Influence of the seed in *affine* preferential attachment trees

DAVID CORLIN MARCHAND[1,2,*] and IOAN MANOLESCU[2,**]

[1]*Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France.*
[2]*Université de Fribourg, 23 Chemin du Musée, CH-1700 Fribourg, Switzerland.*
E-mail: [*]*david.marchand@u-psud.fr;* [**]*ioan.manolescu@unifr.ch*

We study randomly growing trees governed by the affine preferential attachment rule. Starting with a seed tree S, vertices are attached one by one, each linked by an edge to a random vertex of the current tree, chosen with a probability proportional to an affine function of its degree. This yields a one-parameter family of preferential attachment trees $(T_n^S)_{n \geq |S|}$, of which the linear model is a particular case. Depending on the choice of the parameter, the power-laws governing the degrees in T_n^S have different exponents.

We study the problem of the asymptotic influence of the seed S on the law of T_n^S. We show that, for any two distinct seeds S and S', the laws of T_n^S and $T_n^{S'}$ remain at uniformly positive total-variation distance as n increases.

This is a continuation of Curien et al. (*J. Éc. Polytech. Math.* **2** (2015) 1–34), which in turn was inspired by a conjecture of Bubeck et al. (*IEEE Trans. Netw. Sci. Eng.* **2** (2015) 30–39). The technique developed here is more robust than previous ones and is likely to help in the study of more general attachment mechanisms.

Keywords: Barábasi–Albert trees; preferential attachment; seed recognition

1. Introduction

Linear preferential attachment trees – or *Barábasi-Albert trees* – are trees that grow randomly according to the following simple mechanism. Start with a finite tree S with k vertices – we call it the *seed tree* and k its size – and construct inductively a sequence $(T_n^S)_{n \geq k}$ of random trees, where T_{n+1}^S is obtained from T_n^S by adding a new vertex and connecting it to a vertex of T_n^S chosen at random with a probability proportional to its degree. Thus, the tree T_{n+1}^S has one more vertex than T_n^S – namely $n + 1$ – and one more edge – namely n.

The reader might note that this model formalises the adage "the rich get richer" since vertices with high degrees are more likely to receive new connections. This property suits a large class of real networks (see [1]) mainly because of the emergence of power laws for the sequence of degrees of the tree as its size grows, see [4]. Such laws are observed in a wide range of contexts, like social networks (webgraph, citation graph, etc [15,18]) or even in biological networks such as interaction protein networks (see [13]). The degree distribution in linear preferential attachment trees has been deeply investigated; for an extensive overview of these traditional topics in the preferential attachment model's analysis, the reader is directed to [17], Ch. 8.

In the literature, there are numerous variations in the definition of preferential attachment model. Here we will consider one where the new vertex is attached to an old vertex chosen with

probability which is an affine function of its degree, rather than a linear one. Fix a parameter $\alpha > 0$ and S a seed tree of size $k \geq 2$. The α-PA trees grown from the seed S are denoted by $(T_n^S)_{n \geq k}$ and are defined recursively as follows. Given $(T_n^S)_{k \leq n \leq N}$ for some $N \geq k$ with $T_N^S = T$ for some fixed tree T, select randomly a vertex u_N in T with probability:

$$\mathbb{P}\left(u_N = u | T_k^S, \ldots, T_N^S \text{ with } T_N^S = T\right) = \frac{\deg_T(u) - 1 + \alpha}{(1 + \alpha)N - 2} \quad \forall u \in T, \tag{1}$$

and link it to a new vertex by an edge. The resulting tree is T_{N+1}^S.

Note that $(1 + \alpha)N - 2 = \sum_{u \in T} \deg_T(u) - 1 + \alpha$ for all trees T of size N, so (1) defines indeed a probability measure. This model, which was first introduced in [14], is a generalisation of the Barábasi-Albert model, in that the latter is obtained when choosing $\alpha = 1$. The authors of [2] remarked that transitions probabilities (1) provide a Pólya urn representation of α-PA trees; particularly convenient properties follow, such as estimates on the degree growth and a form of exchangeability for the sequence $(u_N)_{N \geq k}$. When the dependence on S is not important, we will drop S from the notation.

<div align="center">Henceforth $\alpha > 0$ is fixed.</div>

The same questions as for the linear model may be asked of the affine one. For instance, one may study the growth of degrees of given vertices in T_n. It may be proved (see [17], Chapters 8.3 & 8.4, or Remark 4.12 of this paper) that the degree of any vertex of the seed increases at polynomial speed, with an exponent that depends on α and varies between 1 and 0 as α ranges from 0 to infinity. It may appear surprising that an apparently insignificant difference in the attachment mechanism leads to different scaling exponents in the power laws governing the degree sequence, and hence to preferential attachment models of different universality classes. More substantial variations of the model will be discussed in Section 5.

The problem of interest in this paper is that of the recognition of the seed. Precisely, we will study whether the seed tree has any influence on the law of the tree obtained after a large number of iterations of the growth procedure. This question was asked (and partly solved) for the linear model in [7]. The complete answer for the linear model was obtained in [8], and for the uniform preferential attachment model in [6]. In both cases, the seed is shown to influence the asymptotic law of the model. Our aim is to generalise the result to the α-PA, of which the uniform model may be perceived as a limit.

Theorem 1.1. *Let S and S' be two finite trees of size $k_1, k_2 \geq 3$. Then for any $\alpha > 0$, the following limit*:

$$d(S, S') = \lim_{n \to +\infty} d_{\mathrm{TV}}\left(T_n^S, T_n^{S'}\right) \tag{2}$$

exists and is non zero when S is not graph-isomorphic to S'.

As a consequence of Theorem 1.1, the function d is a metric on finite trees with at least 3 vertices. It measures the statistical chance of distinguishing between two different seed trees given observations of the α-PA trees grown from them.

Remark 1.2. The existence of the limit in Theorem 1.1 stems from the fact that the sequence $d_{\mathrm{TV}}(T_n^S, T_n^{S'})$ is decreasing in n, which may be proved by a coupling argument.

In proving Theorem 1.1, we will look at the total variation between T_n^S and $T_n^{S'}$ through some integer-valued observable obtained from them. The same is done in all previous studies [6–8].

In [7], where the model is linear (that is $\alpha = 1$) and S and S' are assumed to have different degree sequences, the authors use as observable the maximum degree. Indeed, they show that the maximal degrees of T_n^S and $T_n^{S'}$ have different tail distributions, which then easily leads to $d(S, S') > 0$. The key to this argument is that the degrees of vertices in the Barábasi-Albert model evolve as a Pólya urn. When the degree distributions of S and S' are identical, but their geometry is different, the observable of [7] is unable to distinguish between T_n^S and $T_n^{S'}$. To overcome this difficulty, a more complex observable was introduced in [8]: the number of embeddings of a fixed finite tree τ inside T_n^S, weighted by some function of the degrees in T_n^S of the embedding of τ. The same type of observable was used in [6] and will be used below.

For $n \geq 0$ and $d \geq 1$, set $[n]_d = n(n-1)\ldots(n-d+1)$. Also set $[n]_0 = 1$ for any $n \geq 0$. A *decorated tree* is a couple $\boldsymbol{\tau} = (\tau, \ell)$ where τ is a tree and ℓ is function from the vertices of τ to the set of non-negative integers. For a decorated tree $\boldsymbol{\tau}$ and a (bigger) tree T, set

$$\mathcal{F}_{\boldsymbol{\tau}}(T) := \sum_{\phi:\tau \hookrightarrow T} \prod_{u \in \tau} \big[\deg_T\big(\phi(u)\big) - 1\big]_{\ell(u)}, \tag{3}$$

where the sum is over all graph embeddings of τ in T and the product is over all vertices of τ.

For two distinct seeds S and S', we aim to show that, for a well-chosen decorated tree $\boldsymbol{\tau}$, the difference of the expectations of $\mathcal{F}_{\boldsymbol{\tau}}(T_n^S)$ and $\mathcal{F}_{\boldsymbol{\tau}}(T_n^{S'})$ is of the same order as each of them and as their standard deviation. This allows to control the total variation between T_n^S and $T_n^{S'}$ using the following bound.

Lemma 1.3. *For any two real-valued, square-integrable random variables X and Y,*

$$d_{\mathrm{TV}}(X, Y) \geq \frac{(\mathbb{E}[X] - \mathbb{E}[Y])^2}{(\mathbb{E}[X] - \mathbb{E}[Y])^2 + 2 \cdot (\mathbb{E}[X^2] + \mathbb{E}[Y^2])}.$$

Proof of Lemma 1.3. Let (X', Y') be a coupling of the random variables X and Y. By using Paley-Zigmund's inequality, then Jensen's one, we get:

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

Furthermore, a simple decomposition gives:

$$\mathbb{E}\big[(X' - Y')^2\big] = \mathbb{E}\big[(X' - \mathbb{E}[X'] + \mathbb{E}[X'] - \mathbb{E}[Y'] + \mathbb{E}[Y'] - Y')^2\big]$$
$$\leq 2 \cdot \big(\mathrm{Var}(X') + \mathrm{Var}(Y')\big) + \big(\mathbb{E}[X'] - \mathbb{E}[Y']\big)^2$$
$$\leq 2 \cdot \big(\mathbb{E}[(X')^2] + \mathbb{E}[(Y')^2]\big) + \big(\mathbb{E}[X'] - \mathbb{E}[Y']\big)^2.$$

It finally remains to note that X', resp. Y', has the same moments as X, resp. Y, and that the total variation between X and Y is obtained by taking the infimum of $\mathbb{P}(X' \neq Y')$ over all couplings (X', Y') of X and Y.   □

It comes to no surprise that the evolution of the moments of $\mathcal{F}_\tau(T_n^S)$ is of great importance for the proof. Let us fix τ and S and discuss briefly the first moment of $\mathcal{F}_\tau(T_n^S)$ as $n \to \infty$. There are two competing factors contributing to $\mathcal{F}_\tau(T_n^S)$. First, due to the "rich gets richer" phenomenon, the "oldest" vertices of T_n^S have degree evolving as $n^{\frac{1}{1+\alpha}}$. Thus, the contribution to $\mathbb{E}[\mathcal{F}_\tau(T_n^S)]$ of embeddings included in the seed (or close to it) is of the order $n^{\frac{|\ell|}{1+\alpha}}$ for $|\ell| = \sum_{v \in \tau} \ell(v)$.

Second, one should take into account the embeddings using recently acquired vertices. There are many such vertices, but they have small degrees. Depending on the form of τ, such embeddings contribute to $\mathbb{E}[\mathcal{F}_\tau(T_n^S)]$ by a quantity that may exceed $n^{\frac{|\ell|}{1+\alpha}}$ by some logarithmic or even polynomial factor. Thus we find

$$\mathbb{E}[\mathcal{F}_\tau(T_n)] \approx \log^{\gamma(\tau)}(n) \cdot n^{\frac{\lambda(\tau)}{1+\alpha}}, \tag{4}$$

for some $\gamma(\tau) \geq 0$ and $\lambda(\tau) \geq |\ell|$. Both values $\gamma(\tau)$ and $\lambda(\tau)$ are computed explicitly in Section 3.

Due to the coupling between α-PA started from different seeds (see Section 2), only the first type of embeddings is sensitive to the seed. Thus, for two distinct seeds S, S', we may expect that

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})] \approx n^{\frac{|\ell|}{1+\alpha}}, \tag{5}$$

provided that a difference exists. Let us, for the sake of this explanation, assume that the standard deviations of $\mathcal{F}_\tau(T_n^S)$ and $\mathcal{F}_\tau(T_n^{S'})$ are of the same order as their expectations. Then, in order to successfully apply Lemma 1.3, we should have $\gamma = 0$ and $\lambda(\tau) = |\ell|$.

Such asymptotics (slightly weaker actually) were already observed in [8] for $\alpha = 1$; they were proved using an amenable recursive relation for $\mathbb{E}[\mathcal{F}_\tau(T_n)]$. Rather than showing that τ may be chosen to satisfy (5) and so that $\gamma(\tau) = 0$ and $\lambda(\tau) = |\ell|$, the authors of [8] constructed a linear combination over trees τ of observables $\mathcal{F}_\tau(T_n)$ for which the logarithmic factors cancel out. The resulting observable, properly rescaled, turns out to be a martingale that is bounded in L^2 and whose expectation depends on the seed tree.

We will employ a different, arguably simpler strategy: we will prove that for any two distinct seeds $S \neq S'$, the decorated tree τ may be chosen so as to observe a difference as in (5) and such that $\gamma(\tau) = 0$ and $\lambda(\tau) = |\ell|$. As such, our strategy is more likely to apply to other attachment mechanisms, as will be discussed in Section 5. Additional differences with [8], are the recurrence formula used to prove (4), which is more complex in the affine case, and the fact that the exponent $\lambda(\tau)$ is not always equal to $|\ell|$, as opposed to the linear case, where $\lambda(\tau) = |\ell|$ always.

Let us finally mention that the exact definition (3) of the observable is somewhat arbitrary. Indeed, it is possible to use variations of (3), such as $\sum_{\phi:\tau \hookrightarrow T} \prod_{u \in \tau} (\deg_T \phi(u))^{\ell(u)}$. We chose (3) as it is inspired by the equivalent construction in [8] and has an interpretation in the planar version of the model (see Section 2).

Organisation of the paper

In Section 2, we introduce a planar version of the affine preferential attachment model. This construction is not formally necessary to prove our main result, but we believe it is interesting in its own right and helps clarify the subsequent proofs. In particular, the planar version will give rise to a natural coupling between two α-PA trees starting from two different seeds of same size.

In Section 3, we study the asymptotics of the first and second moments of $\mathcal{F}_{\boldsymbol{\tau}}(T_n^S)$ as $n \to \infty$. The only role of this section within the proof of Theorem 1.1 is to show a bound on the second moment of $\mathcal{F}_{\boldsymbol{\tau}}(T_n^S)$ for certain trees $\boldsymbol{\tau}$. These estimates use a recurrence relation for $\mathcal{F}_{\boldsymbol{\tau}}(T_n^S)$, which is similar to that of [8], but more complicated due to the affine probabilities.

Section 4 contains lower bounds on the first-moment difference of $\mathcal{F}_{\boldsymbol{\tau}}(T_n^S)$ for two distinct seeds S. Precisely, we prove (5) for well-chosen decorated trees $\boldsymbol{\tau}$. This is then used to prove Theorem 1.1.

Finally, certain variations of the model and potential extensions of our result are discussed in Section 5.

Notation

In the rest of the paper, we will use the following notations:

- for two sequences $f, g : \mathbb{N} \to (0, +\infty)$, write $f(n) \approx g(n)$ if there exists some constant $C > 0$ such that $\frac{1}{C} \leq \frac{f(n)}{g(n)} \leq C$;
- also $f(n) = \mathcal{O}(g(n))$ if $\limsup_{n \to +\infty} |\frac{f(n)}{g(n)}| < +\infty$;
- and $f(n) \ll g(n)$ if $\lim_{n \to +\infty} |\frac{f(n)}{g(n)}| = 0$.
- The size of a tree T is written $|T|$ and stands for the number of vertices of T.
- For a graph S, write V_S and E_S for its sets of vertices and edges, respectively. For $v \in V_S$, write $\deg_S(v)$ or $\deg(v)$ for the degree of v in S.

2. Planar affine preferential attachment model

In this section, we first define the planar version of affine preferential attachment trees, then we present a useful coupling between any two α-PA trees derived from different seed trees of same size. The coupling is based on the decomposition of T_n^S into *planted plane trees*, which isolates the roles of the seed and of the growth mechanism, respectively.

We reiterate that this planar version is not formally necessary for any of the results; it is used merely to illustrate the arguments.

2.1. Definition via corners

We first need to introduce the notion of *colouring of corners* in a plane tree. A plane tree is a tree embedded in the plane up to continuous deformation, or equivalently a tree whose vertices are equipped with a cyclic order of their neighbours. A *corner* of a plane tree is an angular sector

of the plane contained between two consecutive half-edges around a vertex. In particular, the number of corners surrounding a vertex is equal to its degree.

Definition 2.1 (Colouring of corners). A colouring of the corners of a plane tree $\mathcal{T}$ is a choice of one distinguished corner for each vertex of the tree. The distinguished corners are called "red"; all other corners are called "blue".

When indexing $\{c_{v,i} : v \in V_T, 1 \leq i \leq \deg(v_i)\}$ the corners of $\mathcal{T}$, we will always suppose that the red vertices are $(c_{v,1})_{v \in V_{\mathcal{T}}}$.

It will be useful to keep in mind that any plane tree with n vertices and coloured corners has n red corners and $n - 2$ blue ones. We are now ready to introduce the planar version of the affine preferential attachment model.

Definition 2.2. Fix a parameter $\alpha > 0$ and a plane tree S – namely the *seed tree* – of size $|S| = k \geq 2$ endowed with a colouring of its corners. The planar α-PA with seed S is a growing sequence of random plane trees $(T_n^S)_{n \geq k}$ (or $(T_n)_{n \geq k}$ when the role of S is clear) with coloured corners, evolving according to the following inductive principle:

   (i) $T_k = S$;
  (ii) Suppose that our sequence of plane trees is built until step $n \geq k$. Then, independently from the past of the process, select at random a corner c_n of T_n with

$$\mathbb{P}(c_n = c_{u,1}) = \frac{\alpha}{n(1+\alpha) - 2} \quad \text{and} \quad \mathbb{P}(c_n = c_{u,i}) = \frac{1}{n(1+\alpha) - 2} \quad \forall 2 \leq i \leq \deg_{T_n}(u),$$

for any vertex $u \in V_{T_n}$. That is, red corners are chosen with probability proportional to α and blue corners with probability proportional to 1.
 (iii) Write u_n for the vertex of T_n of which c_n is a corner. To obtain T_{n+1} add a new vertex v_{n+1} to T_n and we connect it by an edge to u_n through the corner c_n.
 (iv) All corners of T_n except c_n subsist in T_{n+1} and conserve their colour. The unique corner c_n'' of the new vertex v_{n+1} is coloured red, as required by Definition 2.1. The randomly selected corner c_n is divided in two angular areas by the edge connecting u_n to v_{n+1}. If c_n was blue in T_n, then both these corners of T_{n+1} are blue. If c_n was red in T_n, we arbitrarily colour the corner to the right of the edge (u_n, v_{n+1}) in red and the other in blue.

The probabilities appearing in step (ii) indeed sum to 1 since:

$$\underbrace{n}_{\text{number of red corners}} \cdot \frac{\alpha}{n \cdot (1+\alpha) - 2} + \underbrace{(n-2)}_{\text{number of blue corners}} \cdot \frac{1}{n \cdot (1+\alpha) - 2} = 1$$

Observe that, the attachment principle in the planar α-PA is such that the probability for v_{n+1} to be attached to a vertex u of T_n is $\frac{\deg_{T_n}(u) - 1 + \alpha}{n \cdot (1+\alpha) - 2}$. Thus, we obtain the following.

Lemma 2.3. *If $(T_n)_{n \geq k}$ is a planar α-PA started from some plane seed tree S, then the trees $(T_n)_{n \geq k}$ (stripped of their planar embedding) have the law of a α-PA.*

Remark 2.4. Changing the rules of the assignment of labels and colours to the new corners of a planar α-PA affects the law of the plane tree, but only via its embedding.

In the special case $\alpha = 1$, the weights of red and blue corners are equal, and the planar α-PA is simply a uniform embedding of the abstract α-PA. This is not the case when $\alpha \neq 1$. For more on this see [8], Conj. 2.

As mentioned in the introduction, the observables $\mathcal{F}_{\boldsymbol{\tau}}$ have a particular interpretation in terms of the planar model. Let $\boldsymbol{\tau} = (\tau, \ell)$ be a decorated tree and T be a plane tree with coloured corners. Then, for any embedding $\phi : \tau \hookrightarrow T$, the factor $\prod_{u \in \tau} [\deg_T(\phi(u)) - 1]_{\ell(u)}$ is the number of ordered choices of $\ell(u)$ different blue corners around each vertex $\phi(u)$ for $u \in \tau$. As in [8], one may imagine that each vertex $u \in \tau$ is endowed with $\ell(u)$ distinct arrows. Call a decorated embedding an embedding of ϕ of τ in T together with, for each $u \in \tau$ and each arrow of u, a blue corner of $\phi(u)$ to which that arrow points, in such a way that distinct arrows point to distinct corners. Then $\mathcal{F}_{\boldsymbol{\tau}}(T)$ is the number of decorated embeddings of τ in T.

2.2. Decomposition and coupling using planted plane trees

Apart from its intrinsic interest, the decomposition described below will be used in Section 4. We begin with the definition of a planted plane tree.

Definition 2.5. A planted plane tree $T_{\!-\!\circ}$ is a plane tree T with a distinguished vertex called the root and an additional half-edge emerging from the root. This half edge divides the corner of the root delimited by the two edges linking the root to its leftmost and its rightmost children. When $|T| = 1$, the planted plane tree $T_{\!-\!\circ}$ is merely a single vertex with an half-edge attached to it.

It should be noted that there is one more corner in any planted version $T_{\!-\!\circ}$ of a plane tree T. A colouring of corners for a planted plane tree $T_{\!-\!\circ}$ is defined as for plane trees (see Definition 2.1) with the exception that the root is allowed to have no red corners or one red corner, which will always be the corner to the right of the half-edge. In the former situation, we say that $T_{\!-\!\circ}$ is a **blue planted plane tree**, in the latter we say it is a **red planted plane tree**.

Since the recursive procedure used to define the planar α-PA tree is simply based on the colouring of corners, we straightforwardly adapt it to define a planted planar version of the same model. Note by $T_n^{-\!\circ b}$, resp. $T_n^{-\!\circ r}$, a blue planted plane tree, resp. a red planted plane tree, obtained through the preferential attachment algorithm with initial condition $-\!\circ$, which is the tree composed of a single vertex with a half-edge attached to it and a blue corner, resp. a red corner, surrounding it.

Let S be a plane seed tree of size $k \geq 2$ with coloured corners indexed $\{c_{v,i} : v \in V_S, 1 \leq i \leq \deg_S(v)\}$. Fix $n \geq k$ and T_n be a realisation of the planar α-PA tree at step n starting from S. For $v \in V_S$ and $1 \leq i \leq \deg_S(v)$, say that a vertex $u \in V_{T_n} \setminus V_S$ is a descendent of the corner $c_{v,i}$ if the unique path linking u to S arrives at S through $c_{v,i}$. Additionally, say that v is also a descendent of $c_{v,i}$. The tree $T_n^{v,i}$ stemming from $c_{v,i}$ is the planted plane tree containing all vertices of T_n that are descendent of $c_{v,i}$ and all edges of T_n between such vertices; it is rooted at v. The half-edge attached to v is such that it does not split the corner $c_{v,i}$. Call these trees the planted plane subtrees of T_n^S. See Figure 1 for an illustration.
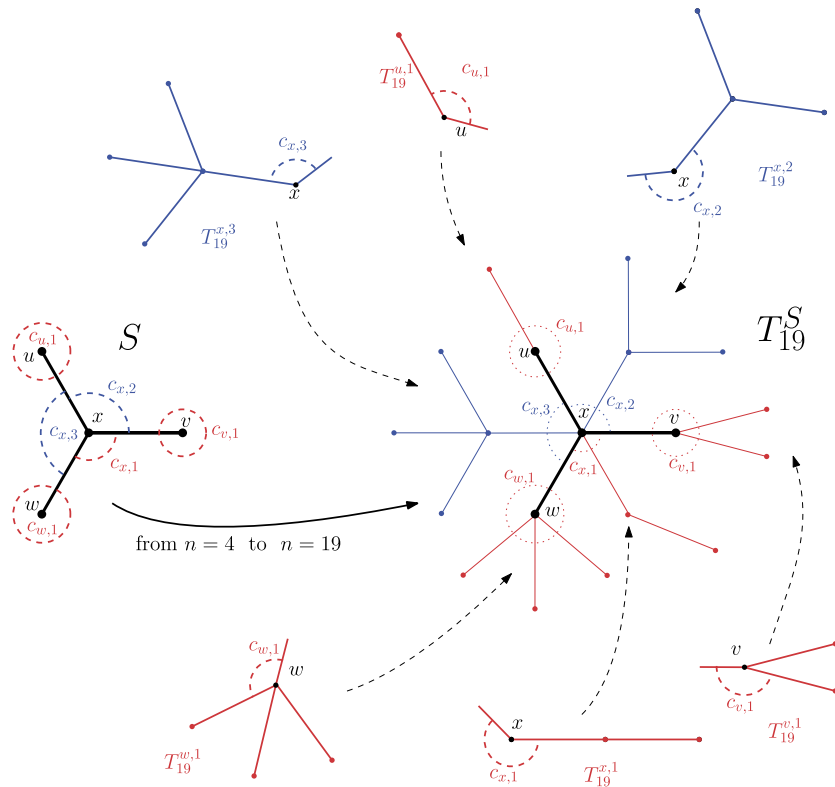
Figure 1. An example of decomposition into red and blue planted plane tree.

Proposition 2.6. *Let S be a plane seed tree of size $k \geq 2$ with coloured corners indexed $\{c_{v,i} : v \in V_S, 1 \leq i \leq \deg_S(v)\}$. For $n \geq k$, let $T_n^{v,i}$ be the planted plane tree stemming from $c_{v,i}$ and let $k_n^{v,i}$ denote its size. Write $x_n^{v,1} = (1+\alpha)k_n^{v,1} - 1$ and $x_n^{v,i} = (1+\alpha)k_n^{v,i} - \alpha$ for all $v \in V_S$ and $i \geq 2$. Then*

- *the vector $(x_n^{v,i})_{v \in V_S, 1 \leq i \leq \deg_S(v)}$ has the distribution of a Pólya urn with $2k-2$ colours and diagonal replacement matrix $(1+\alpha)I_{2k-2}$, starting from $(\alpha \mathbb{1}_{\{i=1\}} + \mathbb{1}_{\{i \neq 1\}})_{v \in V_S, 1 \leq i \leq \deg_S(v)}$;*
- *conditionally on $(x_n^{v,i})_{v \in V_S, 1 \leq i \leq \deg_S(v)}$, the trees $(T_n^{v,i})_{v \in V_S, 1 \leq i \leq \deg_S(v)}$ are independent, with $T_n^{v,i}$ having law $T_{k_n^{v,i}}^{\to r}$ if $i = 1$ and $T_{k_n^{v,i}}^{\to b}$ if $i \neq 1$.*

Proposition 2.6 may be restated as follows. Given a plane seed S with $|S| = k$, we would like to construct T_n^S for some $n \geq k$. This may be done in the following steps.

1. Generate a vector $(x_n^{v,i})_{v \in V_S, 1 \leq i \leq \deg_S(v)} \in \mathbb{R}^{2k-2}$ with the law of the Pólya urn of Proposition 2.6. Define sizes $k_n^{v,1} := \frac{x_{\ell,p}^{(b)} + \alpha}{1+\alpha}$ and $k_n^{v,i} := \frac{x_\ell^{(r)} + 1}{1+\alpha}$ for $v \in V_S$ and $i \geq 2$.

2. Randomly draw independent realisations $\tau_n^{v,1}$ of $T_{k_n^{v,1}}^{\multimap r}$ and $\tau_n^{v,i}$ of $T_{k_n^{v,i}}^{\multimap b}$ for each $v \in V_S$ and $i \geq 2$.

3. Graft each tree $\tau_n^{v,i}$ in the corner $c_{v,i}$ of S.

The resulting tree has the law of a planar α-PA started from S.

This decomposition allows to couple the evolution of two planar α-PA trees emerging from distinct seeds of same size. Indeed, in spite of the notation, the first two steps do not depend on S, only on $k = |S|$: in the first step, the vector $(x_n^{v,i})$ contains $2k - 2$ entries and the Pólya urn generating it starts with k entries α and $k - 2$ entries 1; in the second step, k trees are of type $T^{\multimap r}$ and $k - 2$ of type $T^{\multimap b}$. If S, S' are two seeds of same size k, then T_n^S and $T_n^{S'}$ are coupled as follows. Simulate a common vector for the sizes of the planted plane subtrees of T_n^S and $T_n^{S'}$, then simulate a common set of planted plane trees. Finally, graft these trees to S to obtain T_n^S and the same trees to S' to obtain $T_n^{S'}$.

Proof of Proposition 2.6. Fix S. Let us fist prove that the vector $(x_n^{v,i})_{v \in V_S, 1 \leq i \leq \deg_S(v)}$ has the distribution of the Pólya urn described in the statement. We do this by induction on n. When $n = k$, all planted plane subtrees of $T_k^S = S$ have size 1. The formula relating the sizes of the subtrees to $x_n^{v,i}$ then yields $x_n^{v,1} = \alpha$ and $x_n^{v,i} = 1$ for $i \geq 2$.

Suppose now that the statement is proved up to step n and let T_n^S be a realisation of the planar α-PA tree of size n. Then, each red planted plane subtree $T_n^{v,1}$ has $k_n^{v,1}$ red corners and $k_n^{v,1} - 1$ blue ones. Each blue planted plane subtree $T_n^{v,i}$ with $i \geq 2$ has $k_n^{v,i} - 1$ red corners and $k_n^{v,i}$ blue corners. It follows that, for each v

$$\mathbb{P}(v_{n+1} \text{ descendent of } c_{v,1}) = \frac{(\alpha + 1)k_n^{v,1} - 1}{2n - 2} = \frac{x_n^{v,1}}{2n - 2} \quad \text{and}$$

$$\mathbb{P}(v_{n+1} \text{ descendent of } c_{v,i}) = \frac{(\alpha + 1)k_n^{v,1} - \alpha}{2n - 2} = \frac{x_n^{v,i}}{2n - 2} \quad \text{for all } i \geq 2.$$

Finally, if v_{n+1} is attached to some descendent of a corner $c_{v,i}$, then $x_{n+1}^{v,i} = 1 + \alpha + x_n^{v,i}$. This shows that the vector $(x_{n+1}^{v,i})$ also has the claimed distribution.

Let us now condition on $(x_n^{v,i})_{v,i}$, or equivalently on $(k_n^{v,i})_{v,i}$ for some n. Then, for each $v \in V_s$ and $i \leq \deg_S(v)$, the $k_n^{v,i}$ vertices of $T_n^{v,i}$, independently of the times they join $T_k, \ldots, T_n$, attach themselves to the planted plane subtree of the corner $c_{v,i}$ with a (conditional) distribution that assigns to any red corner a weight proportional to α and to any blue corner a weight proportional to 1. If follows readily that $T_n^{v,i}$ has the law of $T_{k_n^{v,i}}^{\multimap r}$ or $T_{k_n^{v,i}}^{\multimap b}$, depending on whether $c_{v,i}$ is red or blue, respectively. It is also immediate, that the resulting trees in different corners are independent – this independence only holds *conditionally on the size vector* $(k_n^{v,i})_{v,i}$; moreover, we do not claim that the evolution between step k and n is independent in different corners, only the resulting subtrees are. $\square$

3. First and second moment of a class of observables

The ultimate goal of this section is to obtain precise estimates on the second moment of $\mathcal{F}_\tau(T_n)$ for particular decorated trees τ. In doing so, we will also prove a general result on the first moment of $\mathcal{F}_\tau(T_n)$ for any τ. The latter is longer to state, and is deferred to later in the section. Below is the minimal result required in the proof of Theorem 1.1.

Theorem 3.1. *Let τ be a finite decorated tree with $\ell(u) \geq 2$ for any $u \in \tau$ and such that $|\ell| = \sum_{u \in \tau} \ell(u) > 1 + \alpha$. For any seed S, we have the following asymptotic as $n \to \infty$:*

$$\mathbb{E}\big[\mathcal{F}_\tau\big(T_n^S\big)^2\big] \approx n^{\frac{2|\ell|}{1+\alpha}}.$$

To prove the above, we will proceed in several steps, each occupying a subsection below. First, we prove a recurrence relation on the first moment of our observables (see Section 3.1), which is then used in Section 3.2 to obtain an accurate estimate of $\mathbb{E}[\mathcal{F}_\tau(T_n^S)]$ as n goes to infinity (see Theorem 3.7). Finally, in Section 3.3, we use the previous results to prove Theorem 3.1.

3.1. A recurrence formula

Recall from the introduction that the asymptotic of $\mathbb{E}[\mathcal{F}_\tau(T_n^S)]$ as $n \to \infty$ is polynomial with an exponent depending on τ. This exponent will be determined by the weight of τ defined below.

Definition 3.2 (Weight of decorated trees). Let τ be a decorated tree. If τ has size 1 or 2 and all decorations of its vertices are 0, then the weight of τ is 1. In all other cases, the weight of τ is

$$w(\tau) = \sum_{v \in V_\tau} \ell(v) + \mathbb{1}_{\{\deg_\tau(v)=1 \text{ and } \ell(v)=0\}}.$$

Call a vertex $v \in \tau$ with $\ell(v) = 0$ and $\deg_\tau(v) = 1$ a *loose leaf* of τ. Then the weight of τ is $|\ell|$ plus the number of loose leaves.

Definition 3.2 induces a partial order on the set of decorated trees.

Definition 3.3 (Partial order on decorated trees). For any two decorated trees $\tau = (\tau, \ell)$, $\tau' = (\tau', \ell')$ we write $\tau' \prec \tau$ if and only if:

- $w(\tau') < w(\tau)$ and $|\tau'| \leq |\tau|$;
- or if $w(\tau') = w(\tau)$ and $|\tau'| < |\tau|$;
- or if $w(\tau') = w(\tau)$, $|\tau'| = |\tau|$ and $\sum_{v \in V_{\tau'}} \ell'(v) = |\ell'| < |\ell| = \sum_{v \in V_\tau} \ell(v)$.

Clearly, $\prec$ is a strict partial order on the set of decorated trees. We denote by $\preceq$ the associated partial order.

There are three trees τ of weight 1: those with a single vertex and decoration 0 or 1 (they are denoted by ⓪ and ①, respectively) and that with two vertices and decorations 0 for both of them (denoted by ⓪ − ⓪).

Proposition 3.4. *There exists a family of nonnegative real numbers $\{\mathsf{c}(\sigma,\tau)\sigma \prec \tau\}$ such that, for every seed S, every decorated tree τ with $w(\tau) > 1$ and every $n \geq |S|$:*

$$\mathbb{E}\left[\mathcal{F}_\tau\left(T_{n+1}^S\right)|\mathcal{E}_n\right] = \left(1 + \frac{w(\tau)}{(1+\alpha)n - 2}\right) \cdot \mathcal{F}_\tau\left(T_n^S\right)$$

$$+ \frac{1}{(1+\alpha)n - 2} \sum_{\sigma \prec \tau} \mathsf{c}(\sigma,\tau) \cdot \mathcal{F}_\sigma\left(T_n^S\right), \qquad (6)$$

where $\mathcal{E}_n$ is the σ-algebra generated by $T_k^S, \ldots, T_n^S$. In addition, when:

- $\tau = \text{⓪}, \mathcal{F}_\tau(T_n) = n$;
- $\tau = \text{①}, \mathcal{F}_\tau(T_n) = n - 2$;
- $\tau = \text{⓪} - \text{⓪}, \mathcal{F}_\tau(T_n) = 2n - 2$.

It is worth mentioning that the constants $\{c(\sigma,\tau)\ \sigma \prec \tau\}$ do not depend on n or S, but do depend on the parameter α of the model.

The proof below is somewhat algebraic and does not use the planar α-PA. A more visual proof that uses the notion of decorated embedding (as described at the end of Section 2.1) may be given. It is similar to that of [8], Lem. 6, with some additional difficulties due to the colouring of corners.

Proof. Fix the seed tree S and drop it from the notation. First, let us prove the three particular cases.

- For $\tau = \text{⓪}$, then $\mathcal{F}_\tau(T_n)$ is simply the number of vertices of T_n, hence is equal to n.
- For $\tau = \text{①}$, then $\mathcal{F}_\tau(T) = \sum_{u \in T} \deg_T(u) - 1 = 2|E_T| - |V_T| = |V_T| - 2$. Hence $\mathcal{F}_\tau(T_n) = n - 2$.
- For $\tau = \text{⓪} - \text{⓪}$, then $\mathcal{F}_\tau(T) = 2|E_T|$ since the τ may be embedded over any edge of T in one of two directions. Hence $\mathcal{F}_\tau(T_n) = 2n - 2$.

Let us now prove the recurrence formula. Fix $\tau = (\tau, \ell)$ a decorated tree with $w(\tau) > 1$. For a tree T and an embedding $\phi : \tau \hookrightarrow T$, let $\pi(\tau, \phi, T) = \prod_{u \in \tau}[\deg_T(\phi(u)) - 1]_{\ell(u)}$, so that

$$\mathcal{F}_\tau(T) = \sum_{\phi:\tau \hookrightarrow T} \prod_{u \in \tau}[\deg_T(\phi(u)) - 1]_{\ell(u)} = \sum_{\phi:\tau \hookrightarrow T} \pi(\tau, \phi, T). \qquad (7)$$

Recall that, in passing from T_n to T_{n+1}, a new vertex v_n is attached to a randomly chosen vertex u_n of T_n. Our purpose is to compute the sum above over embeddings ϕ of τ in T_{n+1}. We may restrict the sum only to embeddings with $\pi(\tau, \phi, T_{n+1}) > 0$. We separate such embeddings into three categories:

1. those who do not include u_n or v_n in their image;
2. those who include u_n but not v_n in their image;
3. those who include both u_n and v_n in their image.

The embedding cannot contain v_n without u_n in its image. Indeed, we have $\deg_{T_{n+1}}(v_n) = 1$, hence, if ϕ is an embedding that maps a vertex $\mathsf{v} \in \tau$ to v_n, then for $\pi(\boldsymbol{\tau}, \phi, T_{n+1})$ to be non-zero, it is necessary that $\ell(\mathsf{v}) = 0$. By choice of τ, if such a vertex exists, it necessarily has a neighbour u, which then is mapped by ϕ onto u_n. Moreover, the vertex v needs to have a single neighbour in τ, hence needs to be a loose leaf of $\boldsymbol{\tau}$.

Write $\mathcal{F}_\tau^{(i)}(T_{n+1})$ with $i = 1, 2, 3$ for the contribution to (7) of embeddings from each of the categories above. For the first two categories, the embeddings considered are in one to one correspondence with embedding of τ in T_n (although their weights are different whether considered in T_n or T_{n+1}). Thus,

$$\mathcal{F}_\tau^{(1)}(T_{n+1}) = \sum_{\phi: \tau \hookrightarrow T_n} \pi(\boldsymbol{\tau}, \phi, T_n) \cdot \prod_{\mathsf{u} \in \tau} \mathbb{1}_{\phi(\mathsf{u}) \neq u_n} \quad \text{and}$$

$$\mathcal{F}_\tau^{(2)}(T_{n+1}) = \sum_{\phi: \tau \hookrightarrow T_n} \sum_{\mathsf{u} \in \tau} \mathbb{1}_{\phi(\mathsf{u}) = u_n} \left(\prod_{\mathsf{v} \in \tau \setminus \{\mathsf{u}\}} \left[\deg_{T_n} \phi(\mathsf{v}) - 1 \right]_{\ell(\mathsf{v})} \right) \cdot \left[\deg_{T_n} \phi(\mathsf{u}) \right]_{\ell(\mathsf{u})}.$$

Now a basic algebraic manipulation shows that $[k]_\ell = [k-1]_\ell + \ell[k-1]_{\ell-1}$ for all $k \geq 1$ and $\ell \geq 0$. (When $\ell = 0$, the second term in the RHS is not defined. However, its prefactor cancels it out, and we allow this abuse of notation.) Moreover, $\prod_{\mathsf{u} \in \tau} \mathbb{1}_{\phi(\mathsf{u}) \neq u_n} = 1 - \sum_{\mathsf{u} \in \tau} \mathbb{1}_{\phi(\mathsf{u}) = u_n}$. Hence,

$$\mathcal{F}_\tau^{(1)}(T_{n+1}) + \mathcal{F}_\tau^{(2)}(T_{n+1})$$

$$= \sum_{\phi: \tau \hookrightarrow T_n} \left[\pi(\boldsymbol{\tau}, \phi, T_n) \right.$$

$$\left. + \sum_{\mathsf{u} \in \tau} \mathbb{1}_{\phi(\mathsf{u}) = u_n} \ell(\mathsf{u}) \left(\prod_{\mathsf{v} \in \tau \setminus \{\mathsf{u}\}} \left[\deg_{T_n} \phi(\mathsf{v}) - 1 \right]_{\ell(\mathsf{v})} \right) \cdot \left[\deg_{T_n} \phi(\mathsf{u}) - 1 \right]_{\ell(\mathsf{u})-1} \right].$$

Notice that $\sum_{\phi: \tau \hookrightarrow T_n} \pi(\boldsymbol{\tau}, \phi, T_n) = \mathcal{F}_\tau(T_n)$. Moreover, recall that, for any fixed $\mathsf{u} \in \tau$ and embedding $\phi: \tau \hookrightarrow T_n$, we have $\mathbb{P}(\phi(\mathsf{u}) = u_n | \mathcal{E}_n) = \frac{\deg_{T_n} \phi(\mathsf{u}) - 1 + \alpha}{(1+\alpha)n - 2}$. Thus, when taking the expectation in the above we find

$$\mathbb{E}\left[\mathcal{F}_\tau^{(1)}(T_{n+1}) + \mathcal{F}_\tau^{(2)}(T_{n+1}) | \mathcal{E}_n \right] - \mathcal{F}_\tau(T_n)$$

$$= \sum_{\mathsf{u} \in \tau} \ell(\mathsf{u}) \cdot \sum_{\phi: \tau \hookrightarrow T_n} \left(\prod_{\mathsf{v} \in \tau \setminus \{\mathsf{u}\}} \left[\deg_{T_n} \phi(\mathsf{v}) - 1 \right]_{\ell(\mathsf{v})} \right)$$

$$\cdot \left[\deg_{T_n} \phi(\mathsf{u}) - 1 \right]_{\ell(\mathsf{u})-1} \cdot \frac{\deg_{T_n} \phi(\mathsf{u}) - 1 + \alpha}{(1+\alpha)n - 2}.$$

The sum over $\mathsf{u} \in \tau$ in the right-hand side above may be limited to vertices u with $\ell(\mathsf{u}) > 0$.

For $\mathsf{u} \in \tau$ with $\ell(\mathsf{u}) > 0$, let $\boldsymbol{\tau}^{(\mathsf{u}-)}$ be the decorated tree $(\tau, \ell^{(\mathsf{u}-)})$ with decorations identical to those of $\boldsymbol{\tau}$ except at the vertex u for which $\ell^{(\mathsf{u}-)} = \ell(\mathsf{u}) - 1$. Then $\boldsymbol{\tau}^{(\mathsf{u}-)} \prec \boldsymbol{\tau}$. Write $\deg_{T_n} \phi(\mathsf{u}) - $

$1 + \alpha = [\deg_{T_n} \phi(\mathsf{u}) - \ell(\mathsf{u})] + \ell(\mathsf{u}) + \alpha - 1$ to find that the summand in the right-hand side is

$$\frac{1}{(1+\alpha)n - 2} \Big[\pi(\boldsymbol{\tau}, \phi, T_n) + \big(\ell(\mathsf{u}) + \alpha - 1\big) \cdot \pi\big(\boldsymbol{\tau}^{\mathsf{u}-}, \phi, T_n\big) \Big].$$

Thus, we find

$$\mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}^{(1)}(T_{n+1}) + \mathcal{F}_{\boldsymbol{\tau}}^{(2)}(T_{n+1}) | \mathcal{E}_n\big]$$

$$= \mathcal{F}_{\boldsymbol{\tau}}(T_n) + \sum_{\mathsf{u} \in \tau} \frac{\ell(\mathsf{u})}{(1+\alpha)n - 2} \cdot \big[\mathcal{F}_{\boldsymbol{\tau}}(T_n) + \big(\ell(\mathsf{u}) + \alpha - 1\big) \cdot \mathcal{F}_{\boldsymbol{\tau}^{\mathsf{u}-}}(T_n)\big]$$

$$= \Big[1 + \frac{|\ell|}{(1+\alpha)n - 2}\Big] \mathcal{F}_{\boldsymbol{\tau}}(T_n) + \sum_{\mathsf{u} \in \tau} \frac{\ell(\mathsf{u})(\ell(\mathsf{u}) + \alpha - 1)}{(1+\alpha)n - 2} \cdot \mathcal{F}_{\boldsymbol{\tau}^{\mathsf{u}-}}(T_n). \tag{8}$$

Finally, we turn to $\mathcal{F}_{\boldsymbol{\tau}}^{(3)}(T_{n+1})$. Let $\phi : \tau \hookrightarrow T_{n+1}$ be an embedding contributing to $\mathcal{F}_{\boldsymbol{\tau}}^{(3)}(T_{n+1})$, let v be the loose leaf mapped to v_{n+1} and u its only neighbour in τ. Define the following three modifications of $\boldsymbol{\tau}$:

- $\boldsymbol{\tau} \setminus \mathsf{v}$ is the decorated tree obtained from $\boldsymbol{\tau}$ by removing the leaf v and conserving the same decorations for all other vertices;
- $(\boldsymbol{\tau} \setminus \mathsf{v})^-$ is the decorated tree obtained from $\boldsymbol{\tau}$ by removing the leaf v, decreasing the decoration of u by one, and conserving the same decorations for all other vertices (this is only defined when $\ell(\mathsf{u}) > 0$; it will implicitly only appear in such cases in the upcoming formulas);
- $(\boldsymbol{\tau} \setminus \mathsf{v})^+$ is the decorated tree obtained from $\boldsymbol{\tau}$ by removing the leaf v, increasing the decoration of u by one, and conserving the same decorations for all other vertices.

It is immediate to check that all trees above are smaller than $\boldsymbol{\tau}$ for the order $\prec$.

Write $\tau \setminus \mathsf{v}$ for the tree (stripped of decoration) of all of the above. To ϕ associate its restriction $\tilde{\phi} : \tau \setminus \mathsf{v} \to T_n$ to $\tau \setminus \mathsf{v}$. Then, by the same type of computation as above

$$\pi(\boldsymbol{\tau}, \phi, T_{n+1}) = \Big(\prod_{\mathsf{w} \in \tau \setminus \mathsf{v}} \big[\deg_{T_{n+1}} \phi(\mathsf{w}) - 1\big]_{\ell(\mathsf{w})} \Big)$$

$$= \Big(\prod_{\mathsf{v} \in \tau \setminus \{\mathsf{u}, \mathsf{v}\}} \big[\deg_{T_n} \phi(\mathsf{v}) - 1\big]_{\ell(\mathsf{v})} \Big)$$

$$\cdot \big([\deg_{T_n} \phi(\mathsf{u}) - 1]_{\ell(\mathsf{u})} + \ell(\mathsf{u})[\deg_{T_n} \phi(\mathsf{u}) - 1]_{\ell(\mathsf{u})-1}\big). \tag{9}$$

In the first line, since $\ell(\mathsf{v}) = 0$, we removed the term coming from v from the product. The quantity above will be weighted by $\mathbb{P}(\phi(\mathsf{u}) = u_n | \mathcal{E}_n) = \frac{\deg_{T_n} \phi(\mathsf{u}) - 1 + \alpha}{(1+\alpha)n - 2}$. In preparation, observe that

$$(d - 1 + \alpha) \cdot [d - 1]_{\ell(\mathsf{u})} = [d - 1]_{\ell(\mathsf{u})+1} + \big(\ell(\mathsf{u}) + \alpha\big)[d - 1]_{\ell(\mathsf{u})} \quad \text{and}$$

$$(d - 1 + \alpha) \cdot [d - 1]_{\ell(\mathsf{u})-1} = [d - 1]_{\ell(\mathsf{u})} + \big(\ell(\mathsf{u}) - 1 + \alpha\big)[d - 1]_{\ell(\mathsf{u})-1}.$$

Applying the above with $d = \deg_{T_n}(\mathsf{u})$ to (9), multiplied by $\mathbb{P}(\phi(\mathsf{u}) = u_n | \mathcal{E}_n)$, we find

$$\mathbb{P}\big(\phi(\mathsf{u}) = u_n | \mathcal{E}_n\big) \cdot \pi(\boldsymbol{\tau}, \phi, T_{n+1})$$

$$= \frac{1}{(1+\alpha)n - 2} \cdot \Big[\pi\big((\boldsymbol{\tau} \setminus \mathsf{v})^+, \tilde{\phi}, T_n\big) + \big(2\ell(\mathsf{u}) + \alpha\big) \cdot \pi(\boldsymbol{\tau} \setminus \mathsf{v}, \tilde{\phi}, T_n)$$

$$+ \ell(\mathsf{u})\big(\ell(\mathsf{u}) + \alpha - 1\big) \cdot \pi\big((\boldsymbol{\tau} \setminus \mathsf{v})^-, \tilde{\phi}, T_n\big)\Big].$$

Summing over all embeddings and all values of u, v we find

$$\mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}^{(3)}(T_{n+1}) | \mathcal{E}_n\big]$$

$$= \frac{1}{(1+\alpha)n - 2} \sum_{\mathsf{v} \text{ loose leaf}} \mathcal{F}_{(\boldsymbol{\tau} \setminus \mathsf{v})^+}(T_n)$$

$$+ \big(2\ell(\mathsf{u}) + \alpha\big) \cdot \mathcal{F}_{\boldsymbol{\tau} \setminus \mathsf{v}}(T_n) + \ell(\mathsf{u})\big(\ell(\mathsf{u}) + \alpha - 1\big) \cdot \mathcal{F}_{(\boldsymbol{\tau} \setminus \mathsf{v})^-}(T_n), \tag{10}$$

where the sum in the right-hand side is over all loose leaves v of $\boldsymbol{\tau}$ and u denotes their unique neighbour.

By summing (8) and (10), we may obtain a recurrence formula similar to (6), but with one flaw. Indeed, in such a formula the trees $\boldsymbol{\sigma}$ would potentially be of the type $(\boldsymbol{\tau} \setminus \mathsf{v})^+$, hence have same weight as $\boldsymbol{\tau}$. We reduce the contribution of such trees to ones of lower weight via the following lemma.

Lemma 3.5. *Let $\boldsymbol{\tau} = (\tau, \ell)$ be a decorated tree with $w(\boldsymbol{\tau}) \geq 2$, let v be a loose leave of $\boldsymbol{\tau}$ and u its only neighbour in τ. Then, for any tree T:*

$$\mathcal{F}_{(\boldsymbol{\tau} \setminus \mathsf{v})^+}(T) = \mathcal{F}_{\boldsymbol{\tau}}(T) + \big(\deg_{\tau}(\mathsf{u}) - \ell(\mathsf{u}) - 2\big) \cdot \mathcal{F}_{\boldsymbol{\tau} \setminus \mathsf{v}}(T). \tag{11}$$

Proof. [Proof of Lemma 3.5] We proceed in two steps. Let us first express $\mathcal{F}_{\boldsymbol{\tau}}(T)$ using $\mathcal{F}_{\boldsymbol{\tau} \setminus \mathsf{v}}(T)$. To any $\phi : \tau \hookrightarrow T$, associate its restriction $\tilde{\phi} : \tau \setminus \mathsf{v} \to T$ to $\tau \setminus \mathsf{v}$. Conversely, any tree embedding $\tilde{\psi} : \tau \setminus \mathsf{v} \hookrightarrow T$ may be extended to some embedding $\psi : \tau \hookrightarrow T$ in as many different ways as there are neighbouring vertices of $\tilde{\psi}(\mathsf{u})$ not reached by ψ, that is to say $\deg_T(\tilde{\psi}(\mathsf{u})) - \deg_{\tau \setminus \mathsf{v}}(\mathsf{u})$ ways. Therefore:

$$\mathcal{F}_{\boldsymbol{\tau}}(T) = \sum_{\tilde{\psi} : \tau \setminus \mathsf{v} \to T} \big(\deg_T\big(\tilde{\psi}(\mathsf{u})\big) - \deg_{\tau \setminus \mathsf{v}}(\mathsf{u})\big) \cdot \pi(\boldsymbol{\tau} \setminus \mathsf{v}, \tilde{\psi}, T). \tag{12}$$

In the above equation, we further use that $\pi(\boldsymbol{\tau}, \psi, T) = \pi(\boldsymbol{\tau} \setminus \mathsf{v}, \tilde{\psi}, T)$ given that v is a loose leaf.

Next we express $\mathcal{F}_{(\boldsymbol{\tau} \setminus \mathsf{v})^+}(T)$ in terms of $\mathcal{F}_{\boldsymbol{\tau} \setminus \mathsf{v}}(T)$. Recall that $(\boldsymbol{\tau} \setminus \mathsf{v})^+$ is obtained from $\boldsymbol{\tau} \setminus \mathsf{v}$ by increasing the decoration of the vertex u by one. Thus,

$$\mathcal{F}_{(\boldsymbol{\tau} \setminus \mathsf{v})^+}(T) = \sum_{\tilde{\psi} : \tau \setminus \mathsf{v} \to T} \big(\deg_T\big(\tilde{\psi}(\mathsf{u})\big) - 1 - \ell(\mathsf{u})\big) \cdot \pi(\boldsymbol{\tau} \setminus \mathsf{v}, \tilde{\psi}, T). \tag{13}$$

Writing $\deg_T(\tilde{\psi}(\mathsf{u})) - 1 - \ell(u) = [\deg_T(\tilde{\psi}(\mathsf{u})) - \deg_{\tau\backslash\mathsf{v}}(\mathsf{u})] + [\deg_{\tau\backslash\mathsf{v}}(\mathsf{u}) - 1 - \ell(u)]$ and observing that $\deg_{\tau\backslash\mathsf{v}}(\mathsf{u}) = \deg_\tau(\mathsf{u}) - 1$, we deduce Lemma 3.5 by subtracting (12) from (13). □

By Lemma 3.5, equation (10) may be re-expressed as

$$\mathbb{E}\big[\mathcal{F}_\tau^{(3)}(T_{n+1})|\mathcal{E}_n\big] = \frac{1}{(1+\alpha)n - 2} \sum_{\mathsf{v}\text{ loose leaf}} \mathcal{F}_\tau(T_n) + \big(\deg_\tau(\mathsf{u}) + \ell(\mathsf{u}) + \alpha - 2\big) \cdot \mathcal{F}_{\tau\backslash\mathsf{v}}(T_n)$$

$$+ \ell(\mathsf{u})\big(\ell(\mathsf{u}) + \alpha - 1\big) \cdot \mathcal{F}_{(\tau\backslash\mathsf{v})^-}(T_n). \tag{14}$$

Since $w(\tau) > 1$, the vertex u may never be a loose leaf of τ, hence $\deg_\tau(\mathsf{u}) \geq 2$ or $\deg_\tau(\mathsf{u}) = 1$ but $\ell(\mathsf{u}) \geq 1$. In both cases, the multiplicative factor $\deg_\tau(\mathsf{u}) + \ell(\mathsf{u}) + \alpha - 2$ is non-negative. Equations (8) and (14) together yield (6). □

Remark 3.6. Following the proof we find that the constants $c(\sigma, \tau)$ appearing in (6) are nonzero only if σ may be obtained from τ by

  (i) decreasing the value of one decoration by 1 or
 (ii) removing a loose leaf and conserving all other decorations or
(iii) removing a loose leaf and modifying the decoration of its unique neighbour by -1.

It is direct that all the trees σ are smaller than τ for $\prec$. Indeed, in most cases we have $w(\sigma) < w(\tau)$. However, there are two cases where $w(\sigma) = w(\tau)$: when the decoration of a leaf is 1 in τ and decreases to 0 in σ (by the procedure (i)) and when a loose leaf is removed from τ (as in (ii)), with its unique ancestor having decoration 0, hence becoming a loose leaf of σ. In both these cases, $\sigma \prec \tau$ due to the second condition of Definition 3.3.

3.2. The first moment of $\mathcal{F}_\tau(T_n)$

We are ready to state the full estimate of the first moment of $\mathcal{F}_\tau(T_n^S)$.

Theorem 3.7. *For any $\alpha > 0$, any seed tree S of size $k \geq 2$ and any decorated tree τ, we have*:

$$\mathbb{E}\big[\mathcal{F}_\tau\big(T_n^S\big)\big] \approx n^{\max\{1, \frac{w(\tau)}{1+\alpha}\}} \cdot (\log n)^{\gamma(\tau)},$$

where $\gamma(\tau)$ is a nonnegative exponent equal to zero when $w(\tau) < 1 + \alpha$ and otherwise recursively defined by:

$$\gamma(\tau) = \sup_{\sigma\prec\tau, c(\sigma,\tau)>0, w(\sigma)=w(\tau)} \big(\gamma(\sigma) + 1\big) \quad \text{if } w(\tau) > 1 + \alpha,$$

or by (critical case):

$$\gamma(\tau) = \gamma_c(\tau) := \max\bigg\{1, \sup_{\sigma\prec\tau, c(\sigma,\tau)>0, w(\sigma)=w(\tau)} \big(\gamma_c(\sigma) + 1\big)\bigg\} \quad \text{if } w(\tau) = 1 + \alpha$$

with the convention $\sup \varnothing = 0$.

Remark 3.8. Let $\boldsymbol{\tau}$ be a decorated tree with $\ell(u) \geq 2$ for all its leaves. In particular, it contains no loose leaves and $w(\boldsymbol{\tau}) = |\ell| = \sum_{u \in \tau} \ell(u)$. Then Remark 3.6 indicates that $c(\boldsymbol{\sigma}, \boldsymbol{\tau}) > 0$ only for decorated trees $\boldsymbol{\sigma}$ obtained by lowering the decoration of some vertex $u \in \boldsymbol{\tau}$ by 1. Moreover, no such tree $\boldsymbol{\sigma}$ has any loose leaf either. Assuming in addition that $w(\boldsymbol{\tau}) = |\ell| > 1 + \alpha$, we have $w(\boldsymbol{\sigma}) < w(\boldsymbol{\tau})$ and $\gamma(\boldsymbol{\tau}) = 0$.

To prove Theorem 3.7, we proceed by induction on the set of decorated trees (for the partial order $\preceq$) and use the recurrence formula (6).

Proof of Theorem 3.7. We first note that for the three decorated trees $\boldsymbol{\tau}$ with $w(\boldsymbol{\tau}) = 1$ (thus $w(\boldsymbol{\tau}) < 1 + \alpha$ for any $\alpha > 0$) the first moment estimate is satisfied as indicated by the explicit formula of Proposition 3.4.

Consider now a decorated tree $\boldsymbol{\tau}$ with $w(\boldsymbol{\tau}) \geq 2$ and suppose by induction that Theorem 3.7 is valid for all decorated trees $\boldsymbol{\sigma} \prec \boldsymbol{\tau}$. Let us define for any $n \geq k$ the quantity $\omega_{n+1}^{(\tau)}$ equal to:

$$\omega_{n+1}^{(\tau)} := \prod_{\ell=k}^{n} \left(1 + \frac{w(\boldsymbol{\tau})}{(1+\alpha)\ell - 2} \right)^{-1} \approx n^{-\frac{w(\tau)}{1+\alpha}}, \tag{15}$$

where the latter equivalent is obtained by a standard computation. Then, by multiplying (6) by this factor and taking the expectation, we get:

$$\mathbb{E}\big[\omega_{n+1}^{(\tau)} \cdot \mathcal{F}_{\boldsymbol{\tau}}(T_{n+1})\big] = \mathbb{E}\big[\omega_{n}^{(\tau)} \cdot \mathcal{F}_{\boldsymbol{\tau}}(T_n)\big] + \sum_{\boldsymbol{\sigma} \prec \boldsymbol{\tau}} c(\boldsymbol{\sigma}, \boldsymbol{\tau}) \cdot \frac{\omega_{n+1}^{(\tau)}}{(1+\alpha)n - 2} \cdot \mathbb{E}\big[\mathcal{F}_{\boldsymbol{\sigma}}(T_n)\big].$$

By iterating the above over n, we find

$$\mathbb{E}\big[\omega_{n+1}^{(\tau)} \cdot \mathcal{F}_{\boldsymbol{\tau}}(T_{n+1})\big] = \mathcal{F}_{\boldsymbol{\tau}}(S) + \sum_{\boldsymbol{\sigma} \prec \boldsymbol{\tau}} c(\boldsymbol{\sigma}, \boldsymbol{\tau}) \sum_{\ell=k}^{n} \frac{\omega_{\ell+1}^{(\tau)}}{(1+\alpha)\ell - 2} \cdot \mathbb{E}\big[\mathcal{F}_{\boldsymbol{\sigma}}(T_\ell)\big]. \tag{16}$$

Thus, the asymptotic behaviour of $\mathbb{E}[\mathcal{F}_{\boldsymbol{\tau}}(T_{n+1})]$ can be derived from that of $\mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_\ell)]$ for $\boldsymbol{\sigma} \prec \boldsymbol{\tau}$ and $\ell \leq n$. Define the following variables

$$\mathcal{S}_n(\boldsymbol{\sigma}, \boldsymbol{\tau}) := \sum_{\ell=k}^{n} \frac{\omega_{\ell+1}^{(\tau)}}{(1+\alpha)\ell - 2} \cdot \mathbb{E}\big[\mathcal{F}_{\boldsymbol{\sigma}}(T_\ell)\big].$$

The growth rate in n of $\mathcal{S}_n(\boldsymbol{\sigma}, \boldsymbol{\tau})$ depends on $w(\boldsymbol{\sigma})$ through $\mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_\ell)]$. We distinguish three cases according to the value of $w(\boldsymbol{\tau})$.

> **1st case:** $w(\boldsymbol{\tau}) < 1 + \alpha$

In this situation, we necessarily have $w(\boldsymbol{\sigma}) < 1 + \alpha$ for any $\boldsymbol{\sigma} \prec \boldsymbol{\tau}$. Thus, by the induction hypothesis:

$$\mathbb{E}\big[\mathcal{F}_{\boldsymbol{\sigma}}(T_\ell)\big] \approx \ell.$$

The above together with (15) imply that

$$\frac{\omega_{\ell+1}^{(\tau)}}{(1+\alpha)\ell - 2} \cdot \mathbb{E}\big[\mathcal{F}_\sigma(T_\ell)\big] \approx \ell^{-\frac{w(\tau)}{1+\alpha}}.$$

Since $w(\tau) < 1 + \alpha$, the sum over ℓ of the above – which constitutes $\mathcal{S}_n(\sigma, \tau)$ – diverges at rate:

$$\mathcal{S}_n(\sigma, \tau) = \sum_{\ell=k}^{n} \frac{1}{(1+\alpha)\ell - 2} \cdot \omega_{\ell+1}^{(\tau)} \cdot \mathbb{E}\big[\mathcal{F}_\sigma(T_\ell)\big] \approx n^{1-\frac{w(\tau)}{1+\alpha}}.$$

We then sum over every $\sigma \prec \tau$ to get an asymptotic estimate for the quantity on the right of (16):

$$\mathbb{E}\big[\omega_{n+1}^{(\tau)} \cdot \mathcal{F}_\tau(T_{n+1})\big] = \mathcal{F}_\tau(S) + \sum_{\sigma \prec \tau} c(\sigma, \tau) \sum_{\ell=k}^{n} \frac{\omega_{\ell+1}^{(\tau)}}{(1+\alpha)\ell - 2} \cdot \mathbb{E}\big[\mathcal{F}_\sigma(T_\ell)\big] \approx n^{1-\frac{w(\tau)}{1+\alpha}}, \quad (17)$$

since there exists at least one σ with $c(\sigma, \tau) > 0$. Finally, dividing (17) by $\omega_{n+1}^{(\tau)}$ and using (15), we obtain the expected estimate:

$$\mathbb{E}\big[\mathcal{F}_\tau(T_{n+1})\big] \approx n.$$

2nd case: $w(\tau) = 1 + \alpha$

Now, when $\sigma \prec \tau$, we can either have $w(\sigma) < 1 + \alpha$ or $w(\sigma) = 1 + \alpha$. In the former situation, we have

$$\frac{1}{(1+\alpha)n - 2} \cdot \omega_{n+1}^{(\tau)} \cdot \mathbb{E}\big[\mathcal{F}_\sigma(T_n)\big] \approx n^{-1},$$

hence,

$$\mathcal{S}_n(\sigma, \tau) = \sum_{\ell=k}^{n} \frac{1}{(1+\alpha)\ell - 2} \cdot \omega_{\ell+1}^{(\tau)} \cdot \mathbb{E}\big[\mathcal{F}_\sigma(T_\ell)\big] \approx \log(n).$$

In the latter situation, by the induction hypothesis,

$$\mathbb{E}\big[\mathcal{F}_\sigma(T_{n+1})\big] \approx n \cdot (\log n)^{\gamma_c(\sigma)}.$$

Using (15), we find,

$$\frac{1}{(1+\alpha)n - 2} \cdot \omega_{n+1}^{(\tau)} \cdot \mathbb{E}\big[\mathcal{F}_\sigma(T_n)\big] \approx n^{-\frac{w(\tau)}{1+\alpha}} \cdot (\log n)^{\gamma_c(\sigma)} = n^{-1} \cdot (\log n)^{\gamma_c(\sigma)}.$$

As a consequence, $\mathcal{S}_n(\sigma, \tau)$ diverges at rate:

$$\mathcal{S}_n(\sigma, \tau) \approx \sum_{\ell=k}^{n} \ell^{-1} \cdot (\log \ell)^{\gamma_c(\sigma)} \approx (\log n)^{\gamma_c(\sigma)+1}.$$

Thus, only the terms σ with maximal weight and $c(\sigma, \tau) > 0$ contribute to (16) significantly:

$$\mathbb{E}\big[\omega_{n+1}^{(\tau)} \cdot \mathcal{F}_\tau(T_{n+1})\big] \approx \sum_{\substack{\sigma \prec \tau \\ w(\sigma) < 1+\alpha}} c(\sigma, \tau) \cdot \log n + \sum_{\substack{\sigma \prec \tau \\ w(\sigma) = 1+\alpha}} c(\sigma, \tau) \cdot (\log n)^{\gamma_c(\sigma)+1}$$

$$\approx (\log n)^{\gamma_c(\tau)}.$$

where $\gamma_c(\tau)$ is defined as in Theorem 3.7. Dividing the last equation by $\omega_{n+1}^{(\tau)}$ leads to the result.

3rd case: $w(\tau) > 1 + \alpha$

In this case, the trees $\sigma \prec \tau$ can satisfy either $w(\sigma) < 1+\alpha$, $w(\sigma) = 1+\alpha$, $1 + \alpha < w(\sigma) < w(\tau)$, or $1 + \alpha < w(\sigma) = w(\tau)$. In the two first situations, by the induction hypothesis, there exists $\delta > 0$ such that:

$$\frac{1}{(1+\alpha)\ell - 2} \cdot \omega_{\ell+1}^{(\tau)} \cdot \mathbb{E}\big[\mathcal{F}_\sigma(T_\ell)\big] \ll \ell^{-\frac{w(\tau)}{1+\alpha}} \cdot (\log \ell)^\delta.$$

Since $w(\tau) > 1 + \alpha$, the sum over ℓ of the above converges and $\mathcal{S}_n(\sigma, \tau) \approx 1$.

When $1 + \alpha < w(\sigma) < w(\tau)$, the induction hypothesis implies:

$$\frac{1}{(1+\alpha)\ell - 2} \cdot \omega_{\ell+1}^{(\tau)} \cdot \mathbb{E}\big[\mathcal{F}_\sigma(T_\ell)\big] \approx \ell^{\frac{w(\sigma)-w(\tau)}{1+\alpha}-1} \cdot (\log \ell)^{\gamma(\sigma)}.$$

Since $w(\sigma) < w(\tau)$, the sum over ℓ of the above converges again, and $\mathcal{S}_n(\sigma, \tau) \approx 1$.

Finally, when $w(\sigma) = w(\tau)$, the induction hypothesis gives us:

$$\frac{1}{(1+\alpha)\ell - 2} \cdot \omega_{\ell+1}^{(\tau)} \cdot \mathbb{E}\big[\mathcal{F}_\sigma(T_\ell)\big] \approx \ell^{-1} \cdot (\log \ell)^{\gamma(\sigma)}.$$

Consequently, by a direct computation, the divergence rate of $\mathcal{S}_n(\sigma, \tau)$ may be shown to be:

$$\mathcal{S}_n(\sigma, \tau) \approx \sum_{\ell=k}^n \ell^{-1} \cdot (\log \ell)^{\gamma(\sigma)} \approx (\log n)^{\gamma(\sigma)+1}.$$

In conclusion, by considering the asymptotic of $\mathcal{S}_n(\sigma, \tau)$ for all $\sigma \prec \tau$ with $c(\sigma, \tau) > 0$ according to the above, we obtain

$$\mathbb{E}\big[\omega_{n+1}^{(\tau)} \cdot \mathcal{F}_\tau(T_{n+1})\big] \approx \sum_{\substack{\sigma \prec \tau \\ w(\sigma) < w(\tau)}} c(\sigma, \tau) + \sum_{\substack{\sigma \prec \tau \\ w(\sigma) = w(\tau)}} c(\sigma, \tau) \cdot (\log n)^{\gamma(\sigma)+1} \approx (\log n)^{\gamma(\tau)},$$

where the last equivalent is due to how $\gamma(\tau)$ is defined in Theorem 3.7. Divide by $\omega_{n+1}^{(\tau)}$ to obtain the expected result. $\qquad\square$

3.3. The second moment of $\mathcal{F}_\tau(T_n)$

We are now ready to prove the second moment estimate on $\mathcal{F}_\tau(T_n)$ of Theorem 3.1. We will build on the analogous result on the first moment obtained in the previous section as well as on its proof. First, remark that the square of the observables may be written as:

$$\mathcal{F}_\tau(T_n)^2 = \sum_{\phi_1,\phi_2:\tau \hookrightarrow T_n} \pi(\tau,\phi_1,T_n) \cdot \pi(\tau,\phi_2,T_n), \tag{18}$$

where the sum is this time over all pairs ϕ_1, ϕ_2 of graph embeddings of τ in T_n. We will decompose the sum appearing in (18) in two parts, according to whether the embeddings $\phi_1(\tau)$ and $\phi_2(\tau)$ overlap or not. The same is done in [8]. Call $\mathcal{F}_{\tau,\tau}(T_n)$ the first resulting quantity and $\mathcal{F}_{\tau+\tau}(T_n)$ the second one:

$$\mathcal{F}_{\tau,\tau}(T_n) = \sum_{\substack{\phi_1,\phi_2:\tau \to T_n \\ \phi_1(\tau) \cap \phi_2(\tau)=\varnothing}} \pi(\tau,\phi_1,T_n)\pi(\tau,\phi_2,T_n),$$

$$\mathcal{F}_{\tau+\tau}(T_n) = \sum_{\substack{\phi_1,\phi_2:\tau \to T_n \\ \phi_1(\tau) \cap \phi_2(\tau)\neq\varnothing}} \pi(\tau,\phi_1,T_n)\pi(\tau,\phi_2,T_n).$$

The first moments of the two quantities above are bounded separately. First, we estimate $\mathbb{E}[\mathcal{F}_{\tau+\tau}(T_n)]$, which turns out to be the the easiest of the two. Indeed, it may be expressed as a sum of first moments of observables for some decorated trees derived from τ. These are computed using Theorem 3.7. To deal with $\mathbb{E}[\mathcal{F}_{\tau,\tau}(T_n)]$ we will prove a recurrence inequality on such quantities, similar to (6) and using the same techniques.

3.3.1. *An estimate on $\mathbb{E}[\mathcal{F}_{\tau+\tau}(T_n)]$*

The goal of this section is to show the following.

Proposition 3.9. *Let τ be a decorated tree with $\ell(u) \geq 2$ for any $u \in \tau$ and such that $|\ell| = \sum_{u \in \tau} \ell(u) > 1 + \alpha$. Then, for any seed S,*

$$\mathbb{E}\big[\mathcal{F}_{\tau+\tau}\big(T_n^S\big)\big] = \mathcal{O}\big(n^{\frac{2|\ell|}{1+\alpha}}\big). \tag{19}$$

The proposition is based on the following lemma, which we prove below.

Lemma 3.10. *Let $\tau = (\tau,\ell)$ be a decorated tree. There exists a finite set $\mathcal{U}(\tau)$ of decorated trees σ with $w(\sigma) \leq 2w(\tau)$ and positive constants $C(\tau,\sigma)$ for $\sigma \in \mathcal{U}(\tau)$ (see the proof for an explicit description) such that, for any tree T,*

$$\mathcal{F}_{\tau+\tau}(T) = \sum_{\sigma \in \mathcal{U}(\tau)} C(\tau,\sigma) \cdot \mathcal{F}_\sigma(T). \tag{20}$$

In the planar setting the above is very intuitive; we sketch a proof below. If T is plane, then $\mathcal{F}_{\tau+\tau}(T)$ is the number of decorated embeddings (ϕ_1, ϕ_2) of two copies τ_1 and τ_2 of τ in T, which overlap.

Let us first forget about the decorations and focus on graph embeddings. The union of the images of τ_1 and τ_2 via such embeddings is a tree σ; one may see σ as a merger of τ_1 and τ_2. Thus, the pairs of overlapping embeddings of τ_1 and τ_2 in T are in bijection with the embeddings of σ in T where σ ranges over all possible mergers of τ_1 and τ_2.

Now consider overlapping decorated embeddings of τ_1 and τ_2 in T. Then, each corner of T may have no arrow pointing to it, an arrow of τ_1, an arrow of τ_2 or one arrow of τ_1 and one of τ_2 pointing to it. In the first three cases, the arrows pointing to the corner will be considered as arrows of σ; in the last case, the arrow of τ_1 and that of τ_2 merge into a single arrow of σ. Thus, any such pair of decorated embeddings corresponds to a decorated embedding of some decorated tree $\boldsymbol{\sigma} = (\sigma, m)$ obtained as a merger of $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$. In particular $|m| \leq |\ell_1| + |\ell_2| = 2|\ell|$. The constants $C(\boldsymbol{\tau}, \boldsymbol{\sigma})$ are combinatorial factors that account for the different ways to merge arrows of $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$.

The actual proof given below avoids the use of the planar model and is more algebraic.

Proof of Lemma 3.10. Fix $\boldsymbol{\tau}$ and T as in the lemma. Consider two embeddings Φ_1, Φ_2 of τ in T whose images intersect. Since T is a tree, $\Phi_1(\tau) \cup \Phi_2(\tau)$ is itself a subtree of T. Moreover, $\Phi_1(\tau) \cap \Phi_2(\tau)$ is the image via Φ_1 and Φ_2, respectively, of two isomorphic subtrees σ_1 and σ_2 of τ.

For two isomorphic subtrees σ_1 and σ_2 of τ, define the (σ_1, σ_2)-merger of two copies of τ as the tree obtained by "gluing" two copies of τ along σ_1 and σ_2, respectively. Write $\mathcal{M}(\sigma_1, \sigma_2)$ for this tree. To identify the two copies of τ merged to obtain $\mathcal{M}(\sigma_1, \sigma_2)$, call them $\boldsymbol{\tau}_1 = (\tau_1, \ell_1)$ and $\boldsymbol{\tau}_2 = (\tau_2, \ell_2)$.

Each vertex of $\mathcal{M}(\sigma_1, \sigma_2)$ is clearly identified to either one vertex in $\tau_1 \setminus \sigma_1$, a vertex in $\tau_2 \setminus \sigma_2$ or to a vertex in σ_1 and simultaneously to one in σ_2. For a vertex $u \in \mathcal{M}(\sigma_1, \sigma_2)$ write $\ell_1(u)$ for its decoration in $\boldsymbol{\tau}_1$, if it is identified to a vertex of τ_1, otherwise set $\ell_1(u) = 0$. Define $\ell_2(u)$ for $u \in \mathcal{M}(\sigma_1, \sigma_2)$ in the same way.

Then, the pairs of embeddings Φ_1, Φ_2 of τ in T with $\Phi_1(\tau) \cap \Phi_2(\tau) = \Phi_1(\sigma_1) = \Phi_2(\sigma_2)$ are in bijection with the embeddings of $\mathcal{M}(\sigma_1, \sigma_2)$ in T. It follows that

$$\mathcal{F}_{\boldsymbol{\tau}+\boldsymbol{\tau}}(T) = \sum_{(\sigma_1, \sigma_2)} \sum_{\phi: \mathcal{M}(\sigma_1, \sigma_2) \hookrightarrow T} \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} \left[\deg_T\left(\phi(u)\right) - 1\right]_{\ell_1(u)}$$
$$\cdot \left[\deg_T\left(\phi(u)\right) - 1\right]_{\ell_2(u)}, \tag{21}$$

where the first sum is over all pairs of isomorphic subtrees (σ_1, σ_2) of τ. To reduce the above to a formula of the type (3), we use the following combinatorial identity.

Fact 3.11. Fix non-negative integers n, ℓ_1, ℓ_2. Then

$$[n]_{\ell_1} \cdot [n]_{\ell_2} = \sum_{j=0}^{\min\{\ell_1, \ell_2\}} [n]_{\ell_1 + \ell_2 - j} \cdot \frac{\ell_1! \cdot \ell_2!}{(\ell_1 - j)!(\ell_2 - j)!j!}.$$

Proof. First, observe that $[n]_\ell = \binom{n}{\ell} \cdot \ell!$. Now let us inspect the quantity $\binom{n}{\ell_1}\binom{n}{\ell_2}$ which is the number of pairs of subsets A_1, A_2 of $\{1, \ldots, n\}$ with ℓ_1 and ℓ_2 elements, respectively. These may be counted as follows. First establish the set $A_1 \cup A_2$ which can have $\ell_1 + \ell_2 - j$ elements with $j \in \{0, \ldots, \min\{\ell_1, \ell_2\}\}$. Once $A_1 \cup A_2$ is chosen, split it into $A_1 \setminus A_2$, $A_2 \setminus A_1$ and $A_1 \cap A_2$. This forms a partition of $A_1 \cup A_2$ into three sets of cardinality $\ell_1 - j$, $\ell_2 - j$ and j, respectively. Thus,

$$\binom{n}{\ell_1} \cdot \binom{n}{\ell_2} = \sum_{j=0}^{\min\{\ell_1, \ell_2\}} \binom{n}{\ell_1 + \ell_2 - j} \frac{(\ell_1 + \ell_2 - j)!}{(\ell_1 - j)!(\ell_2 - j)!j!}.$$

Multiply by $\ell_1! \cdot \ell_2!$ to find the desired result. $\qquad\square$

Let us return to the proof of Lemma 3.10. A valid decoration m for a tree in $\mathcal{M}(\sigma_1, \sigma_2)$ is one such that, for all $u \in \mathcal{M}(\sigma_1, \sigma_2)$,

$$\max\{\ell_1(u), \ell_2(u)\} \le m(u) \le \ell_1(u) + \ell_2(u). \tag{22}$$

Observe that, if u is not in the images of σ_1 and σ_2, then $m(u)$ is entirely determined by the above. For vertices than are in the overlap of τ_1 and τ_2, $m(u)$ may take one of several values.

Applying Fact 3.11 to the summand in (21), we find

$$\prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} \left[\deg_T\big(\phi(u)\big) - 1\right]_{\ell_1(u)} \cdot \left[\deg_T\big(\phi(u)\big) - 1\right]_{\ell_2(u)}$$

$$= \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} \sum_{j=0}^{\min\{\ell_1(u), \ell_2(u)\}} \left[\deg_T\big(\phi(u)\big) - 1\right]_{\ell_1(u) + \ell_2(u) - j} \cdot \frac{\ell_1(u)! \cdot \ell_2(u)!}{(\ell_1(u) - j)!(\ell_2(u) - j)!j!}$$

$$= \sum_m \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} \frac{[\deg_T(\phi(u)) - 1]_{m(u)} \cdot \ell_1(u)! \cdot \ell_2(u)!}{(m(u) - \ell_2(u))!(m(u) - \ell_1(u))!(\ell_1(u) + \ell_2(u) - m(u))!}$$

$$= \sum_m C\big(\tau, \big(\mathcal{M}(\sigma_1, \sigma_2), m\big)\big) \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} \left[\deg_T\big(\phi(u)\big) - 1\right]_{m(u)},$$

where the sum in the last two terms is over all valid decorations m of $\mathcal{M}(\sigma_1, \sigma_2)$ and

$$C\big(\tau, \big(\mathcal{M}(\sigma_1, \sigma_2), m\big)\big)$$
$$:= \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} \frac{\ell_1(u)! \cdot \ell_2(u)!}{(m(u) - \ell_2(u))!(m(u) - \ell_1(u))!(\ell_1(u) + \ell_2(u) - m(u))!}.$$

Write $\mathcal{M}$ for the tree $\mathcal{M}(\sigma_1, \sigma_2)$ with decoration m. Then

$$\sum_{\phi:\mathcal{M}(\sigma_1, \sigma_2) \hookrightarrow T} \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} \left[\deg_T\big(\phi(u)\big) - 1\right]_{m(u)} = \mathcal{F}_{\mathcal{M}}(T).$$

Inserting this into (21), we find

$$\mathcal{F}_{\tau+\tau}(T) = \sum_{\mathcal{M}} C(\tau, \mathcal{M}) \mathcal{F}_{\mathcal{M}}(T),$$

where the sum is over all trees of the form $\mathcal{M}(\sigma_1, \sigma_2)$ with valid decorations m. These form the set $\mathcal{U}(\tau)$; it is immediate that they have weight at most $2w(\tau)$. $\qquad\square$

We are finally ready to prove Proposition 3.9

Proof of Proposition 3.9. Fix a decorated tree as in the statement. By Theorem 3.7, for any tree $\sigma \in \mathcal{U}(\tau)$ of weight strictly smaller than $2w(\tau)$,

$$\mathbb{E}\big[\mathcal{F}_{\sigma}\big(T_n^S\big)\big] \ll n^{\frac{2w(\tau)}{1+\alpha}}.$$

Moreover, if $\sigma = (\sigma, m) \in \mathcal{U}(\tau)$ is such that $w(\sigma) = 2w(\tau)$ then its decorations satisfy $m(u) = \ell_1(u) + \ell_2(u) \geq 2$ for all $u \in \sigma$ (see (22) for how m is defined). As explained in Remark 3.8, for any such tree

$$\mathbb{E}\big[\mathcal{F}_{\sigma}\big(T_n^S\big)\big] \approx n^{\frac{2w(\tau)}{1+\alpha}}.$$

Using (20) and observing that τ has no loose leaf, we conclude that

$$\mathbb{E}\big[\mathcal{F}_{\tau+\tau}\big(T_n^S\big)\big] = \mathcal{O}\big(n^{\frac{2w(\tau)}{1+\alpha}}\big) = \mathcal{O}\big(n^{\frac{2|\ell|}{1+\alpha}}\big).$$

$\qquad\square$

3.3.2. *An estimate on* $\mathbb{E}[\mathcal{F}_{\tau,\tau}(T_n)]$

We start this section by defining a wider class of observables that will be involved in a recurrence relation which will eventually allow us to estimate $\mathbb{E}[\mathcal{F}_{\tau,\tau}(T_n)]$. Let τ, σ be two decorated trees and T a (bigger) tree. Then we denote by $\mathcal{F}_{\tau,\sigma}(T)$ the following integer-valued observable:

$$\mathcal{F}_{\tau,\sigma}(T) := \sum_{\substack{\phi_1,\phi_2 \\ \phi_1(\tau)\cap\phi_2(\sigma)=\varnothing}} \pi(\tau, \phi_1, T) \cdot \pi(\sigma, \phi_2, T), \tag{23}$$

where the sum is over all graph embeddings ϕ_1, resp. ϕ_2, of τ, resp. σ, in T with no overlap in their image. The quantity of interest to us is that with $\sigma = \tau$ and $T = T_n$. This section is concerned with proving the following bound.

Proposition 3.12. *Let T_n be an α-PA tree. Then it holds that*:

$$\mathbb{E}\big[\mathcal{F}_{\tau,\sigma}(T_n)\big] = \mathcal{O}\big(\mathbb{E}\big[\mathcal{F}_{\tau}(T_n)\big] \cdot \mathbb{E}\big[\mathcal{F}_{\sigma}(T_n)\big]\big). \tag{24}$$

Proposition 3.12 implies directly the bound necessary for the proof of Theorem 3.1.

Corollary 3.13. *Let τ be a decorated tree with $\ell(u) \geq 2$ for any $u \in \tau$ and such that $|\ell| > 1 + \alpha$. Then, for any seed S,*

$$\mathbb{E}\big[\mathcal{F}_{\tau,\tau}\big(T_n^S\big)\big] = \mathcal{O}\big(n^{\frac{2|\ell|}{1+\alpha}}\big). \tag{25}$$

Proof of Corollary 3.13. Apply Proposition 3.12 with $\sigma = \tau$ and observe that, due to Remark 3.8 and the conditions on τ, we have $\mathbb{E}[\mathcal{F}_{\tau}(T_n)] \approx n^{\frac{|\ell|}{1+\alpha}}$. $\square$

Proposition 3.12 is obtained through a recurrence relation, similarly to how Theorem 3.7 follows from Proposition 3.4. Since we only need an upper bound, we state only a (simpler) recursive inequality.

Proposition 3.14. *There exists a family of nonnegative real numbers $\{c^*(\boldsymbol{\theta}', \boldsymbol{\theta}) : \boldsymbol{\theta}' \prec \boldsymbol{\theta}\}$ such that for any two decorated trees $\tau, \sigma \neq \odot$:*

$$\mathbb{E}\big[\mathcal{F}_{\tau,\sigma}(T_{n+1}))|\mathcal{F}_n\big]$$
$$\leq \left(1 + \frac{w(\tau) + w(\sigma)}{(1+\alpha) \cdot n - 2}\right) \cdot \mathcal{F}_{\tau,\sigma}(T_n)$$
$$+ \frac{1}{(1+\alpha) \cdot n - 2}\left[\sum_{t \prec \tau} c^*(t, \tau) \cdot \mathcal{F}_{t,\sigma}(T_n) + \sum_{s \prec \sigma} c^*(s, \sigma) \cdot \mathcal{F}_{\tau,s}(T_n)\right]. \tag{26}$$

In the rest of the section, we show how Proposition 3.14, implies Proposition 3.12, then prove Proposition 3.14. Both proofs follow similar arguments to those in Sections 3.2 and 3.1, respectively.

Proof of Proposition 3.12. We are going to proceed by induction on pairs (τ, σ) of decorated trees, for the partial order induced by $\prec$ on such pairs (precisely $(t, s) \prec (\tau, \sigma)$ if either $t \preceq \tau$ and $s \prec \sigma$ or $t \prec \tau$ and $s \preceq \sigma$).

Base case. We show (24) when τ is any decorated tree and $\sigma = \odot$. If ϕ_1 is a graph embedding of τ in T_n, its image consists of $|\tau|$ vertices of T_n. Hence, the number of ways to embed σ in T_n without overlapping with $\phi_1(\tau)$ is $n - |\tau|$. Thus

$$\mathbb{E}\big[\mathcal{F}_{\tau,\odot}(T_n)\big] = \mathbb{E}\big[\mathcal{F}_{\tau}(T_n)\big] \cdot \big(n - |\tau|\big) = \mathcal{O}\big(\mathbb{E}[\mathcal{F}_{\tau}(T_n)] \cdot \mathbb{E}[\mathcal{F}_{\odot}(T_n)]\big),$$

as required.

Induction step. Let τ, σ be two decorated trees, both different from $\odot$. Assume that (24) holds for all pairs (t, s) with either $t \preceq \tau$ and $s \prec \sigma$ or $t \prec \tau$ and $s \preceq \sigma$. In the following, we set:

$$\omega_n^{(\tau,\sigma)} := \prod_{\ell=k}^{n}\left(1 + \frac{w(\tau) + w(\sigma)}{(1+\alpha) \cdot \ell - 2}\right)^{-1}$$

and

$$\mathcal{S}_n(t, \tau; \sigma) := \sum_{\ell=k}^{n} \frac{1}{(1+\alpha)\ell - 2} \cdot \omega_{\ell+1}^{(\tau,\sigma)} \cdot \mathbb{E}\big[\mathcal{F}_{t,\sigma}(T_\ell)\big]$$

for any decorated tree $t \prec \tau$. Iterating (26), we obtain:

$$\mathbb{E}\big[\omega_{n+1}^{(\tau,\sigma)} \cdot \mathcal{F}_{\tau,\sigma}(T_{n+1})\big] \le \mathcal{F}_{\tau,\sigma}(S) + \sum_{t \prec \tau} c^*(t, \tau) \cdot \mathcal{S}_n(t, \tau; \sigma) + \sum_{s \prec \sigma} c^*(s, \sigma) \cdot \mathcal{S}_n(s, \sigma; \tau).$$

Thus, to prove (24), it suffices to show that $\mathcal{S}_n(t, \tau; \sigma) = \mathcal{O}(\omega_n^{(\tau,\sigma)} \cdot \mathbb{E}[\mathcal{F}_\tau(T_n)] \cdot \mathbb{E}[\mathcal{F}_\sigma(T_n)])$ for all $t \prec \tau$ (by symmetry, the same will also hold for $\mathcal{S}_n(s, \sigma; \tau)$ with $s \prec \sigma$). Recall from Theorem 3.7 the asymptotic

$$\omega_n^{(\tau,\sigma)} \cdot \mathbb{E}\big[\mathcal{F}_\tau(T_n)\big] \cdot \mathbb{E}\big[\mathcal{F}_\sigma(T_n)\big]) \approx n^{\max\{0, 1-\frac{w(\tau)}{1+\alpha}\}+\max\{0, 1-\frac{w(\sigma)}{1+\alpha}\}} \cdot (\log n)^{\gamma(\tau)+\gamma(\sigma)}. \quad (27)$$

Fix $t \prec \tau$. According to the induction hypothesis and the above, the terms of $\mathcal{S}_n(t, \tau; \sigma)$ are bounded as

$$\frac{\omega_{n+1}^{(\tau,\sigma)}}{(1+\alpha)n - 2} \cdot \mathbb{E}\big[\mathcal{F}_{t,\sigma}(T_n)\big]$$
$$= \mathcal{O}\big(n^{\max\{\frac{w(t)-w(\tau)}{1+\alpha}, 1-\frac{w(\tau)}{1+\alpha}\}+\max\{0, 1-\frac{w(\sigma)}{1+\alpha}\}-1}(\log n)^{\gamma(t)+\gamma(\sigma)}\big). \quad (28)$$

The sum of the above has different asymptotics depending on the value of the exponent of n:

> **1st case:** $\max\{\frac{w(t)-w(\tau)}{1+\alpha}, 1 - \frac{w(\tau)}{1+\alpha}\} + \max\{0, 1 - \frac{w(\sigma)}{1+\alpha}\} < 0$

Then the sum of (28) converges, thus $\mathcal{S}_n(t, \tau; \sigma) = \mathcal{O}(1) = \mathcal{O}(\omega_n^{(\tau,\sigma)} \cdot \mathbb{E}[\mathcal{F}_\tau(T_n)] \cdot \mathbb{E}[\mathcal{F}_\sigma(T_n)])$.

> **2nd case:** $\max\{\frac{w(t)-w(\tau)}{1+\alpha}, 1 - \frac{w(\tau)}{1+\alpha}\} + \max\{0, 1 - \frac{w(\sigma)}{1+\alpha}\} > 0$

Then the sum of (28) diverges, and a standard estimate provides the precise rate of growth:

$$\mathcal{S}_n(t, \tau; \sigma) = \mathcal{O}\big(n^{\max\{\frac{w(t)-w(\tau)}{1+\alpha}, 1-\frac{w(\tau)}{1+\alpha}\}+\max\{0, 1-\frac{w(\sigma)}{1+\alpha}\}}(\log n)^{\gamma(t)+\gamma(\sigma)}\big).$$

Compare the above to (27) to find

$$\frac{\mathcal{S}_n(t, \tau; \sigma)}{\omega_n^{(\tau,\sigma)}\mathbb{E}[\mathcal{F}_\tau(T_n)]\mathbb{E}[\mathcal{F}_\sigma(T_n)]}$$
$$= \mathcal{O}\big(n^{\max\{\frac{w(t)-w(\tau)}{1+\alpha}, 1-\frac{w(\tau)}{1+\alpha}\}-\max\{0, 1-\frac{w(\tau)}{1+\alpha}\}} \cdot (\log n)^{\gamma(t)-\gamma(\tau)}\big).$$

Now, recall that

- $w(t) \leq w(\tau)$ always;
- if $w(t) \leq w(\tau) < 1 + \alpha$, then $\gamma(t) = \gamma(\tau) = 0$;
- if $w(t) = w(\tau) \geq 1 + \alpha$, then $\gamma(t) < \gamma(\tau)$.

A separate analysis of the three different situations above shows that $\mathcal{S}_n(t, \tau; \sigma) = \mathcal{O}(\omega_n^{(\tau, \sigma)} \cdot \mathbb{E}[\mathcal{F}_\tau(T_n)] \cdot \mathbb{E}[\mathcal{F}_\sigma(T_n)])$.

3rd case: $\max\{\frac{w(t) - w(\tau)}{1 + \alpha}, 1 - \frac{w(\tau)}{1 + \alpha}\} + \max\{0, 1 - \frac{w(\sigma)}{1 + \alpha}\} = 0$

Then the sum of (28) diverges, and a standard estimate provides the precise rate of growth:

$$\mathcal{S}_n(t, \tau; \sigma) = \mathcal{O}\big((\log n)^{\gamma(t) + \gamma(\sigma) + 1}\big).$$

By (27) and our assumption, we find

$$\frac{\mathcal{S}_n(t, \tau; \sigma)}{\omega_n^{(\tau, \sigma)} \mathbb{E}[\mathcal{F}_\tau(T_n)] \mathbb{E}[\mathcal{F}_\sigma(T_n)]} \tag{29}$$

$$= \mathcal{O}\big(n^{\max\{\frac{w(t) - w(\tau)}{1 + \alpha}, 1 - \frac{w(\tau)}{1 + \alpha}\} - \max\{0, 1 - \frac{w(\tau)}{1 + \alpha}\}} (\log n)^{\gamma(t) - \gamma(\tau) + 1}\big). \tag{30}$$

The power of n in the right-hand side above is negative or null. When it is negative, equation (29) is bounded, as required. It can only be 0 in two cases: when $w(\tau) < 1 + \alpha$ or when $w(t) = w(\tau) \geq 1 + \alpha$. The former is incoherent with the assumption of this 3rd case; when the latter occurs, $\gamma(t) \leq \gamma(\tau) - 1$, hence (29) is bounded. In conclusion, (29) is always bounded, which is to say that $\mathcal{S}_n(t, \tau; \sigma) = \mathcal{O}(\omega_n^{(\tau, \sigma)} \cdot \mathbb{E}[\mathcal{F}_\tau(T_n)] \cdot \mathbb{E}[\mathcal{F}_\sigma(T_n)])$. $\qquad \square$

It remains to prove Proposition 3.14.

Proof of Proposition 3.14. The strategy followed here is the same as for Proposition 3.4. We start by fixing $\tau = (\tau, \ell_\tau)$ and $\sigma = (\sigma, \ell_\sigma)$ two decorated trees, both different from $\circledcirc$. Recall our notation: in passing from T_n to T_{n+1} a new vertex denoted by v_n is attached to a randomly chosen vertex u_n of T_n. Our purpose is to compute the sum (23) over pairs of non-overlapping graph embeddings ϕ_1, ϕ_2 of τ and σ into T_{n+1}. We restrict the sum to embeddings with $\pi(\tau, \phi_1, T_{n+1}) > 0$ and $\pi(\sigma, \phi_2, T_{n+1}) > 0$.

As it has already been noted in the proof of Proposition 3.4, such embeddings fall into three distinct categories: those which do not include u_n nor v_n in their image, those which include u_n but not v_n in their image and those which include both u_n and v_n in their image. Embeddings that include v_n but not u_n in their image are not authorised as they only have positive weight when the decorated tree is $\circledcirc$.

A crucial remark is that two graph embeddings $\phi_1 : \tau \to T_{n+1}$, $\phi_2 : \sigma \to T_{n+1}$ with no overlap cannot be at once in one of the two last categories of embeddings. We thus enumerate three kind of situations:

1. the images of ϕ_1 and ϕ_2 are both in $T_{n+1} \setminus \{u_n, v_n\}$;

2. the image of ϕ_1 contains u_n but not v_n, that of ϕ_2 does not include any of them; or the same situation but reversing the roles of ϕ_1 and ϕ_2;

3. the image of ϕ_1 contains u_n and v_n, that of ϕ_2 does not include any of them; or the same situation but reversing the roles of ϕ_1 and ϕ_2.

Write $\mathcal{F}_{\tau,\sigma}^{(i)}(T_{n+1})$ with $i = 1, 2, 3$ for the contribution to (23) of pairs of embeddings being in the corresponding situation above. As in the proof of Proposition 3.4, the pairs of embeddings in the two first situations are in one-to-one correspondence with non-overlapping embeddings $\phi_1 : \tau \to T_n$, $\phi_2 : \sigma \to T_n$, although their weights $\pi(\tau, \phi_1, T_{n+1})$ and $\pi(\sigma, \phi_2, T_{n+1})$ may differ from that in T_n. The same algebraic manipulations used for $\mathcal{F}_\tau^{(1)}(T_{n+1})$ and $\mathcal{F}_\tau^{(2)}(T_{n+1})$ in the proof of Proposition 3.4 may also be applied here to find

$$
\mathbb{E}\big[\mathcal{F}_{\tau,\sigma}^{(1)}(T_{n+1}) + \mathcal{F}_{\tau,\sigma}^{(2)}(T_{n+1})|\mathcal{E}_n\big] = \left[1 + \frac{|\ell_\tau| + |\ell_\sigma|}{(1+\alpha)n - 2}\right]\mathcal{F}_{\tau,\sigma}(T_n)
$$
$$
+ \sum_{u \in \tau} \frac{\ell_\tau(u)(\ell_\tau(u) + \alpha - 1)}{(1+\alpha)n - 2}\mathcal{F}_{\tau^{u-},\sigma}(T_n)
$$
$$
+ \sum_{v \in \sigma} \frac{\ell_\sigma(v)(\ell_\sigma(v) + \alpha - 1)}{(1+\alpha)n - 2}\mathcal{F}_{\tau,\sigma^{v-}}(T_n) \qquad (31)
$$

Moreover, the quantity $\mathcal{F}_{\tau,\sigma}^{(3)}(T_{n+1})$ may be treated as $\mathcal{F}_\tau^{(3)}(T_{n+1})$ in (10), and we find

$$
\mathbb{E}\big[\mathcal{F}_{\tau,\sigma}^{(3)}(T_{n+1})|\mathcal{E}_n\big]
$$
$$
= \frac{1}{(1+\alpha)n - 2}
$$
$$
\times \Bigg(\sum_{v \text{ l.l of } \tau} \mathcal{F}_{(\tau\setminus v)^+,\sigma}(T_n)
$$
$$
+ \big(2\ell_\tau(u) + \alpha\big)\mathcal{F}_{\tau\setminus v,\sigma}(T_n) + \ell_\tau(u)\big(\ell_\tau(u) + \alpha - 1\big)\mathcal{F}_{(\tau\setminus v)^-,\sigma}(T_n)
$$
$$
+ \sum_{v \text{ l.l of } \sigma} \mathcal{F}_{\tau,(\sigma\setminus v)^+}(T_n)
$$
$$
+ \big(2\ell_\sigma(u) + \alpha\big)\mathcal{F}_{\tau,\sigma\setminus v}(T_n) + \ell_\sigma(u)\big(\ell_\sigma(u) + \alpha - 1\big)\mathcal{F}_{\tau,(\sigma\setminus v)^-}(T_n)\Bigg) \qquad (32)
$$

where the two sums are over the set of loose leaves of τ and σ, respectively. The above equation is a derivative of (10). At the same stage, the last ingredient for the proof of Proposition 3.4 was a combinatorial lemma – Lemma 3.5 – claiming that $\mathcal{F}_{(\tau\setminus v)^+}(T_n)$ is actually a linear combination of $\mathcal{F}_\tau(T_n)$ and $\mathcal{F}_{\tau\setminus v}(T_n)$ when v is a loose leaf of τ. Here that is almost what we do for $\mathcal{F}_{(\tau\setminus v)^+,\sigma}(T_n)$ and of course for $\mathcal{F}_{\tau,(\sigma\setminus v)^+}(T_n)$ by symmetry. The equality (11) is just turned into an inequality.

Lemma 3.15. *Let $\boldsymbol{\tau} = (\tau, \ell_\tau)$ and $\boldsymbol{\sigma} = (\tau, \ell_\sigma)$ be two decorated trees, v a loose leave of $\boldsymbol{\tau}$ and* u *its only neighbour in τ. Then, for any tree T:*

$$\mathcal{F}_{\boldsymbol{\tau},\boldsymbol{\sigma}}(T) + C(\boldsymbol{\tau}, w) \cdot \mathcal{F}_{\boldsymbol{\tau}\backslash\mathsf{v},\boldsymbol{\sigma}}(T) \leq \mathcal{F}_{(\boldsymbol{\tau}\backslash\mathsf{v})^+,\boldsymbol{\sigma}}(T) \leq \mathcal{F}_{\boldsymbol{\tau},\boldsymbol{\sigma}}(T) + C^*(\boldsymbol{\tau}, w) \cdot \mathcal{F}_{\boldsymbol{\tau}\backslash\mathsf{v},\boldsymbol{\sigma}}(T),$$

with $C(\boldsymbol{\tau}, w)$ being the constant defined in Lemma 3.5 and $C^*(\boldsymbol{\tau}, w) := C(\boldsymbol{\tau}, w) + 1$.

Proof. As for Lemma 3.5, we proceed in two steps. First, we express $\mathcal{F}_{\boldsymbol{\tau},\boldsymbol{\sigma}}(T)$ in terms of pairs of non-overlapping embeddings of $\tau \backslash \mathsf{v}$ and σ, respectively, in T.

To any non-overlapping pair of embeddings $\phi_1 : \tau \to T$, $\phi_2 : \sigma \to T$ associate the restriction $\tilde{\phi}_1 : \tau \backslash \mathsf{v} \to T$ of ϕ_1 to the set of vertices of $\boldsymbol{\tau} \backslash \mathsf{v}$ together with the same embedding ϕ_2 of σ. Conversely, if a pair of non-overlapping embeddings $\tilde{\psi}_1 : \tau \backslash \mathsf{v} \to T$ and $\phi_2 : \sigma \to T$ is given, we may extend $\tilde{\psi}_1$ to some $\psi_1 : \tau \to T$ which preserves the non-overlapping property with ϕ_2. The number of such extensions depends on how many neighbours of $\tilde{\psi}_1(u)$ are not contained in the images of $\tilde{\psi}_1$ and ϕ_2. The former occupies $\deg_{\tau\backslash\mathsf{v}}(u)$ neighbours of $\tilde{\psi}_1(u)$; the latter may occupy 0 or 1 neighbour due to the non-overlapping requirement. From this observation, and given that $\pi(\tau, \phi_1, T) = \pi(\tau \backslash \mathsf{v}, \tilde{\phi}_1, T)$ (v being a loose leaf), we deduce that:

$$\mathcal{F}_{\boldsymbol{\tau},\boldsymbol{\sigma}}(T) \leq \sum_{\substack{\tilde{\phi}_1,\phi_2 \\ \tilde{\phi}_1(\tau\backslash\mathsf{v})\cap\phi_2(\sigma)=\varnothing}} \left(\deg_T\big(\tilde{\phi}_1(u)\big) - \deg_{\tau\backslash\mathsf{v}}(u)\right) \cdot \pi(\tau \backslash \mathsf{v}, \tilde{\phi}_1, T) \cdot \pi(\sigma, \phi_2, T) \quad (33)$$

and

$$\mathcal{F}_{\boldsymbol{\tau},\boldsymbol{\sigma}}(T)$$
$$\geq \sum_{\substack{\tilde{\phi}_1,\phi_2 \\ \tilde{\phi}_1(\tau\backslash\mathsf{v})\cap\phi_2(\sigma)=\varnothing}} \left(\deg_T\big(\tilde{\phi}_1(u)\big) - \deg_{\tau\backslash\mathsf{v}}(u) - 1\right) \cdot \pi(\tau \backslash \mathsf{v}, \tilde{\phi}_1, T) \cdot \pi(\sigma, \phi_2, T). \quad (34)$$

The two above equations correspond to (12) in the proof of Lemma 3.5. Finally, similarly to (13), we find

$$\mathcal{F}_{(\boldsymbol{\tau}\backslash\mathsf{v})^+,\boldsymbol{\sigma}}(T)$$
$$= \sum_{\substack{\tilde{\phi}_1,\phi_2 \\ \tilde{\phi}_1(\tau\backslash\mathsf{v})\cap\phi_2(\sigma)=\varnothing}} \left(\deg_T\big(\tilde{\phi}_1(u)\big) - 1 - \ell_\tau(u)\right) \cdot \pi(\tau \backslash \mathsf{v}, \tilde{\phi}_1, T) \cdot \pi(\sigma, \phi_2, T). \quad (35)$$

Subtracting (33) and (34) from (35), we obtain the desired lower and upper bounds, respectively. $\qquad\square$

To finish the proof of Proposition 3.14, insert the upper bound of Lemma 3.15 in (32) to find

$$
\begin{aligned}
&\mathbb{E}\big[\mathcal{F}^{(3)}_{\boldsymbol{\tau},\boldsymbol{\sigma}}(T_{n+1})|\mathcal{E}_n\big]\\
&\quad \leq \frac{|\{\mathsf{v} \text{ loose leaf of } \boldsymbol{\tau}\}| + |\{\mathsf{v} \text{ loose leaf of } \boldsymbol{\sigma}\}|}{(1+\alpha)n-2} \cdot \mathcal{F}_{\boldsymbol{\tau},\boldsymbol{\sigma}}(T_n)\\
&\qquad + \frac{1}{(1+\alpha)n-2} \sum_{\mathsf{v}\text{ l.l. of } \boldsymbol{\tau}} \big(\deg_{\tau}(u) + \ell_{\tau}(u) + \alpha - 1\big) \cdot \mathcal{F}_{\boldsymbol{\tau}\backslash \mathsf{v},\boldsymbol{\sigma}}(T_n)\\
&\qquad + \ell_{\tau}(\mathsf{u})\big(\ell_{\tau}(\mathsf{u}) + \alpha - 1\big) \cdot \mathcal{F}_{(\boldsymbol{\tau}\backslash \mathsf{v})^-,\boldsymbol{\sigma}}(T_n)\\
&\qquad + \frac{1}{(1+\alpha)n-2} \sum_{\mathsf{v}\text{ l.l. of } \boldsymbol{\sigma}} \big(\deg_{\sigma}(u) + \ell_{\sigma}(u) + \alpha - 1\big) \cdot \mathcal{F}_{\boldsymbol{\tau},\boldsymbol{\sigma}\backslash \mathsf{v}}(T_n)\\
&\qquad + \ell_{\sigma}(\mathsf{u})\big(\ell_{\sigma}(\mathsf{u}) + \alpha - 1\big) \cdot \mathcal{F}_{\boldsymbol{\tau},(\boldsymbol{\sigma}\backslash \mathsf{v})^-}(T_n) \qquad\qquad (36)
\end{aligned}
$$

where the two sums are again over the set of loose leaves of $\boldsymbol{\tau}$ and $\boldsymbol{\sigma}$, respectively. Equations (31) and (36) eventually lead to the expected inequality (26). $\qquad\square$

3.3.3. *Proof of Theorem* 3.1

Theorem 3.1 follows directly from the estimates of the two previous sections and from the decomposition

$$
\mathcal{F}_{\boldsymbol{\tau}}(T_n)^2 = \mathcal{F}_{\boldsymbol{\tau},\boldsymbol{\tau}}(T_n) + \mathcal{F}_{\boldsymbol{\tau}+\boldsymbol{\tau}}(T_n). \qquad\qquad (37)
$$

Proof of Theorem 3.1. Take the expectation of (37) and insert (19) and (25). $\qquad\square$

4. Observables and their difference around the seed tree

The goal of this section is to produce decorated trees $\boldsymbol{\tau}$ that can distinguish between two different seeds S and S'. Moreover, we wish the asymptotic of $\mathcal{F}_{\boldsymbol{\tau}}(T_n^S)$ for such trees (both for the first and the square root of the second moment) to be of the type $n^{\frac{w(\tau)}{1+\alpha}}$, with a null logarithmic correction. The main result is the following.

Theorem 4.1. *For any two seed trees $S \neq S'$ of size $k, k' \geq 3$, there exists a decorated tree $\boldsymbol{\tau}$ with $w(\boldsymbol{\tau}) > 1 + \alpha$ and $\ell(u) \geq 2$ for all $u \in \boldsymbol{\tau}$ such that*

$$
\mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S\big)\big] - \mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^{S'}\big)\big] \approx n^{\frac{w(\tau)}{1+\alpha}}.
$$

The next four sections are concerned with the proof of Theorem 4.1. To start, we will consider seeds S, S' of same size. Finally, in Section 4.5, we prove our main result, Theorem 1.1.

4.1. Differences only appear around the seed

We are interested here in the variation of the first moment of (3), depending on the value taken by the seed tree S. We state the following proposition.

Proposition 4.2. *For any two seed trees S, S' of common size $k \geq 3$, any decorated tree τ and any $n \geq k$:*

$$\mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S\big)\big] - \mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^{S'}\big)\big] = \mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S, \{\cdot\} \cap S \neq \varnothing\big)\big] - \mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^{S'}, \{\cdot\} \cap S' \neq \varnothing\big)\big],$$

where:

$$\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S, \{\cdot\} \cap S \neq \varnothing\big) := \sum_{\substack{\phi:\tau \hookrightarrow T_n^S \\ \phi(\tau) \cap S \neq \varnothing}} \prod_{u \in \tau} \big[\deg_{T_n^S}\big(\phi(u)\big) - 1\big]_{\ell(u)},$$

and the same for S'. In other words, $\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S, \{\cdot\} \cap S \neq \varnothing\big)$ is the contribution to $\mathcal{F}_{\boldsymbol{\tau}}$ of embeddings that "intersect the seed".

The reader may be surprised that, in the statement above, we assume the location of the seed inside T_n known, all while trying to prove that the seed may be determined. It should be clear that, while Proposition 4.2 and other steps of the proof of Theorem 4.1 use the knowledge of the seed, their ultimate result (that is Theorem 4.1) does not.

The proof is based on the coupling of Section 2.2, hence we use the planar α-PA formalism.

Proof. Fix seed trees S, S' of equal size $k \geq 3$. We start by decomposing the sum over embeddings ϕ defining (3) according to whether $\phi(\tau) \cap S$ is empty or not:

$$\mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S\big)\big] = \mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S, \{\cdot\} \cap S \neq \varnothing\big)\big] + \mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S, \{\cdot\} \cap S = \varnothing\big)\big].$$

If an embedding ϕ of τ in T_n^S does not intersect S, then $\phi(\tau)$ is necessarily a proper subset of one of the planted plane subtrees $T_n^{v,i}$, and does not intersect the root of said tree. Write $\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^{v,i}\big)$ for the contribution to $\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S\big)$ of all such embeddings. Then

$$\mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S, \{\cdot\} \cap S = \varnothing\big)\big] = \sum_{v \in V_S} \sum_{i=1}^{\deg_S(v)} \mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^{v,i}\big)\big].$$

The right-hand side of the above does not actually depend on S, only on its size. Indeed, the coupling described at the end of Section 2.2 indicates that the same planted plane subtrees may be used to construct T_n^S and $T_n^{S'}$. While it is not necessary for this argument, one may notice that T_n^S and $T_n^{S'}$ may even be coupled so that

$$\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S, \{\cdot\} \cap S = \varnothing\big) = \mathcal{F}_{\boldsymbol{\tau}}\big(T_n^{S'}, \{\cdot\} \cap S' = \varnothing\big) \quad \text{a.s.}$$

The result follows readily. $\qquad\square$

4.2. Blind trees: Differences appear only in the seed

Next, we aim to improve Proposition 4.2 by showing that $\mathbb{E}[\mathcal{F}_{\boldsymbol{\tau}}(T_n^S)] - \mathbb{E}[\mathcal{F}_{\boldsymbol{\tau}}(T_n^{S'})]$ only depends on embeddings totally contained in the seed, not just intersecting it. This will not be true for all trees $\boldsymbol{\tau}$, only for special ones. Some definitions are required.

Definition 4.3 (Perfect embedding). Let $\boldsymbol{\tau} = (\tau, d)$ be a decorated tree and T a (larger) tree. An embedding $\phi : \tau \hookrightarrow T$ is called **perfect** if

$$d(u) = \deg_T(\phi(u)) \quad \forall u \in \tau.$$

Write $D_{\tau,d}(T)$ for the number of perfect embeddings from (τ, d) to T. Note that any embedding of τ in T is perfect for exactly one decoration, hence contributes exactly to one $D_{\tau,d}(T)$.

Definition 4.4 (Blind tree). Let τ, T_1 and T_2 be finite trees with $|T_1| = |T_2|$. We say that τ is a $(\mathbf{T_1}, \mathbf{T_2})$-**blind tree** if for any decoration $d : \tau \to \mathbb{N}^*$:

$$D_{\tau,d}(T_1) = D_{\tau,d}(T_2).$$

Intuitively, τ is (T_1, T_2)-blind if it can not distinguish between T_1 and T_2 using observables involving the number of embeddings and the environment around these embeddings (as are our observables $\mathcal{F}_{\boldsymbol{\tau}}$). For instance, the tree formed of a single vertex is (T_1, T_2)-blind if and only if T_1 and T_2 have the same degree sequence.

Let (τ, d) be a decorated tree, $\sigma \subseteq \tau$ a nonempty subtree of τ, and S be a seed tree. Write $D_{\sigma,\tau,d}(T_n^S)$ for the number of perfect embeddings Φ of (τ, d) in T_n^S with $\Phi(\tau) \cap S = \Phi(\sigma)$. We obviously have:

$$\mathcal{F}_{\boldsymbol{\tau}}(T_n^S, \{\cdot\} \cap S \neq \varnothing) = \sum_{\sigma \subseteq \tau} \sum_{d:\tau \to \mathbb{N}^*} D_{\sigma,\tau,d}(T_n^S) \cdot \prod_{u \in \tau} [d(u) - 1]_{\ell(u)}. \tag{38}$$

The next proposition constitutes the essential step for the upgrade of Proposition 4.2.

Proposition 4.5. *Let S and S' be two seed trees of common size $k \geq 3$. For trees $\sigma \subseteq \tau$ with σ which is (S, S')-blind,*

$$\forall d : \tau \to \mathbb{N}^* \text{ and } n \geq k, \quad \mathbb{E}\big[D_{\sigma,\tau,d}(T_n^S)\big] = \mathbb{E}\big[D_{\sigma,\tau,d}(T_n^{S'})\big].$$

Proof. The idea of the proof is to decompose the embeddings contributing to $D_{\tau,\sigma,d}(T_n^S)$ according to degrees of the vertices belonging to its image *in the seed tree S*. For illustration, we will start with the simpler case when $\sigma = \tau$, then move on to the general case.

> **Particular case $\sigma = \tau$**

For any seed S and $n \geq |S|$,

$$D_{\tau,\tau,d}(T_n^S) = \sum_{\phi:\tau \hookrightarrow S} \prod_{u \in \tau} \mathbf{1}_{\deg_{T_n^S}(\phi(u)) = d(u)}.$$

Due to the coupling of in Section 2.2, the probability of the event $\bigcap_{u \in \tau} \{\deg_{T_n^S}(\phi(u)) = d(u)\}$ only depends on the degrees of the vertices $(\phi(u))_{u \in \tau}$ in S. Write $f_d[\deg_S \phi(u) : u \in \tau]$ for this probability. Then

$$\mathbb{E}[D_{\tau,\tau,d}(T_n^S)] = \sum_{\phi:\tau \hookrightarrow S} f_d[\deg_S \phi(u) : u \in \tau]$$

$$= \sum_{\ell:\tau \to \mathbb{N}^*} D_{\tau,\ell}(S) \cdot f_d[\ell(u) : u \in \tau]. \tag{39}$$

Since τ is assumed (S, S')-blind, the quantity above remains unaltered when S is replaced by S'.

General case $\sigma \subset \tau$

For a seed S and $v \in S$, write $(T_n^S \setminus S)_v$ for the subtree of T_n^S formed of all vertices that may be connected to v without using any edge of S. In terms of vertices $\{(T_n^S \setminus S)_v : v \in S\}$ is a partition of the vertices of T_n^S; in terms of edges, it is a partition of the edges of T_n^S not contained in S. In the language of Section 2.2, $(T_n^S \setminus S)_v$ is the tree obtained by gluing $T_n^{v,1}, \ldots, T_n^{v,\deg_S(v)}$ together at the root. Use the same notation for τ and σ: for $v \in \sigma$ write $(\tau \setminus \sigma)_v$ for the subtree of τ formed of all vertices that may be connected to v in τ without using any edge of σ.

For a decoration d of τ, and vertices $v \in \sigma$ and $v \in S$, write $\mathcal{D}(d, v, v, T_n^S)$ for the number of perfect embeddings ϕ of $((\tau \setminus \sigma)_v, d)$ in $(T_n^S \setminus S)_v$ with $\phi(v) = v$:

$$\mathcal{D}(d, v, v, T_n^S) = \sum_{\phi:(\tau \setminus \sigma)_v \to (T_n^S \setminus S)_v} \mathbb{1}_{\phi(v)=v} \prod_{u \in (\tau \setminus \sigma)_v} \mathbb{1}_{\{\deg_{T_n^S}(\phi(u))=d(u)\}}.$$

It may not be explicit in the above, but the last product does only depend on $(T_n^S \setminus S)_v$ for $u \neq v$, since $\deg_{T_n^S}(\phi(u)) = \deg_{(T_n^S \setminus S)_v}(\phi(u))$. For $v \in S$, it should be noted that $\deg_{T_n^S}(v) = \deg_{(T_n^S \setminus S)_v}(v) + \deg_S(v)$.

Now express $D_{\sigma,\tau,d}(T_n^S)$ as follows:

$$D_{\sigma,\tau,d}(T_n^S) = \sum_{\phi:\sigma \hookrightarrow S} \prod_{v \in \sigma} \mathcal{D}(d, v, \phi(v), T_n^S). \tag{40}$$

Indeed, any embedding contributing $D_{\tau,\sigma,d}(T_n^S)$ comes from an embedding of σ in S and embeddings of the subtrees $(\tau \setminus \sigma)_v$ in the corresponding subtrees of $T_n^S \setminus S$. The conditions that the embedding is perfect may be verified separately in each subtree.

Now, due to the coupling of Section 2.2 and to the expression of $(T_n^S \setminus S)_v$ in terms of $(T_n^{v,i})_{i \geq 1}$, the expectation of the product above only depends on the degrees of the vertices $\phi(v)$ in S:

$$\mathbb{E}\left[\prod_{v \in \sigma} \mathcal{D}(d, v, \phi(v), T_n^S)\right] = f_{\sigma,\tau,d}[\deg_S(\phi(v)) : v \in \sigma],$$

for some explicit function $f_{\sigma,\tau,d}$. Injecting this in (40), we find

$$\mathbb{E}\big[D_{\sigma,\tau,d}\big(T_n^S\big)\big] = \sum_{\phi:\sigma\hookrightarrow S} f_{\sigma,\tau,d}\big[\deg_S(\phi(\mathsf{v})):\mathsf{v}\in\sigma\big] = \sum_{\ell:\sigma\to\mathbb{N}^*} D_{\sigma,\ell}(S)\cdot f_{\sigma,\tau,d}\big[\ell(\mathsf{v}):\mathsf{v}\in\sigma\big].$$

Since σ is assumed (S,S')-blind, the quantity above is equal when S is replaced by S'. □

We are now ready to give a finer version of Proposition 4.2. Consider two distinct seeds S, S' of equal size $k\geq 4$ (non two distinct seeds of smaller size exist). Then there exists at least one tree which is not (S,S')-blind, for instance S or S' have this property. It follows that there exists at least one minimal tree which is not (S,S')-blind, that is a tree τ which is not (S,S')-blind but for which any proper subtree $\sigma\subsetneq\tau$ is (S,S')-blind.

Corollary 4.6. *Let τ be a minimal tree that is not (S,S')-blind. Then for any decoration ℓ of τ*

$$\mathbb{E}\big[\mathcal{F}_\tau\big(T_n^S\big)\big] - \mathbb{E}\big[\mathcal{F}_\tau\big(T_n^{S'}\big)\big] = \mathbb{E}\big[\mathcal{F}_\tau\big(T_n^S,\{\cdot\}\subseteq S\big)\big] - \mathbb{E}\big[\mathcal{F}_\tau\big(T_n^{S'},\{\cdot\}\subseteq S'\big)\big],$$

where

$$\mathcal{F}_\tau\big(T_n^S,\{\cdot\}\subseteq S\big) := \sum_{\phi:\tau\hookrightarrow S}\prod_{u\in\tau}\big[\deg_{T_n^S}(\phi(u))-1\big]_{\ell(u)},$$

and the same for S'.

Proof. We apply (38) together with Proposition 4.5 and use the minimality of τ to obtain

$$\mathbb{E}\big[\mathcal{F}_\tau\big(T_n^S,\{\cdot\}\cap S\neq\varnothing\big)\big] - \mathbb{E}\big[\mathcal{F}_\tau\big(T_n^{S'},\{\cdot\}\cap S'\neq\varnothing\big)\big]$$

$$= \sum_{d:\tau\to\mathbb{N}^*}\bigg(\prod_{u\in\tau}[d(u)-1]_{\ell(u)}\bigg)\cdot\big(\mathbb{E}\big[D_{\tau,\tau,d}\big(T_n^S\big)\big] - \mathbb{E}\big[D_{\tau,\tau,d}\big(T_n^{S'}\big)\big]\big).$$

Furthermore, it is clear that

$$\mathbb{E}\big[\mathcal{F}_\tau\big(T_n^S,\{\cdot\}\subseteq S\big)\big] = \sum_{d:\tau\to\mathbb{N}^*}\bigg(\prod_{u\in\tau}[d(u)-1]_{\ell(u)}\bigg)\cdot\mathbb{E}\big[D_{\tau,\tau,d}\big(T_n^S\big)\big].$$

Of course, the above equality also holds when S is replaced by S'. Hence, the result. □

4.3. Evaluating the difference in the seed

In light of the above, the quantities of interest for the proof of Theorem 4.1 are of the type $\mathbb{E}[\mathcal{F}_\tau(T_n^S,\{\cdot\}\subseteq S)]$. We give below a more convenient expression for them based on (39).

Lemma 4.7. *Let S be a seed tree of size k and $\boldsymbol{\tau} = (\tau, \ell)$ a decorated tree. Then, for any $n \geq k$*

$$\mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n^S, \{\cdot\} \subseteq S\big)\big] = \sum_{d:\tau \to \mathbb{N}^*} \mathbf{f}(k, n; d, \ell) \cdot D_{\tau,d}(S), \tag{41}$$

where $\mathbf{f}(k, n; d, \ell)$ are functions defined as follows (for decorations d such that $|d| < 2k - 2$).
   Let $(y_n(u) : u \in \boldsymbol{\tau} \cup \{\text{other}\}, n \geq k)$ be a Pólya urn with $|\tau| + 1$ colours, replacement matrix $M = (m_{uv})_{u,v \in \tau \cup \{\text{other}\}}$ with

$$m_{uv} = \begin{cases} 1 & \text{if } u = v \neq \text{other}, \\ 1 + \alpha & \text{if } u = v = \text{other}, \\ \alpha & \text{if } u \neq v = \text{other}, \\ 0 & \text{otherwise}, \end{cases}$$

and initial states $y_k(u) = \alpha + d(u) - 1$ for $u \in \tau$ and $y_k(\text{other}) = (1 + \alpha)k - 2 - \sum_{u \in \tau} y_k(u)$. Then

$$\mathbf{f}(k, n; d, \ell) = \mathbb{E}\bigg[\prod_{u \in \tau}\big[y_n(u) - \alpha\big]_{\ell(u)}\bigg].$$

Remark 4.8. The actual definition of the variables $(y_n(u) : u \in \boldsymbol{\tau} \cup \{\text{other}\}, n \geq k)$ is not very important; a more intuitive expression will be used (see (43)). The important aspect of (41) is that $\mathbb{E}[\mathcal{F}_{\boldsymbol{\tau}}(T_n^S, \{\cdot\} \subseteq S)]$ is factorised between a part that depends on n but not the seed structure (namely $\mathbf{f}(k, n; d, \ell)$) and one that depends on the seed structure but not on n (namely $D_{\tau,d}(S)$).

Proof. Fix S and $\boldsymbol{\tau} = (\tau, \ell)$. Recall that

$$\mathbb{E}\big[\mathcal{F}_{\boldsymbol{\tau}}\big(T_n, \{\cdot\} \subseteq S\big)\big] = \sum_{\phi:\tau \hookrightarrow S} \mathbb{E}\bigg[\prod_{u \in \tau}\big[\deg_{T_n^S}\big(\phi(u)\big) - 1\big]_{\ell(u)}\bigg], \tag{42}$$

where the sum is over all embeddings of τ in S.
   Let ϕ be an embedding of τ in S. Then the family $\{\deg_{T_n^S}(\phi(u)) : u \in \tau\}$ has a markovian dynamics as n increases described as follows:

- for each $u \in \tau$, with probability $\frac{\deg_{T_n^S}(\phi(u)) - 1 + \alpha}{(1+\alpha)n - 2}$, we have $\deg_{T_{n+1}^S}(\phi(u)) = \deg_{T_n^S}(\phi(u)) + 1$ and all other entries remain the same;
- otherwise $\deg_{T_{n+1}^S}(\phi(u)) = \deg_{T_n^S}(\phi(u))$ for all $u \in \tau$.

Then, if we set

$$y_n(u) = \deg_{T_n^S}\big(\phi(u)\big) + \alpha - 1, \quad \text{for } u \in \tau \text{ and}$$

$$y_n(\text{other}) = (1 + \alpha)n - 2 - \sum_{u \in \tau} y_n(u) = \sum_{v \in T_n^S \setminus \phi(\tau)} \deg_{T_n^S}(v) + \alpha - 1,$$

we deduce readily that the family $(y_n(u) : u \in \tau \cup \{\text{other}\}, n \geq k)$ has the dynamics of a Pólya urn with the replacement matrix M and initial conditions given in the statement. Thus

$$\mathbb{E}\left[\prod_{u \in \tau}\left[\deg_{T_n^S}(\phi(u)) - 1\right]_{\ell(u)}\right] = \mathbb{E}\left[\prod_{u \in \tau}[y_n(u) - \alpha]_{\ell(u)}\right] = \mathbf{f}(k, n; d, \ell). \tag{43}$$

By inserting the above in (42) and grouping the terms of the sum by the degrees in S of the embedding, we obtain (41). $\qquad\square$

Corollary 4.6 and Lemma 4.7 state that, for seeds $S \neq S'$ of same size and τ which is a minimal non-(S, S')-blind tree, the difference of the observables $\mathcal{F}_\tau$ for T_n^S and $T_n^{S'}$ is a linear combination of functions $\mathbf{f}(k, n; d, \ell)$:

$$\mathbb{E}\left[\mathcal{F}_\tau\left(T_n^S\right)\right] - \mathbb{E}\left[\mathcal{F}_\tau\left(T_n^{S'}\right)\right] = \sum_{d:\tau \to \mathbb{N}^*} \mathbf{f}(k, n; d, \ell) \cdot \left[D_{\tau,d}(S) - D_{\tau,d}\left(S'\right)\right] \tag{44}$$

for any decoration ℓ of τ. Thus, for our proof of Theorem 4.1, it will be of great interest to study the asymptotics of the functions $\mathbf{f}(k, n; d, \ell)$ as $n \to \infty$. The relevant result is the following.

Proposition 4.9. *Let S be a seed tree of size $k \geq 2$, $\boldsymbol{\tau} = (\tau, \ell)$ be a decorated tree and $d : \tau \to \mathbb{N}^*$ be some decoration of τ. Then*

$$n^{-\frac{|\ell|}{1+\alpha}} \cdot \mathbf{f}(k, n; d, \ell) \xrightarrow[n \to \infty]{} C\left(k, |\ell|\right) \cdot \prod_{u \in \tau}\left[d(u) + \ell(u) + \alpha - 2\right]_{\ell(u)}, \tag{45}$$

where $C(k, |\ell|) > 0$ is some constant depending only on k and $|\ell|$.

The above is a technical result based on the study of Pòlya urns; the rest of the section is dedicated to proving it. It is possible to prove (45) using only the abstract quantities $(y_n(u) : u \in \tau \cup \{\text{other}\}, n \geq k)$. We prefer however to use the more visual interpretation of these quantities in terms of the degrees in T_n^S of some embedding of τ in S.

Fix a seed S of size k, R a subtree of S and ℓ a decoration of R.

$$M_n^{(R,\ell)} = \prod_{u \in R}\left[\deg_{T_n^S}(u) + \alpha + \ell(u) - 2\right]_{\ell(u)}$$

and

$$\mathcal{W}_n^{(R,\ell)} = \begin{cases} 1 & \text{if } n = k, \\ \prod_{t=k}^{n-1}\left(1 + \dfrac{|\ell|}{(1+\alpha)t - 2}\right)^{-1} & \text{if } n \geq k + 1. \end{cases}$$

When no ambiguity is possible, the superscript (R, ℓ) will be omitted.

Lemma 4.10. *The sequence $(M_n \cdot \mathcal{W}_n)_{n \geq k}$ is a martingale and is bounded in L^2.*

Proof. We start by proving that $(M_n \cdot \mathcal{W}_n)_{n \geq k}$ is a martingale. We introduce $\mathcal{E}_n$ the σ-algebra generated by $M_k, \ldots, M_n$. When going from n to $n+1$ one of two things can happen: either the new vertex of T_{n+1}^S is attached to a vertex $u \in R$ or it is attached to some other vertex of T_n^S. In the first case M_n is multiplied by $\frac{\deg_{T_n^S}(u)+\alpha+\ell(u)-1}{\deg_{T_n^S}(u)+\alpha-1}$; this occurs with probability $\frac{1}{(1+\alpha)n-2}(\deg_{T_n^S}(u)+\alpha-1)$. In the latter case the value of M_n remains unchanged. Thus

$$\frac{\mathbb{E}(M_{n+1}|\mathcal{E}_n)-M_n}{M_n} = \sum_{u \in R} \frac{\deg_{T_n^S}(u)+\alpha-1}{(1+\alpha)n-2} \cdot \left(\frac{\deg_{T_n^S}(u)+\alpha-1+\ell(u)}{\deg_{T_n^S}(u)+\alpha-1} - 1 \right)$$

$$= \frac{|\ell|}{(1+\alpha)n-2}.$$

It follows that $\mathbb{E}(M_{n+1}|\mathcal{E}_n) = (1 + \frac{|\ell|}{(1+\alpha)n-2}) \cdot M_n$, which is to say that $M_n \cdot \mathcal{W}_n$ is a martingale.

That the martingale $M_n \cdot \mathcal{W}_n$ is bounded in L^2 follows from:

$$\forall u \in R, \quad [d+\alpha+\ell(u)-2]^2_{\ell(u)} \leq [d+\alpha+2\ell(u)-2]_{2\ell(u)}$$

and also:

$$(\mathcal{W}_n^{(R,\ell)})^2 \leq \mathcal{W}_n^{(R,2\ell)},$$

where 2ℓ is the decoration of R such that for every $u \in R$, $(2\ell)(u) = 2 \cdot \ell(u)$. Indeed, the two above inequalities together imply that $(M_n^{(R,\ell)} \cdot \mathcal{W}_n^{(R,\ell)})^2 \leq M_n^{(R,2\ell)} \cdot \mathcal{W}_n^{(R,2\ell)}$ for every $n \geq k$. But since $M_n^{(R,2\ell)} \cdot \mathcal{W}_n^{(R,2\ell)}$ is a martingale according to what we have just proved, it is bounded in L^1, thus $M_n \cdot \mathcal{W}_n$ is for its part bounded in L^2. □

Corollary 4.11. *The following convergence holds almost surely and in L^1:*

$$n^{-\frac{|\ell|}{1+\alpha}} \cdot M_n \xrightarrow[n \to \infty]{} C(k, |\ell|) \cdot \xi(R), \tag{46}$$

where $\xi(R)$ is a random variable of expectation $\mathbb{E}[\xi(R)] = \prod_{u \in R}[\deg_S(u)+\ell(u)+\alpha-2]_{\ell(u)}$ and $C(k, |\ell|) > 0$ is a universal constant depending only on k and $|\ell|$.

Proof. As a martingale that is bounded in L^2, $(M_n \cdot \mathcal{W}_n)$ converges a.s. and in L^1 when $n \to \infty$ to some random variable $\xi(R)$. By the L^1 convergence,

$$\mathbb{E}[\xi(R)] = \mathbb{E}[M_k \cdot \mathcal{W}_k] = \prod_{u \in R}[\deg_S(u)+\ell(u)+\alpha-2]_{\ell(u)}.$$

Finally, as in (15), a straightforward computation proves that $\mathcal{W}_n \cdot n^{\frac{|\ell|}{1+\alpha}}$ converges as $n \to \infty$ to some constant depending only on $|\ell|$ and k. This implies (46). □

Remark 4.12. If we apply Corollary 4.11 to R being formed of a single vertex u with $\ell(u) = 1$, we obtain

$$n^{-\frac{1}{1+\alpha}} \deg_{T_n^S}(u) \xrightarrow[n\to\infty]{} C(k,1) \cdot \xi(u),$$

where $\mathbb{E}[\xi(u)] = \deg_S(u) + \alpha - 1$ and $C(k,1) > 0$ is defined as above. Moreover, it may be shown that $\xi(u) > 0$ a.s. (see [17], Lemma 8.17), which is to say that $\deg_{T_n^S}(u)$ grows as $n^{\frac{1}{1+\alpha}}$.

We are now ready to prove Proposition 4.9.

Proof of Proposition 4.9. Fix some S, τ, ℓ and d as in the proposition. Let ϕ be some perfect embedding of (τ, d) in S. Write $R = \phi(\tau)$ and keep in mind that $\deg_S(\phi(u)) = d(u)$ for all $u \in \tau$. We will use the notation of Corollary 4.11 for R.

Using (43) and the fact that $\deg_{T_n^S}(u) \to \infty$ for all $u \in \tau$, we have

$$\frac{\prod_{u\in\tau}[y_n(u) - \alpha]_{\ell(u)}}{M_n} = \prod_{u\in\tau} \frac{[\deg_{T_n^S}(\phi(u)) - 1]_{\ell(u)}}{[\deg_{T_n^S}(\phi(u)) + \ell(u) + \alpha - 2]_{\ell(u)}} \xrightarrow[n\to\infty]{} 1.$$

Moreover, the ratio is always bounded from above by 1. This may be proved individually for each term in the product: it is clear when $\ell(u) \geq 1$; when $\ell(u) = 0$, the ratio is equal to 1. Thus, by the dominated convergence theorem,

$$\lim_{n\to\infty} n^{-\frac{|\ell|}{1+\alpha}} \cdot \mathbf{f}(k,n;d,\ell) = \lim_{n\to\infty} n^{-\frac{|\ell|}{1+\alpha}} \cdot \mathbb{E}[M_n] = C(k,|\ell|) \cdot \prod_{u\in\tau}[d(u) + \ell(u) + \alpha - 2]_{\ell(u)},$$

where the last equality is due to Corollary 4.11. $\square$

4.4. Proof of Theorem 4.1

In proving Theorem 4.1, we analyse differently the case where the two seeds S, S' have same size and that where they have distinct sizes. We start with the former.

Proof of Theorem 4.1 for seeds of common size. Fix S, S' two distinct seeds of same size $k \geq 4$. Let τ be a minimal tree which is not (S, S')-blind. Then, for any decoration ℓ of τ, equation (44) and Proposition 4.9 imply that

$$\lim_{n\to\infty} n^{-\frac{|\ell|}{1+\alpha}} \cdot \left(\mathbb{E}\big[\mathcal{F}_\tau\big(T_n^S\big)\big] - \mathbb{E}\big[\mathcal{F}_\tau\big(T_n^{S'}\big)\big]\right) = \sum_{d:\tau\to\mathbb{N}^*} \mathbf{f}(\infty; d, \ell) \cdot \big[D_{\tau,d}(S) - D_{\tau,d}(S')\big], \quad (47)$$

where $\mathbf{f}(\infty; d, \ell) := \lim_{n\to\infty} n^{-\frac{|\ell|}{1+\alpha}} \cdot \mathbf{f}(k,n;d,\ell) = C(k,|\ell|) \cdot \prod_{u\in\tau}[d(u) + \ell(u) + \alpha - 2]_{\ell(u)}$. Thus, our goal is to find a decoration $\ell : \tau \to \mathbb{N}$ with $\ell(u) \geq 2$ for all $u \in \tau$, $|\ell| > 1 + \alpha$ and such that the right-hand side of (47) is non-zero.

Fix some arbitrary order $u_1, \ldots, u_r$ for the vertices of τ (where $r := |\tau|$) and write $\prec_\ell$ for the lexicographical order induced on the set of decorations of τ. Consider the set Δ of decorations

d of τ with $D_{\tau,d}(S) \neq D_{\tau,d}(S')$. This set is finite, since $D_{\tau,d}(S) = D_{\tau,d}(S') = 0$ for all large enough decorations. Moreover, it is not void since τ is not (S, S')-blind. Let $d_{\max}$ be the maximal element of Δ for $\prec_\ell$.

Consider some $d \in \Delta$ with $d \neq d_{\max}$ and let $j = \min\{i : d(u_i) \neq d_{\max}(u_i)\}$. Then

$$\frac{\mathbf{f}(\infty; d, \ell)}{\mathbf{f}(\infty; d_{\max}, \ell)} = \prod_{i \geq j} \frac{[d(u_i) + \ell(u_i) + \alpha - 2]_{\ell(u_i)}}{[d_{\max}(u_i) + \ell(u_i) + \alpha - 2]_{\ell(u_i)}} \to 0 \qquad (48)$$

when all $\ell(u_i)$ for $i > j$ are fixed and $\ell(u_j) \to \infty$.

Write Δ_j for the set of decorations $d \in \Delta$ with $d \neq d_{\max}$ and such that $\min\{i : d(u_i) \neq d_{\max}(u_i)\} = j$. Then $\Delta \setminus \{d_{\max}\} = \bigsqcup_{j=1}^r \Delta_j$. Now let us construct a decoration ℓ with the necessary requirements by first choosing $\ell(u_r)$, then $\ell(u_{r-1})$ etc.

First, fix $\ell(u_r) > 1 + \alpha$ large enough so that

$$\left| \sum_{d \in \Delta_r} \frac{\mathbf{f}(\infty; d, \ell)}{\mathbf{f}(\infty; d_{\max}, \ell)} \cdot \left[D_{\tau,d}(S) - D_{\tau,d}(S') \right] \right| \leq \frac{1}{2r}.$$

This is possible by (48). Observe that the values of $\ell(u_i)$ for $i < r$ are irrelevant, as they do not appear in the above. Once $\ell(u_r)$ is fixed, fix $\ell(u_{r-1}) \geq 2$ so that

$$\left| \sum_{d \in \Delta_{r-1}} \frac{\mathbf{f}(\infty; d, \ell)}{\mathbf{f}(\infty; d_{\max}, \ell)} \cdot \left[D_{\tau,d}(S) - D_{\tau,d}(S') \right] \right| \leq \frac{1}{2r}.$$

Again, (48) shows that the values $\{\ell(u_i); i < r - 1\}$ do not appear in the above and that such a choice of $\ell(u_{r-1})$ is possible regardless of the value of $\ell(u_r)$ fixed before.

Continue as such until all values $\{\ell(u) : u \in \tau\}$ are fixed. The resulting decoration ℓ has $\ell(u) \geq 2$ for all $u \in \tau$, it is such that $|\ell| \geq \ell(u_r) > 1 + \alpha$, and satisfies

$$\left| \sum_{d \in \Delta_j} \frac{\mathbf{f}(\infty; d, \ell)}{\mathbf{f}(\infty; d_{\max}, \ell)} \cdot \left[D_{\tau,d}(S) - D_{\tau,d}(S') \right] \right| \leq \frac{1}{2r}, \quad \forall j = 1, \ldots, r.$$

When summing the above, we find

$$\left| \sum_{d \in \Delta \setminus \{d_{\max}\}} \mathbf{f}(\infty; d, \ell) \cdot \left[D_{\tau,d}(S) - D_{\tau,d}(S') \right] \right| \leq \frac{1}{2} \mathbf{f}(\infty; d_{\max}, \ell).$$

Now, since $D_{\tau,d_{\max}}(S) \neq D_{\tau,d_{\max}}(S')$, this implies that

$$\left| \sum_{d:\tau \to \mathbb{N}^*} \mathbf{f}(\infty; d, \ell) \cdot \left[D_{\tau,d}(S) - D_{\tau,d}(S') \right] \right| \geq \frac{1}{2} \mathbf{f}(\infty; d_{\max}, \ell) > 0,$$

which was the desired condition. $\qquad \square$

Proof of Theorem 4.1 for seeds of different sizes. Fix two seeds S, S' of sizes k and k', respectively, with $3 \leq k' < k$. Due to the Markov property, the law of $(T_n^{S'})_{n \geq k}$ is a linear combination

of laws $(T_n^R)_{n \geq k}$ with R ranging over the different values taken by $T_k^{S'}$. Write $\mathcal{T}_k$ for the set of trees of size k with $\mathbb{P}(T_k^{S'} = R) \neq 0$.

First we claim that there exist two trees in $\mathcal{T}_k$ with different maximal degree. Indeed, one possible way of going from S' to a tree in $\mathcal{T}_k$ is to always attach the new vertices to leaves. In this scenario, the maximal degree of the resulting tree is the same as that of S' (since that in S' is at least 2). In conclusion

$$\mathbb{P}\big(\max\{\deg_{T_k^{S'}}(u) : u \in T_k^{S'}\} = \max\{\deg_{S'}(u) : u \in S'\}\big) > 0.$$

Another is to always attach the new vertex to the one of maximal degree. The resulting maximal degree in $T_k^{S'}$ would then be $k - k'$ more than that in S':

$$\mathbb{P}\big(\max\{\deg_{T_k^{S'}}(u) : u \in T_k^{S'}\} = \max\{\deg_{S'}(u) : u \in S'\} + k - k'\big) > 0.$$

Thus, there exist two trees in $\mathcal{T}_k$ with distinct maximal degree, as claimed.

Fix τ to be the tree formed of a single vertex. A decoration for τ is then simply an integer number. This tree has no subtree, hence Proposition 4.5 applies to it. For any decoration $\ell \in \mathbb{N}$ of τ, equation (47) adapts to

$$\lim_{n \to \infty} n^{-\frac{\ell}{1+\alpha}} \big(\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})]\big) = \sum_{d \in \mathbb{N}^*} \mathbf{f}(\infty; d, \ell) \sum_{R \in \mathcal{T}_k} \mathbb{P}(T_k^{S'} = R)[D_d(S) - D_d(R)],$$

where $\mathbf{f}(\infty; d, \ell) = C(k, \ell) \cdot [d + \ell + \alpha - 2]_\ell$ and $D_d(S)$ is simply the number of vertices of degree d in S.

Write $d_{\max}$ for the maximal number $d \in \mathbb{N}^*$ with

$$\sum_{R \in \mathcal{T}_k} \mathbb{P}(T_k^{S'} = R) \cdot [D_d(S) - D_d(R)] \neq 0. \tag{49}$$

First, observe that the set of values d satisfying (49) is finite since $D_d(S) = D_d(R) = 0$ for all $R \in \mathcal{T}_k$ provided d is large enough. Second, notice that there exist at least one d satisfying (49) since the maximal degree of $R \in \mathcal{T}_k$ is not constant.

Now, by (48),

$$\lim_{\ell \to \infty} \sum_{d \in \mathbb{N}^*} \frac{\mathbf{f}(\infty; d, \ell)}{\mathbf{f}(\infty; d_{\max}, \ell)} \sum_{R \in \mathcal{T}_k} \mathbb{P}(T_k^{S'} = R) \cdot [D_d(S) - D_d(R)]$$

$$= \sum_{R \in \mathcal{T}_k} \mathbb{P}(T_k^{S'} = R) \cdot [D_{d_{\max}}(S) - D_{d_{\max}}(R)] \neq 0.$$

In conclusion, one may fix $\ell > 1 + \alpha$ (and implicitly $\ell \geq 2$) so that

$$\lim_{n \to \infty} n^{-\frac{\ell}{1+\alpha}} \big(\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})]\big) \neq 0. \qquad \square$$

4.5. Proof of Theorem 1.1

Proof. Fix S, S' two distinct seeds of sizes at least 3. Let τ be the decorated tree given by Theorem 4.1 for these two seeds. Due to our assumptions on τ, Theorem 3.1 applies to it, and we have

$$\mathbb{E}\big[\mathcal{F}_{\tau}\big(T_n^S\big)^2\big] = \mathcal{O}\big(n^{\frac{2|\ell|}{1+\alpha}}\big) \quad \text{and} \quad \mathbb{E}\big[\mathcal{F}_{\tau}\big(T_n^{S'}\big)^2\big] = \mathcal{O}\big(n^{\frac{2|\ell|}{1+\alpha}}\big).$$

Lemma 1.3 then yields $\liminf_{n\to\infty} d_{\text{TV}}(\mathcal{F}_{\tau}(T_n^S), \mathcal{F}_{\tau}(T_n^{S'})) > 0$, which in turn implies Theorem 1.1. $\qquad\square$

5. Open problems and future research

Other attachment mechanisms

As mentioned in the Introduction, our approach uses the fact that the attachment mechanism is affine to couple the evolution of trees starting from distinct seeds of same size; see the coupling of Section 2.2 and its use in Section 4.1 and 4.2 to deduce Corollary 4.6.

Suppose now that we consider a different attachment model, where the new vertex is attached to a vertex $u \in T_n$ with probability proportional to $g(\deg_{T_n}(u))$ for some function $g : \mathbb{N} \to (0, +\infty)$. Thus, (1) becomes

$$\mathbb{P}\big(u_N = u | T_k^S, \dots, T_N^S \text{ with } T_N^S = T\big) = \frac{g(\deg_T(u))}{\sum_{v \in T} g(\deg_T(v))} \quad \forall u \in T.$$

Notice that, when g is not affine, the denominator above depends on the structure of T, and the attachment probabilities cease to be a local function of u. In other words, the sequence of drawn vertices loses its exchangeability. As a consequence, it is not possible anymore to construct a coupling between sequences $(T_n^S)_{n \geq k}$ and $(T_n^{S'})_{n \geq k}$ as in Section 2.2, where S and S' are two seeds of size k. Nevertheless, if S and S' have same degree sequences, then regardless of the form of g, one may couple the evolution of T_n^S and $T_n^{S'}$ so that the sets of connected components of $T_n^S \setminus S$ and $T_n^{S'} \setminus S'$ are identical (here $T_n^S \setminus S$ stands for the graph obtained from T_n^S by removing all edges of S). The equivalent of Corollary 4.6 may then be proved in the same manner.

This hints to the possibility of seed recognition (*i.e.* Theorem 1.1) for a much larger set of models. One particular example of interest is when $g(k) = k^\beta$ for some $\beta \in (0, 1)$ [10, 11]. In such models, the largest degree in T_n^S is of order $(\log n)^{\frac{1}{1-\beta}}$, hence much smaller than in the affine case [3], Thm. 22. The case $\beta = 1$ is that of the linear preferential attachment model treated in [8]; when $\beta = 0$ we obtain the uniform attachment model of [6]. For $\beta \in (0, 1)$ we expect the same type of result to hold, and plan to investigate this in future work.

Let us also mention that, when $\beta > 1$ a single vertex of T_n^S has degree tending to infinity, all other degrees are a.s. bounded by a constant, as is the diameter of T_n^S [16], Thm. 1.2.

Finding the seed

Our result may be understood as follows: given a large (but uniform in n) number of samples of T_n^S, one may recover S with high precision. A related question is to locate S given a single instance of T_n^S. One may not hope to do this with high probability, but is it possible to do it with uniformly positive probability? Results in this direction were obtained in [5,9,12] for the uniform and linear preferential attachment models.

One may ask whether having a uniformly positive chance of locating the seed is equivalent to Theorem 1.1. Moreover, for specific cases such as the α-PA, is there an explicit algorithm that locates the seed?

Beyond trees

Finally, one may consider randomly growing graphs, rather than trees. Indeed, imagine a model where vertices are added one by one to a growing graph, with each new vertex being attached to each old vertex independently, with a probability depending on the size of the graph and on the degree of the old vertex. This offers great freedom for the choice of the attachment probability, but the resulting graph ceases to be a tree; it may even have multiple connected components. These aspects render the study of such models more delicate. However, we believe that for certain attachment mechanisms – affine for instance – an equivalent of Theorem 1.1 would remain valid.

Acknowledgements

References

[1] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634 https://doi.org/10.1126/science.286.5439.509

[2] Berger, N., Borgs, C., Chayes, J.T. and Saberi, A. (2005). On the spread of viruses on the internet. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 301–310. New York: ACM. MR2298278

[3] Bhamidi, S. (2007). Universal techniques to analyze preferential attachment trees: Global and Local analysis. preprint.

[4] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18** 279–290. MR1824277 https://doi.org/10.1002/rsa.1009

[5] Bubeck, S., Devroye, L. and Lugosi, G. (2017). Finding Adam in random growing trees. *Random Structures Algorithms* **50** 158–172. MR3607120 https://doi.org/10.1002/rsa.20649

[6] Bubeck, S., Eldan, R., Mossel, E. and Rácz, M.Z. (2017). From trees to seeds: On the inference of the seed from large trees in the uniform attachment model. *Bernoulli* **23** 2887–2916. MR3648049 https://doi.org/10.3150/16-BEJ831

[7] Bubeck, S., Mossel, E. and Rácz, M.Z. (2015). On the influence of the seed graph in the preferential attachment model. *IEEE Trans. Netw. Sci. Eng.* **2** 30–39. MR3361606 https://doi.org/10.1109/TNSE.2015.2397592

[8] Curien, N., Duquesne, T., Kortchemski, I. and Manolescu, I. (2015). Scaling limits and influence of the seed graph in preferential attachment trees. *J. Éc. Polytech. Math.* **2** 1–34. MR3326003 https://doi.org/10.5802/jep.15

[9] Devroye, L. and Reddad, T. (2019). On the discovery of the seed in uniform attachment trees. *Internet Math.* 1–29. MR3934290

[10] Krapivsky, P.L. and Redner, S. (2001). Organization of growing random networks. *Phys. Rev. E* **63** 06.

[11] Krapivsky, P.L., Redner, S. and Leyvraz, F. (2000). Connectivity of growing random networks. *Phys. Rev. Lett.* **85** 4629.

[12] Lugosi, G. and Pereira, A.S. (2019). Finding the seed of uniform attachment trees. *Electron. J. Probab.* **24** Paper No. 18, 15. MR3925458 https://doi.org/10.1214/19-EJP268

[13] Middendorf, M., Ziv, E. and Wiggins, C.H. (2005). Inferring network mechanisms: The drosophila melanogaster protein interaction network. *Proc. Natl. Acad. Sci. USA* **102** 3192–3197.

[14] Móri, T.F. (2002). On random trees. *Studia Sci. Math. Hungar.* **39** 143–155. MR1909153 https://doi.org/10.1556/SScMath.39.2002.1-2.9

[15] Newman, M. (2018). *Networks*. Oxford: Oxford Univ. Press. Second edition of [MR2676073]. MR3838417 https://doi.org/10.1093/oso/9780198805090.001.0001

[16] Oliveira, R. and Spencer, J. (2005). Connectivity transitions in networks with super-linear preferential attachment. *Internet Math.* **2** 121–163. MR2193157

[17] van der Hofstad, R. (2017). *Random Graphs and Complex Networks. Vol. 1. Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge: Cambridge Univ. Press. MR3617364 https://doi.org/10.1017/9781316779422

[18] Wang, M., Yu, G. and Yu, D. (2008). Measuring the preferential attachment mechanism in citation networks. *Phys. A* **387** 4692–4698.