/2} \\ &\leq \frac{2(y_i - y_{i-1})}{b_i n^{1/2}}\left(1 - \frac{y_{i-1}}{n-1}\right)^{b_i n^{1/2}/4} .\end{aligned}$$

Now set $y_i = \min(n_0 - 1, 2^{i-1} x n^{1/2})$. For $i \geq 1$, if $y_{i-1} = n_0 - 1$ then $\mathrm{T}_{\rho(y_i)} = \mathrm{T}_{\rho(y_{i-1})}$ so $\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}) = 0$. On the other hand, if $2^{i-1} x n^{1/2} \leq n_0 - 1$, then $y_{i-1} = 2^{i-2} x n^{1/2} < n_0 - 1$. In this case we have $2(y_i - y_{i-1})/(b_i n^{1/2}) \leq 2^{2i+1}/i$ and $1 - y_{i-1}/(n-1) \leq e^{-2^{i-2}x/n^{1/2}}$, so if also $b_i n^{1/2} \geq 4$ then this gives

$$\begin{aligned}\mathbf{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\big) &\leq \frac{2^{2i+1}}{i}\exp\left(-\frac{2^{i-2}x}{n^{1/2}}\frac{b_i n^{1/2}}{4}\right) \\ &= \frac{2^{2i+1}}{i}\exp\left(-\frac{ix^2}{2^6}\right) . \qquad (6)\end{aligned}$$

But if $i \geq 2$ is small enough that $2^{i-1} x n^{1/2} \leq n_0 - 1$, then since $n_0 < n$, we also have $n^{1/2} b_i = n^{1/2} x i / 2^{i+2} \geq x^2 i/8$ and so $b_i n^{1/2} \geq 4$ provided $x \geq 4$. If $i = 1$ then we also have $n^{1/2} b_i = n^{1/2} x/8 \geq 4$ when $x \geq 4$, since we assumed that $n \geq 64$. It follows that (6) holds for all $i \geq 1$.

To handle the $i = 0$ term, note that by the definition of $\rho(y_0)$, the subtree of T consisting of all nodes of $\mathrm{T}_{\rho(y_0)-1}$ and their children contains at most $\lceil y_0 \rceil$ leaves. Since there are no vertices of degree exactly 1, this also implies that $\mathrm{T}_{\rho(y_0)-1}$ has fewer than $\lceil y_0 \rceil$ non-leaf vertices. Therefore, $\mathrm{ht}(\mathrm{T}_{\rho(y_0)-1}) \leq \lceil y_0 \rceil - 1$, so $\mathrm{ht}(\mathrm{T}_{\rho(y_0)}) \leq \lceil y_0 \rceil \leq b_0 n^{1/2}$ and thus

$$\mathbf{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_0)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{-1})}) > b_0 n^{1/2}\big) = \mathbf{P}(\mathrm{ht}(\mathrm{T}_{\rho(y_0)}) > b_0 n^{1/2}) = 0 .$$

Using this bound and (6) in the the telescoping sum bound (5), with $N = \min(i : 2^i x n^{1/2} \geq n_0 - 1) = \min(i : y_i = n_0 - 1)$, we obtain that for all $x \geq 4$,

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}) > \lceil x n^{1/2} \rceil) \leq \sum_{i \geq 1} \frac{2^{2i+1}}{i}\exp\left(-\frac{ix^2}{2^6}\right) = \sum_{i \geq 1}\exp\left((2i+1)\log 2 - \log i - \frac{ix^2}{2^6}\right) .$$

Provided that $x^2/2^6 > 4\log 2$, the exponent in the sum is less than $\log 2 - ix^2/2^7 - \log i$, so for such x the probability we aim to bound is at most

$$\sum_{i \geq 1} 2\exp(-ix^2/2^7) = 2e^{-x^2/2^7}\big(1 - e^{-x^2/2^7}\big)^{-1} < 4e^{-x^2/2^7} .$$

On the other hand, if $x^2/2^6 \leq 4\log 2$ then $4e^{-x^2/2^7} \geq 4e^{-2\log 2} = 1$, which is an upper bound on any probability. It follows that for all $x > 0$,

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}) > \lceil x n^{1/2} \rceil) \leq 4e^{-x^2/2^7} . \qquad (7)$$

This proves (in fact, is stronger than) the necessary bound when T has no vertices of degree 1.

Now suppose that $n_1 > 0$ and write $\delta = (n - n_1(\mathrm{d}))/n$; so $\delta < 1$. By permuting the vertex labels we may assume that all the ones in d are at the end, i.e., $d_i \neq 1$ for $i \leq n - n_1$ and $d_i = 1$ for $n - n_1 < i \leq n$. Let $\mathrm{d}' = (d_1, \ldots, d_{n-n_1})$ be obtained by removing all entries equal to 1 from d, and let $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$. By Lemma 15, for any $h \geq 1$ and any $y \geq 8hn/(n - n_1)$,

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}) > y) \leq \mathbf{P}(\mathrm{ht}(\mathrm{T}') > \lceil h \rceil) + le^{-y/4} \leq \mathbf{P}(\mathrm{ht}(\mathrm{T}') > \lceil h \rceil) + ne^{-y/4} .$$

Take $y = xn^{1/2}$ and $h = y(n-n_1)/(8n) = \lceil x'(n-n_1)^{1/2} \rceil$, where we have set $x' = x(n-n_1)^{1/2}/(8n^{1/2})$. Then the above bound and (7) together yield

$$\begin{aligned}\mathbf{P}(\mathrm{ht}(\mathrm{T}) > xn^{1/2}) &\le \mathbf{P}(\mathrm{ht}(\mathrm{T}') > \lceil x'(n-n_1)^{1/2} \rceil) + ne^{-xn^{1/2}/4} \\ &\le 4e^{-(x')^2/2^7} + ne^{-xn^{1/2}/4} \\ &= 4e^{-\delta x^2/2^{13}} + ne^{-xn^{1/2}/4} \\ &\le 5e^{-\delta x^2/2^{13}},\end{aligned}$$

where the last inequality holds for $8 \le x \le n^{1/2}$, as can be straightforwardly checked. But for $x < 8$ we have $5e^{-\delta x^2/2^{13}} > 1$, and for $x > n^{1/2}$ we have $\mathbf{P}(\mathrm{ht}(\mathrm{T}) > xn^{1/2}) = 0$, so this bound holds in those cases as well. □

In the preceding proof, the first term in the telescoping sum could be essentially "given away" - we just used the deterministic bound $\mathrm{ht}(\mathrm{T}_{\rho(y_0)}) \le \lceil y_0 \rceil$. In the setting of Theorem 7, where we aim for stronger bounds when σ_{d} is large than when it is small, we can not be so casual about this first term. The additional tool we require, to bound the height of the random tree constructed during the early stages of the bijection in the case when σ_{d} is large, is given in the following proposition.

Proposition 16. *Fix a degree sequence* $\mathrm{d} = (d_1, \ldots, d_n)$ *and let* $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$.

Then taking $\alpha = (1-e^{-2})/24$*, with* $\sigma = \sigma_{\mathrm{d}}$*, we have that for all* $b \ge 1$,

$$\mathbf{P}\left(\mathrm{ht}\left(\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}\right) > \tfrac{bn^{1/2}}{2\sigma}\right) \le \exp\left(-\tfrac{3}{32}\tfrac{bn^{1/2}}{\sigma}\right) + \exp\left(-\tfrac{\alpha b}{2}\right).$$

The proof of Proposition 16 appears in Section 2.1. With this proposition in hand, the proof of Theorem 7 proceeds quite similarly to that of Theorem 6 (though it is necessarily somewhat more involved as the choices of the values y_i and b_i for the telescoping sum bound must depend on σ_{d}).

Proof of Theorem 7. In the proof we write $n_i = n_i(\mathrm{d})$ and $\sigma = \sigma_{\mathrm{d}}$ for succinctness. We first assume that $n_1 = 0$, so T contains no degree-1 vertices. In the proof of Theorem 6 we showed that for such a tree, $\mathbf{P}(\mathrm{ht}(\mathrm{T}) > \lceil xn^{1/2} \rceil) \le 4e^{-x^2/2^7}$; see inequality (7). If $\sigma < e$ then this implies that for $x \ge 2^{11}$,

$$\begin{aligned}\mathbf{P}\left(\mathrm{ht}(T) \ge \left\lceil xn^{1/2}\tfrac{\log(1+\sigma)}{\sigma} \right\rceil\right) &\le 4\exp\left(-\frac{(x\log(1+\sigma))^2}{2^7\sigma^2}\right) \\ &< 4\exp\left(-\frac{2^4\log(1+\sigma)}{\sigma^2}(x\log(1+\sigma))\right) < 4e^{-x\log(1+\sigma)}.\end{aligned}$$

Continuing to assume that $n_1 = 0$, we now turn our attention to the case that $\sigma \ge e$. Like in the proof of Theorem 6, we begin by using the telescoping sum inequality (3), and obtain the bound

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}) > \lceil xn^{1/2}\tfrac{\log\sigma}{\sigma} \rceil) \le \sum_{i=0}^{N} \mathbf{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\tfrac{\log\sigma}{\sigma}\big), \tag{8}$$

which holds for any sequence $0 \le y_0 < y_1 < \ldots < y_N = n_0 - 1$ and positive constants $b_0, \ldots, b_N$ with $\sum_{i=0}^N b_i \le x$, and then controlling the contribution of each summand. (We again take $\mathrm{ht}(\mathrm{T}_{\rho(y_{-1})}) := 0$ for convenience.) Also as in the proof of Theorem 6, we take $b_0 = x/2$ and $b_i = x \cdot i/2^{i+2}$, so that $\sum_{i\ge 0} b_i = x$ and $\sum_{i=0}^N b_i < x$ for all $N \in \mathbb{N}$.

For $x \ge 2^{11}$, since $\sigma \ge e$ we have $(b_0 n^{1/2}\log\sigma)/\sigma \ge n^{1/2}/(2\sigma)$, so taking $y_0 = \sigma n^{1/2}(1-e^{-2})/24$, by Corollary 16 applied with $b = x\log\sigma$ we obtain

$$\mathbf{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_0)} > b_0 n^{1/2}\tfrac{\log\sigma}{\sigma}\big) < \exp\left(-\tfrac{3}{32}xn^{1/2}\tfrac{\log\sigma}{\sigma}\right) + \exp\left(-\tfrac{1-e^{-2}}{48}x\log\sigma\right) \le 2e^{-(x\log\sigma)/64}, \tag{9}$$

the last inequality holding since $\sigma \le n^{1/2}$ and $\min(3/32, (1-e^{-2})/48) > 1/64$.

For the remaining summands, we take $y_i = \min(n_0 - 1, 2^i y_0) = \min(n_0 - 1, 2^i \sigma n^{1/2}(1 - e^{-2})/24)$. Provided that $b_i n^{1/2} \frac{\log \sigma}{\sigma} \ge 4$, then Corollary 14 applied with $b = b_i n^{1/2} \frac{\log \sigma}{\sigma} - 1$ and $p = (b-1)/2b$ yields

$$\begin{aligned}\mathbf{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\big) &\le \frac{y_i - y_{i-1}}{(1-p)b}\left(1 - \frac{y_{i-1}}{n-1}\right)^{\lfloor pb \rfloor} \\ &= \frac{2(y_i - y_{i-1})}{b+1}\left(1 - \frac{y_{i-1}}{n-1}\right)^{(b-1)/2} \\ &\le \frac{2(y_i - y_{i-1})}{b_i n^{1/2}} \frac{\sigma}{\log \sigma}\left(1 - \frac{y_{i-1}}{n-1}\right)^{b_i n^{1/2}(\log \sigma)/(4\sigma)} .\end{aligned}$$

To simplify this upper bound, note that

$$\frac{2(y_i - y_{i-1})}{b_i n^{1/2}} \frac{\sigma}{\log \sigma} = \frac{2^{2i+2}(1-e^{-2})}{24ix}\frac{\sigma^2}{\log \sigma} < \frac{2^{2i-2}\sigma^2}{ix \log \sigma},$$

and provided $y_{i-1} < n_0 - 1$ we also have $1 - y_{i-1}/(n-1) \le e^{-2^i \sigma(1-e^{-2})/(24 n^{1/2})}$, so

$$\left(1 - \frac{y_{i-1}}{n-1}\right)^{b_i n^{1/2}(\log \sigma)/(4\sigma)} \le \exp\left(-\frac{2^i \sigma (1-e^{-2})}{24 n^{1/2}} \frac{b_i n^{1/2} \log \sigma}{4\sigma}\right) < \exp(-ix \log \sigma / 2^9) .$$

Combining the three preceding bounds, it follows that when $b_i n^{1/2} \frac{\log \sigma}{\sigma} \ge 4$ we have

$$\begin{aligned}\mathbf{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\big) &\le \frac{2^{2i-2}\sigma^2}{ix \log \sigma} \exp\left(-\frac{ix \log \sigma}{2^9}\right) \\ &= \frac{\sigma^2}{4x \log \sigma} \exp\left((2 \log 2) i - \log i - \frac{ix \log \sigma}{2^9}\right) .\end{aligned}$$

(If $y_{i-1} \ge n_0 - 1$ then $\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}) = 0$ so the bound holds in this case as well.) By the assumption that $x \ge 2^{11}$ and since $\sigma \ge e$, we have $(x \log \sigma)/2^9 > 2(2 \log 2)$, so this yields

$$\mathbf{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\big) < \frac{\sigma^2}{2^{13}} \exp\left(-\frac{ix \log \sigma}{2^{10}}\right) .$$

For the above bound we needed that $b_i n^{1/2} \frac{\log \sigma}{\sigma} \ge 4$, or in other words that $ixn^{1/2} \log \sigma / (2^{i+4}\sigma) \ge 1$. But since $n_0 - 1 < n$, the condition $y_{i-1} < n_0 - 1$ implies that

$$\frac{2^{i-1}\sigma n^{1/2}(1-e^{-2})}{24} < n ,$$

so $ix \log \sigma n^{1/2}/(2^{i+4}\sigma) \ge 1$ provided that $x \log \sigma > 192/(1-e^{-2})$, which holds since $x \ge 2^{11}$ and $\sigma \ge e$. It follows that

$$\begin{aligned}\sum_{i \ge 1} \mathbf{P}\big(\mathrm{ht}(\mathrm{T}_{\rho(y_i)}) - \mathrm{ht}(\mathrm{T}_{\rho(y_{i-1})}) > b_i n^{1/2}\big) &\le \frac{\sigma^2}{2^{13}} \exp\left(-\frac{x \log \sigma}{2^{10}}\right)\left(1 - e^{-(x \log \sigma)/2^{10}}\right)^{-1} \\ &\le \frac{1}{2^{12}} \exp\left(-\frac{x \log \sigma}{2^{11}}\right) ,\end{aligned}$$

the last bound holding since $x \log \sigma \ge x \ge 2^{11}$ and thus $\sigma^2 e^{-(x \log \sigma)/2^{10}} < e^{-(x \log \sigma)/2^{11}}$ and $(1 - e^{-(x \log \sigma)/2^{10}})^{-1} < 2$.

Using this bound and (9) in (8), we thus obtain

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}) > \lceil x n^{1/2} \tfrac{\log \sigma}{\sigma} \rceil) < 3 \exp\left(-\frac{x \log \sigma}{2^{11}}\right) .$$

Combining this with the bound from the start of the proof, which handles the case $\sigma < e$, it follows that when there are no vertices of degree one, for all $x \geq 2^{11}$ we have

$$\mathbf{P}\big(\mathrm{ht}(\mathrm{T}) > \lceil xn^{1/2}\tfrac{\log(\sigma+1)}{\sigma}\rceil\big) \leq 4\exp(-x\log(\sigma+1)/2^{11}). \tag{10}$$

This proves the theorem (in fact, something slightly stronger) when T has no vertices of degree one.

Now suppose that d contains vertices of degree 1. Like in the proof of Theorem 6, we let d′ be obtained from d by removing all entries equal to 1. Let $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and let $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$. Write $\sigma = \sigma_{\mathrm{d}}$. Also write $n' = n - n_1$ for the number of vertices of T′ so that we have that $\sigma_{\mathrm{d}'} = \frac{n}{n'}\sigma = \sigma'$.

By Lemma 15, for any $h \geq 2$ and any $y \geq 8hn/n'$,

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}) > y\right) \leq \mathbf{P}\left(\mathrm{ht}(\mathrm{T}') > \lceil h\rceil\right) + n_0e^{-y/4} \leq \mathbf{P}(\mathrm{ht}(\mathrm{T}') > \lceil h\rceil) + ne^{-y/4}.$$

We take

$$y = xn^{1/2}\frac{\log(\sigma'+1)}{\sigma}, \qquad h = \frac{y}{8}\frac{n'}{n} = \frac{x}{8}(n')^{1/2}\frac{\log(\sigma'+1)}{\sigma'}.$$

Then, if $x/8 \geq 2^{11}$, (10) gives that

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}') > \lceil h\rceil\right) \leq 4\exp\left(-x\log(\sigma'+1)/2^{14}\right).$$

Also, since $\sigma' \leq (n')^{1/2} \leq n^{1/2}$ we have $n^{1/2}\frac{\log(\sigma'+1)}{\sigma} \geq (\log n)/2$, so $y/4 - \log n \geq y/8$ if $x \geq 8$, but this constraint is weaker than $x/8 \geq 2^{11}$. It follows that for $x/8 \geq 2^{11}$,

$$ne^{-y/4} \leq e^{-y/8} \leq \left(-\frac{x}{8}\log(\sigma'+1)\right),$$

so the above bounds together yield that for all $x \geq 2^{14}$,

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}) > xn^{1/2}\frac{\log(\sigma'+1)}{\sigma}\right) \leq 4\exp\left(-x\log(\sigma'+1)/2^{14}\right). \qquad \square$$

2.1. **The postponed proofs.**

This section contains the proofs of the results that were stated without proof earlier in Section 2: namely, Lemmas 13 and 15 and Proposition 16.

Proof of Lemma 13. It is useful to assume d is compressed (and in particular that $d_n = 0$). Write $\mathrm{par}^1_{\mathrm{t}}(u) = \mathrm{par}_{\mathrm{t}}(u)$ for the parent of vertex u in tree t; we define the parent of the root to be the root itself. Inductively set $\mathrm{par}^{b+1}_{\mathrm{t}}(u) = \mathrm{par}_{\mathrm{t}}(\mathrm{par}^b_{\mathrm{t}}(u))$ for $b \geq 1$. Throughout the proof, we take $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and let $\mathrm{V} = \mathrm{v}(\mathrm{T})$, so $\mathrm{V} \in_u \mathcal{S}_{\mathrm{d}}$.

Fix any vector $(i(1), \ldots, i(j))$ of distinct elements of $[m]$ with $\sum_{k=1}^{j}(d_{i(k)} - 1) \geq y$, and let $\mathcal{S}_{\mathrm{d}}(i(1), \ldots, i(j))$ be the set of vectors $\mathrm{v} \in \mathcal{S}_{\mathrm{d}}$ such that, writing $\mathrm{t} = \mathrm{t}(\mathrm{v})$,

$$(i_{\mathrm{t}}(1), \ldots, i_{\mathrm{t}}(j)) = (i(1), \ldots, i(j)).$$

Let $r \in [j]$ be such that

$$\sum_{k=1}^{r-1}(d_{i(k)} - 1) < y \leq \sum_{k=1}^{r}(d_{i(k)} - 1),$$

Note that if $\mathrm{v} \in \mathcal{S}_{\mathrm{d}}(i(1), \ldots, i(j))$ and $\mathrm{t} = \mathrm{t}(\mathrm{v})$ then $v_{\rho_{\mathrm{t}}(y)} = i(r) = i_{\mathrm{t}}(r)$. This allows us to rewrite

$$\begin{aligned}&\mathbf{P}\left(D_{\mathrm{T}}(y,x) > k \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j))\right)\\ &= \mathbf{P}\left(\mathrm{par}^k_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho_{\mathrm{T}}(x)} \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j))\right).\end{aligned}$$

For $\mathrm{V} \in_u \mathcal{S}_{\mathrm{d}}$, the ordering of the integers $\{1,\ldots,n\}$ by their first appearance in V is degree-biased, so

$$\mathbf{P}(i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j)) = (i(1),\ldots,i(j)) = \prod_{k=1}^{j} \frac{d_{i(k)}}{n-1-d_{i(1)}-\ldots-d_{i(k-1)}},$$

or equivalently

$$|\mathcal{S}_{\mathrm{d}}(i(1),\ldots,i(j))| = \binom{n-1}{d_1,\ldots,d_n} \cdot \prod_{k=1}^{j} \frac{d_{i(k)}}{n-1-d_{i(1)}-\ldots-d_{i(k-1)}}.$$

Fix $\ell < r$ and let $\mathrm{d}^\ell = (d^\ell_1,\ldots,d^\ell_{n-1})$ where $d^\ell_{i(\ell)} = d_{i(\ell)} - 1$ and $d^\ell_i = d_i$ for all $i \in [n-1] \setminus \{i(\ell)\}$. Since $d_n = 0$ we have $\sum_{i\in[n-1]} d^\ell_i = n-2$, so d^ℓ is another degree sequence. Now consider the subset $\mathcal{S}^\ell_{\mathrm{d}}(i(1),\ldots,i(j))$ of $\mathcal{S}_{\mathrm{d}}(i(1),\ldots,i(j))$ consisting of those vectors v where the first instance of $i(r)$ in v is the immediate successor of some instance of $i(\ell)$ other than the first. The set $\mathcal{S}^\ell_{\mathrm{d}}(i(1),\ldots,i(j))$ is in bijection with with the set of vectors $\mathrm{v}' \in \mathcal{S}_{\mathrm{d}^\ell}(i(1),\ldots,i(j))$, i.e., the set of vectors $\mathrm{v}' \in \mathcal{S}_{\mathrm{d}^\ell}$ such that

$$(i(1,\mathrm{v}'),\ldots,i(j,\mathrm{v}')) = (i(1),\ldots,i(j)).$$

To see this, fix $\mathrm{v} \in \mathcal{S}^\ell_{\mathrm{d}}(i(1),\ldots,i(j))$, and let v' be the vector obtained from v by deleting the entry with value $i(\ell)$ immediately preceding the first instance of $i(r)$. Then $(i(1,\mathrm{v}),\ldots,i(j,\mathrm{v})) = (i(1,\mathrm{v}'),\ldots,i(j,\mathrm{v}'))$, since the deleted entry was not the first instance of $i(\ell)$ in v, so v' is an element of $\mathcal{S}_{\mathrm{d}^\ell}(i(1),\ldots,i(j))$. To recover v from v', simply insert an entry with value $i(\ell)$ immediately before the first instance of $i(r)$ in v'.

The same computation as for the size of $\mathcal{S}_{\mathrm{d}}(i(1),\ldots,i(j))$ now yields the formula

$$|\mathcal{S}^\ell_{\mathrm{d}}(i(1),\ldots,i(j))| = |\mathcal{S}_{\mathrm{d}^\ell}(i(1),\ldots,i(j))| = \binom{n-2}{d^\ell_1,\ldots,d^\ell_n} \prod_{k=1}^{j} \frac{d^\ell_{i(k)}}{n-2-d^\ell_{i(1)}-\ldots-d^\ell_{i(k-1)}}.$$

Writing $E(r,\ell)$ for the event that the first instance of $i(r,\mathrm{V})$ in V is an immediate successor of some instance of $i(\ell,\mathrm{V})$ other than the first, it follows that

$$\begin{aligned}
&\mathbf{P}\left(E(r,\ell) \mid (i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j)) = (i(1),\ldots,i(j))\right) \\
&= \frac{|\mathcal{S}^\ell_{\mathrm{d}}(i(1),\ldots,i(j))|}{|\mathcal{S}_{\mathrm{d}}(i(1),\ldots,i(j))|} \\
&= \frac{d_{i(\ell)}}{n-1} \cdot \prod_{k=1}^{j} \frac{d^\ell_{i(k)}}{d_{i(k)}} \cdot \prod_{k=1}^{j} \frac{n-1-d_{i(1)}-\ldots-d_{i(k-1)}}{n-2-d^\ell_{i(1)}-\ldots-d^\ell_{i(k-1)}} \\
&= \frac{d_{i(\ell)}-1}{n-1} \cdot \prod_{k=1}^{j} \frac{n-1-d_{i(1)}-\ldots-d_{i(k-1)}}{n-2-d^\ell_{i(1)}-\ldots-d^\ell_{i(k-1)}} \\
&> \frac{d_{i(\ell)}-1}{n-1}.
\end{aligned}$$

Now note that if $E(r,\ell)$ occurs then $i_{\mathrm{T}}(\ell)$ is the parent of $i_{\mathrm{T}}(r)$ in $\mathrm{T} = \mathrm{t}(\mathrm{V})$. Moreover, letting $q \in [j]$ be such that

$$\sum_{k=1}^{q-1} (d_{i(k)} - 1) < x \le \sum_{k=1}^{q} (d_{i(k)} - 1),$$

then on the event $\{(i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))=(i(1),\ldots,i(j))\}$, the tree $\mathrm{T}_{\rho(x)}$ has vertices $i(1),\ldots,i(q)$. It follows that

$$\begin{aligned}
&\mathbf{P}\left(\mathrm{par}_{\mathrm{T}}(v_{\rho_{\mathrm{T}}(y)})\in \mathrm{t}_{\rho(x)}(\mathrm{V})\ \mid (i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))=(i(1),\ldots,i(j))\right)\\
&\mathbf{P}\left(\mathrm{par}_{\mathrm{T}}(i_{\mathrm{T}}(r))\in \mathrm{t}_{\rho(x)}(\mathrm{V})\ \mid (i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))=(i(1),\ldots,i(j))\right)\\
&\geq \sum_{\ell=1}^{q}\mathbf{P}\left(E(r,\ell)\ \mid (i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))=(i(1),\ldots,i(j)\right)\\
&>\sum_{\ell=1}^{q}\frac{d_{i(\ell)}-1}{n-1}\geq\frac{x}{n-1}\,.
\end{aligned}$$

This proves the case $b=1$ of the lemma by summing over $(i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))$. In fact, we have proved something slightly stronger: in the case $b=1$, the necessary bound holds even conditionally given the values of $(i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))$. From this we will prove the full lemma by arguing inductively, by conditioning on the label of the parent of $i_{\mathrm{T}}(r)$.

There are two cases to consider. In either case, we continue to take $\mathrm{T}\in_u \mathcal{T}_{\mathrm{d}}$ and $\mathrm{V}=\mathrm{v}(\mathrm{T})$, and fix a sequence $(i(1),\ldots,i(j))$ of distinct elements of $[m]$. For the first case, fix $\ell\in\{q+1,\ldots,r-1\}$, let $W\in_u \mathcal{S}_{\mathrm{d}^\ell}$, where d^ℓ is as defined above, and suppose that $(i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))=(i(1),\ldots,i(j))$ and that $E(r,\ell)$ occurs. Under this conditioning, we have $\mathrm{V}\in_u\mathcal{S}^{\ell}_{\mathrm{d}}(i(1),\ldots,i(j))$, and in particular the parent of $i(r)=i_{\mathrm{T}}(r)$ in T is $i(\ell)$. Now let V' be obtained from V by deleting the entry immediately preceding the first instance of $i(r)$ in V and let $\mathrm{T}'=\mathrm{t}(\mathrm{V}')$. Under this conditioning, $\mathrm{V}'\in_u \mathcal{S}_{\mathrm{d}^\ell}(i(1),\ldots,i(j))$, which in particular implies that $(i_{\mathrm{T}'}(1),\ldots,i_{\mathrm{T}'}(j))=(i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))$. Moreover, the sequences V and V' agree until after the first instance of $i(\ell)$, so the ancestral line of $i(\ell)$ is the same in both T and T'. Since $q<\ell$, it likewise holds that $\rho_{\mathrm{T}}(x)=\rho_{\mathrm{T}'}(x)$ and, writing $\rho(x)$ for their common value, that $\mathrm{T}_{\rho(x)}=\mathrm{T}'_{\rho(x)}$. It follows that

$$\begin{aligned}
&\mathbf{P}\left(\mathrm{par}^{b}_{\mathrm{T}}(i(r))\notin \mathrm{T}_{\rho(x)}\ \mid (i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))=(i(1),\ldots,i(j)),E(r,\ell)\right)\\
&=\mathbf{P}\left(\mathrm{par}^{b-1}_{\mathrm{T}'}(i(\ell))\notin \mathrm{T}'_{\rho(x)}\ \mid (i_{\mathrm{T}}(1),\ldots,i_{\mathrm{T}}(j))=(i(1),\ldots,i(j)),E(r,\ell)\right)\\
&=\mathbf{P}\left(\mathrm{par}^{b-1}_{\mathrm{t(W)}}(i(\ell))\notin \mathrm{t}(\mathrm{W})_{\rho(x)}\ \Big|\ (i_{\mathrm{t(W)}}(1),\ldots,i_{\mathrm{t(W)}}(j))=(i(1),\ldots,i(j))\right)\\
&\leq\left(1-\frac{x}{n-2}\right)^{b-1}\,. \qquad (11)
\end{aligned}$$

The second equality holds since under the second conditioning $\mathrm{V}'\in_u \mathcal{S}_{\mathrm{d}^\ell}(i(1),\ldots,i(j))$ and under the third conditioning $\mathrm{W}\in_u\mathcal{S}_{\mathrm{d}^\ell}(i(1),\ldots,i(j))$ and $\rho_{\mathrm{t(W)}}(x)=\rho_{\mathrm{T}}(x)$. The last inequality holds by induction (on b or n or both) and since d^ℓ has length $n-1$.

For the second case, consider the subset $\mathcal{S}^*_{\mathrm{d}}$ of $\mathcal{S}_{\mathrm{d}}(i(1),\ldots,i(j))$ consisting of those vectors v where the first instance of $i(r)$ in v is the immediate successor of the first instance of $i(r-1)$. Let $\mathrm{d}^*=(d^*_1,\ldots,d^*_{n-1})$, where

$$d^*_{i(k)}=\begin{cases}d_{i(k)} & \text{if } k<r-1\\ d_{i(r-1)}+d_{i(r)}-1 & \text{if } k=r-1\\ d_{i(k+1)} & \text{if } r\leq k<n\,,\end{cases}$$

and let $(i^*(1),\ldots,i^*(j-1))=(i(1),\ldots,i(r-1),i(r+1),\ldots,i(j-1))$. From a vector $\mathrm{v}\in\mathcal{S}^*_{\mathrm{d}}$, we can obtain a vector $\mathrm{v}'\in\mathcal{S}_{\mathrm{d}^*}(i^*(1),\ldots,i^*(j-1))$ by deleting the first instance of $i(r-1)$, then changing all other instances of $i(r)$ to $i(r-1)$. Conversely, given a vector $\mathrm{v}'\in\mathcal{S}_{\mathrm{d}^*}(i^*(1),\ldots,i^*(j-1))$, there are $\binom{d_{i(r-1)}+d_{i(r)}-2}{d_{i(r)}-1}$ ways to reconstruct a vector $\mathrm{v}\in\mathcal{S}^*_{\mathrm{d}}$: first insert an entry with value $i(r)$ just after the first instance of $i(r-1)$, then replace $d_{i(r)}-1$ of the other instances of $i(r-1)$ by $i(r)$'s. It follows that for $\mathrm{V}\in_u\mathcal{S}_{\mathrm{d}}$ as above, conditionally given that $\mathrm{V}\in\mathcal{S}^*_{\mathrm{d}}$, then $\mathrm{V}' \in_u \mathcal{S}_{\mathrm{d}^*}(i^*(1), \ldots, i^*(j-1))$. Moreover, the ancestral lines of $i(r-1)$ are the same in both $\mathrm{t}(\mathrm{V})$ and $\mathrm{t}(\mathrm{V}')$, and if $\mathrm{par}_{\mathrm{t}(\mathrm{V})}(i(r)) \notin \mathrm{t}_{\rho(x)}(\mathrm{V})$ then we must have $q < r-1$, so also $\rho_{\mathrm{t}(\mathrm{V})}(x) = \rho_{\mathrm{t}(\mathrm{V}')}(x) =: \rho(x)$, and $\mathrm{t}_{\rho(x)}(\mathrm{V}) = \mathrm{t}_{\rho(x)}(\mathrm{V}')$.

Let $F(r)$ be the event that the first instance of $i_{\mathrm{T}}(r)$ is an immediate successor of the first instance of $i_{\mathrm{T}}(r-1)$, and let $W \in_u \mathcal{S}_{\mathrm{d}^*}$. By the conclusions of the preceding paragraph, it follows that

$$
\begin{aligned}
&\mathbf{P}\left(\mathrm{par}^b_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho(x)} \mid (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j)), F(r)\right) \\
&= \mathbf{P}\left(\mathrm{par}^b_{\mathrm{T}'}(i(r)) \notin \mathrm{T}'_{\rho(x)} \mid (i_{\mathrm{T}'}(1), \ldots, i_{\mathrm{T}'}(j)) = (i(1), \ldots, i(j)), F(r)\right) \\
&= \mathbf{P}\left(\mathrm{par}^{b-1}_{\mathrm{t}(\mathrm{W})}(i^*(r-1)) \notin \mathrm{t}(\mathrm{W})_{\rho(x)} \;\middle|\; (i_{\mathrm{t}(\mathrm{W})}(1), \ldots, i_{\mathrm{t}(\mathrm{W})}(j-1) = (i^*(1), \ldots, i^*(j-1))\right) \\
&\le \left(1 - \frac{x}{n-2}\right)^{b-1}, \qquad (12)
\end{aligned}
$$

the second equality holding since $i^*(r-1) = i(r-1)$ and $\rho_{\mathrm{t}(\mathrm{W})}(x) = \rho(x)$, and the final bound again holding by induction.

For $V \in_u \mathcal{S}_{\mathrm{d}}$, on the event that $(i(1, \mathrm{V}), \ldots, i(j, \mathrm{V})) = (i(1), \ldots, i(j))$, if $\mathrm{par}_{\mathrm{t}(\mathrm{V})}(i(r, \mathrm{V})) \notin \mathrm{t}_{\rho(x)}(\mathrm{V})$ then either $F(r)$ occurs or else $E(r, \ell)$ occurs for some $\ell \in \{q+1, \ldots, r-1\}$, so (11) and (12) together imply that

$$
\begin{aligned}
&\mathbf{P}\left(\mathrm{par}^b_{\mathrm{T}}(v_{\rho_{\mathrm{T}}(y)}) \notin \mathrm{T}_{\rho(x)} \;\middle|\; (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j)), \mathrm{par}_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho(x)}\right) \\
&\mathbf{P}\left(\mathrm{par}^b_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho(x)} \;\middle|\; (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j)) = (i(1), \ldots, i(j)), \mathrm{par}_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho(x)}\right) \\
&\le \left(1 - \frac{x}{n-2}\right)^{b-1}.
\end{aligned}
$$

Summing over possible values for $(i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(j))$ yields that

$$
\mathbf{P}\left(\mathrm{par}^b_{\mathrm{T}}(v_{\rho_{\mathrm{T}}(y)}) \notin \mathrm{T}_{\rho(x)} \;\middle|\; \mathrm{par}_{\mathrm{T}}(i(r)) \notin \mathrm{T}_{\rho(x)}\right) \le \left(1 - \frac{x}{n-2}\right)^{b-1},
$$

which combined with the bound for the case $b = 1$ completes the proof. □

Proof of Lemma 15. The bound is obvious for $n_1 = n-1$ so we assume $n_1 < n-1$. We begin with some fairly elementary combinatorics. Given a set $S = \{s_1, \ldots, s_m\}$, by a *labeled composition of* S *with* k *parts* we mean a weak composition $m_1, \ldots, m_k$ of m together with a permutation $(s_{\sigma(1)}, \ldots, s_{\sigma(m)})$ of S. There are $(m+k-1)!/(k-1)!$ labeled compositions of S with k parts.

We now return to the setting of the lemma. By permuting the entries of d, we may assume that the n_1 entries which equal 1 appear at the end; this does not affect the law of the height of T. In this case, we have $\mathrm{d}' = (d_1, \ldots, d_{n-n_1})$.

Next, given a tree $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$, let $\mathrm{t}^- \in \mathcal{T}_{\mathrm{d}'}$ be obtained from t by suppressing all degree-one vertices. More precisely, t^- is formed from t by replacing any path γ in t all of whose internal vertices have exactly one child by a single edge. If this results in the root having degree 1, we remove the root and its adjacent edge and let its only child be the new root; see Figure 3.

For each tree $\mathrm{t}' \in \mathcal{T}_{\mathrm{d}'}$, there are

$$
\frac{(n-1)!}{(n-n_1-1)!}
$$

trees $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$ with $\mathrm{t}^- = \mathrm{t}'$. This is the case since each such tree t is uniquely determined by t' together with a labeled composition of the n_1 degree-one vertices $\{n-n_1+1, \ldots, n\}$ into $n-n_1$ parts, as follows (see Figure 3). Fix an ordering of the vertices of t' as $v_1, \ldots, v_{n-n_1}$. Then for each $i \in [n-n_1]$, assign the vertices of the i'th part of the composition to v_i's ancestral edge, in the

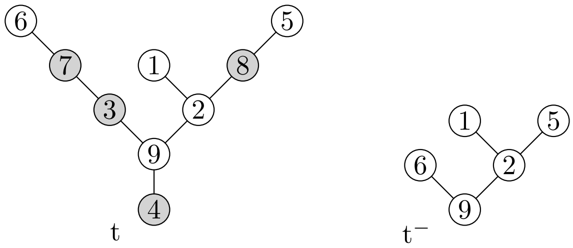


Figure 3. Left: a tree t. Right: the tree t^- obtained from t by suppressing degree-one vertices. Considering the vertices in the order $(1, 2, 5, 6, 9)$, the corresponding labeled composition of the set $\{3, 4, 7, 8\}$ of degree-one vertices is $(), (), (8), (7, 3), (4)$.

same order they appear in the composition. (When v_i is the root, these vertices are simply attached as ancestors of v_i.)

It follows that if $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ then $\mathrm{T}^- \in_u \mathcal{T}_{\mathrm{d}'}$. Moreover, given that $\mathrm{T}^- = \mathrm{t}'$, the conditional distribution of T may be described as follows. Choose any fixed order of the vertices of t'. Then choose a uniformly random labeled composition of $\{n-n_1+1, \ldots, n\}$ with $n-n_1$ parts, and then assign degree-one vertices to the ancestral edges of vertices of t' as specified by the composition, as above.

For any *root-to-leaf path* $U = (u_1, \ldots, u_m)$ in t', it follows from the above construction that conditionally given $\mathrm{T}^- = \mathrm{t}'$, the total number of vertices lying along the corresponding root-to-leaf path in T has the same distribution as the random variable X_m described in the following experiment. Consider an urn with $n - 1 - n_1$ white balls and n_1 black balls. Repeatedly sample without replacement from the urn until it is empty. For $0 \le i \le n-1$ write W_i for the total number of white balls drawn after the i'th sample, and set $W_n = n - n_1$. Then let $X_m = \min(i : W_i = m)$.

For $0 \le i \le n-1$ let W_i^* be Binomial$(i, (n-1-n_1)/(n-1))$; so W_i^* can be thought of as the number of white balls drawn after i samples *with* replacement from the urn of the previous paragraph. By [6, Proposition 20.6], W_i^* is a dilation of W_i, which is to say that there is a coupling (W, W^*) of W_i and W_i^* so that $\mathbf{E}(W^*|W) = W$. This implies that $\mathbf{E}(\phi(W)) \le \mathbf{E}(\phi(W^*))$ for all continuous convex functions of $\phi : \mathbb{R} \to \mathbb{R}$. In particular, concentration inequalities for W_i^* which are proved by bounding exponential moments $\mathbf{E}(e^{\lambda W})$ of W apply without change to W_i. By a Chernoff bound (see, e.g., [11, Theorem 2.1]). It follows that that for $x \ge 2$,

$$\begin{aligned}\mathbf{P}(X_m > xm(n-1)/(n-n_1-1)) &= \mathbf{P}(W_{\lceil xm(n-1)/(n-n_1-1)\rceil} < m) \\ &\le \mathbf{P}(W_{\lceil xm(n-1)/(n-n_1-1)\rceil} \le \mathbf{E}W_{\lceil xm(n-1)/(n-n_1-1)\rceil}/2) \\ &\le e^{-xm(n-1)/(2(n-n_1-1))}\end{aligned}$$

Since $n_1 < n-1$ we have $x(n-1)/(n-n_1-1) \le 2xn/(n-n_1)$, so the preceding bound implies that for $y \ge 4mn/(n-n_1)$,

$$\mathbf{P}(X_m > y) \le e^{-y/4}\,.$$

Write r' for the root of t' and r for the root of T. Since the number of vertices on a fixed root-to-leaf path is one larger than the distance of that leaf from the root, it follows that for any leaf v of t', for any $y \ge 4(\mathrm{dist}_{\mathrm{t}'}(v, r') + 1)n/(n-n_1)$,

$$\mathbf{P}\left(\mathrm{dist}_{\mathrm{T}}(v, r) > y \;\middle|\; \mathrm{T}^- = \mathrm{t}'\right) \le e^{-(y+1)/4}.$$

If t' has height at most $\lceil h \rceil$ then since $h \ge 2$ we always have $4(\mathrm{dist}_{\mathrm{t}'}(v, r') + 1)n/(n-n_1) \le 8hn/(n-n_1)$, so if T' has n_0 leaves and height at most $\lceil h \rceil$ then, applying the above inequality to all root-to-leaf paths in t', a union bound yields that for any $y \ge h \cdot 8n/(n-n_1)$,

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}') > y \mid \mathrm{T}^- = \mathrm{t}') \le n_0 e^{-(y+1)/4} < n_0 e^{-y/4}\,.$$

Finally, since $\mathrm{T}^- \overset{\mathrm{d}}{=} \mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$, it follows that

$$\begin{aligned}\mathbf{P}(\mathrm{ht}(\mathrm{T}) > y) &\le \mathbf{P}(\mathrm{ht}(\mathrm{T}^-) > \lceil h \rceil) + \sum_{\{\mathrm{t}' \in \mathcal{T}_{\mathrm{d}'} : \mathrm{ht}(\mathrm{t}') \le \lceil h \rceil\}} \mathbf{P}(\mathrm{ht}(\mathrm{T}) > y \mid \mathrm{T}^- = \mathrm{t}')\mathbf{P}(\mathrm{T}^- = \mathrm{t}') \\ &\le \mathbf{P}(\mathrm{ht}(\mathrm{T}') > \lceil h \rceil) + n_0 e^{-y/4} . \end{aligned}$$ □

The proof of Proposition 16 requires an auxiliary lemma, which itself has an auxiliary lemma.[3] We first state and prove these two lemmas, then proceed to the proof of Proposition 16.

The auxiliary auxiliary lemma, which is proved by a Chernoff bound-type argument, is similar to [8, Lemma 15].

Lemma 17. *Fix positive constants $x_1, \ldots, x_m$ and let $E_1, \ldots, E_m$ be independent with $E_m \sim \mathrm{Exp}(x_i)$. Fix $t > 0$ and write $S = x_1 \mathbf{1}_{[E_1 \le t]} + \ldots + x_m \mathbf{1}_{[E_m \le t]}$. Then*

$$\mathbf{P}(S < \mathbf{E}S/2) \le \exp(-t\mathbf{E}S/4).$$

Proof. For readability we take $t = 1$, and explain how to handle general $t > 0$ at the end of the proof. Let $X_i = x_i \mathbf{1}_{[E_i \le 1]}$, so $\mathbf{E}X_i = x_i(1 - e^{-x_i})$ and

$$\overline{X}_i := X_i - \mathbf{E}X_i = x_i e^{-x_i} - x_i \mathbf{1}_{[E_i > 1]} ,$$

so that $\overline{S} := S - \mathbf{E}S = \sum_{i=1}^m \overline{X}_i$. An easy computation gives

$$\mathbf{E}\left[e^{-\overline{X}_i}\right] = e^{-x_i e^{-x_i}} \mathbf{E}\left[e^{x_i \mathbf{1}_{[E_i > 1]}}\right] = e^{-x_i e^{-x_i}} (2 - e^{-x_i}) ,$$

and a delicate but elementary computation shows that this quantity is at most $\exp\left(x_i(1 - e^{-x_i})/4\right) = \exp(\mathbf{E}X_i/4)$. Markov's inequality then gives

$$\mathbf{P}(S < \mathbf{E}S/2) \le \mathbf{E}e^{-\overline{S}} e^{-\mathbf{E}S/2} = \prod_{i=1}^m \mathbf{E}e^{-\overline{X}_i} e^{-\mathbf{E}X_i/2} \le \prod_{i=1}^m e^{-\mathbf{E}X_i/4} = e^{-\mathbf{E}S/4}.$$

This proves the lemma in the case $t = 1$. For general t, the proof works identically by instead showing that $\mathbf{E}e^{-t\overline{X}_i} \le e^{t\mathbf{E}X_i/4}$. □

We use Lemma 17 to prove a lower tail bound for sums of the form $\sum_{1 \le j \le k} (d_{i_{\mathrm{T}}(j)} - 1)$.

Lemma 18. *Fix a degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$ and let $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$. Then taking $c = (1 - e^{-2})/8$, for all $t \in [0, 1]$, with $\sigma = \sigma_{\mathrm{d}} = (n^{-1} \sum_{j \in [n]} d_j(d_j - 1))^{1/2}$, we have*

$$\mathbf{P}\Big(\sum_{1 \le j \le nt} (d_{i_{\mathrm{T}}(j)} - 1) \le \frac{c\sigma^2 nt}{1 + \sigma n^{1/2} t} \Big) \le \exp(-\tfrac{3}{16} tn) + \exp\left(-\frac{c}{4} \frac{n\sigma^2 t^2}{1 + \sigma n^{1/2} t} \right).$$

Proof. We assume d is compressed, so that $m := |\{i \in [n] : d_i \ne 0\}| = \max(i \in [n] : d_i \ne 0)$ is the number of non-leaf vertices of T. Let $E_1, \ldots, E_m$ be independent with $E_i \sim \mathrm{Exp}(d_i)$, and let $(I_1, \ldots, I_m)$ be the permutation of $(1, \ldots, m)$ for which $E_{I_1} < \ldots < E_{I_m}$. Then $d_{I_1}, \ldots, d_{I_m}$ is a size-biased permutation of $(d_1, \ldots, d_m)$. In other words, $(I_1, \ldots, I_m)$ has the same distribution as $(i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(m))$ for $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$. For the remainder of the proof we may therefore assume that T and $E_1, \ldots, E_m$ are coupled so that

$$(I_1, \ldots, I_m) = (i_{\mathrm{T}}(1), \ldots, i_{\mathrm{T}}(m)).$$

[3]"So, Nat'ralists observe, a Flea/Hath smaller Fleas that on him prey/And these have smaller yet to bite 'em/And so proceed *ad infinitum*." From *On Poetry: A Rapsody*, Johnathan Swift, 1733.

Now let $s = t/2$ and define

$$N_s := |\{i \in [m] : E_i \le s\}| = \sum_{i\in[m]} \mathbf{1}_{[E_i \le s]} = \max(j : E_{I_j} \le s)\,,$$

$$X_s := \sum_{i\in[m]} (d_i - 1)\mathbf{1}_{[E_i\le s]} = \sum_{j\in[N_s]} (d_{I_j} - 1)$$

and

$$M = \frac{1-e^{-2}}{4}\frac{n\sigma^2 s}{1+\sigma n^{1/2}s} > \frac{1-e^{-2}}{8}\frac{n\sigma^2 t}{1+\sigma n^{1/2}t} = \frac{cn\sigma^2 t}{1+\sigma n^{1/2}t}$$

Under the above coupling,

$$\sum_{1\le j\le nt} (d_{i_{\mathrm{T}}(j)} - 1) = \sum_{1\le j\le nt} (d_{I_j} - 1).$$

It follows that if $\sum_{1\le j\le nt}(d_{i_{\mathrm{T}}(j)} - 1) \le M$ then either $nt \le N_s$ or $X_s \le M$, so

$$\begin{aligned}\mathbf{P}\Big(\sum_{1\le j\le nt} (d_{i_{\mathrm{T}}(j)} - 1) \le \frac{c\sigma^2 nt}{1+\sigma n^{1/2}t}\Big) &\le \mathbf{P}\Big(\sum_{1\le j\le nt} (d_{i_{\mathrm{T}}(j)} - 1) \le M\Big)\\ &\le \mathbf{P}(N_s \ge nt) + \mathbf{P}(X_s \le M), \qquad (13)\end{aligned}$$

where in the first line we have used our lower bound on M.

To make use of this inequality, we first bound $\mathbf{E}X_s = \sum_{i\in[m]}(d_i - 1)(1 - e^{-d_i s})$ from below. We consider two cases, depending on whether $\sigma^2 = n^{-1}\sum_{i\in[m]} d_i(d_i - 1)$ is dominated by the contributions from small or large summands.

Fix $r \in (0, 1)$. First suppose that $\sum_{i\in[m]:d_i s\le 2} d_i(d_i - 1) \ge rn\sigma^2$. For $d_i s \le 2$ we have $(1 - e^{-d_i s}) \ge (1 - e^{-2})d_i s/2$, so

$$\sum_{i\in[m]:d_i s\le 2} (d_i - 1)(1 - e^{-d_i s}) \ge (1 - e^{-2})\frac{s}{2} \sum_{i\in[m]:d_i s\le 2} d_i(d_i - 1) \ge (1 - e^{-2})\frac{s}{2} rn\sigma^2\,.$$

Next suppose that $\sum_{i\in[m]:d_i s\le 2} d_i(d_i - 1) < rn\sigma^2$. For i such that $d_i s > 2$ we have $1 - e^{-d_i s} > 1 - e^{-2}$ and $(d_i - 1)/d_i > (2 - s)/2$, so using that $\sum_i x_i > (\sum_i x_i^2)^{1/2}$ for non-negative x_i, we obtain

$$\begin{aligned}\sum_{i\in[m]:d_i s>2} (d_i - 1)(1 - e^{-d_i s}) &\ge (1 - e^{-2})\Big(\sum_{i\in[m]:d_i s>2} (d_i - 1)^2\Big)^{1/2}\\ &\ge (1 - e^{-2})\Big(\frac{2-s}{2}\sum_{i\in[m]:d_i s>2} d_i(d_i - 1)\Big)^{1/2}\\ &> (1 - e^{-2})\frac{2-s}{2}(1 - r)n^{1/2}\sigma\,.\end{aligned}$$

Taking $r = (2 - s)/(2 - s + \sigma n^{1/2}s)$ makes the lower bounds on $\sum_{i\in[m]:d_i s\le 2} d_i(d_i - 1)$ and on $\sum_{i\in[m]:d_i s>2} d_i(d_i - 1)$ in the two cases equal, and yields the bound

$$\mathbf{E}X_s = \sum_{i\in[m]} (d_i - 1)(1 - e^{-d_i s}) \ge \frac{1-e^{-2}}{2}\frac{(2-s)n\sigma^2 s}{2-s+\sigma n^{1/2}s} \ge \frac{1-e^{-2}}{2}\frac{n\sigma^2 s}{1+\sigma n^{1/2}s} = 2M\,,$$

the last inequality holding since $s = t/2 \le 1/2$. By Lemma 17 it follows that

$$\mathbf{P}(X_s \le M) \le \mathbf{P}(X_s \le \mathbf{E}X_s/2) \le \exp(-s\mathbf{E}X_s/4) \le \exp(-tM/4)\,. \qquad (14)$$

Next, note that

$$\mathbf{E}N_s = \sum_{i\in[m]} \mathbf{P}(E_i \le s) = \sum_{i\in[m]} (1 - e^{-d_i s}) \le \sum_{i\in[m]} d_i s = s(n - 1) < sn\,.$$

Since N_s is a sum of $[0,1]$-valued random variables, Bernstein's inequality [7, 9] then gives

$$\mathbf{P}(N_s \geq tn) = \mathbf{P}(N_s \geq 2sn) \leq \exp(-\tfrac{3}{8}sn) = \exp(-\tfrac{3}{16}tn)\,. \tag{15}$$

Using the bounds (14) and (15) in (13) now yields

$$\mathbf{P}\Big(\sum_{1\leq j\leq nt}(d_{i_{\mathrm{T}}(j)}-1) \leq \frac{c\sigma^2 nt}{1+\sigma n^{1/2}t}\Big) \leq \exp(-\tfrac{3}{16}tn) + \exp(-tM/4).$$

In view of the lower bound on M, this proves the bound claimed in the lemma. □

Proof of Proposition 16. First recall that the non-leaf vertices contained in $\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}$ are precisely $i(1),\ldots,i(k)$, with k minimal such that $\sum_{j=1}^{k}(d_{i(j)}-1) \geq \alpha\sigma n^{1/2}$. Now, if $\mathrm{ht}\left(\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}\right) > \frac{bn^{1/2}}{2\sigma}$, then $\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}$ must contain more than $\frac{1}{2}bn^{1/2}/\sigma$ non-leaf vertices. This implies that $k > \frac{1}{2}bn^{1/2}/\sigma$, which by the definition of k yields that $\sum_{1\leq j\leq \frac{1}{2}bn^{1/2}/\sigma}\left(d_{i_{\mathrm{T}}(j)}-1\right) < \alpha\sigma n^{1/2}$. It follows from this inclusion of events that

$$\mathbf{P}\left(\mathrm{ht}\left(\mathrm{T}_{\rho(\alpha\sigma n^{1/2})}\right) > \tfrac{bn^{1/2}}{2\sigma}\right) \leq \mathbf{P}\left(\sum_{1\leq j\leq \frac{1}{2}bn^{1/2}/\sigma}\left(d_{i_{\mathrm{T}}(j)}-1\right) < \alpha\sigma n^{1/2}.\right).$$

Now fix $b \geq 1$ and set $t = b/(2\sigma n^{1/2})$ so that $tn = bn^{1/2}/(2\sigma)$. Then with $c = (1-e^{-2})/8$ as in the statement of Lemma 18, we have $c = 3\alpha$, so since $b \geq 1$,

$$\alpha\sigma n^{1/2} = \frac{c\sigma n^{1/2}}{3} \leq c\sigma n^{1/2}\frac{b/2}{1+b/2} = \frac{c\sigma^2 nt}{1+\sigma n^{1/2}t}\,.$$

It follows that

$$\mathbf{P}\left(\sum_{1\leq j\leq \frac{1}{2}bn^{1/2}/\sigma}\left(d_{i_{\mathrm{T}}(j)}-1\right) < \alpha\sigma n^{1/2}\right) \leq \mathbf{P}\left(\sum_{1\leq j\leq nt}(d_{i_{\mathrm{T}}(j)}-1) \leq \frac{c\sigma^2 nt}{1+\sigma n^{1/2}t}\right),$$

and Lemma 18 now yields the result. □

3. Bienaymé trees and simply generated trees: proofs of Theorems 1, 2 and 3.

In this section we use Theorems 6 and 7 to prove Theorems 1, 2 and 3, and additionally to prove two conjectures from [10] on *simply generated trees*; this class of trees is defined as follows. Fix non-negative real weights $\mathrm{w} = (w_k, k \geq 0)$ with $w_0 > 0$. Given a finite plane tree t, we define the weight of t to be

$$\mathrm{w}(\mathrm{t}) = \prod_{v\in v(\mathrm{t})} w_{d_{\mathrm{t}}(v)}.$$

Next, for positive integers n, let

$$Z_n = Z_n(\mathrm{w}) = \sum_{\{\text{plane trees t}:|v(\mathrm{t})|=n\}} \mathrm{w}(\mathrm{t}).$$

If $Z_n > 0$, then we define a random tree $\mathrm{T}_{\mathrm{w},n}$ by setting

$$\mathbf{P}(\mathrm{T}_{\mathrm{w},n} = \mathrm{t}) = \frac{\mathrm{w}(\mathrm{t})}{Z_n}$$

for plane trees t with $|v(\mathrm{t})| = n$. The random tree $\mathrm{T}_{\mathrm{w},n}$ is called a simply generated tree of size n with weight sequence w. Note that if $\sum_{k\geq 0} w_k = 1$, then w is an offspring distribution, and $\mathrm{T}_{\mathrm{w},n}$ is indeed distributed as a Bienaymé tree with offspring distribution w, conditioned to have size $n-$

so this notation agrees with (but generalizes) that from earlier in the paper. From this we also see that conditioned Bienaymé trees are a subclass of simply generated trees; we will use this fact below.

Let $\Phi(z) = \Phi_{\mathrm{w}}(z) = \sum_{k \geq 0} w_k z^k$ be the generating function of the weight sequence w, and let $\rho = \rho_{\mathrm{w}} = \sup(s \geq 0 : \Phi(s) < \infty)$ be its radius of convergence. For $s > 0$ such that $\Phi(s) < \infty$, we let

$$\Psi(s) = \Psi_{\mathrm{w}}(s) = \frac{t\Phi'(s)}{\Phi(s)} = \frac{\sum_{k \geq 0} k w_k s^k}{\sum_{k \geq 0} w_k s^k}.$$

If $\Phi(\rho) = \infty$ then also define

$$\Psi(\rho) = \Psi_{\mathrm{w}}(\rho) = \lim_{s \uparrow \rho} \Psi(s);$$

by Lemma 3.1(i) in [10], Ψ is strictly increasing on $[0, \rho)$ so this limit exists. Let $\nu = \nu_{\mathrm{w}} = \Psi_{\mathrm{w}}(\rho)$, and define

$$\tau = \tau_{\mathrm{w}} := \begin{cases} \rho & \text{if } \nu_{\mathrm{w}} \leq 1 \\ \Psi^{-1}(1) & \text{if } \nu_{\mathrm{w}} > 1\,. \end{cases} \tag{16}$$

Note that if $\nu_{\mathrm{w}} > 1$ then $\tau \in [0, \rho)$. Define $\hat{\sigma}^2 = \hat{\sigma}^2_{\mathrm{w}} = \tau\Psi'(\tau)$. For later use, we note that if $\mathrm{w} = (w_k, k \geq 0)$ is a probability distribution with finite mean, so $\sum_{k \geq 0} w_k = 1$ and $|\mathrm{w}|_1 = \sum_{k \geq 0} k w_k < \infty$, then we always have $\rho \geq 1$, so

$$\nu = \Psi_{\mathrm{w}}(\rho) \geq \Psi_{\mathrm{w}}(1) = \sum_{k \geq 0} k w_k = |\mathrm{w}|_1, \text{ and } \Psi'_{\mathrm{w}}(1) = \sum_{k \geq 0} k(k-1) w_k = |\mathrm{w}|_2^2 - |\mathrm{w}|_1\,. \tag{17}$$

The following theorem resolves Conjecture 21.5 and Problem 21.7 from [10].

Theorem 19. *Let* $\mathrm{w} = (w_k, k \geq 0)$ *be a weight sequence with* $w_0 > 0$ *and with* $w_k > 0$ *for some* $k \geq 2$. *If either* $\hat{\sigma}^2 = \infty$ *or* $\nu < 1$ *then* $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) \to 0$, *where the convergence is in both probability and expectation, as* $n \to \infty$ *along integers* n *such that* $Z_n(\mathrm{w}) > 0$.

We also prove the following, more quantitative theorem.

Theorem 20. *Fix* $\epsilon \in [0, 1)$. *Then, there exist constants* $c_1 = c_1(\epsilon)$ *and* $c_2 = c_2(\epsilon)$ *such that the following holds. Let* $\mathrm{w} = (w_k, k \geq 0)$ *be a weight sequence with* $w_0 > 0$ *and* $w_k > 0$ *for some* $k \geq 2$ *and* $\frac{w_1 \tau}{\Phi(\tau)} < 1 - \epsilon$. *For any* $t > 1$ *and any* n *large enough with* $Z_n(\mathrm{w}) > 0$ *and*

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > t n^{1/2}\right) \leq c_1 \exp(-c_2 t^2).$$

This theorem has the following immediate corollary.

Corollary 21. *For any weight sequence* w *on* $\mathbb{N}$ *with* $w_0 > 0$ *and* $w_k > 0$ *for some* $k \geq 2$, $\mathbf{E}[\mathrm{ht}(\mathrm{T}_{\mathrm{w},n})] = O(n^{1/2})$, *and, more generally, for any fixed* $r < \infty$, $\mathbf{E}[\mathrm{ht}(\mathrm{T}_{\mathrm{w},n})^r] = O(n^{r/2})$ *as* $n \to \infty$ *over all* n *such that* $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$. □

In this section we will make use of the following notation. For w a weight sequence, let $\mathrm{D} = \mathrm{D}(\mathrm{w}, n) = (D_1, \ldots, D_n)$ be a random degree sequence with the following law. Let T be a simply generated tree of size n with weight sequence w. Conditionally given T, let $\hat{\mathrm{T}}$ be the random tree obtained as follows: label the vertices of T by a uniformly random permutation of $[n]$, then forget about the plane structure. Let D_i be equal to the degree of i in $\hat{\mathrm{T}}$. Then, for $k \in \mathbb{N}$, let $N_k = N_k(\mathrm{w}, n) = |\{i \in [n] : D_i = k\}|$ be the number of entries of D which equal k, so $N_k = n_k(\mathrm{D}(\mathrm{w}, n))$ and $\sum_{k=0}^{\infty} N_k = n$.

Our proofs will exploit two distributional identities. The first is the following. Let $\mathrm{T}_{\mathrm{D}(\mathrm{w},n)} \in_u \mathcal{T}_{\mathrm{D}(\mathrm{w},n)}$, by which we mean that, conditionally given that $\mathrm{D}(\mathrm{w}, n) = \mathrm{d}$, we have $\mathrm{T}_{\mathrm{D}(\mathrm{w},n)} \in_u \mathcal{T}_{\mathrm{d}}$. Then $\hat{\mathrm{T}} \stackrel{d}{=} \mathrm{T}_{\mathrm{D}(\mathrm{w},n)}$ by Proposition 8, so $\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) \stackrel{d}{=} \mathrm{ht}(\mathrm{T}_{\mathrm{D}(\mathrm{w},n)})$. The second is the fact that for any weight sequence $\mathrm{w} = (w_k, k \geq 0)$ and any constants $a, b > 0$, the weight sequence $\hat{\mathrm{w}}$ with $\hat{w}_k = a b^k w_k$ is *equivalent* to w, i.e., it satisfies that $\mathrm{T}_{\mathrm{w},n} \stackrel{\mathrm{d}}{=} \mathrm{T}_{\hat{\mathrm{w}},n}$ for all n for which either (and thus both) of the random trees are defined. This is an immediate consequence of the formula (1) for the distribution of $\mathrm{T}_{\mathrm{w},n}$.

These two equalities in distribution imply that if we obtain good control over the asymptotic behaviour of $(N_k(\hat{w}, n), k \geq 0)$ for some weight sequence $\hat{w}$ which is equivalent to w, then we can use Theorems 6 and 7, on the heights of trees with given degree sequences, to prove tail bounds for $\mathrm{ht}(\mathrm{T}_{\mathrm{w},n})$. To obtain such control, we rely on the following result of Svante Janson.

Theorem 22 ([10], Theorem 11.4)**.** *Let* $\mathrm{w} = (w_k, k \geq 0)$ *be a weight sequence with* $w_0 > 0$ *and* $w_k > 0$ *for some* $k \geq 2$*. Write* $\tau = \tau_{\mathrm{w}}$ *and define*

$$\hat{\mu}_k = \hat{\mu}_k(\mathrm{w}) = \frac{w_k \tau^k}{\Phi(\tau)}$$

for $k \in \mathbb{N}$*. Then* $\hat{\mu} = (\hat{\mu}_k, k \geq 0)$ *is a probability distribution with mean* $m = 1 \wedge \nu$ *and variance* $\hat{\sigma}^2$*. For* $k \geq 0$ *define* $N_k = n_k(\mathrm{D}(\mathrm{w}, n))$*, as above. Then for every* $\epsilon > 0$ *there exists* $c(\epsilon) > 0$ *such that for all* n *sufficiently large, for every integer* $k \geq 0$*,*

$$\mathbf{P}\left(\left|\frac{N_k(\mathrm{w}, n)}{n} - \hat{\mu}_k\right| > \epsilon\right) < e^{-cn}.$$

The exponentially small error bounds stated above are not made explicit in the statement of [10, Theorem 11.4], but are recorded in the course of its proof (see [10, pages 163-164]).

Before we prove Theorems 19, 20, 1 and 3, we first illustrate that the requirements in Theorem 3, that $1 - \mu_0 - \mu_1$ and $\mu_0/(\mu_0 + \mu_1)$ are bounded from below, are necessary. To accomplish this we will consider probability distributions $\mu = \mu^{p,q}$ of the form

$$\mu_0 = q(1-p) \qquad \mu_1 = (1-q)(1-p) \qquad \mu_2 = p\,.$$

Claim 23. *For any* $x > 0$ *we have that for any* $q > 0$, $\lim_{p \downarrow 0} \liminf_{n \to \infty} \mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mu^{p,q},n}) > x n^{1/2}) = 1$ *and for any* $p > 0$, $\lim_{q \downarrow 0} \liminf_{n \to \infty} \mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mu^{p,q},n}) > x n^{1/2}) = 1$

Proof. We apply Theorem 22 with $\mathrm{w} = \mathrm{w}^{p,q} = \mu^{p,q}$. Elementary computation shows that the probability distribution $\hat{\mu} = \hat{\mu}^{p,q}$ from Theorem 22 is given by

$$\hat{\mu}_1 = \hat{\mu}_1^{p,q} = \frac{(1-q)\sqrt{1-p}}{(1-q)\sqrt{1-p} + 2\sqrt{p}\sqrt{q}}$$

and $\hat{\mu}_0 = \hat{\mu}_2 = (1 - \hat{\mu}_1)/2$. Since $\hat{\mu}$ is equivalent to μ, it follows that that $T_{\mu^{p,q},n} \overset{\mathrm{d}}{=} T_{\hat{\mu}^{p,q},n}$ for all n. Write $\sigma^{p,q} = |\hat{\mu}^{p,q}|_2 - 1$ for the standard deviation of $\hat{\mu}^{p,q}$. Using this equivalence in distribution together with the convergence of the search-depth process for large critical random Bienaymé trees [5, Theorem 23], it follows that

$$\frac{2}{n^{1/2}}\mathrm{ht}(T_{\mu^{p,q}}, n) \overset{\mathrm{d}}{=} \frac{2}{n^{1/2}}\mathrm{ht}(T_{\hat{\mu}^{p,q}}, n) \overset{\mathrm{d}}{\to} \frac{1}{\sigma^{p,q}} \sup_{0 \leq t \leq 1} B(t)\,,$$

where B is a standard Brownian excursion. Finally, since $\hat{\mu}_1^{p,q} \to 1$ (and thus $\sigma^{p,q} \to 0$) as either $p \to 0$ or $q \to 0$, and $\mathbf{P}(\sup_{0 \leq t \leq 1} B(t) > 0) = 1$, it follows that $(\sigma^{p,q})^{-1} \sup_{0 \leq t \leq 1} B(t) \to \infty$ in probability as either $p \to 0$ or $q \to 0$. The result follows. □

In our proof of Theorem 19, we make use of the following consequence of Theorem 22.

Corollary 24. *If* $\nu < 1$*, or if* $\hat{\sigma}^2 = \infty$*, then for each* $C > 0$ *there exists* $c = c(\mathrm{w}, C) > 0$ *such that for all* n *sufficiently large,*

$$\mathbf{P}\left(\sigma_{\mathrm{D}(\mathrm{w},n)} \leq C\right) \leq e^{-cn}\,.$$

Proof. First suppose that $\nu < 1$; in this case $\sum_{i=1}^{\infty} \hat{\mu}_k = \nu$. Suppose without loss of generality that $C \geq 1$ and set $K = \frac{2C}{1-\nu} + 1$. Then, $\sum_{k=1}^{K-1} \hat{\mu}_k \leq \nu$, so by Theorem 22, there is $c = c(\mathrm{w}, C)$ such that with $N_k = n_k(\mathrm{D}(\mathrm{w}, n))$,

$$\mathbf{P}\left(\sum_{k=0}^{K-1} N_k > \frac{1+\nu}{2} n\right) \leq e^{-cn}$$

for all n sufficiently large. But on the event that $\sum_{k=0}^{K-1} N_k \leq \frac{1+\nu}{2} n$, we have $\#\{1 \leq i \leq n : D_i \geq K\} \geq \frac{1-\nu}{2} n$, which implies that

$$\sum_{i=1}^{n} D_i(D_i - 1) \geq \tfrac{1-\nu}{2} K(K-1)n \geq C^2 n,$$

so $\sigma_{\mathrm{D}(\mathrm{w},n)} > C$. This proves the claim in the case that $\nu < 1$.

Next suppose that $\hat{\sigma}^2 = \infty$. In this case there exists $K \in \mathbb{N}$ such that $\sum_{k=0}^{K} \hat{\mu}_k k(k-1) > 2C$. Since $\sigma_{\mathrm{D}(\mathrm{w},n)} = n^{-1} \sum_{i=1}^{n} D_i(D_i - 1) \geq n^{-1} \sum_{k=0}^{K} N_k k(k-1)$, Theorem 22 then implies that there exists $c = c(\mathrm{w}, C)$ such that

$$\mathbf{P}\left(\sigma_{\mathrm{D}(\mathrm{w},n)} \leq C\right) \leq \mathbf{P}\left(\sum_{k=0}^{K} N_k k(k-1) \leq Cn\right) \leq e^{-cn}$$

for all n sufficiently large. This proves the claim in the case that $\hat{\sigma}^2 = \infty$. □

In the next proof we write $x \vee y := \max(x, y)$.

Proof of Theorem 19. Fix $0 < \epsilon \leq 2^{-14}$. Let w be a weight sequence with $\nu = 1$ and $\hat{\sigma}^2 = \infty$, or with $\nu < 1$, and fix n such that $Z_n(\mathrm{w}) > 0$. Let $\hat{\mu}_1 = \hat{\mu}_1(\mathrm{w})$ and let K be large enough such that $2\log(k+1)/k < \epsilon^2$ and such that $\log(k+1) \geq (\frac{1}{2} \log \frac{2}{(1-\hat{\mu}_1)}) \vee 2^{14}$ for all $k \geq K$. Then let

$$\mathcal{D}_n = \left\{\text{degree sequences } \mathrm{d} = (d_1, \ldots, d_n) : n_1(\mathrm{d}) \leq \frac{n(1+\hat{\mu}_1)}{2}, \sigma_{\mathrm{d}} \geq K\right\}.$$

By Theorem 22 and Corollary 24, there exists $c = c(\mathrm{w}) > 0$ such that for all n sufficiently large,

$$\mathbf{P}(\mathrm{D} \notin \mathcal{D}_n) \leq e^{-cn}.$$

Moreover, for any $\mathrm{d} \in \mathcal{D}_n$, with $(\sigma')^2 = \sigma_{\mathrm{d}}^2 n/(n - n_1(\mathrm{d}))$ as in Theorem 7, we have $\log(\sigma' + 1) \leq \log(\sigma_{\mathrm{d}} + 1) + \frac{1}{2} \log \frac{2}{(1-\hat{\mu}_1)} < 2\log(\sigma_{\mathrm{d}} + 1)$, so $\log(\sigma' + 1)/\sigma_{\mathrm{d}} \leq \epsilon^2$. Also, $\sigma' \geq \sigma_{\mathrm{d}}$, so $\log(\sigma' + 1) \geq 2^{14}$.

Fix any $t \geq \epsilon$. We apply Theorem 7 with $x = \epsilon^{-2} t \geq 2^{14}$ to obtain that for any n and any $\mathrm{d} \in \mathcal{D}_n$,

$$\begin{aligned}
\mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > tn^{1/2} \,\middle|\, \mathrm{D} = \mathrm{d}\right) &= \mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > x\epsilon^2 n^{1/2} \,\middle|\, \mathrm{D} = \mathrm{d}\right) \\
&\leq \mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > xn^{1/2} \frac{\log(\sigma' + 1)}{\sigma_{\mathrm{d}}} \,\middle|\, \mathrm{D} = \mathrm{d}\right) \\
&= \mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{d}}) > xn^{1/2} \frac{\log(\sigma' + 1)}{\sigma_{\mathrm{d}}}\right) \\
&\leq 4\exp(-x\log(\sigma' + 1)/2^{14}) \leq 4\exp(-\epsilon^{-2} t).
\end{aligned}$$

Since

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{D}}) > tn^{1/2} \,\middle|\, \mathrm{D} \in \mathcal{D}_n\right) = \sum_{\mathrm{d} \in \mathcal{D}_n} \mathbf{P}\left(\mathrm{ht}(\mathrm{T}_n) > tn^{1/2} \,\middle|\, \mathrm{D} = \mathrm{d}\right) \mathbf{P}(\mathrm{D} = \mathrm{d} | \mathrm{d} \in \mathcal{D}_n),$$

it follows that for all n sufficiently large,

$$\begin{aligned}\mathbf{E}\left[\mathrm{ht}(\mathrm{T}_{\mathrm{w},n})\mathbb{1}_{\{\mathrm{ht}(\mathrm{T}_{\mathrm{w},n})>\epsilon n^{1/2}\}}\right] &\le \mathbf{E}\left[\mathrm{ht}(\mathrm{T}_{\mathrm{D}})\mathbb{1}_{\{\mathrm{ht}(\mathrm{T}_{\mathrm{D}})>\epsilon n^{1/2}\}}\middle|\,\mathrm{D}\in\mathcal{D}_n\right] + n\mathbf{P}(\mathrm{D}\notin\mathcal{D}_n)\\ &\le n^{1/2}\left[\epsilon + \int_\epsilon^\infty \mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{D}}) > tn^{1/2}\middle|\,\mathrm{D}\in\mathcal{D}_n\right)dt\right] + o(1)\\ &\le n^{1/2}\left[\epsilon + 4\int_\epsilon^\infty \exp(-\epsilon^{-2}t)dt\right] + o(1)\\ &= n^{1/2}\left[\epsilon + 4\epsilon^2\exp(-\epsilon^{-1}) + o(1)\right].\end{aligned}$$

We can pick ϵ arbitrary small, so the statement follows. □

Proof of Theorem 20. Let $\mathrm{D} = \mathrm{D}(\mathrm{w}, n)$ and, as above, write $N_k = n_k(\mathrm{D})$. Then, by Theorem 22, there is $c = c(\epsilon) > 0$ such that for all n sufficiently large,

$$\mathbf{P}\left(1 - \frac{N_1}{n} < \epsilon/2\right) < e^{-cn}.$$

Let $\mathcal{D}_n$ be the set of degree sequences d with $\mathbf{P}(\mathrm{D} = \mathrm{d}) > 0$ such that $1 - n_1(\mathrm{d})/n \ge \epsilon/2$; so $\mathbf{P}(\mathrm{D} \notin \mathcal{D}_n) < e^{-cn}$. By Theorem 6, there are $c_1 = c_1(\epsilon)$ and $c_2 = c_2(\epsilon)$ such that for any $\mathrm{d} \in \mathcal{D}_n$, for all $t > 1$,

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}_n) > tn^{1/2}\middle|\,\mathrm{D} = \mathrm{d}\right) \le c_1\exp\left(-c_2t^2\right).$$

The theorem now follows from the observation that

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{w},n}) > tn^{1/2}\right) \le \mathbf{P}(\mathrm{D}\notin\mathcal{D}_n) + \mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{D}}) > tn^{1/2}\middle|\,\mathrm{D}\in\mathcal{D}_n\right)$$

and the fact that

$$\mathbf{P}\left(\mathrm{ht}(\mathrm{T}_{\mathrm{D}}) > tn^{1/2}\middle|\,\mathrm{D}\in\mathcal{D}_n\right) = \sum_{\mathrm{d}\in\mathcal{D}_n}\mathbf{P}\left(\mathrm{ht}(\mathrm{T}_n) > tn^{1/2}\middle|\,\mathrm{D} = \mathrm{d}\right)\mathbf{P}(\mathrm{D} = \mathrm{d}|\mathrm{d}\in\mathcal{D}_n).$$ □

Proof of Theorems 1 and 2. We use the fact that conditioned Bienaymé trees are a special case of simply generated trees.

First, if $|\mu|_1 \le 1$ and $|\mu|_2 = \infty$ then $\mu_0 > 0$ and $\mu_k > 0$ for some $k \ge 2$. Next, since $|\mu|_2 = \infty$, for all $t > 0$ we have $\sum_{k\ge0} e^{tk}\mu_k = \infty$. This implies that the generating function $\Phi = \Phi_\mu$ has radius of convergence $\rho = \rho_\mu = 1$, so $\nu_\mu = \Psi(\rho) = \Psi(1) = |\mu|_1 \le 1$. This implies that $\tau = \tau_\mu$ defined by (16) satisfies $\tau = \rho = 1$ and so by (17) we have

$$\hat\sigma^2 = \hat\sigma^2_\mu := \tau\Psi'(\tau) = \Psi'(1) = |\mu|_2^2 - |\mu|_1 = \infty.$$

Theorem 19 now implies that $n^{1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ in probability and expectation along integers n such that $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$. This proves Theorem 1.

Similarly, if $|\mu|_1 < 1$ and $\sum_{k\ge0} e^{tk}\mu_k = \infty$ for all $t > 0$ then $\mu_0 > 0$ and $\mu_k > 0$ for some $k \ge 2$. Moreover, Φ_μ again has radius of convergence $\rho_\mu = 1$, and so $\nu = \Psi(1) = |\mu|_1 < 1$. In this case Theorem 19 also implies that $n^{-1/2}\mathrm{ht}(\mathrm{T}_{\mu,n}) \to 0$ in probability and expectation along integers n such that $\mathbf{P}(|\mathrm{T}_\mu| = n) > 0$. This proves Theorem 2. □

Proof of Theorem 3. We again use the fact that Bienaymé trees are special cases of simply generated trees. We aim to apply Theorem 20, so proceed to verify the assumptions of that result. The assumptions on μ_0 and μ_1 in particular imply that $\mu_0 > 0$ and that $\mu_k > 0$ for some $k \ge 2$, so that requirement of the theorem is satisfied.

Recall from (16) that $\tau = \rho$ if $\Psi(\rho) \le 1$, and $\tau = \Psi^{-1}(1)$ if $\Psi(\rho) > 1$. With $\hat\mu_k = \mu_k\tau^k/\Phi(\tau)$ as in Theorem 22, we then have

$$\hat\mu_1 = \frac{\mu_1\tau}{\Phi(\tau)} = \frac{\mu_1\tau}{\sum_{k\ge0}\mu_k\tau^k}.$$

If $\tau \geq 1$ then the denominator is at least $\mu_1\tau + (1-\mu_1-\mu_0)\tau > \mu_1\tau + \epsilon\tau$, since by assumption $(1-\mu_1-\mu_0) > \epsilon$. This yields that

$$\hat{\mu}_1 < \frac{\mu_1\tau}{\mu_1\tau + \epsilon\tau} < \frac{(1-\epsilon)\tau}{(1-\epsilon)\tau + \epsilon\tau} = 1-\epsilon\,,$$

the second inequality holding since $\mu_1 < 1-\epsilon$ and $x/(x+\epsilon\tau)$ is increasing in x. On the other hand, if $\tau < 1$ then

$$\hat{\mu}_1 = \frac{\mu_1\tau}{\Phi(\tau)} < \frac{\mu_1\tau}{\mu_0 + \mu_1\tau} < \frac{\mu_1}{\mu_0+\mu_1} < 1-\epsilon\,,$$

the last inequality holding by the assumptions of the theorem. In either case we have $\hat{\mu}_1 < 1-\epsilon$, so the result follows by Theorem 20. □

4. **Stochastic domination results**

This section presents the proof of Theorem 9. The following decomposition is a key input to the proof. Given a tree t, let $f(\mathrm{t})$ be the unordered set of rooted trees obtained from t by removing all edges from vertices 1 and 2 to their children. Also, write t^{12} for the tree obtained from t by swapping the labels of vertices 1 and 2. Then we say that $\mathrm{t} \sim \mathrm{t}'$ if either t and t′ have the same root and $f(\mathrm{t}) = f(\mathrm{t}')$ or t^{12} and t′ have the same root and $f(\mathrm{t}^{12}) = f(\mathrm{t}')$. (See Figures 4 and 5) for examples.)

We prove Theorem 9 using the following two propositions.

Proposition 25. *Let $\mathcal{C}$ be an equivalence class for the equivalence relation $\sim$. Fix a degree sequence $\mathrm{d} = (d_1,\ldots,d_n)$ with $d_2 \geq 1$ and let $\mathrm{d}' = (d_1+1, d_2-1,\ldots,d_n)$. Then with $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$,*

$$\mathbf{P}(\mathrm{T} \in \mathcal{C}) = \mathbf{P}(\mathrm{T}' \in \mathcal{C}).$$

Proposition 26. *Fix a degree sequence $\mathrm{d} = (d_1,\ldots,d_n)$ with $d_2 \geq 1$ and let $\mathrm{d}' = (d_1+1, d_2-1,\ldots,d_n)$. Then for any $\sim$-equivalence class $\mathcal{C}$ with $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}} \neq \emptyset$, letting $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{T}_{\mathrm{d}} \cap \mathcal{C}$ and $\mathrm{T}'_{\mathcal{C}} \in_u \mathcal{T}_{\mathrm{d}'} \cap \mathcal{C}$, we have*

$$\mathrm{ht}(\mathrm{T}_{\mathcal{C}})' \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathcal{C}})\,.$$

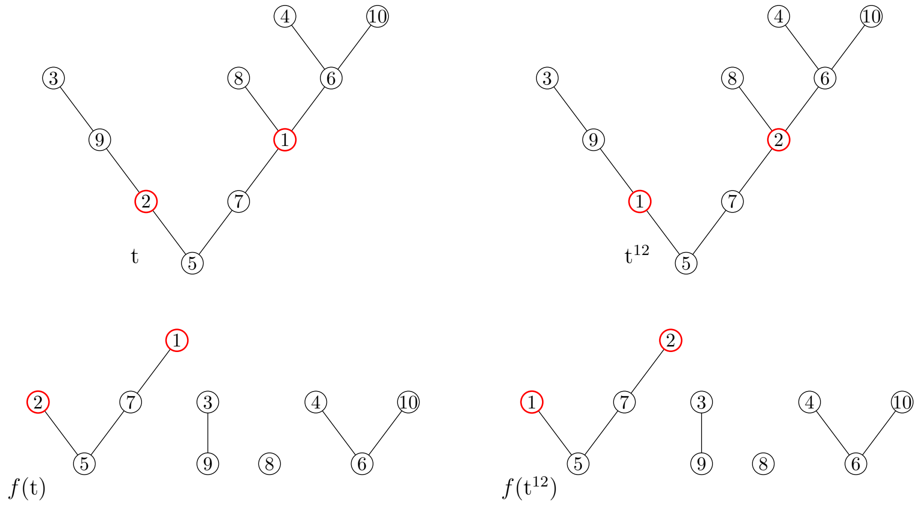
Moreover, if d contains at least three non-zero entries then there exists at least one $\sim$-equivalence class $\mathcal{C}$ for which the preceding stochastic domination is strict.

Before proving the propositions, we show how they straightforwardly imply Theorem 9.

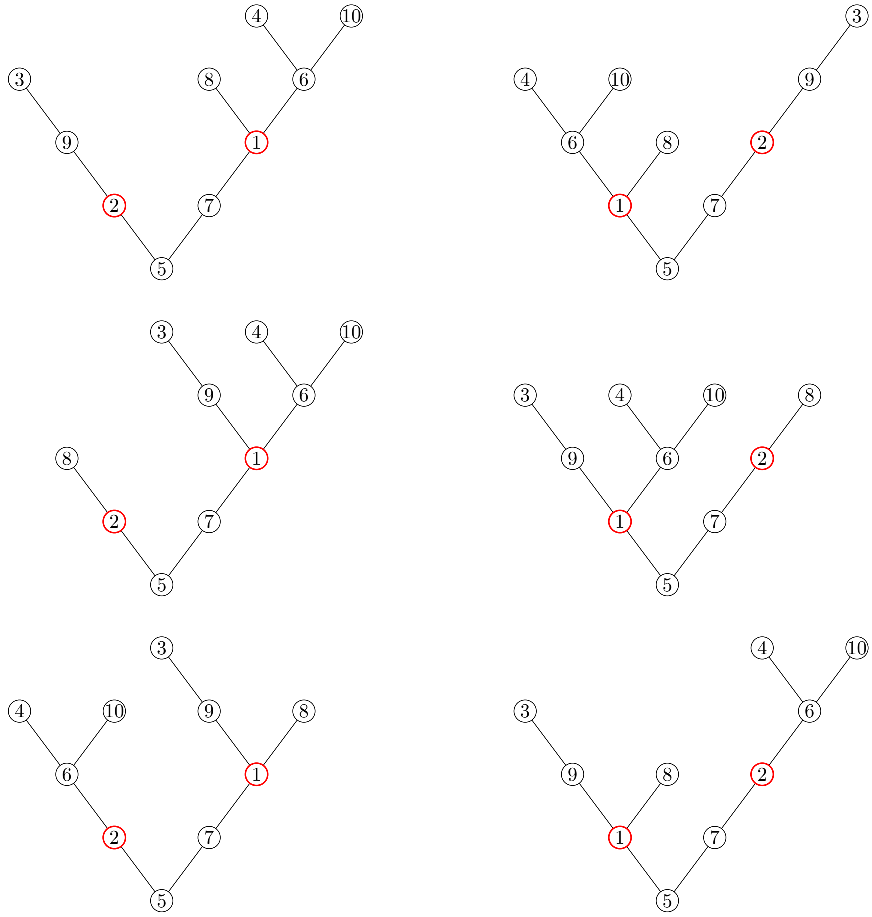
Proof of Theorem 9. First, by relabeling the vertices it suffices to show that for any degree sequence $\mathrm{d} = (d_1,\ldots,d_n)$ with $d_1 \geq d_2 \geq 1$, if $\mathrm{d}' = (d_1+1, d_2-1, d_3,\ldots,d_n)$ then for $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$ we have $\mathrm{ht}(\mathrm{T}') \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T})$, and that this stochastic domination is strict if d contains at least three non-zero entries.

Fix degree sequences d and d′ related as in the previous paragraph, and let $\mathrm{T} \in_u \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}' \in_u \mathcal{T}_{\mathrm{d}'}$. For a $\sim$-equivalence class $\mathcal{C}$ such that $\mathcal{T}_{\mathrm{d}} \cap \mathcal{C} \neq \emptyset$ we will also use the notation $\mathrm{T}_{\mathcal{C}}$ and $\mathrm{T}'_{\mathcal{C}}$ to denote uniformly random elements of $\mathcal{T}_{\mathrm{d}} \cap \mathcal{C}$ and $\mathcal{T}_{\mathrm{d}'} \cap \mathcal{C}$, respectively. For any $x > 0$, writing $\sum_{\mathcal{C}}$ to denote a summation over all $\sim$-equivalence classes $\mathcal{C}$, we have

$$\begin{aligned}
\mathbf{P}(\mathrm{ht}(\mathrm{T}) \leq x) &= \sum_{\mathcal{C}} \mathbf{P}\left(\mathrm{ht}(\mathrm{T}) \leq x \;\middle|\; \mathrm{T} \in \mathcal{C}\right) \mathbf{P}(\mathrm{T} \in \mathcal{C}) \\
&= \sum_{\mathcal{C}} \mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) \leq x) \mathbf{P}(\mathrm{T}' \in \mathcal{C}) \\
&\leq \sum_{\mathcal{C}} \mathbf{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) \leq x) \mathbf{P}(\mathrm{T}' \in \mathcal{C}) \\
&= \sum_{\mathcal{C}} \mathbf{P}\left(\mathrm{ht}(\mathrm{T}') \leq x \;\middle|\; \mathrm{T}' \in \mathcal{C}\right) \mathbf{P}(\mathrm{T}' \in \mathcal{C}) \\
&= \mathbf{P}(\mathrm{ht}(\mathrm{T}') \leq x)
\end{aligned}$$

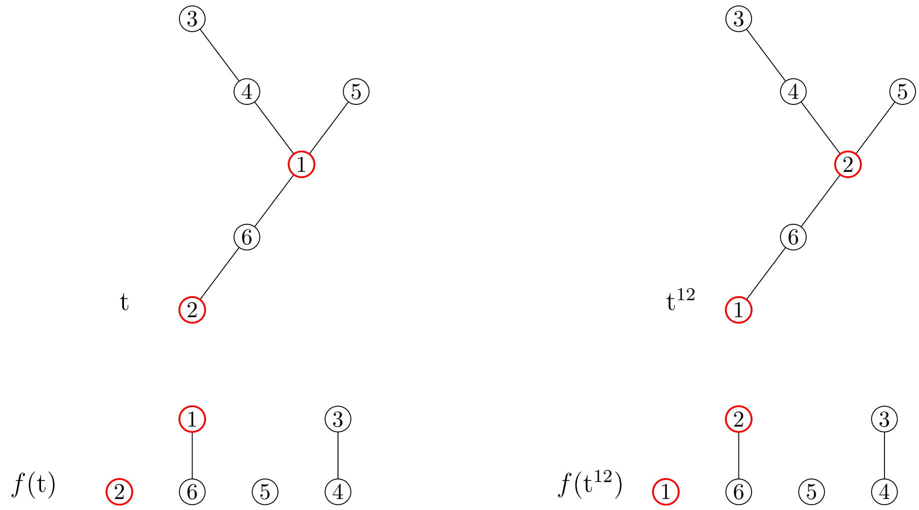


(a) For a tree t, we obtain $f(\mathrm{t})$ by removing the edges to the children of vertex 1 and 2. We obtain t^{12} by swapping the labels of 1 and 2 in t.

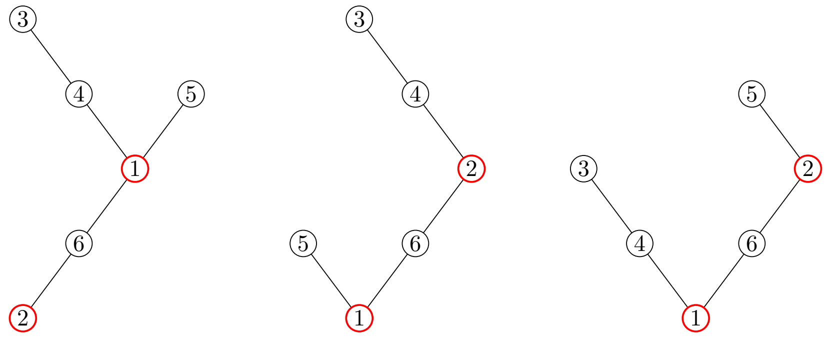


(b) The trees t' such that $\mathrm{t}' \sim \mathrm{t}$ and $d_{\mathrm{t}'}(1) = 2$ and $d_{\mathrm{t}'}(2) = 1$. For each tree t' on the left, $f(\mathrm{t}') = f(\mathrm{t})$. For each tree t' on the right, $f(\mathrm{t}') = f(\mathrm{t}^{12})$.

Figure 4. We show a subset of the equivalence class of tree t. In total, t is equivalent to 2^4 trees: to specify a tree t' such that $\mathrm{t}' \sim \mathrm{t}$ we must choose, for each element in $\{6, 8, 9\}$, whether its parent in t' is 1 or 2, and we must choose whether or not to swap the labels of 1 and 2.



(a) For a tree t, we obtain $f(\mathrm{t})$ by removing the edges to the children of vertex 1 and 2. We obtain t^{12} by swapping the labels of 1 and 2 in t.



(b) The trees in the equivalence class of t in which $d_1 = 2$ and $d_2 = 1$. For the leftmost tree t', we have that $f(\mathrm{t}') = f(\mathrm{t})$ and t' and t have the same root. For each other tree t'', we have that $f(\mathrm{t}'') = f(\mathrm{t}^{12})$ and t'' and t^{12} have the same root .

Figure 5. We show a subset of the equivalence class of tree t. In total, t is equivalent to 2^3 trees: to specify a tree t' such that $\mathrm{t}' \sim \mathrm{t}$ we must choose, for each element in $\{4, 5\}$, whether its parent in t' is 1 or 2, and we must choose whether or not to swap the labels of 1 and 2.

The second equality holds by Proposition 25 and since the conditional distribution of T given that $\mathrm{T} \in \mathcal{C}$ is precisely that of $\mathrm{T}_{\mathcal{C}}$. The inequality holds by Proposition 26. The third equality again holds since the conditional distribution of T' given that $\mathrm{T}' \in \mathcal{C}$ is precisely that of $\mathrm{T}'_{\mathcal{C}}$. This shows that $\mathrm{ht}(\mathrm{T}') \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T})$.

Finally, if d has at least three non-zero entries then by Proposition 26 there exists at least one equivalence class $\mathcal{C}$ and some $x > 0$ for which $\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) < \mathbf{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x)$. For such x the above chain of inequalities then yields that $\mathbf{P}(\mathrm{ht}(\mathrm{T}) \leq x) < \mathbf{P}(\mathrm{ht}(\mathrm{T}') \leq x)$, so in this case in fact $\mathrm{ht}(\mathrm{T}') \prec_{\mathrm{st}} \mathrm{ht}(\mathrm{T})$. □

Proof of Proposition 25. Let t be a tree in $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$. We first suppose that neither vertex 1 nor vertex 2 is an ancestor of the other in t.

The forest $f(\mathrm{t})$ contains $d_1 + d_2 + 1$ trees; list their roots as $r_1, \ldots, r_{d_1+d_2+1}$ so that $r_{d_1+d_2+1}$ is the root of t. Then both 1 and 2 lie in the tree rooted at $r_{d_1+d_2+1}$. In this case, there are $\binom{d_1+d_2}{d_1}$ trees $\hat{\mathrm{t}} \in \mathcal{C}$ with $f(\hat{\mathrm{t}}) = f(\mathrm{t})$: these are precisely the trees obtained from $f(\mathrm{t})$ as follows.

- Select a set $S \subset [d_1 + d_2]$ of size d_1.

- Add edges from vertex 1 to the vertices in the set $\{r_i, i \in S\}$, and add edges from vertex 2 to the vertices in the set $\{r_i, i \in [d_1 + d_2] \setminus S\}$.

Likewise, there are $\binom{d_1+d_2}{d_2}$ trees $\hat{\mathrm{t}} \in \mathcal{C}$ with $f(\hat{\mathrm{t}}) = f(\mathrm{t}^{12})$; these are obtained from $f(\mathrm{t})$ as follows.

- Select a set $S \subset [d_1 + d_2]$ of size d_2.
- Add edges from vertex 1 to the vertices in the set $\{r_i, i \in S\}$, add edges from vertex 2 to the vertices in the set $\{r_i, i \in [d_1 + d_2] \setminus S\}$, then swap the labels of vertices 1 and 2.

It follows that $|\mathcal{C} \cap \mathcal{T}_d| = 2\binom{d_1+d_2}{d_1}$. Likewise, we have $|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}'}| = 2\binom{d_1+d_2}{d_1+1}$, since any element of $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}'}$ may be constructed by selecting a size-$(d_1 + 1)$ subset S' of $\{r_i, i \in [d_1 + d_2]\}$, then either (a) attaching the roots in S' to 1 and the remaining roots to 2, or (b) attaching the roots in S' to 2 and the remaining roots to 1 and switching the labels of vertices 1 and 2.

Recalling the formula (2) for $|\mathcal{T}_{\mathrm{d}}|$, the preceding computations yield that

$$\frac{|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}|}{|\mathcal{T}_{\mathrm{d}}|} = 2\binom{d_1 + d_2}{d_1}\binom{n-1}{d_1, \ldots, d_n}^{-1} = 2\binom{d_1 + d_2}{d_1 + 1}\binom{n-1}{d_1 + 1, d_2 - 1, d_3, \ldots, d_n}^{-1} = \frac{|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}'}|}{|\mathcal{T}_{\mathrm{d}'}|},$$

so in this case $\mathbf{P}(\mathrm{T} \in \mathcal{C}) = \mathbf{P}(\mathrm{T}' \in \mathcal{C})$, as required.

Next suppose that vertex 1 is an ancestor of vertex 2 in t. The forest $f(\mathrm{t})$ contains $d_1 + d_2 + 1$ trees; list their roots as $r_1, \ldots, r_{d_1+d_2+1}$ so that $r_{d_1+d_2}$ and $r_{d_1+d_2+1}$ are the roots of the trees containing vertices 2 and 1, respectively. (This also means that $r_{d_1+d_2+1}$ is the root of t, and that $r_{d_1+d_2}$ is a child of 1 in t.) In this case there are $\binom{d_1+d_2-1}{d_2}$ trees $\hat{\mathrm{t}} \in \mathcal{C}$ with $f(\hat{\mathrm{t}}) = f(\mathrm{t})$: these are precisely the trees obtained from $f(\mathrm{t})$ as follows.

- Select a set $S \subset [d_1 + d_2 - 1]$ of size d_2.
- Add edges from vertex 2 to the vertices in the set $\{r_i, i \in S\}$, and add edges from vertex 1 to the vertices in the set $\{r_i, i \in [d_1 + d_2] \setminus S\}$.

Similarly, there are $\binom{d_1+d_2-1}{d_1}$ trees $\hat{\mathrm{t}} \in \mathcal{C}$ with $f(\hat{\mathrm{t}}) = f(\mathrm{t}^{12})$; these are obtained from $f(\mathrm{t})$ as follows.

- Select a set $S \subset [d_1 + d_2 - 1]$ of size d_1.
- Add edges from vertex 2 to the vertices in the set $\{r_i, i \in S\}$, add edges from vertex 2 to the vertices in the set $\{r_i, i \in [d_1 + d_2] \setminus S\}$, then swap the labels of vertices 1 and 2.

It follows that

$$|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}| = \binom{d_1 + d_2 - 1}{d_2} + \binom{d_1 + d_2 - 1}{d_1} = \binom{d_1 + d_2}{d_1},$$

and likewise $|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}'}| = \binom{d_1+d_2}{d_1+1}$. (We omit the details for this last identity as they are so similar to the previous arguments.) It follows that in this case we also have $|\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}|/|\mathcal{T}_{\mathrm{d}}| = |\mathcal{C} \cap \mathcal{T}_{\mathrm{d}'}|/|\mathcal{T}_{\mathrm{d}'}|$, so again $\mathbf{P}(\mathrm{T} \in \mathcal{C}) = \mathbf{P}(\mathrm{T}' \in \mathcal{C})$.

Finally, if vertex 2 is an ancestor of vertex 1 in t, then in t^{12} vertex 1 is an ancestor of vertex 2, so since $\mathrm{t} \sim \mathrm{t}^{12}$, this situation is already handled by the previous case. □

For the proof of Proposition 26 we require an additional lemma, which although fairly straightforward we find independently pleasing. Write $\binom{[n]}{k} = \{S \subset [n] : |S| = k\}$. Below we use the convention that $\max \emptyset = 0$.

Lemma 27 (Eggs-in-one-basket lemma). *Fix non-negative real numbers $0 < a_1 \leq \ldots \leq a_n$ and integers k, ℓ with $n/2 \leq k < \ell \leq n$.*

(1) Let $\mathrm{A} \in_u \binom{[n]}{k} \cup \binom{[n]}{n-k}$ and $\mathrm{A}' \in_u \binom{[n]}{\ell} \cup \binom{[n]}{n-\ell}$. Then $\max(a_i, i \in \mathrm{A}') \preceq_{\mathrm{st}} \max(a_i, i \in \mathrm{A})$.
(2) Let $\mathrm{B} \in_u \binom{[n-1]}{k} \cup \binom{[n-1]}{n-k}$ and $\mathrm{B}' \in_u \binom{[n-1]}{\ell} \cup \binom{[n-1]}{n-\ell}$. Then $\max(a_i, i \in \mathrm{B}') \preceq_{\mathrm{st}} \max(a_i, i \in \mathrm{B})$.

The proverb "don't put all your eggs in one basket" means "don't put all your resources into a single endeavour" or, more pithily, "diversify your portfolio". (Its origins are obscure but an Italian equivalent, "non mettere tutte le uova in un solo cesto", has been traced to at least 1666 [13].) To understand our use of this phrase, note that if the "portfolio" is the random set A or the set B from the lemma, and the payoff of a portfolio is the value of its largest element, then the lemma implies that larger-entropy portfolios have stochastically higher payoffs. To our knowledge, this lemma is the first rigorous proof of of the wisdom of the proverb.

Proof of Lemma 27. If $a_1, \ldots, a_\ell$ are all equal then the result is obvious so we hereafter assume that this is not the case. It suffices to prove the lemma with $\ell = k+1$; the general case follows by induction. It's useful to set $a_0 = 0$. Then for any $0 \le i < n$ and $a_i < x \le a_{i+1}$,

$$\begin{aligned}\mathbf{P}\left(\max\{a_j : j \in \mathrm{A}\} < x\right) &= \mathbf{P}\left(\mathrm{A} \cap \{i+1, \ldots, n\} = \emptyset\right) \\ &= \frac{1}{2\binom{n}{k}}\left[\binom{i}{k} + \binom{i}{n-k}\right].\end{aligned}$$

To prove the first claim of the lemma, that $\max(a_i, i \in \mathrm{A}') \preceq_{\mathrm{st}} \max(a_i, i \in \mathrm{A})$, it thus suffices to show that

$$\frac{1}{2\binom{n}{k}}\left[\binom{i}{k} + \binom{i}{n-k}\right] \le \frac{1}{2\binom{n}{k+1}}\left[\binom{i}{k+1} + \binom{i}{n-k-1}\right].$$

It is possible that some of the binomial coefficients above are zero; regardless, multiplying through by $2\binom{n}{n-i}$ and rearranging terms yields that this is equivalent to showing that

$$\binom{n-k}{n-i} - \binom{n-k-1}{n-i} \le \binom{k+1}{n-i} - \binom{k}{n-i},$$

which by the addition rule of binomials reduces to

$$\binom{n-k-1}{n-i-1} \le \binom{k}{n-i-1}.$$

This holds, because $k > n-k-1$.

For the second claim of the lemma, note that B has the law of A conditional on the event $\{n \notin \mathrm{A}\}$, which, by symmetry, has probability $1/2$. This implies that for any $x \le a_n$,

$$\begin{aligned}\mathbf{P}\left(\max\{a_i : i \in \mathrm{A}\} < x\right) &= \mathbf{P}\left(n \notin \mathrm{A}\right)\mathbf{P}\left(\max\{a_i : i \in \mathrm{A}\} < x \mid n \notin \mathrm{A}\right) \\ &= \frac{1}{2}\mathbf{P}\left(\max\{a_i : i \in \mathrm{B}\} < x\right),\end{aligned}$$

and similarly,

$$\mathbf{P}\left(\max\{a_i : i \in \mathrm{A}'\} < x\right) = \frac{1}{2}\mathbf{P}\left(\max\{a_i : i \in \mathrm{B}'\} < x\right).$$

Therefore,

$$\begin{aligned}&\mathbf{P}\left(\max\{a_i : i \in \mathrm{A}\} < x\right) - \mathbf{P}\left(\max\{a_i : i \in \mathrm{A}'\} < x\right) \\ &= \frac{1}{2}\left(\mathbf{P}\left(\max\{a_i : i \in \mathrm{B}\} < x\right) - \mathbf{P}\left(\max\{a_i : i \in \mathrm{B}'\} < x\right)\right),\end{aligned}$$

so the second claim of the lemma follows from the first. □

Proof of Proposition 26. Fix a $\sim$-equivalence class $\mathcal{C}$ and a tree $\mathrm{t} \in \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$. We first suppose that neither vertex 1 nor vertex 2 is an ancestor of the other in t. The forest $f(\mathrm{t})$ contains $d_1 + d_2 + 1$ trees; list their roots as $r_1, \ldots, r_{d_1+d_2+1}$ so that $r_{d_1+d_2+1}$ is the root of t, and let t_i be the tree with root r_i. Then both 1 and 2 lie in the tree $\mathrm{t}_{d_1+d_2+1}$ rooted at $r_{d_1+d_2+1}$. Write h_1 and h_2 for the distance from $r_{d_1+d_2+1}$ to 1 and 2, respectively.

By the definition of $\mathcal{C}$, starting from $f(\mathrm{t})$ we may sample $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ as follows.

- Let $(\mathrm{C}, \mathrm{A}) \in_u \left\{(1, S) : S \in \binom{[d_1+d_2]}{d_1}\right\} \cup \left\{(2, S) : S \in \binom{[d_1+d_2]}{d_2}\right\}$.

- Add edges from vertex 1 to the roots $\{r_i, i \in \mathrm{A}\}$ and from vertex 2 to the roots $\{r_i, i \in [d_1+d_2] \setminus A\}$.
- If $\mathrm{C}=2$ then swap the labels of vertices 1 and 2.

Note that $\mathrm{A} \in_u \binom{[d_1+d_2]}{d_1} \cup \binom{[d_1+d_2]}{d_2}$. For $1 \le i \le d_1+d_2$ letting $a_i = 1+\mathrm{ht}(\mathrm{t}_i)$, with the above construction of $\mathrm{T}_{\mathcal{C}}$ we then have

$$\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) = \max(\mathrm{ht}(\mathrm{t}_{d_1+d_2+1}), h_1 + \max(a_i, i \in \mathrm{A}), h_2 + \max(a_i, i \in [d_1+d_2] \setminus \mathrm{A}))\,.$$

Next, again starting from $f(\mathrm{t})$, apply the same procedure (with d_1 and d_2 replaced by d_1+1 and d_2-1, respectively) to sample $\mathrm{T}'_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$. We obtain

$$\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) = \max(\mathrm{ht}(\mathrm{t}_{d_1+d_2+1}), h_1 + \max(a_i, i \in \mathrm{A}'), h_2 + \max(a_i, i \in [d_1+d_2] \setminus \mathrm{A}'))\,.$$

where $\mathrm{A}' \in_u \binom{[d_1+d_2]}{d_1+1} \cup \binom{[d_1+d_2]}{d_2-1}$.

Since A and $[d_1+d_2] \setminus \mathrm{A}$ have the same distribution, as do A' and $[d_1+d_2] \setminus \mathrm{A}'$, we may assume without loss of generality that $h_1 \ge h_2$. It then follows that both the above maxima are at least $M^- := \max(\mathrm{ht}(\mathrm{t}_{d_1+d_2+1}), h_2 + \max(a_i, i \in [d_1+d_2]))$ and at most $M^+ := \max(\mathrm{ht}(\mathrm{t}_{d_1+d_2+1}), h_1 + \max(a_i, i \in [d_1+d_2]))$, so for $x \le M^-$ we have

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) = 0 = \mathbf{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x)$$

while for $x > M^+$ we have

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) = 1 = \mathbf{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x).$$

For $M^- < x \le M^+$, we have

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) = \mathbf{P}(h_1 + \max(a_i, i \in \mathrm{A}) < x)$$

and

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x) = \mathbf{P}(h_1 + \max(a_i, i \in \mathrm{A}') < x)\,,$$

so the first part of the eggs-in-one-basket lemma yields that

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) \le \mathbf{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x).$$

This establishes that $\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathcal{C}})$ when neither 1 nor 2 is an ancestor of the other for trees in $\mathcal{C}$.

We next suppose that either 1 is an ancestor of 2 in t or vice-versa. Note that $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ contains a tree in which 1 is an ancestor of 2 if and only if it contains a tree in which 2 is an ancestor of 1. It follows that, by replacing t by another element of $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ if necessary, we may assume that in fact 1 is an ancestor of 2 in t.

List the roots of the trees in $f(\mathrm{t})$ as $r_1, \ldots, r_{d_1+d_2+1}$ so that $r_{d_1+d_2}$ and $r_{d_1+d_2+1}$ are the roots of the trees containing vertices 2 and 1, respectively, and write t_i for the tree of $f(\mathrm{t})$ with root r_i. Necessarily $r_{d_1+d_2+1}$ is also the root of t, and $r_{d_1+d_2}$ is a child of 1 in t. Write h_1 and h_2 for the distance from $r_{d_1+d_2+1}$ to 1 and from $r_{d_1+d_2}$ to 2, respectively.

By the definition of the equivalence class $\mathcal{C}$, starting from the forest $\mathrm{t}_0, \ldots, \mathrm{t}_{d_1+d_2}$, we may sample $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ as follows.

- Let

$$(\mathrm{C}, \mathrm{B}) \in_u \left\{(1, S) : S \in \binom{[d_1+d_2-1]}{d_2}\right\} \cup \left\{(2, S) : S \in \binom{[d_1+d_2-1]}{d_1} : \right\}.$$

- Add edges from vertex 2 to the roots $\{r_i, i \in \mathrm{B}\}$ and from vertex 1 to the roots $\{r_i, i \in [d_1+d_2] \setminus \mathrm{B}\}$.
- If $\mathrm{C}=2$ then swap the labels of vertices 1 and 2.

Note that $\mathrm{B} \in_u \{S \subset [d_1+d_2-1] : |S| \in \{d_1, d_2\}\}$. Moreover, letting $a_i = 1 + \mathrm{ht}(\mathrm{t}_i)$ for $i \in [d_1+d_2-1]$, and letting $H = \max(\mathrm{ht}(\mathrm{t}_{d_1+d_2+1}), h_1 + 1 + \mathrm{ht}(\mathrm{t}_{d_1+d_2}))$, with the above construction of $\mathrm{T}_{\mathcal{C}}$ we then have

$$\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) = \max(H, h_1 + \max(a_i, i \in [d_1+d_2] \setminus \mathrm{B}), h_1 + 1 + h_2 + \max(a_i, i \in \mathrm{B}))\,.$$

The term H accounts for the possibility that the height of $\mathrm{T}_{\mathcal{C}}$ is achieved by a vertex of either $\mathrm{t}_{d_1+d_2}$ or $\mathrm{t}_{d_1+d_2+1}$.

Next, again starting from $f(\mathrm{t})$, apply the same procedure (with d_1 and d_2 replaced by d_1+1 and d_2-1, respectively) to sample $\mathrm{T}'_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$. We obtain

$$\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) = \max(H, h_1 + \max(a_i, i \in [d_1+d_2] \setminus \mathrm{B}'), h_1 + 1 + h_2 + \max(a_i, i \in \mathrm{B}'))\,.$$

where $\mathrm{B}' \in_u \{S \subset [d_1+d_2-1] : |S| \in \{d_1+1, d_2-1\}\}$.

The heights of $\mathrm{T}_{\mathcal{C}}$ and $\mathrm{T}'_{\mathcal{C}}$ both lie between

$$M^- := \max(H, h_1 + \max(a_i, i \in [d_1+d_2-1]))$$

and

$$M^+ := \max(H, h_1 + 1 + h_2 + \max(a_i, i \in [d_1+d_2-1]))$$

so for $x \notin (M^-, M^+]$ we have $\mathbf{P}(\mathrm{ht}(\mathrm{T}_C) < x) = \mathbf{P}(\mathrm{ht}(\mathrm{T}'_C) < x)$. For $M^- < x \le M^+$, we have $\mathrm{ht}(\mathrm{T}_C) < x$ if and only if $h_1 + 1 + h_2 + \max(a_i, i \in \mathrm{B}) < x$, and likewise $\mathrm{ht}(\mathrm{T}'_C) < x$ if and only if $h_1 + 1 + h_2 + \max(a_i, i \in \mathrm{B}') < x$. It thus follows by the second part of the eggs-in-one-basket lemma that

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) < x) \le \mathbf{P}(\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) < x).$$

This establishes that $\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) \preceq_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathcal{C}})$ when 1 is an ancestor of 2 in t. (As already noted, this also handles the case where 2 is an ancestor of 1.)

It remains to establish strict stochastic inequality when there are at least three non-leaf vertices. We accomplish this by showing that in this case there exists $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$ such that for $\mathcal{C}$ the $\sim$-equivalence class of t, if $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}'_{\mathcal{C}} \in_u \mathcal{T}_{\mathrm{d}'}$ then $\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) \prec_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathcal{C}})$.

By relabeling we may assume that vertex 3 is not a leaf. (We also still assume that $d_1 \ge d_2 \ge 1$.) Consider a tree $\mathrm{t} \in \mathcal{T}_{\mathrm{d}}$ with root 1, such that 3 is a child of 2 and 2 is a child of 1, and such that all other children of vertices 1 and 2 are leaves (see Figure 6). Then the $\sim$-equivalence class $\mathcal{C}$ of t contains $\binom{d_1+d_2-1}{d_1}$ trees with root 2 and $\binom{d_1+d_2-1}{d_2}$ trees with root 1, so $\binom{d_1+d_2}{d_1}$ trees in total.

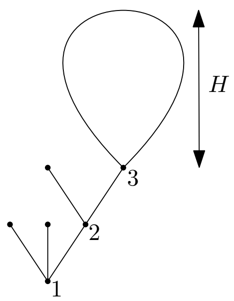


Figure 6. A schematic depiction of a tree in an equivalence class $\mathcal{C}$ for which, if $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ and $\mathrm{T}'_{\mathcal{C}} \in_u \mathcal{T}_{\mathrm{d}'}$ then $\mathrm{ht}(\mathrm{T}'_{\mathcal{C}}) \prec_{\mathrm{st}} \mathrm{ht}(\mathrm{T}_{\mathcal{C}})$.

Let H be the height of the subtree of t rooted at vertex 3 (this is at least 1 by the assumption that 3 is not a leaf). Then the height of $\mathrm{T}_{\mathcal{C}} \in_u \mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ is either $H+1$ or $H+2$, and is $H+2$ precisely if either 1 is the root and 3 is a child of 2, or if 2 is the root and 3 is a child of 1. The total number of trees in $\mathcal{C} \cap \mathcal{T}_{\mathrm{d}}$ with height $H+2$ is thus $2\binom{d_1+d_2-2}{d_1-1}$, so

$$\mathbf{P}(\mathrm{ht}(\mathrm{T}_{\mathcal{C}}) = H+2) = 2\binom{d_1+d_2-2}{d_1-1}\binom{d_1+d_2}{d_1}^{-1} = \frac{2d_1d_2}{(d_1+d_2-1)(d_1+d_2)}.$$

This probability decreases if d_1 and d_2 are replaced by $d_1 + 1$ and $d_2 - 1$, respectively, which establishes the required strict stochastic domination. □

References

[1] Louigi Addario-Berry. Most trees are short and fat. *Probability theory and related fields*, 173(1-2): 1–26, 2019. ISSN 0178-8051. 1, 2

[2] Louigi Addario-Berry, Luc Devroye, and Svante Janson. Sub-Gaussian tail bounds for the width and height of conditioned Galton-Watson trees. *Ann. Probab.*, 41(2):1072–1087, 2013. ISSN 0091-1798. doi: 10.1214/12-AOP758. URL `https://doi.org/10.1214/12-AOP758`. 1, 2

[3] Louigi Addario-Berry, Anna Brandenberger, Jad Hamdan, and Céline Kerriou. Universal height and width bounds for random trees. preprint arXiv:2105.03195, 2021. URL `https://arxiv.org/abs/2105.03195`. 1

[4] Louigi Addario-Berry, Serte Donderwinkel, Mickaël Maazoun, and James B. Martin. A new proof of Cayley's formula. preprint arXiv:2107.09726, 2021. URL `https://arxiv.org/abs/2107.09726`. 1, 5, 7

[5] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993. ISSN 0091-1798. URL `http://links.jstor.org/sici?sici=0091-1798(199301)21:1<248:TCRTI>2.0.CO;2-1&origin=MSN`. 23

[6] David J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985. doi: 10.1007/BFb0099421. URL `https://doi-org.proxy3.library.mcgill.ca/10.1007/BFb0099421`. 18

[7] S.N. Bernstein. *Theory of Probability (Russian)*. Moscow, 1927. 21

[8] Arthur Blanc-Renaudie. Limit of trees with fixed degree sequence. arXiv:2110.03378 [math.PR], 2021. URL `https://arxiv.org/abs/2110.03378`. 5, 19

[9] Cecil C Craig. On the Tchebychef inequality of Bernstein. *The Annals of Mathematical Statistics*, 4(2):94–102, 1933. ISSN 0003-4851. doi: 10.1214/aoms/1177732803. URL `https://www.doi.org/10.1214/aoms/1177732803`. 21

[10] Svante Janson. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probab. Surv.*, 9:103–252, 2012. doi: 10.1214/11-PS188. URL `https://doi.org/10.1214/11-PS188`. 1, 2, 21, 22, 23

[11] Colin McDiarmid. Concentration. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 195–248, New York, 1998. Springer Verlag. 18

[12] Richard P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. ISBN 0-521-56069-1; 0-521-78987-7. doi: 10.1017/CBO9780511609589. URL `https://doi.org/10.1017/CBO9780511609589`. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin. 8

[13] Giovanni Torriano. *Piazza Universale Di Proverbi Italiani: Or, A Common Place Of Italian Proverbs And Proverbial Phrases. Digested in Alphabetical Order by way of Dictionary. Interpreted, and occasionally Illustrated with Notes. Together with A Supplement of Italian Dialogves*. F. and T.W. for the Author, 1666. URL `https://leme.library.utoronto.ca/lexicons/1115/details`. 30

Department of Mathematics and Statistics, McGill University, Montréal, Canada

Email address: `louigi.addario@mcgill.ca`

URL: `http://problab.ca/louigi/`

Department of Statistics, Oxford, UK

Email address: `serte.donderwinkel@st-hughs.ox.ac.uk`