From trees to seeds: on the inference of the seed from large trees in the uniform attachment model

Sébastien Bubeck * Ronen Eldan † Elchanan Mossel ‡ Miklós Z. Rácz §

October 20, 2014

Abstract

We study the influence of the seed in random trees grown according to the uniform attachment model, also known as uniform random recursive trees. We show that different seeds lead to different distributions of limiting trees from a total variation point of view. To do this, we construct statistics that measure, in a certain well-defined sense, global "balancedness" properties of such trees. Our paper follows recent results on the same question for the preferential attachment model.

1 Introduction

We consider one of the simplest models of a randomly growing graph: starting from a single node, each arriving new node connects uniformly at random to one of the existing nodes. We denote by $\mathrm{UA}(n)$ the corresponding tree on n vertices, also known as the uniform random recursive tree. We investigate whether a snapshot of the tree at a finite time can give information about the asymptotic future evolution of the tree. More precisely, we want to know if conditioning on the value of $\mathrm{UA}(k)$ for some finite k affects the limiting distribution of $\mathrm{UA}(n)$ as n goes to infinity. Equivalently, we study the uniform attachment model starting from an initial seed, and we want to know if the limiting distribution depends on the seed.

For $n \geq k \geq 2$ and a tree S on k vertices, we define the random tree $\mathrm{UA}(n, S)$ by induction. First, $\mathrm{UA}(k, S) = S$. Then, given $\mathrm{UA}(n, S)$, $\mathrm{UA}(n+1, S)$ is formed from $\mathrm{UA}(n, S)$ by adding a new vertex u and adding a new edge uv where the vertex v is chosen uniformly at random among vertices of $\mathrm{UA}(n, S)$, independently of all past choices. For two seed trees S and T, we are interested in studying the quantity

$$\delta(S, T) = \lim_{n \to \infty} \mathrm{TV}\left(\mathrm{UA}(n, S), \mathrm{UA}(n, T)\right),$$

a limit which is well-defined[1], where we recall that the total variation distance between two random variables X and Y taking values in a finite space $\mathcal{X}$ with laws μ and ν is defined as $\mathrm{TV}(X, Y) = \frac{1}{2}\sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|$. Our main result shows that each seed leads to a unique limiting distribution of the uniform attachment tree.

*Microsoft Research and Princeton University; sebubeck@microsoft.com.

†Microsoft Research; roneneldan@gmail.com.

‡University of Pennsylvania and University of California, Berkeley; mossel@wharton.upenn.edu.

§University of California, Berkeley; racz@stat.berkeley.edu.

[1]This is because $\mathrm{TV}(\mathrm{UA}(n, S), \mathrm{UA}(n, T))$ is non-increasing in n (since one can simulate the future evolution of the process) and always nonnegative.

Theorem 1 *For any S and T non-isomorphic with at least 3 vertices, we have that $\delta(S,T) > 0$.*

In some cases our method can say even more. As a proof of concept, we prove the following result, which states that the distance between a fixed tree and a star can be arbitrarily close to 1 if the star is large enough.

Theorem 2 *Let S_k denote the k-vertex star. For any fixed tree T one has*

$$\lim_{k\to\infty} \delta\left(S_k, T\right) = 1.$$

1.1 Our approach and comparison with the preferential attachment model

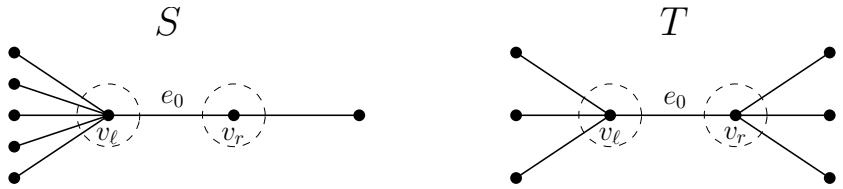
The theoretical study of the influence of the seed graph in growing random graph models was initiated, to the best of our knowledge, by Bubeck et al. [2014] for the particular case of preferential attachment trees. They showed that seeds with different degree profiles lead to different distributions of limiting trees from a total variation point of view. Curien et al. [2014], using a different but related approach, then showed that this also holds for any two non-isomorphic seeds, i.e., that the analogue of Theorem 1 above holds for the preferential attachment model.

The main idea in both papers, motivated by the rich-get-richer property of the preferential attachment model, is to consider various statistics based on large degree nodes, and show that the initial seed influences the distribution of these statistics. Consider, for instance, the problem of differentiating between the two seed trees in Figure 1. On the one hand, in S the degree of v_ℓ is greater than that of v_r, and this unbalancedness in the degrees likely remains as the tree grows according to preferential attachment. On the other hand, in T the degrees of v_ℓ and v_r are the same, so they will have the same distribution at larger times as well. This difference in the balancedness vs. unbalancedness of the degrees of v_ℓ and v_r is at the root of why the seed trees S and T are distinguishable in the preferential attachment model. A precise understanding of the balancedness properties of the degrees relies on the classical theory of Pólya urns.

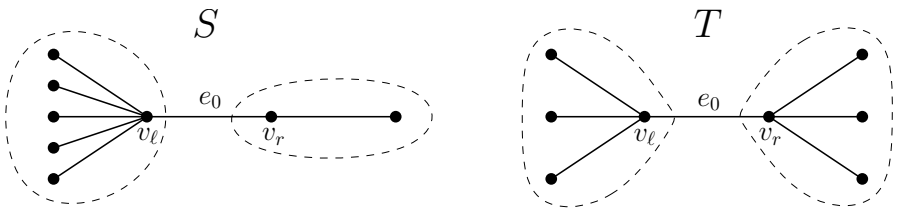
In the uniform attachment model the degrees of vertices do not play a special role. In particular, in the example of Figure 1, v_ℓ and v_r will have approximately similar degrees in a large tree grown according to the uniform attachment model, irrespective of whether the seed tree is S or T. Nonetheless, we are able to distinguish the seed trees S and T, but the statistics we use to do this are based on more global balancedness properties of these trees.

An edge of a tree partitions the tree into two parts on either side of the edge. For most edges in a tree, this partition has very unbalanced sizes; for instance, if an edge is adjacent to a leaf, then one part contains only a single vertex. On the other hand, for edges that are in some sense "central" the partition is more balanced, in the sense that the sizes of the two parts are comparable. Intuitively, the edges of the seed tree will be among the "most central" edges of the uniform attachment tree at large times, and so we expect that the seed should influence the global balancedness properties of such trees.

Consider again the example of the two seed trees S and T in Figure 1. The edge e_0 partitions the tree into two parts: a subtree under v_ℓ and a subtree under v_r. In S these subtree sizes are unbalanced, and this likely remains the case as the tree grows according to uniform attachment. On the other hand, in T the subtree sizes are equal, and they will likely remain balanced as the tree grows. Again, Pólya urns play an important role, since the subtree sizes evolve according to a classical Pólya urn initialized by the subtree sizes in the seed tree. The difference in the balancedness



(a) **Preferential attachment.** The degrees of v_ℓ and v_r are unbalanced in S but balanced in T, and this likely remains the case as the trees grow according to preferential attachment. This is at the root of why S and T are distinguishable as seed trees in the preferential attachment model.



(b) **Uniform attachment.** The sizes of the subtrees under v_ℓ and v_r are unbalanced in S but balanced in T, and this likely remains the case as the trees grow according to uniform attachment. This is at the root of why S and T are distinguishable as seed trees in the uniform attachment model.

Figure 1: Distinguishing between two trees requires different approaches for the uniform and the preferential attachment models.

vs. unbalancedness of the subtree sizes is at the root of why S and T are distinguishable in the uniform attachment model. To prove Theorem 1 we need to analyze statistics based on more general global balancedness properties of such trees, but the underlying intuition is what is described in the preceding paragraphs.

To formalize this intuition we essentially follow the proof scheme developed in Curien et al. [2014]. However, the devil is in the details: since the underlying statistics are markedly different—in particular, statistics based on degrees are local, whereas those based on balancedness properties of subtree sizes are global—some of the essential steps of the proof become different. We provide a more detailed comparison to the work of Curien et al. [2014] in Section 4.1, after we present our proof.

1.2 Further related work

A tree with node set $[n] := \{1, \dots, n\}$ is called recursive if the node numbers along the unique path from 1 to j increase for every $j \in \{2, \dots, n\}$. A stochastic process of random growing trees where nodes are labeled according to the time they are born is thus a sequence of random recursive trees. If we choose a recursive tree with node set $[n]$ uniformly at random, then the resulting tree has the same distribution as $\mathrm{UA}(n)$. A random tree grown according to the preferential attachment process starting from a single node is also known as a random plane-oriented recursive tree, see Mahmoud [1992].

There is a large literature on random recursive trees and their various statistics; we refer the reader to the book by Drmota [2009]. Of particular interest to the question we study here are recent works on a boundary theory approach to the convergence of random recursive trees in the limit as the tree size goes to infinity, see Evans et al. [2012] and Grübel and Michailow [2014]. The main difference between these and the current work is that they consider labeled and rooted trees,

whereas we are interested in what can be said about the seed given an unlabeled and unrooted copy of the tree.

2 Partitions and their balancedness: simple examples

In this section we show on a simple example how to formalize the intuition described in Section 1.1. We define a simple statistic based on this intuition, and after collecting some preliminary facts in Section 2.1, we show in Section 2.2 that $\delta(P_4, S_4) > 0$, where P_4 and S_4 are the path and the star on four vertices, respectively. We conclude the section by proving Theorem 2 in Section 2.3. The goal of this section is thus to provide a gentle introduction into the methods and statistics used to distinguish different seed trees, before analyzing more general statistics in the proof of Theorem 1 in Section 3.

For a tree T and an edge $e \in E(T)$, let T_1 and T_2 be the two connected components of $T \setminus \{e\}$. Define

$$g(T, e) = |T_1|^2 |T_2|^2 / |T|^4,$$

where $|T|$ denotes the number of vertices of T. Clearly, $0 \leq g(T, e) \leq 1/16$, and for "peripheral" edges e, $g(T, e)$ is closer to 0, while for more "central" edges e, $g(T, e)$ is closer to $1/16$. Define the following statistic:

$$G(T) = \sum_{e \in E(T)} g(T, e).$$

The statistic $G(T)$ thus measures in a particular way the global balancedness properties of the tree T, and "central" edges contribute the most to this statistic.

2.1 Preliminary facts

For all $\alpha, \beta, n \in \mathbb{N}$, let $B_{\alpha,\beta,n}$ be a random variable such that $B_{\alpha,\beta,n} - \alpha$ has the beta-binomial distribution with parameters (α, β, n), i.e., it is a random variable satisfying

$$\mathbb{P}(B_{\alpha,\beta,n} = \alpha + k) = \frac{(k+\alpha-1)!(n-k+\beta-1)!(\alpha+\beta-1)!}{(n+\alpha+\beta-1)!(\alpha-1)!(\beta-1)!}\binom{n}{k}, \quad \forall k \in \{0, 1, \ldots, n\}.$$

The key to understanding the statistic G is the following distributional identity:

$$g(\mathrm{UA}(n, S), e) \stackrel{d}{=} \frac{1}{n^4} B^2_{|T_1|,|T_2|,n-|S|} (n - B_{|T_1|,|T_2|,n-|S|})^2, \quad \forall e \in E(S), \tag{1}$$

where T_1 and T_2 are defined, given e, as above, and $\stackrel{d}{=}$ denotes equality in distribution. This is an immediate consequence of the characterization of $(B_{\alpha,\beta,n}, n + (\alpha+\beta) - B_{\alpha,\beta,n})$ as the distribution of a classical Pólya urn with replacement matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and starting state (α, β) after n draws. Similarly, for edges not in the seed S we have

$$g(\mathrm{UA}(n, S), e_j) \stackrel{d}{=} \frac{1}{n^4} B^2_{1,j,n-j-1} \big(n - B_{1,j,n-j-1}\big)^2, \tag{2}$$

where $e_j \in E(\mathrm{UA}(j+1, S)) \setminus E(\mathrm{UA}(j, S))$, $j \in \{|S|, \ldots, n-1\}$.

We use the following elementary facts about the beta-binomial distribution, which we prove in Appendix A:

Fact 1 *For every $p \geq 1$ there exists a constant $C(p)$ such that for all α, β, n such that $n \geq \alpha + \beta$, we have*

$$\left(\mathbb{E}[B^p_{\alpha,\beta,n-\alpha-\beta}]\right)^{1/p} \leq C(p) n \frac{\alpha}{\alpha+\beta}. \tag{3}$$

Fact 2 *There exists a universal constant $C > 0$ such that whenever $\alpha, \beta \geq 1$, $n \geq \alpha+\beta$, and $t \geq 0$, we have*

$$\mathbb{P}\left(B_{\alpha,\beta,n-\alpha-\beta} < tn\frac{\alpha}{\alpha+\beta}\right) \leq Ct. \tag{4}$$

2.2 A simple example

After these preliminaries we are now ready to show that $\delta(P_4, S_4) > 0$. To abbreviate notation, in the following we write simply $P \equiv P_4$ and $S \equiv S_4$. In order to show that $\delta(P, S) > 0$, it is enough to show two things:

$$\liminf_{n\to\infty} |\mathbb{E}[G(\mathrm{UA}(n, P))] - \mathbb{E}[G(\mathrm{UA}(n, S))]| > 0, \tag{5}$$

and

$$\limsup_{n\to\infty} (\mathbb{V}\mathrm{ar}[G(\mathrm{UA}(n, P))] + \mathbb{V}\mathrm{ar}[G(\mathrm{UA}(n, S))]) < \infty. \tag{6}$$

The proof can then be concluded using the Paley-Zigmund inequality (for more detail on this point, see the proof of Theorem 1 in Section 3.2).

For $j \geq 4$, let e_j^P denote the edge in $\mathrm{UA}(j+1, P) \setminus \mathrm{UA}(j, P)$, and define e_j^S similarly. Towards (5), we first observe that

$$g\left(\mathrm{UA}(n, P), e_j^P\right) \stackrel{d}{=} g\left(\mathrm{UA}(n, S), e_j^S\right), \quad \forall j \in \{4, \ldots, n-1\}.$$

Consequently, we have

$$\mathbb{E}[G(\mathrm{UA}(n, P))] - \mathbb{E}[G(\mathrm{UA}(n, S))] = \sum_{e\in P} \mathbb{E}[g(\mathrm{UA}(n, P), e)] - \sum_{e\in S} \mathbb{E}[g(\mathrm{UA}(n, S), e)].$$

Moreover, note that P has two edges, e_1 and e_2, such that $P \setminus \{e_i\}$ has two connected components of sizes 1 and 3 for $i = 1, 2$. Since this is true for all edges of the star S, we conclude that

$$\mathbb{E}[G(\mathrm{UA}(n, P))] - \mathbb{E}[G(\mathrm{UA}(n, S))] = \mathbb{E}[g(\mathrm{UA}(n, P), e_3)] - \mathbb{E}[g(\mathrm{UA}(n, S), e_1)],$$

where e_3 is the remaining edge of P (i.e., the middle edge of the path). Using (1), we thus have

$$\begin{aligned}
&\mathbb{E}[G(\mathrm{UA}(n, P))] - \mathbb{E}[G(\mathrm{UA}(n, S))] \\
&\qquad = \frac{1}{n^4}\left(\mathbb{E}\left[B^2_{2,2,n-4}\left(n - B_{2,2,n-4}\right)^2\right] - \mathbb{E}\left[B^2_{1,3,n-4}\left(n - B_{1,3,n-4}\right)^2\right]\right) \\
&\qquad = \frac{2n^3 + 5n^2 + 8n + 5}{140n^3},
\end{aligned}$$

where the last equality is attained via a straightforward calculation using explicit formulae for the first four moments of the beta-binomial distribution. We see that

$$\lim_{n\to\infty} (\mathbb{E}[G(\mathrm{UA}(n, P))] - \mathbb{E}[G(\mathrm{UA}(n, S))]) = \frac{1}{70} \neq 0,$$

which establishes (5).

It remains to prove (6). We show now that $\limsup_{n\to\infty} \mathbb{V}\mathrm{ar}\left[G(\mathrm{UA}(n,P))\right] < \infty$; the proof that $\limsup_{n\to\infty} \mathbb{V}\mathrm{ar}\left[G(\mathrm{UA}(n,S))\right] < \infty$ is identical. To abbreviate notation, write T_n for $\mathrm{UA}\left(n,P\right)$. Similarly as above, for $j \geq 4$ let e_j be the edge in $T_{j+1} \setminus T_j$, and let e_1, e_2, and e_3 be the edges of P in some arbitrary order. Using Cauchy-Schwarz we have that

$$\mathbb{V}\mathrm{ar}\left[G(T_n)\right] \leq \left(\sum_{j=1}^{n-1} \sqrt{\mathbb{V}\mathrm{ar}\left[g(T_n,e_j)\right]}\right)^2. \tag{7}$$

For any edge e_i we clearly have $0 \leq g(T_n,e_i) \leq 1$, and so

$$\sum_{j=1}^{3} \sqrt{\mathbb{V}\mathrm{ar}\left[g(T_n,e_j)\right]} \leq 3. \tag{8}$$

Next, fix $4 \leq j \leq n-1$. Using formula (2) we know that

$$g(T_n,e_j) \stackrel{d}{=} \frac{1}{n^4} B^2_{1,j,n-j-1}(n - B_{1,j,n-j-1})^2.$$

The estimate (3) yields $\mathbb{E}[B^4_{1,j,n-j-1}] \leq Cn^4/j^4$, where $C > 0$ is a universal constant. Consequently, we have

$$\mathbb{E}[g(T_n,e_j)^2] \leq C/j^4, \quad \forall j \in \{4,\ldots,n-1\},$$

which, in turn, implies that

$$\sqrt{\mathbb{V}\mathrm{ar}\left[g(T_n,e_j)\right]} \leq C/j^2.$$

Plugging this inequality and (8) into (7) establishes (6). This completes the proof of $\delta\left(P_4,S_4\right) > 0$.

The statistic $G\left(\cdot\right)$ cannot distinguish between all pairs of non-isomorphic trees; however, appropriate generalizations of it can. An alternative description of $G\left(\cdot\right)$ is as follows. Let τ be a tree consisting of two vertices connected by a single edge. Up to normalization, the quantity $G\left(T\right)$ is equal to the sum over all embeddings $\varphi : \tau \to T$ of the product of the squares of the connected components of $T \setminus \varphi\left(\tau\right)$. A natural generalization of this definition is to take τ to be an arbitrary finite tree. Moreover, we can assign natural numbers to each vertex of τ, which determines the power to which we raise the size of the respective connected components. In this way we obtain a family of statistics associated with so-called *decorated trees*. It turns that this generalized family of statistics can indeed distinguish between any pair of non-isomorphic trees; for details see Section 3.

2.3 Distinguishing large stars: a proof of Theorem 2

In the following we first give an upper bound on the probability that $G\left(\mathrm{UA}\left(n,T\right)\right)$ is small, and then we give an upper bound on the probability that $G\left(\mathrm{UA}\left(n,S_k\right)\right)$ is not too small. The two together will prove Theorem 2.

First, fix a tree T and choose an arbitrary edge $e_1 \in E\left(T\right)$. Let T'_n and T''_n be the two connected components of $\mathrm{UA}(n,T) \setminus \{e_1\}$, defined consistently such that $T'_j \subset T'_{j+1}$ and $T''_j \subset T''_{j+1}$ (and otherwise the order is chosen arbitrarily). We have

$$g(\mathrm{UA}(n,T),e_1) = \frac{1}{n^4}|T'_n|^2|T''_n|^2 = \frac{1}{n^4}|T'_n|^2(n-|T'_n|)^2.$$

By equation (1) we have

$$|T'_n| \stackrel{d}{=} B^2_{a,|T|-a,n-|T|},$$

where $a := |T'_{|T|}| \geq 1$. Using Fact 2 we then have that for all $t > 0$,

$$\mathbb{P}\left(\frac{|T'_n|^2}{n^2} < t\frac{1}{|T|^2}\right) \leq \mathbb{P}\left(\frac{|T'_n|^2}{n^2} < t\frac{a^2}{|T|^2}\right) = \mathbb{P}\left(\frac{B^2_{a,|T|-a,n-|T|}}{n^2} < t\frac{a^2}{|T|^2}\right) \leq C\sqrt{t}. \tag{9}$$

Consider the event $E_n = \{|T'_n| \leq n/2\}$ and note that if E_n holds then $g(\mathrm{UA}(n,T), e_1) \geq |T'_n|^2 / (4n^2)$. This, together with (9), gives

$$\mathbb{P}\left(E_n \cap \left\{g(\mathrm{UA}(n,T), e_1) < t\frac{1}{4|T|^2}\right\}\right) \leq C\sqrt{t}, \quad \forall t > 0.$$

By repeating the above argument with T''_n instead of T'_n, we also have

$$\mathbb{P}\left(E_n^C \cap \left\{g(\mathrm{UA}(n,T), e_1) < t\frac{1}{4|T|^2}\right\}\right) \leq C\sqrt{t}, \quad \forall t > 0,$$

and thus we conclude that

$$\mathbb{P}\left(g(\mathrm{UA}(n,T), e_1) < t\frac{1}{4|T|^2}\right) \leq 2C\sqrt{t}, \quad \forall t > 0.$$

Now since $G(\mathrm{UA}(n,T)) \geq g(\mathrm{UA}(n,T), e_1)$, and by setting $z = t/\left(4|T|^2\right)$, we finally get that

$$\mathbb{P}(G(\mathrm{UA}(n,T)) < z) \leq 4C\sqrt{z}\,|T|, \quad \forall z > 0. \tag{10}$$

In order to understand the distribution of $G(\mathrm{UA}(n,S_k))$ we first estimate its mean:

$$\begin{aligned}
&\mathbb{E}[G(\mathrm{UA}(n,S_k))] \\
&\quad = \frac{k-1}{n^4}\mathbb{E}\left[B^2_{1,k-1,n-k}(n - B_{1,k-1,n-k})^2\right] + \sum_{j=k}^{n-1}\frac{1}{n^4}\mathbb{E}\left[B^2_{1,j,n-j-1}(n - B_{1,j,n-j-1})^2\right] \\
&\quad \leq \frac{k-1}{n^2}\mathbb{E}\left[B^2_{1,k-1,n-k}\right] + \frac{1}{n^2}\sum_{j=k}^{n-1}\mathbb{E}\left[B^2_{1,j,n-j-1}\right] \overset{(3)}{\leq} \frac{C'}{k} + \sum_{j=k}^{n-1}\frac{C'}{j^2} \leq \frac{3C'}{k}
\end{aligned}$$

for some absolute constant C'. Now using Markov's inequality with this estimate, and also taking $z = 3C'/\sqrt{k}$ in the inequality (10), we get that

$$\mathbb{P}\left(G(\mathrm{UA}(n,S_k)) \geq 3C'/\sqrt{k}\right) \leq \frac{1}{\sqrt{k}} \qquad \text{and} \qquad \mathbb{P}\left(G(\mathrm{UA}(n,T)) < 3C'/\sqrt{k}\right) \leq \frac{C''}{k^{1/4}}$$

for some absolute constant C''. This then immediately implies that $\delta(S_k, T) \to 1$ as $k \to \infty$.

3 Proof of Theorem 1

After the intuition and simple examples provided in Section 2, in this section we fully prove Theorem 1. As mentioned in Section 1.1, the proof shares some features with the proof in Curien et al. [2014] for preferential attachment, but is different in several ways. These differences are discussed in detail after the proof, in Section 4.1.

Notation. For a graph G, denote by $V(G)$ the set of its vertices, by $E(G)$ the set of its edges, and by $\operatorname{diam}(G)$ its diameter. For brevity, we often write $v \in G$ instead of $v \in V(G)$. For integers $k, j \geq 1$, define the descending factorial $[k]_j = k(k-1)\dots(k-j+1)$, and also let $[k]_0 = 1$. For two sequences of real numbers $\{a_n\}_{n\geq 0}$ and $\{b_n\}_{n\geq 0}$, we write $a_n \lesssim b_n$ (to be read as a_n is less than b_n up to log factors) if there exist constants $c > 0$, $\gamma \in \mathbb{R}$, and n_0 such that $|a_n| \leq c(\log(n))^{\gamma} |b_n|$ for all $n \geq n_0$. For a sequence $\{a_n\}_{n\geq 0}$ of real numbers, define $\Delta_n a = a_{n+1} - a_n$ for $n \geq 0$.

3.1 Decorated trees

A *decorated tree* is a pair $\underline{\tau} = (\tau, \ell)$ consisting of a tree τ and a family of nonnegative integers $(\ell(v); v \in \tau)$, called labels, associated with its vertices; see Figure 2 for an illustration. Let $\mathcal{D}$ denote the set of all decorated trees, $\mathcal{D}_+$ the set of all decorated trees where every label is positive, $\mathcal{D}_0$ the set of all decorated trees where there exists a label which is zero, and finally let $\mathcal{D}_0^*$ denote the set of all decorated trees where there exists a leaf which has label zero.

Define $|\underline{\tau}|$ to be the number of vertices of τ, and let $w(\underline{\tau}) := \sum_{v\in\tau} \ell(v)$ denote the total weight of $\underline{\tau}$. For $\underline{\tau}, \underline{\tau}' \in \mathcal{D}$, let $\underline{\tau} \prec \underline{\tau}'$ if $|\underline{\tau}| < |\underline{\tau}'|$ and $w(\underline{\tau}) \leq w(\underline{\tau}')$ or $|\underline{\tau}| \leq |\underline{\tau}'|$ and $w(\underline{\tau}) < w(\underline{\tau}')$. This defines a strict partial order $\prec$, and let $\preccurlyeq$ denote the associated partial order, i.e., $\underline{\tau} \preccurlyeq \underline{\tau}'$ if and only if $\underline{\tau} \prec \underline{\tau}'$ or $\underline{\tau} = \underline{\tau}'$.

For $\underline{\tau} \in \mathcal{D}$, let $L(\underline{\tau})$ denote the set of leaves of τ, let $L_0(\underline{\tau}) := \{v \in L(\underline{\tau}) : \ell(v) = 0\}$, $L_1(\underline{\tau}) := \{v \in L(\underline{\tau}) : \ell(v) = 1\}$, and $L_{0,1}(\underline{\tau}) := L_0(\underline{\tau}) \cup L_1(\underline{\tau})$. For $\underline{\tau} \in \mathcal{D}$ and $v \in L(\underline{\tau})$, define $\underline{\tau}_v$ to be the same as $\underline{\tau}$ except the leaf v and its label are removed. For $\underline{\tau} \in \mathcal{D}$ and a vertex $v \in \underline{\tau}$ such that $\ell(v) \geq 2$, define $\underline{\tau}'_v$ to be the same as $\underline{\tau}$ except the label of v is decreased by one, i.e., $\ell_{\underline{\tau}'_v}(v) = \ell_{\underline{\tau}}(v) - 1$.

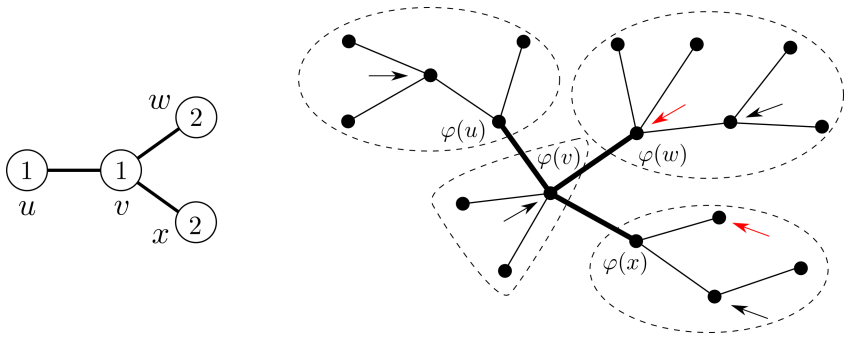


Figure 2: **A decorated tree and a decorated embedding.** On the left is a decorated tree $\underline{\tau} = (\tau, \ell)$ with four vertices, two of them having label 1, and two of them having label 2. On the right is a larger tree T, and an embedding $\varphi : \tau \to T$ depicted in bold. The connected components of the forest $\widehat{T}(T, \tau, \varphi)$ are circled with dashed lines, with the component sizes being $f_{\varphi(u)}(T) = 5$, $f_{\varphi(v)}(T) = 3$, $f_{\varphi(w)}(T) = 6$, and $f_{\varphi(x)}(T) = 4$. A decorated embedding $\underline{\varphi}$ is also depicted, which consists of the embedding φ together with the mapping of $w(\underline{\tau}) = 6$ arrows to vertices of T. The arrows in each subtree are distinguishable, which is why they are depicted using different colors.

3.2 Statistics and distinguishing martingales

Given two trees τ and T, a map $\varphi : \tau \to T$ is called an *embedding* if φ is an injective graph homomorphism. That is, φ is an injective map from $V(\tau)$ to $V(T)$ such that $\{u, v\} \in E(\tau)$ implies that $\{\varphi(u), \varphi(v)\} \in E(T)$.

For two trees τ and T, and an embedding $\varphi : \tau \to T$, denote by $\widehat{T} = \widehat{T}(T, \tau, \varphi)$ the forest obtained from T by removing the images of the edges of τ under the embedding φ; see Figure 2 for an illustration. Note that the forest $\widehat{T}$ consists of exactly $|\tau|$ trees, and each tree contains exactly one vertex which is the image of a vertex of τ under φ. For $v \in \tau$, denote by $f_{\varphi(v)}(T)$ the number of vertices of the tree in $\widehat{T}$ which contains the vertex $\varphi(v)$. Using this notation, for a decorated tree $\underline{\tau} \in \mathcal{D}$ define

$$F_{\underline{\tau}}(T) = \sum_{\varphi} \prod_{v \in \tau} \left[f_{\varphi(v)}(T) \right]_{\ell(v)}, \tag{11}$$

where the sum is over all embeddings $\varphi : \tau \to T$. If there are no such embeddings then $F_{\underline{\tau}}(T) = 0$ by definition. Note that if $\underline{\tau}$ consists of a single vertex with label $k \geq 0$, then $F_{\underline{\tau}}(T) = |T| \times [|T|]_k$.

The quantity $F_{\underline{\tau}}(T)$ has a combinatorial interpretation which is useful: it is the number of *decorated embeddings* of $\underline{\tau}$ in T, defined as follows. Imagine that for each vertex $v \in \tau$ there are $\ell(v)$ distinguishable (i.e., ordered) arrows pointing to v. A decorated embedding $\underline{\varphi}$ is an embedding φ of τ in T, together with a mapping of the arrows to vertices of T in such a way that each arrow pointing to $v \in \tau$ is mapped to a vertex in the tree of $\widehat{T}$ that contains $\varphi(v)$, with distinct arrows mapped to distinct vertices. See Figure 2 for an illustration.

The quantities $F_{\underline{\tau}}$ are also more amenable to analysis than other statistics, because their expectations satisfy recurrence relations, as described in Section 3.3. Using the statistics $F_{\underline{\tau}}$ it is possible to create martingales that distinguish between different seeds.

Proposition 1 *Let $\underline{\tau} \in \mathcal{D}_+$. There exists a family of constants $\{c_n(\underline{\tau}, \underline{\tau}') : \underline{\tau}' \in \mathcal{D}_+, \underline{\tau}' \preccurlyeq \underline{\tau}, n \geq 2\}$ with $c_n(\underline{\tau}, \underline{\tau}) > 0$ such that for every seed tree S, the process $\left\{ M_{\underline{\tau}}^{(S)}(n) \right\}_{n \geq |S|}$ defined by*

$$M_{\underline{\tau}}^{(S)}(n) = \sum_{\underline{\tau}' \in \mathcal{D}_+ : \underline{\tau}' \preccurlyeq \underline{\tau}} c_n\left(\underline{\tau}, \underline{\tau}'\right) F_{\underline{\tau}'}\left(\mathrm{UA}(n, S)\right)$$

is a martingale with respect to the natural filtration $\mathcal{F}_n = \sigma\{\mathrm{UA}(|S|, S), \ldots, \mathrm{UA}(n, S)\}$ and is bounded in L^2.

Note that in the construction of these martingales we only use decorated trees where every label is positive. As we shall see, we analyze decorated trees having a label which is zero in order to show that the martingales above are bounded in L^2. See Sections 3.4 and 4.1 for more details and discussion on this point.

We now prove Theorem 1 using Proposition 1, which we then prove in the following subsections.

Proof of Theorem 1 Let S and T be two non-isomorphic trees with at least three vertices, and let $n_0 := \max\{|S|, |T|\}$. First we show that there exists $\underline{\tau} \in \mathcal{D}_+$ such that

$$\mathbb{E}\left[F_{\underline{\tau}}(\mathrm{UA}(n_0, S))\right] \neq \mathbb{E}\left[F_{\underline{\tau}}(\mathrm{UA}(n_0, T))\right]. \tag{12}$$

Assume without loss of generality that $|S| \leq |T|$, and let $\underline{\tau}$ be equal to T with labels $\ell(v) = 1$ for all $v \in T$. Then for every tree T' with $|T'| = |T|$ we have $F_{\underline{\tau}}(T') = F_{\underline{\tau}}(T) \times \mathbf{1}_{\{T' = T\}}$, since if T' and T are non-isomorphic then there is no embedding of T in T'. Note also that $F_{\underline{\tau}}(T)$ is the number of automorphisms of T, which is positive. Consequently we have

$$\mathbb{E}\left[F_{\underline{\tau}}(\mathrm{UA}(n_0, S))\right] = F_{\underline{\tau}}(T) \times \mathbb{P}\left[\mathrm{UA}(n_0, S) = T\right].$$

When $|S| = |T|$, we have $\mathbb{P}[\mathrm{UA}(n_0, S) = T] = 0$. When $|S| < |T|$ it is easy to see that the isomorphism class of $\mathrm{UA}(n_0, S)$ is nondeterministic (here we use the fact that $|S| \geq 3$), and so $\mathbb{P}[\mathrm{UA}(n_0, S) = T] < 1$. In both cases we have that (12) holds.

Now let $\underline{\tau} \in \mathcal{D}_+$ be a minimal (for the partial order $\preccurlyeq$ on $\mathcal{D}_+$) decorated tree for which (12) holds. By definition we then have that $\mathbb{E}\left[F_{\underline{\tau}'}(\mathrm{UA}(n_0, S))\right] = \mathbb{E}\left[F_{\underline{\tau}'}(\mathrm{UA}(n_0, T))\right]$ for every $\underline{\tau}' \in \mathcal{D}_+$ such that $\underline{\tau}' \prec \underline{\tau}$. By the construction of the martingales in Proposition 1 we then have that

$$\mathbb{E}\left[M_{\underline{\tau}}^{(S)}(n_0)\right] \neq \mathbb{E}\left[M_{\underline{\tau}}^{(T)}(n_0)\right].$$

Clearly for any $n \geq n_0$ we have that $\mathrm{TV}(\mathrm{UA}(n, S), \mathrm{UA}(n, T)) \geq \mathrm{TV}\left(M_{\underline{\tau}}^{(S)}(n), M_{\underline{\tau}}^{(T)}(n)\right)$. Now let (X, Y) be a coupling of $\left(M_{\underline{\tau}}^{(S)}(n), M_{\underline{\tau}}^{(T)}(n)\right)$. We need to bound from below $\mathbb{P}(X \neq Y)$ in order to obtain a lower bound on $\mathrm{TV}\left(M_{\underline{\tau}}^{(S)}(n), M_{\underline{\tau}}^{(T)}(n)\right)$. Using Paley-Zigmund's inequality we have that

$$\mathbb{P}(X \neq Y) \geq \frac{(\mathbb{E}[|X - Y|])^2}{\mathbb{E}\left[(X - Y)^2\right]}.$$

By Jensen's inequality we have that $(\mathbb{E}[|X - Y|])^2 \geq (\mathbb{E}[X] - \mathbb{E}[Y])^2$, and furthermore

$$\begin{aligned}
\mathbb{E}\left[(X - Y)^2\right] &= \mathbb{E}\left[(X - \mathbb{E}[X] + \mathbb{E}[X] - \mathbb{E}[Y] + \mathbb{E}[Y] - Y)^2\right] \\
&= \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y) + (\mathbb{E}[X] - \mathbb{E}[Y])^2 + 2\mathbb{E}[(X - \mathbb{E}[X])(\mathbb{E}[Y] - Y)] \\
&\leq 2\mathbb{V}\mathrm{ar}(X) + 2\mathbb{V}\mathrm{ar}(Y) + (\mathbb{E}[X] - \mathbb{E}[Y])^2.
\end{aligned}$$

Thus we have shown that

$$\begin{aligned}
&\mathrm{TV}(\mathrm{UA}(n, S), \mathrm{UA}(n, T)) \\
&\qquad \geq \frac{\left(\mathbb{E}\left[M_{\underline{\tau}}^{(S)}(n)\right] - \mathbb{E}\left[M_{\underline{\tau}}^{(T)}(n)\right]\right)^2}{2\mathbb{V}\mathrm{ar}\left(M_{\underline{\tau}}^{(S)}(n)\right) + 2\mathbb{V}\mathrm{ar}\left(M_{\underline{\tau}}^{(T)}(n)\right) + \left(\mathbb{E}\left[M_{\underline{\tau}}^{(S)}(n)\right] - \mathbb{E}\left[M_{\underline{\tau}}^{(T)}(n)\right]\right)^2}.
\end{aligned}$$

Since $M_{\underline{\tau}}^{(S)}$ and $M_{\underline{\tau}}^{(T)}$ are martingales, for every $n \geq n_0$ we have that $\mathbb{E}\left[M_{\underline{\tau}}^{(S)}(n)\right] - \mathbb{E}\left[M_{\underline{\tau}}^{(T)}(n)\right] = \mathbb{E}\left[M_{\underline{\tau}}^{(S)}(n_0)\right] - \mathbb{E}\left[M_{\underline{\tau}}^{(T)}(n_0)\right] \neq 0$. Also, since the two martingales are bounded in L^2, we have that $\mathbb{V}\mathrm{ar}\left(M_{\underline{\tau}}^{(S)}(n)\right) + \mathbb{V}\mathrm{ar}\left(M_{\underline{\tau}}^{(T)}(n)\right)$ is bounded as $n \to \infty$. We conclude that $\delta(S, T) > 0$. ■

3.3 Recurrence relation

The following recurrence relation for the conditional expectations of $F_{\underline{\tau}}(\mathrm{UA}(n, S))$ is key to estimating the moments of $F_{\underline{\tau}}(\mathrm{UA}(n, S))$.

Lemma 1 *Let $\underline{\tau} \in \mathcal{D}$ such that $|\underline{\tau}| \geq 2$. Then for every seed tree S and for every $n \geq |S|$ we have*

$$\begin{aligned}
\mathbb{E}\left[F_{\underline{\tau}}(\mathrm{UA}(n+1, S)) \,\middle|\, \mathcal{F}_n\right] &= \left(1 + \frac{w(\underline{\tau})}{n}\right) F_{\underline{\tau}}(\mathrm{UA}(n, S)) \\
&\quad + \frac{1}{n}\left\{\sum_{v \in \tau : \ell(v) \geq 2} \ell(v)(\ell(v) - 1) F_{\underline{\tau}_v'}(\mathrm{UA}(n, S)) + \sum_{v \in L_{0,1}(\underline{\tau})} F_{\underline{\tau}_v}(\mathrm{UA}(n, S))\right\}. \quad (13)
\end{aligned}$$

Proof Fix $\underline{\tau} \in \mathcal{D}$ with $|\underline{\tau}| \geq 2$, fix a seed tree S, and let $n \geq |S|$. To simplify notation we omit the dependence on S and write T_n instead of $\mathrm{UA}(n, S)$. When evaluating $\mathbb{E}\left[F_{\underline{\tau}}(T_{n+1}) \,\middle|\, \mathcal{F}_n\right]$ we work conditionally on $\mathcal{F}_n$, so we may consider T_n as being fixed.

Let u_{n+1} denote the vertex present in T_{n+1} but not in T_n, and let u_n denote its neighbor in T_{n+1}. Let $\mathcal{E}_{n+1}$ denote the set of all embeddings $\varphi : \tau \to T_{n+1}$; we can write $\mathcal{E}_{n+1}$ as the disjoint union of the set of those using only vertices of T_n, denoted by $\mathcal{E}_n$, and the set of those using the new vertex u_{n+1}, denoted by $\mathcal{E}_{n+1} \setminus \mathcal{E}_n$. To simplify notation, if $\underline{\tau} \in \mathcal{D}$, T is a tree, and $\varphi : \tau \to T$ is an embedding, write $\mathcal{W}_\varphi(T) = \prod_{v \in \tau} \left[f_{\varphi(v)}(T)\right]_{\ell(v)}$ for the number of decorated embeddings of $\underline{\tau}$ in T that use the embedding φ. We then have $F_{\underline{\tau}}(T_{n+1}) = \sum_{\varphi \in \mathcal{E}_n} \mathcal{W}_\varphi(T_{n+1}) + \sum_{\varphi \in \mathcal{E}_{n+1} \setminus \mathcal{E}_n} \mathcal{W}_\varphi(T_{n+1})$ and we deal with the two sums separately.

First let $\varphi \in \mathcal{E}_n$. For $v \in \tau$, denote by E_v the event that u_n is in the same tree of $\widehat{T_n}$ as $\varphi(v)$ (recall the definition of $\widehat{T_n}$ from Section 3.2). Clearly $\mathbb{P}(E_v \,|\, \mathcal{F}_n) = f_{\varphi(v)}(T_n)/n$. Under the event E_v we have that $f_{\varphi(v)}(T_{n+1}) = f_{\varphi(v)}(T_n) + 1$, while for every $v' \in \tau \setminus \{v\}$ we have $f_{\varphi(v')}(T_{n+1}) = f_{\varphi(v')}(T_n)$. Now using the identities $[d+1]_\ell = [d]_\ell + \ell \times [d]_{\ell-1}$ and $d \times [d]_{\ell-1} = [d]_\ell + (\ell-1) \times [d]_{\ell-1}$, which hold for every $d, \ell \geq 1$, and also using $[d+1]_0 = [d]_0$, we have that

$$
\begin{aligned}
&\mathbb{E}\left[\mathcal{W}_\varphi(T_{n+1}) \,|\, \mathcal{F}_n\right] \\
&= \sum_{v \in \tau} \frac{f_{\varphi(v)}(T_n)}{n} \left[f_{\varphi(v)}(T_n) + 1\right]_{\ell(v)} \prod_{v' \in \tau \setminus \{v\}} \left[f_{\varphi(v')}(T_n)\right]_{\ell(v')} \\
&= \mathcal{W}_\varphi(T_n) + \frac{1}{n} \sum_{v \in \tau : \ell(v) \geq 1} \ell(v)\, f_{\varphi(v)}(T_n) \left[f_{\varphi(v)}(T_n)\right]_{\ell(v)-1} \prod_{v' \in \tau \setminus \{v\}} \left[f_{\varphi(v')}(T_n)\right]_{\ell(v')} \\
&= \left(1 + \frac{w(\underline{\tau})}{n}\right) \mathcal{W}_\varphi(T_n) + \frac{1}{n} \sum_{v \in \tau : \ell(v) \geq 2} \ell(v)\,(\ell(v)-1) \left[f_{\varphi(v)}(T_n)\right]_{\ell(v)-1} \prod_{v' \in \tau \setminus \{v\}} \left[f_{\varphi(v')}(T_n)\right]_{\ell(v')} \\
&= \left(1 + \frac{w(\underline{\tau})}{n}\right) \mathcal{W}_\varphi(T_n) + \frac{1}{n} \sum_{v \in \tau : \ell(v) \geq 2} \ell(v)\,(\ell(v)-1)\, \mathcal{W}_{\varphi'_v}(T_n),
\end{aligned}
$$

where φ'_v is the embedding equal to φ of the decorated tree $\underline{\tau}'_v$. Now as φ runs through the embeddings of $\underline{\tau}$ in T_n, φ'_v runs exactly through the embeddings of $\underline{\tau}'_v$. So we have that

$$
\sum_{\varphi \in \mathcal{E}_n} \mathbb{E}\left[\mathcal{W}_\varphi(T_{n+1}) \,|\, \mathcal{F}_n\right] = \left(1 + \frac{w(\underline{\tau})}{n}\right) F_{\underline{\tau}}(T_n) + \frac{1}{n} \sum_{v \in \tau : \ell(v) \geq 2} \ell(v)\,(\ell(v)-1)\, F_{\underline{\tau}'_v}(T_n). \tag{14}
$$

Now fix T_{n+1} and consider $\varphi \in \mathcal{E}_{n+1} \setminus \mathcal{E}_n$. Let $w \in \tau$ be such that $\varphi(w) = u_{n+1}$. Since φ is an embedding, we must have $w \in L(\underline{\tau})$. Note that if $w \notin L_{0,1}(\underline{\tau})$ then $\mathcal{W}_\varphi(T_{n+1}) = 0$. If $w \in L_{0,1}(\underline{\tau})$ then denote by $\mathcal{E}_w$ the set of all embeddings $\varphi \in \mathcal{E}_{n+1} \setminus \mathcal{E}_n$ such that $\varphi(w) = u_{n+1}$. Now fix $w \in L_{0,1}(\underline{\tau})$ and $\varphi \in \mathcal{E}_w$. Note that φ restricted to $\tau \setminus \{w\}$ is an embedding of $\underline{\tau}_w$ in T_n; call this φ_w. Let x be the neighbor of w in τ. We then must have $\varphi(x) = u_n$, and also $\left[f_{\varphi(w)}(T_{n+1})\right]_{\ell(w)} = 1$ (irrespective of whether $w \in L_0(\underline{\tau})$ or $w \in L_1(\underline{\tau})$). Furthermore, for every $w' \in \tau \setminus \{w\}$ we have $f_{\varphi(w')}(T_{n+1}) = f_{\varphi(w')}(T_n)$. Thus we have

$$
\mathcal{W}_\varphi(T_{n+1}) = \mathcal{W}_{\varphi_w}(T_n)\, \mathbf{1}_{\{\varphi(w) = u_{n+1}\}}.
$$

For fixed $w \in L_{0,1}(\underline{\tau})$, as φ runs through $\mathcal{E}_w$, φ_w runs through all the embeddings of $\underline{\tau}_w$ in T_n. So summing over $w \in L$ we obtain

$$
\sum_{\varphi \in \mathcal{E}_{n+1} \setminus \mathcal{E}_n} \mathcal{W}_\varphi(T_{n+1}) = \sum_{w \in L_{0,1}(\underline{\tau})} \sum_{\varphi_w : \underline{\tau}_w \to T_n} \mathcal{W}_{\varphi_w}(T_n)\, \mathbf{1}_{\{\varphi(w) = u_{n+1}\}} = \sum_{w \in L_{0,1}(\underline{\tau})} F_{\underline{\tau}_w}(T_n)\, \mathbf{1}_{\{\varphi(w) = u_{n+1}\}}.
$$

Now taking conditional expectation given $\mathcal{F}_n$, we get that

$$\mathbb{E}\left[\sum_{\varphi\in\mathcal{E}_{n+1}\setminus\mathcal{E}_n}\mathcal{W}_\varphi\left(T_{n+1}\right)\middle|\mathcal{F}_n\right]=\frac{1}{n}\sum_{w\in L_{0,1}(\underline{\tau})}F_{\underline{\tau}_w}\left(T_n\right). \tag{15}$$

Summing (14) and (15) we obtain (13). ■

3.4 Moment estimates

Using the recurrence relation of Lemma 1 proved in the previous subsection, we now establish moment estimates on the number of decorated embeddings $F_{\underline{\tau}}\left(\mathrm{UA}\left(n,S\right)\right)$. These are then used in the next subsection to show that the martingales of Proposition 1 are bounded in L^2.

The first moment estimates are a direct corollary of Lemma 1.

Corollary 1 *Let $\underline{\tau}\in\mathcal{D}$ be a decorated tree and let S be a seed tree.*

(a) We have that $n^{w(\underline{\tau})}\lesssim\mathbb{E}\left[F_{\underline{\tau}}\left(\mathrm{UA}\left(n,S\right)\right)\right]$.

(b) If $|\underline{\tau}|\geq 2$ and $\underline{\tau}\in\mathcal{D}\setminus\mathcal{D}_0^*$, then we have that $\mathbb{E}\left[F_{\underline{\tau}}\left(\mathrm{UA}\left(n,S\right)\right)\right]\lesssim n^{w(\underline{\tau})}$.

(c) If $|\underline{\tau}|=1$ or $\underline{\tau}\in\mathcal{D}_0^*$, then we have that $\mathbb{E}\left[F_{\underline{\tau}}\left(\mathrm{UA}\left(n,S\right)\right)\right]\lesssim n^{w(\underline{\tau})+1}$.

Proof Fix a seed tree S and, as before, write T_n instead of $\mathrm{UA}\left(n,S\right)$ in order to simplify notation. First, recall that if $|\underline{\tau}|=1$, then $F_{\underline{\tau}}\left(T_n\right)=n\times[n]_{w(\underline{\tau})}$, so the statements of part (a) and (c) hold in this case. In the following we can therefore assume that $|\underline{\tau}|\geq 2$.

Lemma 1 then implies that for every $n\geq|S|$ we have

$$\mathbb{E}\left[F_{\underline{\tau}}\left(T_{n+1}\right)\right]\geq\left(1+\frac{w\left(\underline{\tau}\right)}{n}\right)\mathbb{E}\left[F_{\underline{\tau}}\left(T_n\right)\right].$$

Since there exists n_0 such that $\mathbb{E}\left[F_{\underline{\tau}}\left(T_{n_0}\right)\right]>0$ (one can take, e.g., $n_0=|S|+|\underline{\tau}|$), this immediately implies part (a) (see Appendix B for further details).

We prove parts (b) and (c) by induction on $\underline{\tau}$ for the partial order $\preccurlyeq$. We have already checked that the statement holds when $|\underline{\tau}|=1$, so the base case of the induction holds. Now fix $\underline{\tau}$ such that $|\underline{\tau}|\geq 2$, and assume that (b) and (c) hold for all $\underline{\tau}'$ such that $\underline{\tau}'\prec\underline{\tau}$. There are two cases to consider: either $\underline{\tau}\in\mathcal{D}\setminus\mathcal{D}_0^*$ or $\underline{\tau}\in\mathcal{D}_0^*$.

If $\underline{\tau}\in\mathcal{D}\setminus\mathcal{D}_0^*$ then $L_0\left(\underline{\tau}\right)=\emptyset$. Note that for every $v\in L_1\left(\underline{\tau}\right)$ we have $w\left(\underline{\tau}_v\right)=w\left(\underline{\tau}\right)-1$, and also for every $v\in\tau$ such that $\ell\left(v\right)\geq 2$ we have $w\left(\underline{\tau}'_v\right)=w\left(\underline{\tau}\right)-1$. Therefore by induction for every $v\in L_1\left(\underline{\tau}\right)$ we have $\mathbb{E}\left[F_{\underline{\tau}_v}\left(T_n\right)\right]\lesssim n^{w(\underline{\tau})}$ and also for every $v\in\tau$ such that $\ell\left(v\right)\geq 2$ we have $\mathbb{E}\left[F_{\underline{\tau}'_v}\left(T_n\right)\right]\lesssim n^{w(\underline{\tau})}$. So by Lemma 1 there exist constants $C,\gamma>0$ such that

$$\mathbb{E}\left[F_{\underline{\tau}}\left(T_{n+1}\right)\right]\leq\left(1+\frac{w\left(\underline{\tau}\right)}{n}\right)\mathbb{E}\left[F_{\underline{\tau}}\left(T_n\right)\right]+C\left(\log\left(n\right)\right)^\gamma n^{w(\underline{\tau})-1}. \tag{16}$$

This then implies that $\mathbb{E}\left[F_{\underline{\tau}}\left(T_n\right)\right]\lesssim n^{w(\underline{\tau})}$; see Appendix B for details.

If $\underline{\tau}\in\mathcal{D}_0^*$ then $L_0\left(\underline{\tau}\right)\neq\emptyset$ and note that for every $v\in L_0\left(\underline{\tau}\right)$ we have $w\left(\underline{\tau}_v\right)=w\left(\underline{\tau}\right)$. If for every $v\in L_0\left(\underline{\tau}\right)$ we have $\underline{\tau}_v\in\mathcal{D}\setminus\mathcal{D}_0^*$, then the same argument as in the previous paragraph

goes through, and we have that $\mathbb{E}\left[F_{\underline{\tau}}(T_n)\right] \lesssim n^{w(\underline{\tau})}$. However, if there exists $v \in L_0(\underline{\tau})$ such that $\underline{\tau}_v \in \mathcal{D}_0^*$, then (16) does not hold in this case; instead we have from Lemma 1 that there exist constants $C, \gamma > 0$ such that

$$\mathbb{E}\left[F_{\underline{\tau}}(T_{n+1})\right] \leq \left(1 + \frac{w(\underline{\tau})}{n}\right) \mathbb{E}\left[F_{\underline{\tau}}(T_n)\right] + C(\log(n))^{\gamma} n^{w(\underline{\tau})}.$$

Similarly as before, this then implies that $\mathbb{E}\left[F_{\underline{\tau}}(T_n)\right] \lesssim n^{w(\underline{\tau})+1}$; see Appendix B for details. ■

The second moment estimates require additional work. First, recall again that if $|\underline{\tau}| = 1$, then $F_{\underline{\tau}}(\mathrm{UA}(n,S)) = n \times [n]_{w(\underline{\tau})}$. Consequently, we have that $F_{\underline{\tau}}(\mathrm{UA}(n,S))^2 \lesssim n^{2w(\underline{\tau})+2}$ and also that $\left(F_{\underline{\tau}}(\mathrm{UA}(n+1,S)) - F_{\underline{\tau}}(\mathrm{UA}(n,S))\right)^2 \lesssim n^{2w(\underline{\tau})}$.

Lemma 2 *Let $\underline{\tau} \in \mathcal{D}_+$ with $|\underline{\tau}| \geq 2$ and let S be a seed tree.*

(a) We have that $\mathbb{E}\left[F_{\underline{\tau}}(\mathrm{UA}(n,S))^2\right] \lesssim n^{2w(\underline{\tau})}$.

(b) We have that $\mathbb{E}\left[\left(F_{\underline{\tau}}(\mathrm{UA}(n+1,S)) - F_{\underline{\tau}}(\mathrm{UA}(n,S))\right)^2\right] \lesssim n^{2w(\underline{\tau})-2}$.

We note that part (a) of the lemma follows in a short and simple way once part (b) is proven. However, for expository reasons, we first prove part (a) directly, and then prove part (b), whose proof is similar to, and builds upon, the proof of part (a).

Proof Fix $\underline{\tau} \in \mathcal{D}_+$ with $|\underline{\tau}| \geq 2$ and a seed tree S. Define $K \equiv K(\underline{\tau}) := \max\{4(|\underline{\tau}| + w(\underline{\tau})), 20\}$. We always have $F_{\underline{\tau}}(\mathrm{UA}(n,S)) \leq n^{K/4}$, since the number of embeddings of τ in $\mathrm{UA}(n,S)$ is at most $n^{|\underline{\tau}|}$, and the product of the subtree sizes raised to appropriate powers is at most $n^{w(\underline{\tau})}$. By Lemma 4 in Appendix C there exists a constant $C(S)$ such that $\mathbb{P}(\mathrm{diam}(\mathrm{UA}(n,S)) > K\log(n)) \leq C(S)\, n^{-K/2}$. Therefore we have

$$\mathbb{E}\left[F_{\underline{\tau}}(\mathrm{UA}(n,S))^2 \mathbf{1}_{\{\mathrm{diam}(\mathrm{UA}(n,S)) > K\log(n)\}}\right] \leq C(S),$$

and similarly

$$\mathbb{E}\left[\left(F_{\underline{\tau}}(\mathrm{UA}(n+1,S)) - F_{\underline{\tau}}(\mathrm{UA}(n,S))\right)^2 \mathbf{1}_{\{\mathrm{diam}(\mathrm{UA}(n,S)) > K\log(n)\}}\right] \lesssim 1.$$

Therefore in the remainder of the proof we may, roughly speaking, assume that $\mathrm{diam}(\mathrm{UA}(n,S)) \leq K\log(n)$; this will be made precise later.

To simplify notation, write simply T_n instead of $\mathrm{UA}(n,S)$. Our proof is combinatorial and uses the notion of decorated embeddings as described in Section 3.2. We start with the proof of (a) which is simpler. We say that $\underline{\varphi} = \underline{\varphi}_1 \times \underline{\varphi}_2$ is a *decorated map* if it is a map such that both $\underline{\varphi}_1$ and $\underline{\varphi}_2$ are decorated embeddings from $\underline{\tau}$ to T_n. Note that $\underline{\varphi}$ is not necessarily injective. If $\underline{\varphi}$ is a decorated embedding or a decorated map, we denote by φ the map of the tree without the choices of vertices associated with the arrows.

Now observe that $F_{\underline{\tau}}(T_n)^2$ is exactly the number of decorated maps $\underline{\varphi} = \underline{\varphi}_1 \times \underline{\varphi}_2$. We partition the set of such decorated maps into two parts: let $\mathcal{E}^1_{\underline{\tau}}(T_n)$ denote the set of all such decorated maps where $\varphi_1(\tau) \cap \varphi_2(\tau) \neq \emptyset$, and let $\mathcal{E}^2_{\underline{\tau}}(T_n)$ denote the set of all such decorated maps where $\varphi_1(\tau) \cap \varphi_2(\tau) = \emptyset$. Clearly $F_{\underline{\tau}}(T_n)^2 = \left|\mathcal{E}^1_{\underline{\tau}}(T_n)\right| + \left|\mathcal{E}^2_{\underline{\tau}}(T_n)\right|$. This partition is not necessary for the proof, but it is helpful for exposition.

We first estimate $\left|\mathcal{E}^1_{\underline{\tau}}(T_n)\right|$. To do this, we associate to each decorated map $\underline{\varphi} \in \mathcal{E}^1_{\underline{\tau}}(T_n)$ a decorated tree $\underline{\sigma}$ and a decorated embedding $\underline{\psi}$ of it in T_n, in the following way; see also Figure 3 for an illustration. We take simply the union of the images of the decorated embeddings $\underline{\varphi}_1$ and $\underline{\varphi}_2$, and if these share any vertices, edges, or arrows, then we identify them (i.e., we take only a single copy). The resulting union is the image of a decorated tree $\underline{\sigma}$ under a decorated embedding $\underline{\psi}$; note that $\underline{\sigma}$ is uniquely defined, and $\underline{\psi}$ is uniquely defined up to the ordering of the arrows associated with $\underline{\sigma}$. To define $\underline{\psi}$ uniquely, we arbitrarily define the ordering of the arrows associated with $\underline{\sigma}$ to be the concatenation of the orderings associated with $\underline{\varphi}_1$ and $\underline{\varphi}_2$. Here we used the fact that $\underline{\varphi} \in \mathcal{E}^1_{\underline{\tau}}(T_n)$, since when $\varphi_1(\tau) \cap \varphi_2(\tau) = \emptyset$, the union of the two decorated embeddings cannot be the image of a single decorated tree under a decorated embedding.

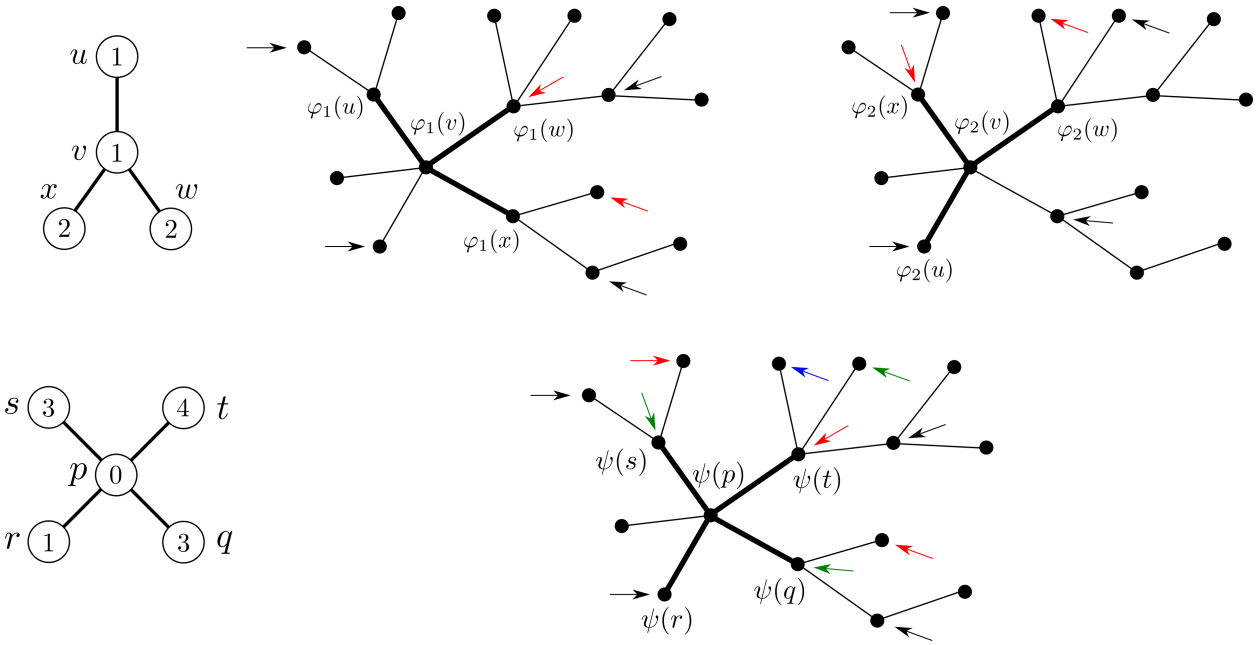


Figure 3: **A decorated map and an associated decorated embedding.** The top row depicts a decorated tree $\underline{\tau}$ and two decorated embeddings, $\underline{\varphi}_1$ and $\underline{\varphi}_2$, of it into a larger tree T. The bottom row depicts the associated decorated tree $\underline{\sigma}$, and the decorated embedding $\underline{\psi}$ of it into T.

Note that when taking the union of the decorated embeddings we do not introduce any new arrows, so we must have $w(\underline{\sigma}) \leq 2w(\underline{\tau})$. Note also that, due to the nonlocality of the decorations, $\underline{\sigma}$ might have vertices having a label being zero, see, e.g., Figure 3. However, importantly, the construction implies that all leaves of $\underline{\sigma}$ have positive labels, i.e., $\underline{\sigma} \in \mathcal{D} \setminus \mathcal{D}^*_0$.

Let $\mathcal{U}(\underline{\tau})$ denote the set of all decorated trees $\underline{\sigma}$ that can be obtained in this way. The cardinality of $\mathcal{U}(\underline{\tau})$ is bounded above by a constant depending only on $\underline{\tau}$, as we now argue. The number of ways that two copies of τ can be overlapped clearly depends only on τ. Once the union σ of the two copies of τ is fixed, only the arrows need to be associated with vertices of σ. There are at most $2w(\underline{\tau})$ arrows, and σ has at most $2|\underline{\tau}|$ vertices, so there are at most $(2|\underline{\tau}|)^{2w(\underline{\tau})}$ ways to associate arrows to vertices.

The function $\underline{\varphi} \mapsto (\underline{\sigma}, \underline{\psi})$ is not necessarily one-to-one. However, there exists a constant $c(\underline{\tau})$ depending only on $\underline{\tau}$ such that any pair $(\underline{\sigma}, \underline{\psi})$ is associated with at most $c(\underline{\tau})$ decorated maps $\underline{\varphi}$. To see this, note that given $(\underline{\sigma}, \underline{\psi})$, in order to recover $\underline{\varphi}$, it is sufficient to know the following:

(i) for every edge of $\psi(\sigma)$, whether it is a part of $\varphi_1(\tau)$, a part of $\varphi_2(\tau)$, or a part of both, (ii) for every arrow of $\underline{\psi}(\underline{\sigma})$, whether it is a part of $\underline{\varphi}_1(\underline{\tau})$, a part of $\underline{\varphi}_2(\underline{\tau})$, or a part of both, and (iii) the ordering of the arrows for $\underline{\varphi}_1$ and $\underline{\varphi}_2$. Since $|\underline{\sigma}| \leq 2|\underline{\tau}|$ and $w(\underline{\sigma}) \leq 2w(\underline{\tau})$, we can take $c(\underline{\tau}) = 3^{2|\underline{\tau}|+2w(\underline{\tau})}(w(\underline{\tau})!)^2$.

We have thus shown that

$$\left|\mathcal{E}^1_{\underline{\tau}}(T_n)\right| \leq c(\underline{\tau}) \sum_{\underline{\sigma} \in \mathcal{U}(\underline{\tau})} F_{\underline{\sigma}}(T_n).$$

For every $\underline{\sigma} \in \mathcal{U}(\underline{\tau})$ we have that $\underline{\sigma} \in \mathcal{D} \setminus \mathcal{D}^*_0$, $|\underline{\sigma}| \geq |\underline{\tau}| \geq 2$, and $w(\underline{\sigma}) \leq 2w(\underline{\tau})$, and so by Corollary 1 we have that $\mathbb{E}\left[F_{\underline{\sigma}}(T_n)\right] \lesssim n^{2w(\underline{\tau})}$. Since the cardinality of $\mathcal{U}(\underline{\tau})$ depends only on $\underline{\tau}$, this implies that $\mathbb{E}\left[\left|\mathcal{E}^1_{\underline{\tau}}(T_n)\right|\right] \lesssim n^{2w(\underline{\tau})}$.

Now we turn to estimating $\left|\mathcal{E}^2_{\underline{\tau}}(T_n)\right|$. Again, we associate to each decorated map $\underline{\varphi} \in \mathcal{E}^2_{\underline{\tau}}(T_n)$ a decorated tree $\underline{\sigma}$ and a decorated embedding $\underline{\psi}$ of it in T_n. This is done by first, just as before, taking the union of the images of the decorated embeddings $\underline{\varphi}_1$ and $\underline{\varphi}_2$, and if these share any arrows, identifying them (i.e., we take only a single copy). (Note that, since $\varphi_1(\tau) \cap \varphi_2(\tau) = \emptyset$, the decorated embeddings do not share any vertices or edges; however, due to the nonlocality of decorations they might share arrows.) We then take the union of this with the unique path in T_n that connects $\varphi_1(\tau)$ and $\varphi_2(\tau)$. The result of this is a tree in T_n, together with a set of at most $2w(\underline{\tau})$ arrows associated with vertices of T_n; this is thus the image of a decorated tree $\underline{\sigma}$ under a decorated embedding $\underline{\psi}$, and this is how we define $(\underline{\sigma}, \underline{\psi})$. See Figure 4 for an illustration.

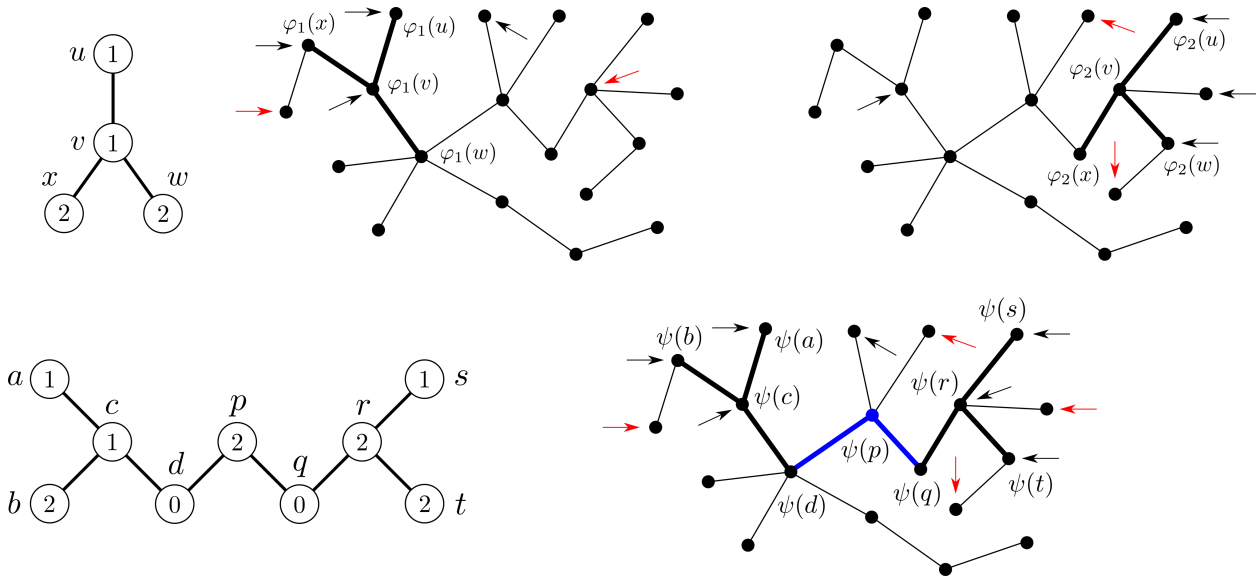


Figure 4: **Another decorated map and an associated decorated embedding.** The top row depicts a decorated tree $\underline{\tau}$ and two decorated embeddings, $\underline{\varphi}_1$ and $\underline{\varphi}_2$, of it into a larger tree T, where now $\varphi_1(\tau) \cap \varphi_2(\tau) = \emptyset$. The bottom row depicts the associated decorated tree $\underline{\sigma}$, and the decorated embedding $\underline{\psi}$ of it into T. The path connecting $\varphi_1(\tau)$ and $\varphi_2(\tau)$ is depicted in blue.

Again we have that any such $\underline{\sigma}$ must satisfy $w(\underline{\sigma}) \leq 2w(\underline{\tau})$ and $\underline{\sigma} \in \mathcal{D} \setminus \mathcal{D}^*_0$. The important difference now is that a priori we have no bound on $|\underline{\sigma}|$. This is where we use that $\operatorname{diam}(T_n) \leq K \log(n)$ with high probability.

Let $\mathcal{U}_2^{(n)}(\underline{\tau})$ denote the set of all decorated trees $\underline{\sigma}$ of diameter at most $K\log(n)$ that can be obtained in this way. The cardinality of $\mathcal{U}_2^{(n)}(\underline{\tau})$ cannot be bounded above by a constant depending only on $\underline{\tau}$, but it is at most polylogarithmic in n, as we now argue. There are at most $|\underline{\tau}|^2$ ways to choose which vertices of $\varphi_1(\tau)$ and $\varphi_2(\tau)$ are closest to each other, and the path connecting them has length at most $K\log(n)$. So the number of trees σ that can be obtained is at most $|\underline{\tau}|^2 K\log(n)$. Once the tree σ is fixed, only the arrows need to be associated with vertices of σ. There are at most $2w(\underline{\tau})$ arrows, and σ has at most $K\log(n)+2|\underline{\tau}|$ vertices, which shows that

$$\left|\mathcal{U}_2^{(n)}(\underline{\tau})\right| \leq |\underline{\tau}|^2 K\log(n)\left(K\log(n)+2|\underline{\tau}|\right)^{2w(\underline{\tau})} \lesssim 1.$$

The function $\underline{\varphi} \mapsto \left(\underline{\sigma}, \underline{\psi}\right)$ is not one-to-one. However, there exists a constant $c_2(\underline{\tau})$ depending only on $\underline{\tau}$ such that any pair $\left(\underline{\sigma}, \underline{\psi}\right)$ is associated with at most $c_2(\underline{\tau})$ decorated maps $\underline{\varphi}$, as we now show. First, given $\left(\underline{\sigma}, \underline{\psi}\right)$, we know that $\varphi_1(\tau)$ and $\varphi_2(\tau)$ are at the two "ends" of $\psi(\sigma)$. The two "ends" of $\psi(\sigma)$ are well-defined: an edge e of $\psi(\sigma)$ is part of the path connecting $\varphi_1(\tau)$ and $\varphi_2(\tau)$ (and hence not part of an "end") if and only if there are at least $|\underline{\tau}|$ vertices on both sides of the cut defined by e. In order to recover $\underline{\varphi}$, we also need to know for each arrow of $\underline{\psi}(\underline{\sigma})$, whether it is a part of $\underline{\varphi}_1(\underline{\tau})$, a part of $\underline{\varphi}_2(\underline{\tau})$, or a part of both. Finally, we need to know the ordering of the arrows for $\underline{\varphi}_1$ and $\underline{\varphi}_2$. Since $w(\underline{\sigma}) \leq 2w(\underline{\tau})$, we can thus take $c_2(\underline{\tau}) = 2\times 3^{2w(\underline{\tau})}\left(w(\underline{\tau})!\right)^2$.

We have thus shown that

$$\left|\mathcal{E}_{\underline{\tau}}^2(T_n)\right| \mathbf{1}_{\{\mathrm{diam}(T_n)\leq K\log(n)\}} \leq c_2(\underline{\tau}) \sum_{\underline{\sigma}\in\mathcal{U}_2^{(n)}(\underline{\tau})} F_{\underline{\sigma}}(T_n).$$

For every $\underline{\sigma} \in \mathcal{U}_2^{(n)}(\underline{\tau})$ we have that $\underline{\sigma} \in \mathcal{D}\setminus\mathcal{D}_0^*$, $|\underline{\sigma}| \geq |\underline{\tau}| \geq 2$, and $w(\underline{\sigma}) \leq 2w(\underline{\tau})$, and so by Corollary 1 we have that $\mathbb{E}\left[F_{\underline{\sigma}}(T_n)\right] \lesssim n^{2w(\underline{\tau})}$. Since we have $\left|\mathcal{U}_2^{(n)}(\underline{\tau})\right| \lesssim 1$, we thus have

$$\mathbb{E}\left[\left|\mathcal{E}_{\underline{\tau}}^2(T_n)\right| \mathbf{1}_{\{\mathrm{diam}(T_n)\leq K\log(n)\}}\right] \lesssim n^{2w(\underline{\tau})}.$$

This concludes the proof of (a).

For the proof of (b) we work conditionally on $\mathcal{F}_n$. As in the proof of Lemma 1, let u_{n+1} denote the vertex present in T_{n+1} but not in T_n, and let u_n denote its neighbor in T_{n+1}. Observe that $F_{\underline{\tau}}(T_{n+1}) - F_{\underline{\tau}}(T_n)$ is equal to the number of decorated embeddings of $\underline{\tau}$ in T_{n+1} that use the new vertex u_{n+1}. There are two ways that this may happen, and we call such decorated embeddings "type A" and "type B" accordingly (see Figure 5 for an illustration):

- **Type A.** The decorated embedding maps a vertex $v \in \tau$ to u_{n+1}. Since $\underline{\tau} \in \mathcal{D}_+$ and the arrows are mapped to different vertices, we must then have $\ell(v) = 1$, and the arrow pointing to v in $\underline{\tau}$ must be mapped to u_{n+1}.
- **Type B.** The decorated embedding maps τ in T_n, but there exists an arrow of $\underline{\tau}$ which it maps to u_{n+1}.

Consequently $\left(F_{\underline{\tau}}(T_{n+1}) - F_{\underline{\tau}}(T_n)\right)^2$ is equal to the number of decorated maps $\underline{\varphi} = \underline{\varphi}_1 \times \underline{\varphi}_2$ such that $\underline{\varphi}_1$ is either of type A or of type B, and the same holds for $\underline{\varphi}_2$. We denote by $\widetilde{\mathcal{E}}_{\underline{\tau}}^1(T_{n+1})$ the set of all such decorated maps where $\varphi_1(\tau) \cap \varphi_2(\tau) \neq \emptyset$, and let $\widetilde{\mathcal{E}}_{\underline{\tau}}^2(T_{n+1})$ denote the set

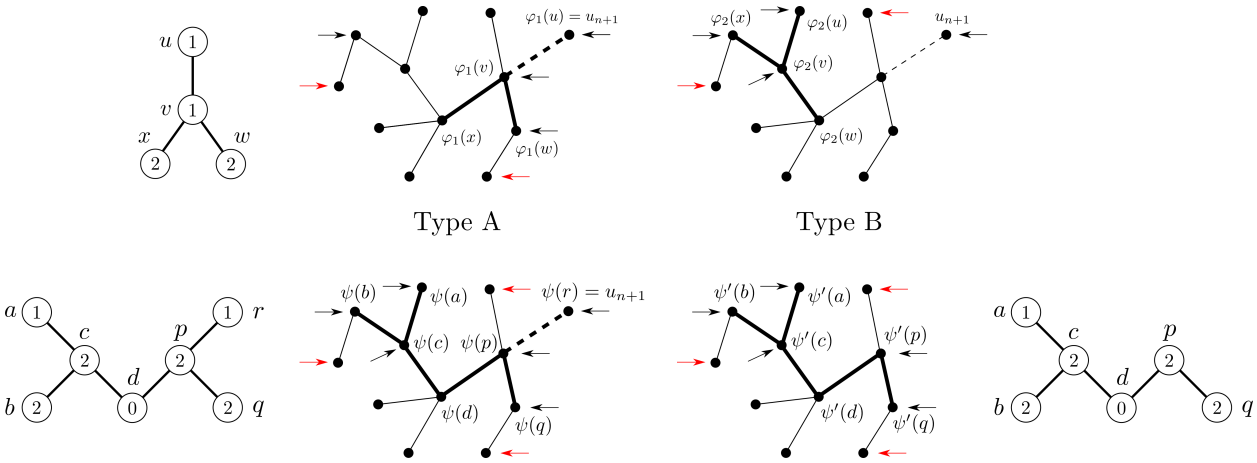


Figure 5: **Type A and type B decorated embeddings.** The top row depicts a decorated tree $\underline{\tau}$ and two decorated embeddings, $\underline{\varphi}_1$ and $\underline{\varphi}_2$, of it into a larger tree T_{n+1}. Here $\underline{\varphi}_1$ is of type A, and $\underline{\varphi}_2$ is of type B. In the bottom left is the associated decorated tree $\underline{\sigma}$, together with the decorated embedding $\underline{\psi}$ of it into T_{n+1}. In the bottom right is the pair $\left(\underline{\sigma}', \underline{\psi}'\right)$.

of all such decorated maps where $\varphi_1(\tau) \cap \varphi_2(\tau) = \emptyset$. Thus we have $\left(F_{\underline{\tau}}(T_{n+1}) - F_{\underline{\tau}}(T_n)\right)^2 = \left|\widetilde{\mathcal{E}}^1_{\underline{\tau}}(T_{n+1})\right| + \left|\widetilde{\mathcal{E}}^2_{\underline{\tau}}(T_{n+1})\right|$. Again, this partition is not necessary for the proof, but it helps the exposition.

We first estimate $\left|\widetilde{\mathcal{E}}^1_{\underline{\tau}}(T_{n+1})\right|$. In the same way as in part (a), we associate to each decorated map $\underline{\varphi} \in \widetilde{\mathcal{E}}^1_{\underline{\tau}}(T_{n+1})$ a pair $\left(\underline{\sigma}, \underline{\psi}\right)$. Note that both $\underline{\varphi}_1$ and $\underline{\varphi}_2$ map an arrow to u_{n+1}, so $w(\underline{\sigma}) \leq 2w(\underline{\tau}) - 1$, and also there exists an arrow $\underline{a}^* \in \underline{\sigma}$ that is mapped to u_{n+1}, denoted by $\underline{\psi}(\underline{a}^*) = u_{n+1}$. We again have $\underline{\sigma} \in \mathcal{D} \setminus \mathcal{D}_0^*$. As before, the set $\widetilde{\mathcal{U}}(\underline{\tau})$ of all decorated trees $\underline{\sigma}$ that can be obtained in this way has cardinality bounded above by a constant depending only on $\underline{\tau}$. Furthermore, there exists a constant $\widetilde{c}(\underline{\tau})$ depending only on $\underline{\tau}$ such that any pair $\left(\underline{\sigma}, \underline{\psi}\right)$ is associated with at most $\widetilde{c}(\underline{\tau})$ decorated maps $\underline{\varphi}$.

We partition $\widetilde{\mathcal{E}}^1_{\underline{\tau}}(T_{n+1})$ further into two parts. Let $\widetilde{\mathcal{E}}^{1,A}_{\underline{\tau}}(T_{n+1})$ denote the set of decorated maps $\underline{\varphi} \in \widetilde{\mathcal{E}}^1_{\underline{\tau}}(T_{n+1})$ such that at least one of $\underline{\varphi}_1$ and $\underline{\varphi}_2$ is of type A, and let $\widetilde{\mathcal{E}}^{1,B}_{\underline{\tau}}(T_{n+1}) := \widetilde{\mathcal{E}}^1_{\underline{\tau}}(T_{n+1}) \setminus \widetilde{\mathcal{E}}^{1,A}_{\underline{\tau}}(T_{n+1})$. That is, $\widetilde{\mathcal{E}}^{1,B}_{\underline{\tau}}(T_{n+1})$ consists of those decorated maps $\underline{\varphi} \in \widetilde{\mathcal{E}}^1_{\underline{\tau}}(T_{n+1})$ such that both $\underline{\varphi}_1$ and $\underline{\varphi}_2$ is of type B.

We first estimate $\left|\widetilde{\mathcal{E}}^{1,A}_{\underline{\tau}}(T_{n+1})\right|$. We associate to each $\underline{\varphi} \in \widetilde{\mathcal{E}}^{1,A}_{\underline{\tau}}(T_{n+1})$ a pair $\left(\underline{\sigma}, \underline{\psi}\right)$ as above. Let $v \in \sigma$ denote the vertex such that $\psi(v) = u_{n+1}$, and let $v' \in \sigma$ denote the vertex such that $\psi(v') = u_n$ (these vertices exist because $\underline{\varphi} \in \widetilde{\mathcal{E}}^{1,A}_{\underline{\tau}}(T_{n+1})$). Define the decorated tree $\underline{\sigma}'$ from $\underline{\sigma}$ by removing the vertex v from $\underline{\sigma}$, as well as the arrow $\underline{a}^*$ pointing to it. Define also the decorated embedding $\underline{\psi}' : \underline{\sigma}' \to T_n$ to be equal to $\underline{\psi}$ on $\underline{\sigma}'$, i.e., $\underline{\psi}' = \underline{\psi}|_{\underline{\sigma}'}$; see Figure 5 for an illustration. We have that $w(\underline{\sigma}') = w(\underline{\sigma}) - 1 \leq 2w(\underline{\tau}) - 2$, it can be checked that $\underline{\sigma}' \in \mathcal{D} \setminus \mathcal{D}_0^*$, and we also have $\psi'(v') = u_n$. Let $\widetilde{\mathcal{U}}'(\underline{\tau})$ denote the set of all decorated trees $\underline{\sigma}'$ that can be obtained in this way, and note that the cardinality of $\widetilde{\mathcal{U}}'(\underline{\tau})$ is bounded from above by a constant depending only on $\underline{\tau}$. Since the map $\left(\underline{\sigma}, \underline{\psi}\right) \mapsto \left(\underline{\sigma}', \underline{\psi}', v'\right)$ is one-to-one, we have obtained that

$$\left|\widetilde{\mathcal{E}}^{1,A}_{\underline{\tau}}(T_{n+1})\right| \leq \sum_{\underline{\sigma}' \in \widetilde{\mathcal{U}}'(\underline{\tau})} \sum_{v' \in \sigma'} \sum_{\underline{\psi}' : \underline{\sigma}' \to T_n} \widetilde{c}(\underline{\tau}) \mathbf{1}_{\{\psi'(v') = u_n\}}.$$

Since u_n is uniform, we obtain

$$\mathbb{E}\left[\left|\widetilde{\mathcal{E}}_{\underline{\tau}}^{1,A}\left(T_{n+1}\right)\right|\right]=\mathbb{E}\left[\mathbb{E}\left[\left|\widetilde{\mathcal{E}}_{\underline{\tau}}^{1,A}\left(T_{n+1}\right)\right|\,\middle|\,\mathcal{F}_n\right]\right]\leq\sum_{\underline{\sigma}'\in\widetilde{\mathcal{U}}'(\underline{\tau})}\sum_{v'\in\sigma'}\frac{\widetilde{c}\left(\underline{\tau}\right)}{n}\mathbb{E}\left[F_{\underline{\sigma}'}\left(T_n\right)\right]\lesssim n^{2w(\underline{\tau})-3},$$

where in the last inequality we used that for every $\underline{\sigma}'\in\widetilde{\mathcal{U}}'\left(\underline{\tau}\right)$ we have $w\left(\underline{\sigma}'\right)\leq 2w\left(\underline{\tau}\right)-2$ and $\underline{\sigma}'\in\mathcal{D}\setminus\mathcal{D}_0^*$, and so by Corollary 1 we have that $\mathbb{E}\left[F_{\underline{\sigma}'}\left(T_n\right)\right]\lesssim n^{2w(\underline{\tau})-2}$.

We now turn to estimating $\left|\widetilde{\mathcal{E}}_{\underline{\tau}}^{1,B}\left(T_{n+1}\right)\right|$. Let

$$v^*:=\operatorname*{argmin}_{v\in\sigma}\operatorname{dist}_{T_{n+1}}\left(\psi\left(v\right),u_{n+1}\right)=\operatorname*{argmin}_{v\in\sigma}\operatorname{dist}_{T_n}\left(\psi\left(v\right),u_n\right),$$

where dist_G denotes graph distance in a graph G. Note that the arrow $\underline{a}^*\in\underline{\sigma}$ is associated with v^* in $\underline{\sigma}$. Define the decorated map $\underline{\sigma}^*$ from $\underline{\sigma}$ by removing the arrow $\underline{a}^*$ from $\underline{\sigma}$. We have that $w\left(\underline{\sigma}^*\right)=w\left(\underline{\sigma}\right)-1\leq 2w\left(\underline{\tau}\right)-2$. Furthermore, either $\underline{\sigma}^*\in\mathcal{D}\setminus\mathcal{D}_0^*$ or v^* is the only leaf of $\underline{\sigma}$ that has label zero. Let $\widetilde{\mathcal{U}}^*\left(\underline{\tau}\right)$ denote the set of all decorated trees $\underline{\sigma}^*$ that can be obtained in this way, and note that the cardinality of $\widetilde{\mathcal{U}}^*\left(\underline{\tau}\right)$ is bounded from above by a constant depending only on $\underline{\tau}$. Define also the decorated embedding $\underline{\psi}^*:\underline{\sigma}^*\to T_n$ to be equal to $\underline{\psi}$ on $\underline{\sigma}^*$, i.e., $\underline{\psi}^*=\underline{\psi}|_{\underline{\sigma}^*}$. Define furthermore z^* to be the neighbor of $\underline{\psi}\left(\underline{a}^*\right)$ in T_{n+1}; we thus have $z^*=u_n$. Due to the ordering of the arrows, the map $\left(\underline{\sigma},\underline{\psi}\right)\mapsto\left(\underline{\sigma}^*,\underline{\psi}^*,z^*\right)$ is not necessarily one-to-one, but any triple $\left(\underline{\sigma}^*,\underline{\psi}^*,z^*\right)$ is associated with at most $w\left(\underline{\tau}\right)$ pairs $\left(\underline{\sigma},\underline{\psi}\right)$. Thus, defining $\widetilde{c}'\left(\underline{\tau}\right):=\widetilde{c}\left(\underline{\tau}\right)w\left(\underline{\tau}\right)$, we have that

$$\begin{aligned}
&\left|\widetilde{\mathcal{E}}_{\underline{\tau}}^{1,B}\left(T_{n+1}\right)\right|\\
&\quad\leq\sum_{\underline{\sigma}^*\in\widetilde{\mathcal{U}}^*(\underline{\tau})}\sum_{\underline{\psi}^*:\underline{\sigma}^*\to T_n}\sum_{z^*\in T_n}\widetilde{c}'\left(\underline{\tau}\right)\mathbf{1}_{\{z^*=u_n\}}\mathbf{1}_{\left\{\underline{\sigma}^*\in\mathcal{D}\setminus\mathcal{D}_0^*\right\}\cup\left\{\underline{\sigma}^*\in\mathcal{D}_0^*,\operatorname{argmin}_{v\in\sigma^*}\operatorname{dist}_{T_n}(\psi(v),z^*)\in L_0(\underline{\sigma}^*)\right\}}\\
&\quad\leq\widetilde{c}'\left(\underline{\tau}\right)\sum_{\underline{\sigma}^*\in\widetilde{\mathcal{U}}^*(\underline{\tau})}\sum_{\underline{\psi}^*:\underline{\sigma}^*\to T_n}\mathbf{1}_{\left\{\underline{\sigma}^*\in\mathcal{D}\setminus\mathcal{D}_0^*\right\}\cup\left\{\underline{\sigma}^*\in\mathcal{D}_0^*,\operatorname{argmin}_{v\in\sigma^*}\operatorname{dist}_{T_n}(\psi(v),u_n)\in L_0(\underline{\sigma}^*)\right\}}.
\end{aligned}$$

Now if $\underline{\sigma}^*\in\widetilde{\mathcal{U}}^*\left(\underline{\tau}\right)\cap\left(\mathcal{D}\setminus\mathcal{D}_0^*\right)$, then the sum over embeddings $\underline{\psi}^*:\underline{\sigma}^*\to T_n$ becomes $F_{\underline{\sigma}^*}\left(T_n\right)$, and by Corollary 1 we have that $\mathbb{E}\left[F_{\underline{\sigma}^*}\left(T_n\right)\right]\lesssim n^{2w(\underline{\tau})-2}$. If $\underline{\sigma}^*\in\widetilde{\mathcal{U}}^*\left(\underline{\tau}\right)\cap\mathcal{D}_0^*$, then, as mentioned above, $L_0\left(\underline{\sigma}^*\right)=\{v^*\}$, and we have

$$\mathbb{P}\left(\operatorname*{argmin}_{v\in\sigma^*}\operatorname{dist}_{T_n}\left(\psi\left(v\right),u_n\right)=v^*\,\middle|\,\mathcal{F}_n\right)=\frac{f_{\psi(v^*)}\left(T_n\right)}{n}.$$

So by summing over $\underline{\psi}^*:\underline{\sigma}^*\to T_n$, if $\underline{\sigma}^*\in\widetilde{\mathcal{U}}^*\left(\underline{\tau}\right)\cap\mathcal{D}_0^*$, then

$$\mathbb{E}\left[\sum_{\underline{\psi}^*:\underline{\sigma}^*\to T_n}\mathbf{1}_{\left\{\underline{\sigma}^*\in\mathcal{D}_0^*,\operatorname{argmin}_{v\in\sigma^*}\operatorname{dist}_{T_n}(\psi(v),u_n)\in L_0(\underline{\sigma}^*)\right\}}\,\middle|\,\mathcal{F}_n\right]=\frac{1}{n}F_{\underline{\sigma}}\left(T_n\right).$$

Since $\underline{\sigma}\in\mathcal{D}\setminus\mathcal{D}_0^*$ and $w\left(\underline{\sigma}\right)\leq 2w\left(\underline{\tau}\right)-1$, by Corollary 1 we have that $\mathbb{E}\left[F_{\underline{\sigma}}\left(T_n\right)\right]\lesssim n^{2w(\underline{\tau})-1}$ and thus $\mathbb{E}\left[n^{-1}F_{\underline{\sigma}}\left(T_n\right)\right]\lesssim n^{2w(\underline{\tau})-2}$. Putting everything together we thus obtain that

$$\mathbb{E}\left[\left|\widetilde{\mathcal{E}}_{\underline{\tau}}^{1,B}\left(T_{n+1}\right)\right|\right]\lesssim n^{2w(\underline{\tau})-2}.$$

To estimate $\left|\widetilde{\mathcal{E}}^2_{\underline{\tau}}(T_{n+1})\right|$ we can do the same thing as in part (a), and we obtain the same bound as for $\left|\widetilde{\mathcal{E}}^1_{\underline{\tau}}(T_{n+1})\right|$ up to polylogarithmic factors in n. We omit the details. This concludes the proof of part (b). ■

3.5 Constructing the martingales

We now construct the martingales of Proposition 1 with the help of the recurrence relation of Lemma 1. In order to show that these martingales are bounded in L^2, we use the moment estimates of Section 3.4.

Proof of Proposition 1 Fix a seed tree S with $|S| = n_0 \geq 2$. For a decorated tree $\underline{\tau} \in \mathcal{D}_+$ and $n \geq 2$, define

$$\beta_n(\underline{\tau}) := \prod_{j=2}^{n-1} \left(1 + \frac{w(\underline{\tau})}{j}\right)^{-1}, \quad \text{when } |\underline{\tau}| \geq 2; \quad \text{and} \quad \beta_n(\underline{\tau}) := \left(n \times [n]_{w(\underline{\tau})}\right)^{-1}, \quad \text{when } |\underline{\tau}| = 1.$$

Note that when $|\underline{\tau}| \geq 2$, we have $n^{-w(\underline{\tau})} \lesssim \beta_n(\underline{\tau}) \lesssim n^{-w(\underline{\tau})}$.

We now construct, by induction on the order $\preccurlyeq$ on decorated trees, coefficients

$$\left\{a_n\left(\underline{\tau}, \underline{\tau}'\right) : \underline{\tau}, \underline{\tau}' \in \mathcal{D}_+, \underline{\tau}' \prec \underline{\tau}, n \geq n_0\right\}$$

such that

$$a_n\left(\underline{\tau}, \underline{\tau}'\right) \lesssim 1, \qquad \Delta_n a\left(\underline{\tau}, \underline{\tau}'\right) \lesssim 1/n, \tag{17}$$

and

$$M^{(S)}_{\underline{\tau}}(n) = \beta_n(\underline{\tau}) \left(F_{\underline{\tau}}(\mathrm{UA}(n, S)) - \sum_{\underline{\tau}' \in \mathcal{D}_+ : \underline{\tau}' \prec \underline{\tau}} a_n\left(\underline{\tau}, \underline{\tau}'\right) F_{\underline{\tau}'}(\mathrm{UA}(n, S))\right) \tag{18}$$

is a martingale. Importantly, we shall see that the coefficients $a_n(\underline{\tau}, \underline{\tau}')$ do not depend on S. To simplify notation, in the following we omit dependence on S and write $M_{\underline{\tau}}(n)$ for $M^{(S)}_{\underline{\tau}}(n)$. Also, as before, we write T_n for $\mathrm{UA}(n, S)$.

First, when $|\underline{\tau}| = 1$, we have $M_{\underline{\tau}}(n) = \beta_n(\underline{\tau}) F_{\underline{\tau}}(T_n) = 1$, which is a martingale. Now fix $\underline{\tau} \in \mathcal{D}_+$ with $|\underline{\tau}| \geq 2$. Assume that the coefficients $a_n(\underline{\sigma}, \underline{\sigma}')$ have been constructed for every $\underline{\sigma}, \underline{\sigma}' \in \mathcal{D}_+$ such that $\underline{\sigma}' \prec \underline{\sigma} \prec \underline{\tau}$ and every $n \geq n_0$, and that they have the desired properties. We first claim that there exist constants $\{b_n(\underline{\sigma}, \underline{\sigma}') : \underline{\sigma}' \prec \underline{\sigma} \prec \underline{\tau}, n \geq n_0\}$ such that $b_n(\underline{\sigma}, \underline{\sigma}') \lesssim 1$ and

$$F_{\underline{\sigma}}(T_n) = \frac{1}{\beta_n(\underline{\sigma})} M_{\underline{\sigma}}(n) + \sum_{\underline{\sigma}' \in \mathcal{D}_+ : \underline{\sigma}' \prec \underline{\sigma}} \frac{b_n(\underline{\sigma}, \underline{\sigma}')}{\beta_n(\underline{\sigma}')} M_{\underline{\sigma}'}(n) \tag{19}$$

for $n \geq n_0$. To see this, define the matrix $A_n = (A_n(\underline{\sigma}, \underline{\sigma}'))_{\underline{\sigma}, \underline{\sigma}' \prec \underline{\tau}}$ by $A_n(\underline{\sigma}, \underline{\sigma}') = -a_n(\underline{\sigma}, \underline{\sigma}')$ if $\underline{\sigma}' \prec \underline{\sigma}$, $A_n(\underline{\sigma}, \underline{\sigma}') = 1$ if $\underline{\sigma} = \underline{\sigma}'$, and $A_n(\underline{\sigma}, \underline{\sigma}') = 0$ otherwise. Then, using (18), we have for every $n \geq n_0$ the following equality of vectors indexed by $\underline{\sigma} \in \mathcal{D}_+$ such that $\underline{\sigma} \prec \underline{\tau}$:

$$\left(\frac{1}{\beta_n(\underline{\sigma})} M_{\sigma}(n)\right)_{\underline{\sigma} \prec \underline{\tau}} = A_n \cdot \left(F_{\underline{\sigma}}(T_n)\right)_{\underline{\sigma} \prec \underline{\tau}}. \tag{20}$$

We can write $\{\underline{\sigma} \in \mathcal{D}_+ : \underline{\sigma} \prec \underline{\tau}\} = \{\underline{\sigma}_1, \ldots, \underline{\sigma}_K\}$ in such a way that $\underline{\sigma}_i \prec \underline{\sigma}_j$ implies $i < j$. With this convention, A_n is a lower triangular matrix with all diagonal entries equal to 1 and all entries

satisfying $A_n(\underline{\sigma},\underline{\sigma}') \lesssim 1$. Therefore A_n is invertible, and its inverse also satisfies these properties. That is, if we write $A_n^{-1} = (b_n(\underline{\sigma},\underline{\sigma}'))_{\underline{\sigma},\underline{\sigma}' \prec \underline{\tau}}$, then A_n^{-1} is a lower triangular matrix that satisfies $b_n(\underline{\sigma},\underline{\sigma}) = 1$ for all $\underline{\sigma} \prec \underline{\tau}$, and $b_n(\underline{\sigma},\underline{\sigma}') \lesssim 1$ for all $\underline{\sigma},\underline{\sigma}' \prec \underline{\tau}$. So (19) follows directly from (20).

Note that we can write equation (13) of Lemma 1 more compactly as follows:

$$\mathbb{E}\left[F_{\underline{\tau}}(\mathrm{UA}(n+1,S)) \,\middle|\, \mathcal{F}_n\right] = \left(1 + \frac{w(\underline{\tau})}{n}\right) F_{\underline{\tau}}(\mathrm{UA}(n,S)) + \frac{1}{n} \sum_{\underline{\tau}' \in \mathcal{D}: \underline{\tau}' \prec \underline{\tau}} c\left(\underline{\tau},\underline{\tau}'\right) F_{\underline{\tau}'}(\mathrm{UA}(n,S)), \tag{21}$$

for appropriately defined constants $\{c(\underline{\tau},\underline{\tau}') : \underline{\tau},\underline{\tau}' \in \mathcal{D}, \underline{\tau}' \prec \underline{\tau}\}$, and note that since $\underline{\tau} \in \mathcal{D}_+$, we have $c(\underline{\tau},\underline{\tau}') = 0$ if $\underline{\tau}' \notin \mathcal{D}_+$. Therefore, using (21) and (19), together with the identities $\beta_{n+1}(\underline{\tau})(1 + w(\underline{\tau})/n) = \beta_n(\underline{\tau})$ and $\beta_{n+1}(\underline{\tau}) n^{-1} = \beta_n(\underline{\tau})(n + w(\underline{\tau}))^{-1}$, we have for $n \geq n_0$ that

$$\mathbb{E}\left[\beta_{n+1}(\underline{\tau}) F_{\underline{\tau}}(T_{n+1}) \,\middle|\, \mathcal{F}_n\right] = \beta_n(\underline{\tau}) F_{\underline{\tau}}(T_n) + \frac{\beta_n(\underline{\tau})}{n + w(\underline{\tau})} \sum_{\underline{\tau}' \in \mathcal{D}_+: \underline{\tau}' \prec \underline{\tau}} c\left(\underline{\tau},\underline{\tau}'\right) F_{\underline{\tau}'}(T_n)$$

$$= \beta_n(\underline{\tau}) F_{\underline{\tau}}(T_n) + \frac{1}{n + w(\underline{\tau})} \sum_{\underline{\sigma} \in \mathcal{D}_+: \underline{\sigma} \prec \underline{\tau}} \left(c(\underline{\tau},\underline{\sigma}) + \sum_{\underline{\tau}' \in \mathcal{D}_+: \underline{\sigma} \prec \underline{\tau}' \prec \underline{\tau}} c\left(\underline{\tau},\underline{\tau}'\right) b_n\left(\underline{\tau}',\underline{\sigma}\right) \right) \frac{\beta_n(\underline{\tau})}{\beta_n(\underline{\sigma})} M_{\underline{\sigma}}(n).$$

For $n \geq n_0$ define

$$\overline{a}_n(\underline{\tau},\underline{\sigma}) = \sum_{j=n_0}^{n-1} \frac{1}{j + w(\underline{\tau})} \left(c(\underline{\tau},\underline{\sigma}) + \sum_{\underline{\tau}' \in \mathcal{D}_+: \underline{\sigma} \prec \underline{\tau}' \prec \underline{\tau}} c\left(\underline{\tau},\underline{\tau}'\right) b_j\left(\underline{\tau}',\underline{\sigma}\right) \right) \frac{\beta_j(\underline{\tau})}{\beta_j(\underline{\sigma})}.$$

We thus have

$$\mathbb{E}\left[\beta_{n+1}(\underline{\tau}) F_{\underline{\tau}}(T_{n+1}) \,\middle|\, \mathcal{F}_n\right] = \beta_n(\underline{\tau}) F_{\underline{\tau}}(T_n) + \sum_{\underline{\sigma} \in \mathcal{D}_+: \underline{\sigma} \prec \underline{\tau}} \left(\overline{a}_{n+1}(\underline{\tau},\underline{\sigma}) - \overline{a}_n(\underline{\tau},\underline{\sigma})\right) M_{\underline{\sigma}}(n).$$

By our induction hypothesis, $\left\{M_{\underline{\sigma}}(n)\right\}_{n \geq n_0}$ is an $(\mathcal{F}_n)$-martingale for every $\underline{\sigma} \prec \underline{\tau}$, and consequently $\beta_n(\underline{\tau}) F_{\underline{\tau}}(T_n) - \sum_{\underline{\sigma} \in \mathcal{D}_+: \underline{\sigma} \prec \underline{\tau}} \overline{a}_n(\underline{\tau},\underline{\sigma}) M_{\underline{\sigma}}(n)$ is also an $(\mathcal{F}_n)$-martingale. By (18) we have

$$\beta_n(\underline{\tau}) F_{\underline{\tau}}(T_n) - \sum_{\underline{\sigma} \in \mathcal{D}_+: \underline{\sigma} \prec \underline{\tau}} \overline{a}_n(\underline{\tau},\underline{\sigma}) M_{\underline{\sigma}}(n)$$

$$= \beta_n(\underline{\tau}) F_{\underline{\tau}}(T_n) - \sum_{\underline{\sigma} \in \mathcal{D}_+: \underline{\sigma} \prec \underline{\tau}} \overline{a}_n(\underline{\tau},\underline{\sigma}) \beta_n(\underline{\sigma}) \left(F_{\underline{\sigma}}(T_n) - \sum_{\underline{\tau}' \in \mathcal{D}_+: \underline{\tau}' \prec \underline{\sigma}} a_n\left(\underline{\sigma},\underline{\tau}'\right) F_{\underline{\tau}'}(T_n) \right)$$

$$= \beta_n(\underline{\tau}) \left(F_{\underline{\tau}}(T_n) - \sum_{\underline{\sigma} \in \mathcal{D}_+: \underline{\sigma} \prec \underline{\tau}} \left[\overline{a}_n(\underline{\tau},\underline{\sigma}) \frac{\beta_n(\underline{\sigma})}{\beta_n(\underline{\tau})} - \sum_{\underline{\tau}' \in \mathcal{D}_+: \underline{\sigma} \prec \underline{\tau}' \prec \underline{\tau}} \overline{a}_n\left(\underline{\tau},\underline{\tau}'\right) a_n\left(\underline{\tau}',\underline{\sigma}\right) \frac{\beta_n(\underline{\tau}')}{\beta_n(\underline{\tau})} \right] F_{\underline{\sigma}}(T_n) \right).$$

So if we set

$$a_n(\underline{\tau},\underline{\sigma}) := \overline{a}_n(\underline{\tau},\underline{\sigma}) \frac{\beta_n(\underline{\sigma})}{\beta_n(\underline{\tau})} - \sum_{\underline{\tau}' \in \mathcal{D}_+: \underline{\sigma} \prec \underline{\tau}' \prec \underline{\tau}} \overline{a}_n\left(\underline{\tau},\underline{\tau}'\right) a_n\left(\underline{\tau}',\underline{\sigma}\right) \frac{\beta_n(\underline{\tau}')}{\beta_n(\underline{\tau})},$$

then it is clear that $\left\{M_{\underline{\tau}}(n)\right\}_{n \geq n_0}$ defined as in (18) is a martingale.

Now let us establish that the coefficients are of the correct order, i.e., let us show (17). First note that $(n+w(\underline{\tau}))^{-1} \lesssim 1/n$, and that when $|\underline{\tau}| \geq 2$, $\beta_n(\underline{\tau})\, n^{w(\underline{\tau})}$ has a positive and finite limit as $n \to \infty$. Therefore a simple computation shows that for $\underline{\sigma}, \underline{\sigma}' \in \mathcal{D}_+$ with $|\underline{\sigma}|, |\underline{\sigma}'| \geq 2$, we have

$$\frac{\beta_n(\underline{\sigma})}{\beta_n(\underline{\sigma}')} \lesssim n^{w(\underline{\sigma}')-w(\underline{\sigma})} \qquad \text{and} \qquad \Delta_n \frac{\beta(\underline{\sigma})}{\beta(\underline{\sigma}')} \lesssim n^{w(\underline{\sigma}')-w(\underline{\sigma})-1}.$$

Furthermore, by the induction hypothesis we have that $b_n(\underline{\sigma}, \underline{\sigma}') \lesssim 1$ for every $\underline{\sigma}, \underline{\sigma}' \prec \underline{\tau}$. From the definition of $\overline{a}_n(\underline{\tau}, \underline{\sigma})$ we then immediately get that $\Delta_n \overline{a}(\underline{\tau}, \underline{\sigma}) \lesssim n^{w(\underline{\sigma})-w(\underline{\tau})-1}$, and consequently also $\overline{a}_n(\underline{\tau}, \underline{\sigma}) \lesssim n^{w(\underline{\sigma})-w(\underline{\tau})}$, for every $\underline{\sigma} \in \mathcal{D}_+$ such that $\underline{\sigma} \prec \underline{\tau}$ and $|\underline{\sigma}| \geq 2$. So for every $\underline{\sigma} \in \mathcal{D}_+$ such that $\underline{\sigma} \prec \underline{\tau}$ and $|\underline{\sigma}| \geq 2$ we have that

$$\overline{a}_n(\underline{\tau}, \underline{\sigma}) \frac{\beta_n(\underline{\sigma})}{\beta_n(\underline{\tau})} \lesssim 1 \qquad \text{and} \qquad \Delta_n \left(\overline{a}(\underline{\tau}, \underline{\sigma}) \frac{\beta(\underline{\sigma})}{\beta(\underline{\tau})} \right) \lesssim \frac{1}{n}. \tag{22}$$

One can easily check that (22) holds also when $|\underline{\sigma}| = 1$. Now combining all of these estimates with the definition of $a_n(\underline{\tau}, \underline{\sigma})$, we get that (17) holds. This completes the induction.

Finally, what remains to show is that the martingales $M_{\underline{\tau}}$ are bounded in L^2. Since $M_{\underline{\tau}}$ is a martingale, its increments are orthogonal in L^2, and so

$$\mathbb{E}\left[M_{\underline{\tau}}(n)^2\right] = \sum_{j=n_0}^{n-1} \mathbb{E}\left[\left(M_{\underline{\tau}}(j+1) - M_{\underline{\tau}}(j)\right)^2\right] + \mathbb{E}\left[M_{\underline{\tau}}(n_0)^2\right].$$

Clearly $\mathbb{E}\left[M_{\underline{\tau}}(n_0)^2\right] < \infty$ and so it suffices to show that

$$\sum_{n=n_0}^{\infty} \mathbb{E}\left[\left(M_{\underline{\tau}}(n+1) - M_{\underline{\tau}}(n)\right)^2\right] < \infty.$$

Recalling the definition of $M_{\underline{\tau}}$ from (18) we have

$$\mathbb{E}\left[\left(\Delta_n\left(M_{\underline{\tau}}\right)\right)^2\right] = \mathbb{E}\left[\left(\Delta_n\left(\beta_{\cdot}(\underline{\tau})\, F_{\underline{\tau}}(T_{\cdot})\right) - \sum_{\underline{\tau} \in \mathcal{D}_+ : \underline{\tau}' \prec \underline{\tau}} \Delta_n\left(\beta_{\cdot}(\underline{\tau})\, a_{\cdot}(\underline{\tau}, \underline{\tau}')\, F_{\underline{\tau}'}(T_{\cdot})\right)\right)^2\right],$$

where the dots in the subscripts denote dependence on n, on which the difference operator Δ_n acts. By the Cauchy-Schwarz inequality, there exists a positive and finite constant c that depends only on $\underline{\tau}$ such that for every $n \geq n_0$, the quantity $c \times \mathbb{E}\left[\left(\Delta_n\left(M_{\underline{\tau}}\right)\right)^2\right]$ is bounded from above by

$$\mathbb{E}\left[\left(\Delta_n\left(\beta_{\cdot}(\underline{\tau})\, F_{\underline{\tau}}(T_{\cdot})\right)\right)^2\right] + \sum_{\underline{\tau} \in \mathcal{D}_+ : \underline{\tau}' \prec \underline{\tau}} \mathbb{E}\left[\left(\Delta_n\left(\beta_{\cdot}(\underline{\tau})\, a_{\cdot}(\underline{\tau}, \underline{\tau}')\, F_{\underline{\tau}'}(T_{\cdot})\right)\right)^2\right] \tag{23}$$

Since $\Delta_n\left(\beta_{\cdot}(\underline{\tau})\, F_{\underline{\tau}}(T_{\cdot})\right) = \beta_{n+1}(\underline{\tau})\, \Delta_n\left(F_{\underline{\tau}}(T_{\cdot})\right) + \left(\Delta_n\left(\beta_{\cdot}(\underline{\tau})\right)\right) F_{\underline{\tau}}(T_n)$, we have that

$$\mathbb{E}\left[\left(\Delta_n\left(\beta_{\cdot}(\underline{\tau})\, F_{\underline{\tau}}(T_{\cdot})\right)\right)^2\right] \leq 2\left(\beta_{n+1}(\underline{\tau})\right)^2 \mathbb{E}\left[\left(\Delta_n\left(F_{\underline{\tau}}(T_{\cdot})\right)\right)^2\right] + 2\left(\Delta_n\left(\beta_{\cdot}(\underline{\tau})\right)\right)^2 \mathbb{E}\left[F_{\underline{\tau}}(T_n)^2\right].$$

We have seen that $\left(\beta_{n+1}(\underline{\tau})\right)^2 \lesssim n^{-2w(\underline{\tau})}$ and $\left(\Delta_n\left(\beta_{\cdot}(\underline{\tau})\right)\right)^2 \lesssim n^{-2w(\underline{\tau})-2}$, and by Lemma 2 we have that $\mathbb{E}\left[F_{\underline{\tau}}(T_n)^2\right] \lesssim n^{2w(\underline{\tau})}$ and $\mathbb{E}\left[\left(\Delta_n\left(F_{\underline{\tau}}(T_{\cdot})\right)\right)^2\right] \lesssim n^{2w(\underline{\tau})-2}$. Putting these together we thus have

that $\mathbb{E}\left[\left(\Delta_n\left(\beta_{\cdot}\left(\underline{\tau}\right)F_{\underline{\tau}}\left(T_{\cdot}\right)\right)\right)^2\right] \lesssim n^{-2}$. For the other terms in (23) we similarly have

$$\begin{aligned}&\mathbb{E}\left[\left(\Delta_n\left(\beta_{\cdot}\left(\underline{\tau}\right)a_{\cdot}\left(\underline{\tau},\underline{\tau}'\right)F_{\underline{\tau}'}\left(T_{\cdot}\right)\right)\right)^2\right]\\ &\quad\leq 2\left(a_{n+1}\left(\underline{\tau},\underline{\tau}'\right)\right)^2\mathbb{E}\left[\left(\Delta_n\left(\beta_{\cdot}\left(\underline{\tau}\right)F_{\underline{\tau}'}\left(T_{\cdot}\right)\right)\right)^2\right]+2\left(\Delta_n\left(a_{\cdot}\left(\underline{\tau},\underline{\tau}'\right)\right)\right)^2\mathbb{E}\left[\left(\beta_n\left(\underline{\tau}\right)F_{\underline{\tau}'}\left(T_n\right)\right)^2\right].\end{aligned}$$

We have seen that $\left(a_{n+1}\left(\underline{\tau},\underline{\tau}'\right)\right)^2 \lesssim 1$ and $\left(\Delta_n\left(a_{\cdot}\left(\underline{\tau},\underline{\tau}'\right)\right)\right)^2 \lesssim n^{-2}$. Furthermore, by Lemma 2 we have that $\mathbb{E}\left[\left(\beta_n\left(\underline{\tau}\right)F_{\underline{\tau}'}\left(T_n\right)\right)^2\right] \lesssim n^{2w(\underline{\tau}')-2w(\underline{\tau})} \leq 1$, and similarly to the computation above we have that $\mathbb{E}\left[\left(\Delta_n\left(\beta_{\cdot}\left(\underline{\tau}\right)F_{\underline{\tau}'}\left(T_{\cdot}\right)\right)\right)^2\right] \lesssim n^{2w(\underline{\tau}')-2w(\underline{\tau})-2} \leq n^{-2}$. Putting everything together we get that

$$\mathbb{E}\left[\left(M_{\underline{\tau}}\left(n+1\right)-M_{\underline{\tau}}\left(n\right)\right)^2\right] \lesssim n^{-2},$$

which is summable, so M_τ is indeed bounded in L^2. ■

4 Discussion

We conclude with a comparison of our proof to that of Curien et al. [2014], and with a list of open problems.

4.1 Comparison to Curien et al. [2014]

As discussed in Section 1.1, the key difference in our proof for uniform attachment compared to the proof of Curien et al. [2014] for preferential attachment is the underlying family of statistics. For preferential attachment these are based on the degrees of the nodes, whereas for uniform attachment they are based on partition sizes when embedding a given tree, i.e., they are based on global balancedness properties of the tree.

The statistics $F_{\underline{\tau}}\left(T\right)$ are defined in this specific way in order to make the analysis simpler. In particular, it is useful that $F_{\underline{\tau}}\left(T\right)$ has a combinatorial interpretation as the number of decorated embeddings of $\underline{\tau}$ in T, similarly to the statistics of Curien et al. [2014]. However, the notion of a decorated embedding is different in the two settings. In Curien et al. [2014], arrows associated with the decorated tree $\underline{\tau}$ are mapped by $\underline{\varphi}$ to corners around the vertices of $\varphi\left(\tau\right)$, or in other words, the decorations are local. In contrast, in the notion of a decorated embedding as defined in this paper, arrows associated with a decorated tree $\underline{\tau}$ can be mapped to any vertex in the graph T, or in other words, the decorations are nonlocal/global.

While the general structure of our proof is identical to that of Curien et al. [2014], this local vs. global difference in the underlying statistics manifests itself in the details. In particular, the main challenge is the second moment estimate provided in Lemma 2. Here, we associate to each decorated map $\underline{\varphi} = \underline{\varphi}_1 \times \underline{\varphi}_2$ a decorated tree $\underline{\sigma}$ and a decorated embedding $\underline{\psi}$ of it in $\mathrm{UA}\left(n,S\right)$. In the case of preferential attachment, the decorated tree $\underline{\sigma}$ necessarily has all labels be positive, due to the decorations being local. However, in the case of uniform attachment, it might happen that a vertex of $\underline{\sigma}$ has a label being zero, due to the global nature of decorations. This is the reason why we need to deal with decorated trees having labels being zero, in contrast with the preferential attachment model, where it suffices to consider decorated trees with positive labels. The recurrence relation and the subsequent moment estimates show that there is a subtlety in

dealing with decorated trees having labels being zero, as it matters whether the vertices with label zero are leaves or not.

Finally, in our proof of the second moment estimate we also use the fact that the diameter of $\mathrm{UA}(n, S)$ is on the order of $\log(n)$ with high probability (see Lemma 4). This is again due to the global nature of decorations, and such an estimate is not necessary in the case of preferential attachment.

4.2 Open problems

Bubeck et al. [2014] list several open problems for the preferential attachment model, and these questions can also be asked for the uniform attachment model. We list here a few of them.

1. What can be said about general uniform attachment graphs, where multiple edges are added at each step?

2. Under what conditions on two tree sequences (T_k), (R_k) do we have $\lim_{k\to\infty} \delta(T_k, R_k) = 1$?

3. Is it possible to give a combinatorial description of the metric δ?

4. A simple model that interpolates between uniform attachment and (linear) preferential attachment is to consider probabilities of connection proportional to the degree of the vertex raised to some power α. The results of this paper show that for $\alpha = 0$ different seeds are distinguishable, while for $\alpha = 1$ this is shown by Bubeck et al. [2014] and Curien et al. [2014]. What about for $\alpha \in (0, 1)$? Is $\delta_\alpha(S, T) > 0$ whenever S and T are non-isomorphic and have at least three vertices? What can be said about $\delta_\alpha(S, T)$ as a function of α? Is it monotone in α? Is it convex?

Acknowledgements

This work was done while R.E. and E.M. were visiting researchers and M.Z.R. was an intern at the Theory Group of Microsoft Research. They thank the Theory Group for their kind hospitality. S.B. thanks Luc Devroye for enlightening discussions on the uniform random recursive tree. E.M. was supported by NSF grants DMS 1106999 and CCF 1320105, by DOD ONR grant N00014-14-1-0823, and by grant 328025 from the Simons Foundation.

References

Sébastien Bubeck, Elchanan Mossel, and Miklós Z. Rácz. On the influence of the seed graph in the preferential attachment model. *arXiv preprint arXiv:1401.4849*, 2014.

Nicolas Curien, Thomas Duquesne, Igor Kortchemski, and Ioan Manolescu. Scaling limits and influence of the seed graph in preferential attachment trees. *arXiv preprint arXiv:1406.1758*, 2014.

Luc Devroye and Svante Janson. Long and short paths in uniform random recursive dags. *Arkiv för Matematik*, 49(1):61–77, 2011.

Michael Drmota. *Random trees: an interplay between combinatorics and probability.* Springer, 2009.

Steven N. Evans, Rudolf Grübel, and Anton Wakolbinger. Trickle-down processes and their boundaries. *Electronic Journal of Probability*, 17(1):1–58, 2012.

Rudolf Grübel and Igor Michailow. Random recursive trees: A boundary theory approach. *arXiv preprint arXiv:1406.7614*, 2014.

Norman L. Johnson, Samuel Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, volume 2. Wiley, second edition, 1995.

Hosam M. Mahmoud. Distances in random plane-oriented recursive trees. *Journal of Computational and Applied Mathematics*, 41(1-2):237–245, 1992.

A Facts about the beta-binomial distribution

We prove here Facts 1 and 2 stated in Section 2.1. Let M_k, $k = \alpha+\beta, \alpha+\beta+1, ...$, be the martingale associated with the standard Pólya urn process with starting state (α, β). In other words, the martingale M_k is defined by $M_{\alpha+\beta} = \frac{\alpha}{\alpha+\beta}$ and

$$(k+1)M_{k+1} = \begin{cases} kM_k + 1 & \text{with probability } M_k \\ kM_k & \text{with probability } 1 - M_k \end{cases}$$

independently for different values of k. Note that for $n \geq \alpha+\beta$, $nM_n \stackrel{d}{=} B_{\alpha,\beta,n-\alpha-\beta}$, so all results for the martingale M_n transfer to results for $B_{\alpha,\beta,n-\alpha-\beta}$. Define $M_\infty = \lim_{k\to\infty} M_k$, and note that this limit exists almost surely by the martingale convergence theorem. It is a well-known fact about Pólya urns that M_∞ has a beta distribution with parameters α and β, i.e., the density of M_∞ with respect to the Lebesgue measure is

$$h(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbf{1}_{\{x\in[0,1]\}}.$$

By the formula for the moments of M_∞ (see, e.g., [Johnson et al., 1995, Chapter 21]), we have

$$\mathbb{E}[M_\infty^p] = \prod_{j=0}^{p-1} \frac{\alpha+j}{\alpha+\beta+j} \leq \left(\frac{\alpha+p}{\alpha+\beta}\right)^p \leq (p+1)^p \left(\frac{\alpha}{\alpha+\beta}\right)^p, \quad \forall p \in \mathbb{N}.$$

Moreover, since M_n is a martingale, M_n^p is a submartingale for all $p \geq 1$, and thus $\mathbb{E}[M_n^p] \leq \mathbb{E}[M_\infty^p]$. So we have that

$$\mathbb{E}[(nM_n)^p] \leq n^p (p+1)^p \left(\frac{\alpha}{\alpha+\beta}\right)^p,$$

which establishes Fact 1 with $C(p) = p+1$.

Next, in order to prove Fact 2, we first use the formula for the negative first moment of M_∞ (see, e.g., [Johnson et al., 1995, Chapter 21]): for every $\alpha > 1$ we have $\mathbb{E}\left[M_\infty^{-1}\right] = (\alpha+\beta-1)/(\alpha-1)$. Thus by Markov's inequality we have that $\mathbb{P}(M_\infty < z) \leq z(\alpha+\beta-1)/(\alpha-1)$ for every $z > 0$, and thus

$$\mathbb{P}\left(M_\infty < t\frac{\alpha}{\alpha+\beta}\right) \leq 2t. \tag{24}$$

In the case that $\alpha = 1$, we have $h(x) \leq \beta$ which implies that $\int_0^{\frac{t}{\beta+1}} h(x)dx \leq t$. We conclude that (24) holds for any $\alpha, \beta \geq 1$. Since M_k is a positive martingale, we have

$$\mathbb{P}\left(M_\infty \leq 2z \,|\, M_n \leq z\right) \geq 1/2, \quad \forall z \in (0,1).$$

Combining this inequality with (24) gives

$$\mathbb{P}\left(M_n \leq t\frac{\alpha}{\alpha+\beta}\right) \leq 8t,$$

and formula (4) then follows with $C = 8$.

B Estimates on sequences

Lemma 3 *Suppose that $\{a_n\}_{n\geq 1}$ is a sequence of nonnegative real numbers and that there exists n_0 such that $a_{n_0} > 0$. Let α be a positive integer.*

(a) If there exists N such that $a_{n+1} \geq (1+\alpha/n)\, a_n$ for every $n \geq N$, then $\liminf_{n\to\infty} a_n/n^\alpha > 0$.

(b) If there exist constants c, γ, and N such that for every $n \geq N$,

$$a_{n+1} \leq (1+\alpha/n)\, a_n + c\left(\log(n)\right)^\gamma n^{\alpha-1},$$

then $a_n \lesssim n^\alpha$.

(c) If there exist constants c, γ, and N such that for every $n \geq N$,

$$a_{n+1} \leq (1+\alpha/n)\, a_n + c\left(\log(n)\right)^\gamma n^{\alpha},$$

then $a_n \lesssim n^{\alpha+1}$.

Proof (a) By the assumption we have that $a_n \geq a_{n_0} \exp\left(\sum_{j=n_0}^{n-1} \log(1+\alpha/j)\right)$, where $a_{n_0} > 0$. For $0 \leq x \leq 1$ we have that $\log(1+x) \geq x - x^2$, and so using the fact that $\sum_{j=1}^{\infty} 1/j^2 < \infty$, we have that there exists $c > 0$ such that $a_n \geq c \exp\left(\alpha \sum_{j=n_0\vee\alpha}^{n-1} 1/j\right)$. To conclude, recall that $\sum_{j=1}^{n-1} 1/j > \log(n)$.

(b) Let $b_n := a_n/n^\alpha$. We then have that $b_{n+1} \leq (1+\alpha/n)\left(n/(n+1)\right)^\alpha b_n + c\left(\log(n)\right)^\gamma/n$. There exists a constant $c'' = c''(\alpha)$ such that $\left(n/(n+1)\right)^\alpha \leq 1 - \alpha/n + c''/n^2$ for every $n \geq 1$. Therefore there exists a constant $c' = c'(\alpha)$ such that $(1+\alpha/n)\left(n/(n+1)\right)^\alpha \leq 1 + c'/n^2$ for every $n \geq 1$. Thus we have that $b_{n+1} \leq \left(1 + c'/n^2\right) b_n + c\left(\log(n)\right)^\gamma/n$, and iterating this we get that

$$b_n \leq b_1 \prod_{j=1}^{n-1}\left(1 + c'/j^2\right) + \sum_{j=1}^{n-1}\left(\prod_{i=j+1}^{n-1}\left(1 + c'/i^2\right)\right) c\left(\log(j)\right)^\gamma/j.$$

Since $\prod_{j=1}^{\infty}\left(1 + c'/j^2\right) < \infty$, we immediately get that $b_n \lesssim 1$, and so $a_n \lesssim n^\alpha$.

(c) This is similar to (b) so we do not repeat the argument. ■

C Tail behavior of the diameter

We reproduce a simple argument of Devroye and Janson [2011] to obtain a tail bound for the diameter $\mathrm{diam}\left(\mathrm{UA}\left(n,S\right)\right)$ of a uniform attachment tree.

Lemma 4 *For every seed tree S there exists a constant $C=C\left(S\right)$ such that for every $K>20$ we have*

$$\mathbb{P}\left(\mathrm{diam}\left(\mathrm{UA}\left(n,S\right)\right)>K\log\left(n\right)\right)\leq\frac{C\left(S\right)}{n^{K/2}}.$$

Proof First, if we set $C\left(S\right):=\left(\mathbb{P}\left(\mathrm{UA}\left(\left|S\right|,S_2\right)=S\right)\right)^{-1}$, then we have

$$\begin{aligned}\mathbb{P}\left(\mathrm{diam}\left(\mathrm{UA}\left(n,S\right)\right)>K\log\left(n\right)\right)&=\mathbb{P}\left(\mathrm{diam}\left(\mathrm{UA}\left(n,S_2\right)\right)>K\log\left(n\right)\mid\mathrm{UA}\left(\left|S\right|,S_2\right)=S\right)\\&\leq C\left(S\right)\mathbb{P}\left(\mathrm{diam}\left(\mathrm{UA}\left(n,S_2\right)\right)>K\log\left(n\right)\right),\end{aligned}$$

so it remains to bound the tail of $\mathrm{diam}\left(\mathrm{UA}\left(n,S_2\right)\right)$.

For notational convenience, shift the names of the vertices so that they consist of the set $\{0,1,\ldots,n-1\}$ (instead of $\{1,2,\ldots,n\}$), and call vertex 0 the root. With this convention, the label of the parent of vertex j is distributed as $\lfloor jU\rfloor$ where U is uniform on $[0,1]$. Similarly, an ancestor ℓ generations back has a label distributed like $\lfloor\ldots\lfloor\lfloor jU_1\rfloor U_2\rfloor\ldots U_\ell\rfloor$, where the U_i's are i.i.d. uniform on $[0,1]$.

Define R_j to be the distance from vertex j to the root. By the triangle inequality we have that $\mathrm{diam}\left(\mathrm{UA}\left(n,S\right)\right)\leq 2\max_{1\leq j\leq n-1}R_j$, so it suffices to bound the tail of this latter quantity. Using a union bound, it then suffices to bound the tail of R_j for each j. Now notice that $R_j\leq\min\left\{t:jU_1\ldots U_t<1\right\}$. Consequently, for any $\lambda>0$ we have

$$\mathbb{P}\left(R_j>t\right)\leq\mathbb{P}\left(jU_1\ldots U_t\geq 1\right)\leq\mathbb{E}\left[\left(jU_1\ldots U_t\right)^\lambda\right]=j^\lambda\left(\lambda+1\right)^{-t}.$$

This is optimized by choosing $\lambda=t/\log\left(j\right)-1$ (provided $t>\log\left(j\right)$) to obtain

$$\mathbb{P}\left(R_j>t\right)\leq\exp\left(t-\log\left(j\right)-t\log\left(\frac{t}{\log\left(j\right)}\right)\right)\leq\exp\left(t-t\log\left(\frac{t}{\log\left(n\right)}\right)\right),$$

when $j\leq n$. Putting everything together we get that

$$\begin{aligned}\mathbb{P}\left(\mathrm{diam}\left(\mathrm{UA}\left(n,S_2\right)\right)>K\log\left(n\right)\right)&\leq\mathbb{P}\left(\max_{1\leq j\leq n-1}R_j>\frac{K}{2}\log\left(n\right)\right)\leq\sum_{j=1}^{n-1}\mathbb{P}\left(R_j>\frac{K}{2}\log\left(n\right)\right)\\&\leq n^{-\frac{K}{2}\log\left(\frac{K}{2}\right)+\frac{K}{2}+1}\leq n^{-K/2},\end{aligned}$$

where the last inequality holds when $K>20$. ■