# Scaling limits and influence of the seed graph in preferential attachment trees

Nicolas Curien♠ Thomas Duquesne♥ Igor Kortchemski♦ and Ioan Manolescu♣

Abstract

We are interested in the asymptotics of random trees built by linear preferential attachment, also known in the literature as Barabási–Albert trees or plane-oriented recursive trees. We first prove a conjecture of Bubeck, Mossel & Rácz [8] concerning the influence of the seed graph on the asymptotic behavior of such trees. Separately we study the geometric structure of nodes of large degrees in a plane version of Barabási–Albert trees *via* their associated looptrees. As the number of nodes grows, we show that these looptrees, appropriately rescaled, converge in the Gromov–Hausdorff sense towards a random compact metric space which we call the Brownian looptree. The latter is constructed as a quotient space of Aldous' Brownian Continuum Random Tree and is shown to have almost sure Hausdorff dimension 2.
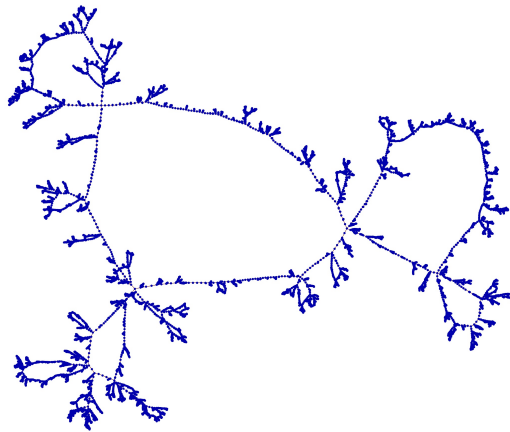
Figure 1: The looptree associated with a large plane Barabási–Albert tree.

♠CNRS and LPMA, Université Pierre et Marie Curie (Paris 6).   nicolas.curien@gmail.com
♥LPMA, Université Pierre et Marie Curie (Paris 6).   thomas.duquesne@upmc.fr
♦DMA, École Normale Supérieure.   igor.kortchemski@normalesup.org
♣Département de Mathématiques, Université de Genève.   ioan.manolescu@unige.ch

1 Introduction

Random graphs constructed recursively by preferential attachment rules have attracted a lot of attention in the last decade. They are sensible models for many real-world networks, and have the remarkable scale-free property, meaning that their degree distribution exhibits a power law behavior. The literature on the subject is extremely vast, and we refer to [29] for an overview and references.

In this work, we focus on the simplest and the best known of these models, the linear preferential attachment model (LPAM in short). Starting with a finite tree T_1 (i.e. a finite connected graph without cycles, considered up to graph isomorphisms), one constructs recursively a sequence of random trees $T_1, T_2, \ldots$ by requiring that for $i \geqslant 1$, the tree T_{i+1} is obtained from the tree T_i by joining with an edge a new vertex with a random vertex of T_i, chosen proportionally to its degree. These trees are also known in the literature as plane-oriented recursive trees. This model was introduced by Szymánski [28], and generalized and popularized by Albert & Barabási [5] and Bollobás, Riordan, Spencer & Tusnády [7].

This work concerns two related aspects of the LPAM. First we investigate the influence of the initial tree (also called the seed) on the behavior of T_n as $n \to \infty$. Next we study the graph structure of T_n as $n \to \infty$ by studying its associated looptree (see below for the definition of a looptree associated with a tree).

Influence of the seed graph.   Bubeck, Mossel and Rácz [8] recently raised the question of the influence of the initial tree on the large time behavior of the LPAM. More precisely, given a tree S with $|S| = n_0 \geqslant 2$ vertices, consider the sequence of trees $(T_n^{(S)})_{n \geqslant n_0}$ constructed by using the previously mentioned preferential attachment rule and starting with $T_{n_0} = S$. The tree S is called the *seed graph*. Informally, the question is whether the seed graph can be determined from the law of T_n for large values of n. Following [8], for finite trees S_1 and S_2, set

$$d(S_1, S_2) = \lim_{n \to \infty} d_{\mathrm{TV}}(T_n^{(S_1)}, T_n^{(S_2)}),$$

where d_{TV} denotes the total variation distance for random variables taking values in the space of finite trees. Bubeck, Mossel and Rácz [8] have observed that d is a pseudo-metric and have conjectured that d is a metric in non-trivial cases. We confirm this conjecture:

**Theorem 1.** *The function d is a metric on trees with at least 3 vertices.*

Observe that Theorem 1 means that $d_{\mathrm{TV}}(T_n^{(S_1)}, T_n^{(S_2)})$ remains bounded away from 0 as $n \to \infty$, as soon as the two seeds S_1, S_2 are different and consist of at least 3 vertices. In [8] this is proved for seeds with different degree sequences by studying the asymptotic behavior of the tail of the degrees of the vertices of $T_n^{(S)}$, and the authors notice that additional information concerning the graph structure has to be incorporated to solve the general case. To this end, they suggest to study the maximum of the sum of the degrees over all embeddings of a fixed tree in $T_n^{(S)}$.

In order to establish Theorem 1, we design another family of "observables" of $T_n^{(S)}$, indexed by finite trees τ which roughly correspond to the total number of possible embeddings of a given tree τ into $T_n^{(S)}$. Using these variables, we then construct a family of martingales such that their laws differ asymptotically for different seed graphs.

Although rather implicit in our proof of Theorem 1, the underlying key feature of the LPAM is the geometric structure induced by the nodes of large degree in $T_n^{(S)}$. It is known that the maximal degree in $T_n^{(S)}$ is of order $\sqrt{n}$ (see e.g. [24]) and that there is a tight number of vertices with degree of this order. Roughly speaking, the geometric tree structure induced by these vertices is captured by the martingales constructed for the proof of Theorem 1. In this spirit, our second main result is devoted to giving a precise sense to the continuous scaling limit of this structure through the looptree associated with $T_n^{(S)}$. As we will see below, the looptree of a plane tree encodes in a natural way the geometric structure of nodes with exceptionally large degree. Let us mention that this approach differs with that of recent works on local limits of preferential attachment graphs describing the (sparse) neighborhoods of typical vertices [6, 15].

Scaling limits of looptrees. For our next results, we consider the planar version of the LPAM. For a plane (i.e. embedded in the plane) tree S with n_0 vertices, consider the sequence of random plane trees $(\overline{T}_n^{(S)})_{n \geqslant n_0}$ defined by $\overline{T}_{n_0}^{(S)} = S$ and, for $n \geqslant n_0$, conditionally on $\overline{T}_{n_0}^{(S)}, \ldots, \overline{T}_n^{(S)}$, $\overline{T}_{n+1}^{(S)}$ is obtained by grafting an edge leading to a new vertex inside a uniformly chosen corner of $\overline{T}_n^{(S)}$ (by definition, a corner is an angular sector in the plane formed by two consecutive half-edges around a vertex). Since the number of corners around a vertex is equal to its degree, it is immediate that the tree structure of $(\overline{T}_n^{(S)})_{n \geqslant n_0}$ is that of a LPAM. Thus we no longer distinguish between $\overline{T}_n^{(S)}$ and $T_n^{(S)}$.

The plane embedding of the LPAM allows us to consider its associated looptree. The notion of looptree was introduced in [14] (see also [13] for the appearance of looptrees in the context of random maps). Informally speaking, the looptree $\mathrm{Loop}(\tau)$ of a plane tree τ is the graph constructed by replacing each vertex $u \in \tau$ by a discrete cycle of length given by the degree of u in τ and gluing these cycles according to the tree structure of τ, see Fig. 2. See [14] for a formal definition. One may view $\mathrm{Loop}(\tau)$ as a compact metric space by endowing the set of its vertices with the graph distance.
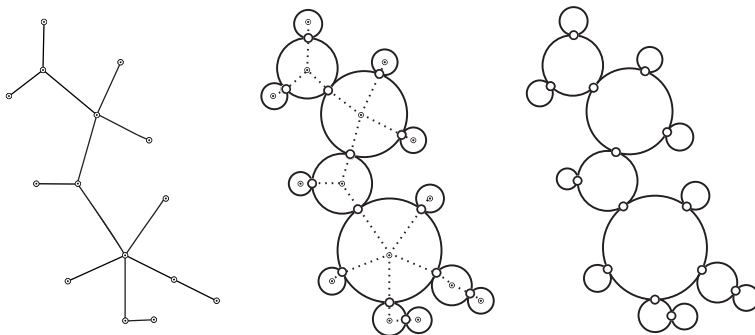
Figure 2: An example of the looptree associated with a plane tree.

We will show that, for a plane tree S, the sequence of compact metric spaces $(\mathrm{Loop}(T_n^{(S)}))_{n\geqslant|S|}$, suitably rescaled by a factor $n^{-1/2}$, converges towards a random compact metric space. The latter convergence is almost sure with respect to the Gromov–Hausdorff topology of compact metric spaces; see Section 3.3 for background.

It will be useful to consider for a start the case of the particular seed graph $\multimap$ consisting of a single vertex with a unique corner. Formally, $\multimap$ is a *planted* tree. By definition, a tree τ is planted if a distinguished half-edge is attached to a vertex of τ (thus increasing the degree of this vertex by one and adding a corner to it). One defines the sequence of random planted plane trees $(T_n^{\multimap})_{n\geqslant 1}$ by the preferential attachment rule described above, starting with the seed graph $\multimap$ (to simplify notation we write $T_n^{\multimap}$ instead of $T_n^{(\multimap)}$). See Fig. 3 for an illustration.
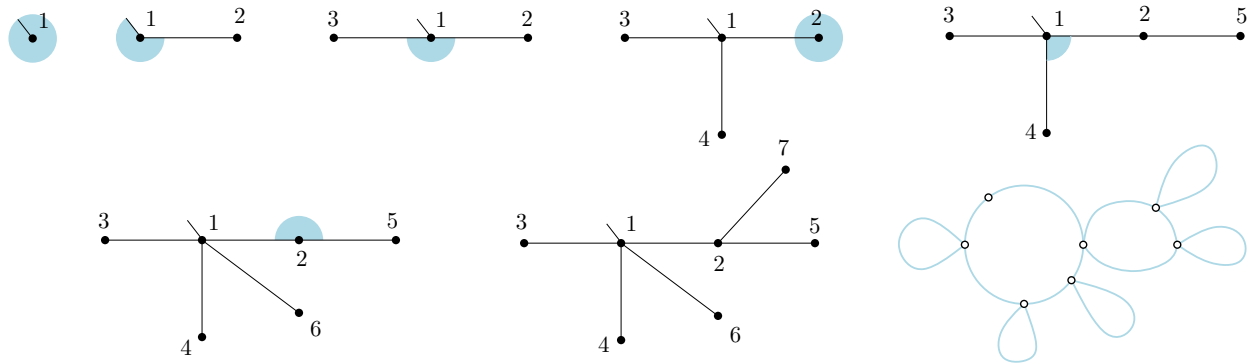
Figure 3: Illustration of the growth mechanism of the plane LPAM started from $\multimap$. At each step, the corner in which a new edge is grafted is highlighted in light blue. The last diagram is the looptree of the last planted tree displayed.

The looptree associated with a planted tree τ is the looptree of the tree obtained by adding a new vertex to the endpoint of the half-edge of τ, but where the self-loop surrounding this new vertex is removed, see Fig. 3 for an example. If $\mathcal{M}$ is a metric space, we write $c \cdot \mathcal{M}$ for the metric space obtained from $\mathcal{M}$ by multiplying all distances by $c > 0$. Our second main result is the following.

Theorem 2. *The following convergence holds almost surely in the Gromov–Hausdorff topology*

$$ n^{-1/2} \cdot \mathrm{Loop}(T_n^{\multimap}) \ \xrightarrow[n\to\infty]{a.s.} \ 2\sqrt{2} \cdot \mathcal{L}, $$

where $\mathcal{L}$ is a random compact metric space called the Brownian looptree.

Remark. It is natural to scale $\mathrm{Loop}(T_n^{\multimap})$ by a factor $n^{-1/2}$ in order to obtain a non-degenerate limiting compact metric space. Indeed, lengths of loops in $\mathrm{Loop}(T_n^{\multimap})$ correspond to vertex degrees of $T_n^{\multimap}$, and it is well-known that the maximum degree of $T_n^{\multimap}$ is of order $\sqrt{n}$ (see e.g. [24]). Moreover the diameter of $T_n^{\multimap}$ is of much lower order, namely of order $\log(n)$ (see e.g. [29, Sec. 11]). In light of the above, it is not surprising that looptrees associated with $T_n^{\multimap}$ admit a nontrivial scaling limit, while the trees $T_n^{\multimap}$ themselves do not.

The metric space $\mathcal{L}$ is constructed as a quotient of the Brownian Continuum Random Tree (in short the CRT) which was introduced by Aldous in [2]. Let us give for the moment a heuristic construction of $\mathcal{L}$. Denote by $\mathcal{T}_\mathbf{e}$ the CRT obtained from a Brownian excursion $\mathbf{e}$ (see e.g. [19, Sec. 2]). This random tree supports a natural mass measure μ. This is a probability measure on $\mathcal{T}_\mathbf{e}$ and is supported by the leaves of $\mathcal{T}_\mathbf{e}$. Denote by $(X_i)_{i \geqslant 0}$ a sequence of i.i.d. points sampled according to μ. For every $n \geqslant 2$, consider the subtree $\mathsf{Span}(\mathcal{T}_\mathbf{e}; X_0, \ldots, X_n)$ of $\mathcal{T}_\mathbf{e}$ spanned by $X_0, X_1, \ldots, X_n$ (see Sec. 3.2 for a precise definition). Denote by P_n the point in $\mathsf{Span}(\mathcal{T}_\mathbf{e}; X_0, \ldots, X_{n-1})$ which is the closest to X_n, see Fig. 4. Set also $P_1 = X_0$. Informally, the compact metric space $\mathcal{L}$ is obtained from $\mathcal{T}_\mathbf{e}$ by making the point identifications $X_n \sim P_n$ for every $n \geqslant 1$. See Section 3 for the rigorous construction. Let us also mention that the Brownian looptree $\mathcal{L}$ may be constructed by a line-breaking cycle-creating procedure, similar to Aldous' line-breaking construction of the Brownian CRT (see Remark 12).
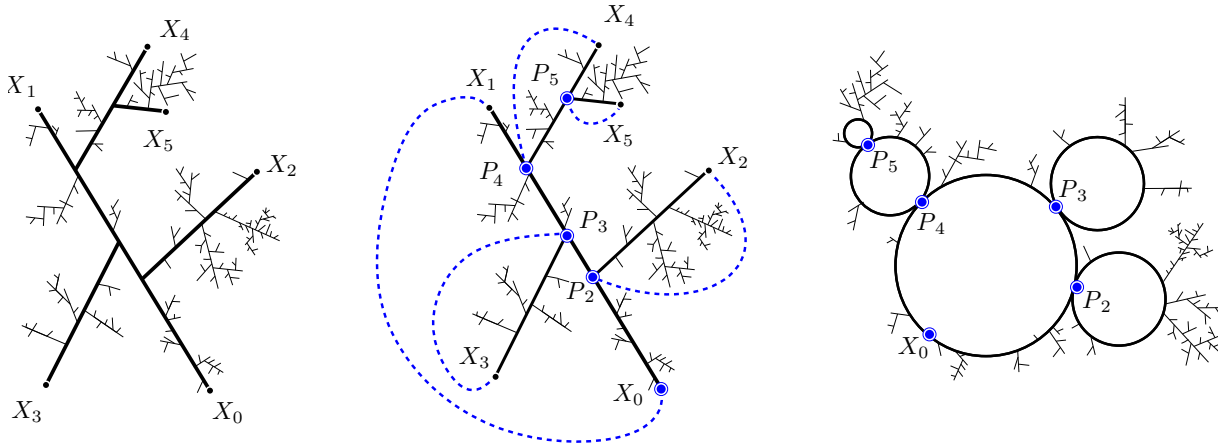
Figure 4: The metric space obtained from $\mathcal{T}_\mathbf{e}$ by making the identifications $X_n \sim P_n$ for $1 \leqslant n \leqslant 5$.

Since the sizes of loops in $\mathsf{Loop}(T_n^{-\circ})$ correspond to vertex degrees in $T_n^{-\circ}$, $\mathcal{L}$ contains the limiting joint distribution of the scaled degrees in $T_n^{-\circ}$. This distribution has been computed in [25] and asymptotic estimates on its tails studied in [8]. But $\mathcal{L}$ incorporates additional information concerning the graph structure of $T_n^{-\circ}$.

An important tool is a coupling between the LPAM and Rémy's algorithm [27] which appeared in [25]. The proof of Theorem 2 combines this coupling with the convergence of scaled uniform binary trees towards the Brownian CRT.

For planar LPAM's starting with a generic seed graph S, we obtain as consequence of Theorem 1 a convergence similar to that for the seed $-\circ$.

Corollary 3. *For any plane tree S there exists a random compact metric space $\mathcal{L}^{(S)}$ such that following convergence holds almost surely for the Gromov–Hausdorff topology*

$$n^{-1/2} \cdot \mathsf{Loop}(T_n^{(S)}) \xrightarrow[n \to \infty]{a.s.} 2\sqrt{2} \cdot \mathcal{L}^{(S)}.$$

The limiting metric space $\mathcal{L}^{(S)}$ is constructed by gluing weighted i.i.d. copies of $\mathcal{L}$; see Section 3.4 for details. In light of Theorem 1, we expect that if $S_1 \neq S_2$ are different seed graphs with at least three vertices, then the laws of $\mathcal{L}^{(S_1)}$ and of $\mathcal{L}^{(S_2)}$ are different and we further conjecture (see Section 3.4) that the distance $d(\cdot, \cdot)$ appearing in Theorem 1 can be expressed as

$$d(S_1, S_2) = d_{\mathrm{TV}}(\mathcal{L}^{(S_1)}, \mathcal{L}^{(S_2)}). \tag{1}$$

Several other random compact metric spaces have been constructed as quotients of the Brownian CRT and appear as limits of discrete structures. For example, the scaling limit of the connected components of the Erdős-Rényi random graph is described by a tilted Brownian CRT with a finite number of point identifications [1]. These conserve many of the properties of the CRT (such as the Hausdorff dimension equal to 2). Another example is the Brownian map, obtained from the CRT by gluing a continuum number of points using additional randomness involving Brownian motion indexed by the CRT, see [20]. In this case, the structure of the metric space is drastically altered by the identifications, and it is known that the Brownian map is almost surely homeomorphic to the sphere and has Hausdorff dimension 4 (see [20]). The Brownian looptree is, in some sense, in-between the two examples above, since it involves a countable number of point identifications in the CRT, which change completely its topological structure, but conserve the Hausdorff dimension of the CRT.

Proposition 4. *Almost surely, the Hausdorff dimension of $\mathcal{L}$ is 2.*

We mention that in [14], a related one-parameter family of random compact metric spaces $(\mathscr{L}_\alpha)_{\alpha \in (1,2)}$ has been constructed. They are called stable looptrees, and appear as scaling limits of discrete loop-trees associated with large critical Galton–Watson trees whose offspring distribution belongs to the domain of attraction of an α-stable law. The Brownian looptree introduced in this work differs substantially from stable looptrees. For example, in the Brownian looptree, large loops are adjacent, while in $\mathscr{L}_\alpha$ large loops are connected through infinitely many microscopic loops. In addition, in [14] it is shown that the Hausdorff dimension of $\mathscr{L}_\alpha$ is almost surely $\alpha < 2$.

We believe that, as illustrated by Theorem 2, looptrees are an interesting means to give a sense to scaling limits of highly dense random trees. See in particular, Section 4.1 for a conjecture concerning affine preferential attachment models and random trees built by Ford's algorithm. We hope to pursue this line of research in future work.

Outline. The paper is organized as follows. Sections 2 and 3 establish Theorems 1 and 2, respectively. In Section 2, we first define the observables that we use, then prove Theorem 1. In Section 3 we start by presenting the connection between the plane LPAM and Rémy's algorithm, then construct the Brownian looptree from the Brownian CRT and prove Theorem 2 and its corollary. We end the section with the computation of the Hausdorff dimension of the Brownian looptree. These two sections are largely independent. Finally, in Section 4, we propose several extensions and generalizations.

2 Influence of the seed graph

In this section, we assume that the LPAM is started from a seed graph S, which is a (non planted) tree with at least two vertices. In particular, the total degree of $T_n^{(S)}$ (that is the sum of the degrees of all its vertices) is always equal to $2n - 2$.

2.1 Decorated trees

A decorated tree is a pair $\boldsymbol{\tau} = (\tau, \ell)$ consisting of a tree τ and a family of positive integers $(\ell(u); u \in \tau)$ carried by its vertices. We denote by $|\boldsymbol{\tau}|$ the total number of vertices of τ and set $w(\boldsymbol{\tau}) := \sum_{u \in \tau} \ell(u)$ to be the total weight of $\boldsymbol{\tau}$. We insist on the fact that $\ell(u) > 0$ for every $u \in \tau$.

Let $\mathcal{D}$ be the set of all decorated trees. For $\boldsymbol{\tau}, \boldsymbol{\tau}' \in \mathcal{D}$, we write $\boldsymbol{\tau} \prec \boldsymbol{\tau}'$ if $w(\boldsymbol{\tau}) < w(\boldsymbol{\tau}')$ and $|\boldsymbol{\tau}| \leqslant |\boldsymbol{\tau}'|$ or if $w(\boldsymbol{\tau}) = w(\boldsymbol{\tau}')$ and $|\boldsymbol{\tau}| < |\boldsymbol{\tau}'|$. Thus $\prec$ is a strict partial order on $\mathcal{D}$ and we denote by $\preccurlyeq$ the associated partial order.

We now define the observables which will be used to identify the seed of a LPAM. For $k, j \geqslant 1$, write $[k]_j = k(k-1) \cdots (k-j+1)$. If τ, T are trees, we say that a map $\phi : \tau \to T$ is an embedding if ϕ is an injective graph homomorphism. For a decorated tree $\boldsymbol{\tau}$, set

$$D_{\boldsymbol{\tau}}(T) = \sum_{\phi} \prod_{u \in \tau} [\deg_T \phi(u)]_{\ell(u)},$$

where the sum is taken over all embeddings $\phi : \tau \to T$ and where $\deg_T(x)$ denotes the degree of a vertex $x \in T$. When $\boldsymbol{\tau} = \textcircled{1}$ is the decorated tree formed of a single vertex with label one, $D_{\textcircled{1}}(T)$ is just the total degree of T. Theorem 1 is a consequence of the following proposition.

Proposition 5. *Let $\boldsymbol{\tau}$ be a decorated tree. There exist constants $\{c_n(\boldsymbol{\tau}, \boldsymbol{\tau}') : \boldsymbol{\tau}' \preccurlyeq \boldsymbol{\tau}, n \geqslant 2\}$ with $c_n(\boldsymbol{\tau}, \boldsymbol{\tau}) > 0$ such that, for every seed S, the process $(M_{\boldsymbol{\tau}}^{(S)}(n))_{n \geqslant n_0}$ defined by*

$$M_{\boldsymbol{\tau}}^{(S)}(n) = \sum_{\boldsymbol{\tau}' \preccurlyeq \boldsymbol{\tau}} c_n(\boldsymbol{\tau}, \boldsymbol{\tau}') \cdot D_{\boldsymbol{\tau}'}(T_n^{(S)})$$

is a martingale with respect to the filtration $\mathcal{F}_n = \sigma(T_{n_0}^{(S)}, \ldots, T_n^{(S)})$ and is bounded in $\mathbb{L}^2$.

Remark. Rather than the quantities $D_{\boldsymbol{\tau}}$ defined above, a more natural family of observables to consider are the number $E_{\boldsymbol{\tau}}$ of embeddings of a tree τ inside T. These observables could indeed be used to distinguish between seeds of the LPAM (the martingales M of Proposition 5 could be written in terms of $E_{\boldsymbol{\tau}}$ only). However, as we will see, the main advantage of the observables $D_{\boldsymbol{\tau}}$ is that they are more amenable for recurrence relations (see Lemma 6).

The quantity $D_{\boldsymbol{\tau}}(T)$ has a special interpretation for plane trees T. Imagine that there are $\ell(u)$ distinguishable arrows pointing to each vertex $u \in \tau$. Then $D_{\boldsymbol{\tau}}(T)$ is the number of ways to embed τ in T in such a way that each arrow pointing to a vertex of τ is associated with a corner of T adjacent

to the corresponding vertex, with distinct arrows associated with distinct corners. We call this type of embeddings decorated embeddings.

Proposition 5 is the main ingredient in the proof of Theorem 1; its proof occupies the following subsections. Before, let us explain how to deduce Theorem 1 from Proposition 5.

Proof of Theorem 1. Let $S_1 \neq S_2$ be two distinct trees with at least 3 vertices. We claim that if $n_0 = \max(|S_1|, |S_2|)$, then there exists a deterministic decorated tree τ such that

$$\mathbb{E}\left[D_{\tau}(T_{n_0}^{(S_1)})\right] \quad \neq \quad \mathbb{E}[D_{\tau}(T_{n_0}^{(S_2)})]. \tag{2}$$

To see this, suppose by symmetry that $|S_1| \leqslant |S_2|$ and set $S_1' = T_{|S_2|}^{(S_1)}$. Thus S_1' is a random tree when $|S_1| < |S_2|$. If we take $\tau = S_2$ with labels $\ell(u) = \deg(u)$, then, for every tree T with $|T| = |S_2|$, we have $D_{\tau}(T) = D_{\tau}(S_2) \cdot \mathbb{1}_{\{T=S_2\}}$. Consequently, for this particular value of τ,

$$\mathbb{E}\left[D_{\tau}(S_1')\right] = D_{\tau}(S_2) \cdot \mathbb{P}[S_1' = S_2].$$

When $|S_1| = |S_2|$ the above probability is 0. When $|S_1| < |S_2|$ it may easily be checked that S_1' is non-deterministic (here it is essential that $|S_1| \geqslant 3$), hence the probability above is strictly less than 1. In both cases (2) holds for this choice of τ.

Let τ be a minimal (for the partial order $\preccurlyeq$) decorated tree for which (2) holds. Then $\mathbb{E}[D_{\tau'}(T_{n_0}^{(S_1)})] = \mathbb{E}[D_{\tau'}(T_{n_0}^{(S_2)})]$ for all $\tau' \prec \tau$ and it follows that

$$\mathbb{E}\left[M_{\tau}^{(S_1)}(n_0)\right] \neq \mathbb{E}\left[M_{\tau}^{(S_2)}(n_0)\right],$$

where $M_{\tau}^{(S_1)}$ and $M_{\tau}^{(S_2)}$ are martingales as in Proposition 5. To simplify notation, set $M_1(n) = M_{\tau}^{(S_1)}(n)$ and $M_2(n) = M_{\tau}^{(S_2)}(n)$. For $n \geqslant n_0$, we may bound the distance in total variation between $T_n^{(S_1)}$ and $T_n^{(S_2)}$ as follows (see for instance [9, p.8])

$$d_{TV}\left(T_n^{(S_1)}, T_n^{(S_2)}\right) \geqslant d_{TV}\left(M_1(n), M_2(n)\right) \geqslant \frac{\left(\mathbb{E}\left[M_1(n) - M_2(n)\right]\right)^2}{2\left(\mathrm{Var}\left(M_1(n)\right) + \mathrm{Var}\left(M_2(n)\right)\right) + \left(\mathbb{E}\left[M_1(n) - M_2(n)\right]\right)^2}.$$

Since M_1 and M_2 are martingales, we have $\mathbb{E}\left[M_1(n)\right] - \mathbb{E}\left[M_2(n)\right] = \mathbb{E}\left[M_1(n_0)\right] - \mathbb{E}\left[M_2(n_0)\right] \neq 0$ and $\mathrm{Var}(M_1(n)) + \mathrm{Var}(M_2(n))$ is bounded as $n \to \infty$ since the two martingales are bounded in $\mathbb{L}^2$. Thus the quantity $d_{TV}\left(T_n^{(S_1)}, T_n^{(S_2)}\right)$ is uniformly bounded away from 0 as $n \to \infty$ as desired.   □

2.2 The recurrence relation

In this section, we present a recurrence relation for the conditional expectations of $D_{\tau}(T_n^{(S)})$. This relation is the key to Theorem 1 since it is used to build the martingales of Proposition 5 and get moment estimates on them.

Lemma 6. *There exists a family of nonnegative real numbers $\{c(\tau, \tau') : \tau' \prec \tau\}$ such that, for every seed S, every decorated tree τ with $w(\tau) > 1$ and every $n \geq |S|$ we have*

$$\mathbb{E}\left[D_\tau(T_{n+1}^{(S)}) | \mathcal{F}_n\right] = \left(1 + \frac{w(\tau)}{2n-2}\right) D_\tau(T_n^{(S)}) + \frac{1}{2n-2} \sum_{\tau' \prec \tau} c(\tau, \tau') D_{\tau'}(T_n^{(S)}). \tag{3}$$

When $\tau = ①$ we have $D_①(T_n^{(S)}) = 2n - 2$.

Proof. Fix a tree S with $|S| \geq 2$ and $n > |S|$. To simplify notation, we omit the dependence on S and write T_n instead of $T_n^{(S)}$. It will be clear by construction that the coefficients $c(\tau, \tau')$ do not depend on S. We have already noticed that when $\tau = ①$, $D_\tau(T_n)$ is the total degree of T_n, which is indeed $2n - 2$.

Now fix a decorated tree τ with $w(\tau) \geq 2$. We denote by u_{n+1} the vertex present in T_{n+1} but not in T_n, and by v_n its neighbour in T_n. We write the set all embeddings $\phi : \tau \to T_{n+1}$ as the disjoint union of the set of those using only vertices of T_n, denoted by $\mathcal{E}_n$, and the set of those using the new vertex u_{n+1}, which is denoted by $\mathcal{E}_{n+1} \backslash \mathcal{E}_n$. If T is a tree and $\phi : \tau \to T$ is an embedding, we write $\mathcal{W}_\phi(T) = \prod_{u \in \tau} [\deg_T \phi(u)]_{\ell(u)}$.

Let us evaluate $\mathbb{E}\left[D_\tau(T_{n+1}) \mid \mathcal{F}_n\right]$. Since we work conditionally on $\mathcal{F}_n$, we may consider T_n as being fixed. Then $D_\tau(T_{n+1}) = \sum_{\phi \in \mathcal{E}_{n+1}} \mathcal{W}_\phi(T_{n+1})$, and we split the sum into two, depending on whether $\phi \in \mathcal{E}_n$ or $\phi \in \mathcal{E}_{n+1} \backslash \mathcal{E}_n$. First of all, it is a simple matter to check that for every $\ell, d \geq 1$,

$$(\star) \qquad [d+1]_\ell = [d]_\ell + \ell \cdot [d]_{\ell-1},$$
$$(\star\star) \qquad d \cdot [d]_{\ell-1} = [d]_\ell + (\ell - 1) \cdot [d]_{\ell-1},$$
$$(\star\star\star) \qquad d \cdot [d+1]_\ell = [d]_{\ell+1} + 2\ell \cdot [d]_\ell + \ell(\ell-1) \cdot [d]_{\ell-1}.$$

First assume that $\phi \in \mathcal{E}_n$. Since $\deg_{T_{n+1}}(v_n) = \deg_{T_n}(v_n) + 1$, it follows that

$$\mathbb{E}\left[\mathcal{W}_\phi(T_{n+1}) \mid \mathcal{F}_n\right]$$

$$\underset{(\star)}{=} \mathcal{W}_\phi(T_n) + \mathbb{E}\left[\sum_{w \in \tau} \mathbb{1}_{\{\phi(w) = v_n\}} \cdot \ell(w) \cdot [\deg_{T_n} \phi(w)]_{\ell(w)-1} \cdot \prod_{w' \in \tau \backslash \{w\}} [\deg_{T_n}(\phi(w'))]_{\ell(w')} \mid \mathcal{F}_n\right]$$

$$= \mathcal{W}_\phi(T_n) + \sum_{w \in \tau} \frac{\deg_{T_n} \phi(w)}{2n-2} \cdot \ell(w) \cdot [\deg_{T_n} \phi(w)]_{\ell(w)-1} \cdot \prod_{w' \in \tau \backslash \{w\}} [\deg_{T_n}(\phi(w))]_{\ell(w')}$$

$$\underset{(\star\star)}{=} \left(1 + \frac{w(\tau)}{2n-2}\right) \cdot \mathcal{W}_\phi(T_n) + \sum_{w \in \tau} \frac{\ell(w)(\ell(w)-1)}{2n-2} \cdot [\deg_{T_n} \phi(w)]_{\ell(w)-1} \cdot \prod_{w' \in \tau \backslash \{w\}} [\deg_{T_n} \phi(w)]_{\ell(w')}$$

$$= \left(1 + \frac{w(\tau)}{2n-2}\right) \cdot \mathcal{W}_\phi(T_n) + \sum_{w \in \tau} \frac{\ell(w)(\ell(w)-1)}{2n-2} \cdot \mathcal{W}_{\phi_w}(T_n),$$

where ϕ_w is the embedding equal to ϕ of the decorated tree τ_w identical to τ except for the label of w which is $\ell_{\tau_w}(w) = \ell_\tau(w) - 1$. Note that such trees appear in the expression only if $\ell_{\tau_w}(w) > 0$.

When ϕ runs through the embeddings of τ in T_n, ϕ_w runs exactly through the embeddings of τ_w in T_n. Thus

$$\mathbb{E}\left[\sum_{\phi \in \mathcal{E}_n} \mathcal{W}_\phi(T_{n+1}) \,\middle|\, \mathcal{F}_n\right] = \left(1 + \frac{w(\tau)}{2n-2}\right) D_\tau(T_n) + \sum_{w \in \tau} \frac{\ell(w)(\ell(w)-1)}{2n-2} \cdot D_{\tau_w}(T_n). \quad (4)$$

Notice that the trees τ_w for $w \in \tau$ may not be distinct but all have the property $\tau_w \prec \tau$.

Denote by L the set of all leaves (i.e. vertices of τ of degree 1) $w \in \tau$ such that $\ell(w) = 1$. If $\phi \in \mathcal{E}_{n+1} \setminus \mathcal{E}_n$, consider $w \in \tau$ such that $\phi(w) = u_{n+1}$. Note that if $w \notin L$, then necessarily $\mathcal{W}_\phi(T_{n+1}) = 0$. Since $\tau \neq \mathbb{1}$, we may assume that $|\tau| \geqslant 2$. If $w \in L$, we denote by $\mathcal{E}_w$ the set of all embeddings $\phi \in \mathcal{E}_{n+1} \setminus \mathcal{E}_n$ such that $\phi(w) = u_{n+1}$. Now fix $w \in L$, $\phi \in \mathcal{E}_w$ and let a be the neighbour of w in τ. Since $[\deg_{T_{n+1}}(\phi(w))]_{\ell(w)} = 1$ and $\phi : \tau \setminus \{w\} \to T_n$ restricted to $\tau \setminus \{w\}$ is an embedding, we have

$$\begin{aligned}
\mathbb{E}\left[\mathcal{W}_\phi(T_{n+1}) \mid \mathcal{F}_n\right] &= \mathbb{E}\left[\mathbb{1}_{\{\phi(a)=v_n\}} \cdot \prod_{w' \in \tau \setminus w} \left[\deg_{T_{n+1}}(\phi(w'))\right]_{\ell(w')} \,\middle|\, \mathcal{F}_n\right] \\
&= \frac{\deg_{T_n} \phi(a)}{2n-2} \cdot \left[\deg_{T_n} \phi(a) + 1\right]_{\ell(a)} \cdot \prod_{w' \in \tau \setminus \{w,a\}} \left[\deg_{T_n}(\phi(w')).\right]_{\ell(w')}
\end{aligned}$$

Write τ_w^1, τ_w^2 and τ_w^3 for the decorated trees obtained from τ by removing w and respectively increasing by one the label of a, leaving it unchanged and decreasing it by one. Let ϕ_w^1, ϕ_w^2 and ϕ_w^3 be the respective embeddings of τ_w^1, τ_w^2 and τ_w^3 in T_n obtained by restricting ϕ to $\tau \setminus \{w\}$. Then, using the previous display and $(\star\star\star)$, we obtain

$$\mathbb{E}\left[\mathcal{W}_\phi(T_{n+1}) \mid \mathcal{F}_n\right] = \frac{1}{2n-2}\left(\mathcal{W}_{\phi_w^1}(T_n) + 2\ell(a) \cdot \mathcal{W}_{\phi_w^2}(T_n) + \ell(a)(\ell(a)-1) \cdot \mathcal{W}_{\phi_w^3}(T_n)\right).$$

Now note that, for fixed $w \in L$, as ϕ runs through $\mathcal{E}_w$, the embeddings ϕ_w^1, ϕ_w^2 and ϕ_w^3 run respectively through all the embeddings of τ_w^1, τ_w^2 and τ_w^3 in T_n. Thus, summing over all $w \in L$, we obtain

$$\sum_{\phi \in \mathcal{E}_{n+1} \setminus \mathcal{E}_n} \mathbb{E}\left[\mathcal{W}_\phi(T_{n+1}) \mid \mathcal{F}_n\right] = \frac{1}{2n-2} \sum_{w \in L} \left(D_{\tau_w^1}(T_n) + 2\ell(a) \cdot D_{\tau_w^2}(T_n) + \ell(a)(\ell(a)-1) \cdot D_{\tau_w^3}(T_n)\right). \quad (5)$$

After adding (4) and (5), one gets an expression for $\mathbb{E}\left[D_\tau(T_{n+1}) \mid \mathcal{F}_n\right]$ as a function of $D_\tau(T_n)$, $D_{\tau_w}(T_n)$, $D_{\tau_w^1}(T_n)$, $D_{\tau_w^2}(T_n)$ and $D_{\tau_w^3}(T_n)$, with coefficients depending on τ only. Finally, we mention that $\tau_w, \tau_w^1, \tau_w^2, \tau_w^3$ are all smaller than τ for the order $\prec$. $\qquad \square$

We now sketch another possible argument to prove Lemma 6 relying on decorated embeddings, which were defined just after the statement of Proposition 5. We mention this approach since a similar one will be used later.

First note that T_{n+1} contains two more corners than T_n: one around the vertex to which the new edge is grafted, and one around the new vertex which is added in the transition from T_n to T_{n+1}. Call

these corners respectively c'_n and c''_n and let c_n be the corner of T_n in which the additional edge of T_{n+1} is grafted. Then c_n corresponds to one of the neighbouring corners of c'_n in T_{n+1}. The $D_\tau(T_{n+1})$ decorated embeddings of τ in T_{n+1} may be split between those using at least one of the corners c'_n, c''_n, and those using none of them. There are $D_\tau(T_n)$ embeddings of the third type. With each decorated embedding ϕ of the first or second type, associate an embedding ϕ_σ of a decorated tree $\sigma \preccurlyeq \tau$ in T_n obtained by conserving the arrows associated by ϕ with corners common to T_n and T_{n+1} and adding an arrow pointing to c_n (if such an arrow did not already exist).

Depending on which of the corners c_n, c'_n, c''_n are used by ϕ, σ takes different values. Thus $D_\tau(T_{n+1}) - D_\tau(T_n)$ may be expressed as a linear combination of numbers of decorated embeddings of trees $\sigma \preccurlyeq \tau$ in T_n with an arrow in the corner c_n. But since c_n is uniform among the corners of T_n, in expectation these numbers are $\frac{1}{2n-2}D_\sigma(T_n)$, which leads to (3).

The proof of Lemma 8 illustrates in more detail the use of these ideas.

2.3 Moment estimates

Relying on Lemma 6, we now establish moment estimates on the number of decorated embeddings, which will be used to check that the martingales of Proposition 5 are bounded in $\mathbb{L}^2$. In the following, if $(a_n)_{n\geqslant 0}$ and $(b_n)_{n\geqslant 0}$ are two sequences of real numbers, we write $a_n \ll b_n$ if there exist $c > 0$ and $\gamma \in \mathbb{R}$ such that $|a_n| \leqslant c \cdot \log(n)^\gamma \cdot |b_n|$ for n large enough.

In this section, we fix a tree S with $|S| \geqslant 2$, and write T_n for $T_n^{(S)}$ to simplify notation.

Corollary 7. *Let τ be a decorated tree with $w(\tau) > 1$. Then, for every seed S, we have*

$$n^{w(\tau)/2} \ll \mathbb{E}\left[D_\tau(T_n)\right] \ll n^{w(\tau)/2}.$$

Proof of Corollary 7. The first bound is immediate because Lemma 6 implies that, for τ with $w(\tau) > 1$,

$$\mathbb{E}\left[D_\tau(T_n)\right] \geqslant C \cdot \prod_{j=|S|}^{n-1} \left(1 + \frac{w(\tau)}{2j-2}\right) \geqslant C' \cdot n^{w(\tau)/2},$$

for constants $C, C' > 0$ depending on S and τ.

We prove the second bound by induction on τ (for the partial order $\preccurlyeq$). Fix τ with $w(\tau) > 1$ and assume that $\mathbb{E}\left[D_{\tau'}(T_n)\right] \ll n^{w(\tau')/2}$ for every $\tau' \prec \tau$ with $w(\tau') > 1$. Since $D_\mathbb{O}(T_n) = 2n-2$, it follows that $\mathbb{E}\left[D_{\tau'}(T_n)\right] \ll n^{w(\tau)/2}$ for every $\tau' \prec \tau$. Then, by Lemma 6, there exist constants $C, \gamma > 0$ such that

$$\mathbb{E}\left[D_\tau(T_{n+1})\right] \leqslant \left(1 + \frac{w(\tau)}{2n-2}\right) \cdot \mathbb{E}\left[D_\tau(T_n)\right] + C \cdot (\log n)^\gamma \cdot n^{\frac{w(\tau)}{2}-1}.$$

It is then a simple matter to show that this implies $\mathbb{E}\left[D_\tau(T_n)\right] \ll n^{w(\tau)/2}$. $\qquad\square$

We now turn to second moment estimates on $D_\tau(T_n)$ which will be useful in the proof of Proposition 5.

Lemma 8. *Let $\boldsymbol{\tau}$ be a decorated tree with $w(\boldsymbol{\tau}) > 1$. Then*

(i) $\mathbb{E}\left[D_{\boldsymbol{\tau}}(T_n)^2\right] \ll n^{w(\boldsymbol{\tau})},$

(ii) $\mathbb{E}\left[\left(D_{\boldsymbol{\tau}}(T_{n+1}) - D_{\boldsymbol{\tau}}(T_n)\right)^2\right] \ll n^{-3/2} \cdot n^{w(\boldsymbol{\tau})}.$

To establish these results, we will need to estimate the number of embeddings in T_n of pairs of decorated trees. If $\boldsymbol{\tau}$ and $\boldsymbol{\tau}'$ are decorated trees, set

$$D_{\boldsymbol{\tau},\boldsymbol{\tau}'}(T) = \sum_{\phi} \prod_{u \in \boldsymbol{\tau} \sqcup \boldsymbol{\tau}'} [\phi(u)]_{\ell(u)}, \tag{6}$$

where the sum is taken over all injective graph homomorphisms from $\boldsymbol{\tau} \sqcup \boldsymbol{\tau}'$ to T (in particular, $\phi(u) \neq \phi(u')$ if $u \in \boldsymbol{\tau}$ and $u' \in \boldsymbol{\tau}'$).

Lemma 9. *Let $\boldsymbol{\tau}, \boldsymbol{\tau}'$ be two decorated trees. Then $\mathbb{E}\left[D_{\boldsymbol{\tau},\boldsymbol{\tau}'}(T_n)\right] \ll \mathbb{E}\left[D_{\boldsymbol{\tau}}(T_n)\right] \cdot \mathbb{E}\left[D_{\boldsymbol{\tau}'}(T_n)\right].$*

Sketch of proof of Lemma 9. The proof follows the same lines as that of Lemma 6 and Corollary 7. For this reason, we only lay out the main steps without giving additional detail. As in Lemma 6, one starts by writing a recurrence relation for $\mathbb{E}\left[D_{\boldsymbol{\tau},\boldsymbol{\tau}'}(T_n)\right]$ of the following form:

$$
\begin{aligned}
\mathbb{E}\left[D_{\boldsymbol{\tau}}(T_{n+1})\big|\mathcal{F}_n\right] &= \left(1 + \frac{w(\boldsymbol{\tau}) + w(\boldsymbol{\tau}')}{2n - 2}\right) D_{\boldsymbol{\tau},\boldsymbol{\tau}'}(T_n) \\
&\quad + \frac{1}{2n - 2}\left(\sum_{\sigma \prec \boldsymbol{\tau}} c(\boldsymbol{\tau}, \boldsymbol{\tau}', \sigma) D_{\sigma,\boldsymbol{\tau}'}(T_n) + \sum_{\sigma' \prec \boldsymbol{\tau}'} c'(\boldsymbol{\tau}, \boldsymbol{\tau}', \sigma') D_{\boldsymbol{\tau},\sigma'}(T_n)\right),
\end{aligned}
$$

for certain nonnegative real numbers $c(\boldsymbol{\tau}, \boldsymbol{\tau}', \sigma), c'(\boldsymbol{\tau}, \boldsymbol{\tau}', \sigma)$ and $n \geqslant |S|$. We stress that in the previous equation, the decorated trees σ and σ' may also take the value $\emptyset$, in which case $D_{\sigma,\boldsymbol{\tau}'}(T_n)$ and $D_{\boldsymbol{\tau},\sigma'}(T_n)$ are equal to respectively $D_{\boldsymbol{\tau}'}(T_n)$ and $D_{\boldsymbol{\tau}}(T_n)$. The same inductive argument as that of Corollary 7 leads to the conclusion. □

Proof of Lemma 8. One could try to write a formal proof in the spirit of the one of Lemma 6. However, to simplify the presentation, we use a planar embedding of T_n and the interpretation of $D_{\boldsymbol{\tau}}(T_n)$ as the number of decorated embeddings of $\boldsymbol{\tau}$ in T_n, as explained after the statement of Proposition 5. Let $\boldsymbol{\tau}'$ be a disjoint copy of $\boldsymbol{\tau}$. By definition, a decorated map $\boldsymbol{\phi} : \boldsymbol{\tau} \cup \boldsymbol{\tau}' \to T_n$ is a map such that both $\boldsymbol{\phi}|_{\boldsymbol{\tau}}$ and $\boldsymbol{\phi}|_{\boldsymbol{\tau}'}$ are decorated embeddings. We insist on the fact that $\boldsymbol{\phi}$ is not necessarily injective. If $\boldsymbol{\phi}$ is a decorated embedding or a decorated map, ϕ will denote the map without the choice of corners associated with arrows.

For the first assertion, observe that $D_{\boldsymbol{\tau}}(T_n)^2$ is the number of decorated maps $\boldsymbol{\phi} : \boldsymbol{\tau} \cup \boldsymbol{\tau}' \to T_n$. We denote by $\mathcal{E}^1_{\boldsymbol{\tau}}(T_n)$ the set of all such decorated maps with $\phi(\boldsymbol{\tau}) \cap \phi(\boldsymbol{\tau}') = \emptyset$ (as in the definition of $D_{\boldsymbol{\tau},\boldsymbol{\tau}}(T_n)$), and by $\mathcal{E}^2_{\boldsymbol{\tau}}(T_n)$ the set of all such decorated maps with $\phi(\boldsymbol{\tau}) \cap \phi(\boldsymbol{\tau}') \neq \emptyset$. The cardinality of $\mathcal{E}^1_{\boldsymbol{\tau}}(T_n)$ is $D_{\boldsymbol{\tau},\boldsymbol{\tau}}(T_n)$, and Lemma 9 applies.

If $\boldsymbol{\phi} \in \mathcal{E}^2_{\boldsymbol{\tau}}(T_n)$ is a decorated map, we may associate with $\boldsymbol{\phi}$ a decorated embedding $\boldsymbol{\phi}'$ of a decorated tree $\sigma_{\boldsymbol{\phi}}$ obtained by overlapping two copies of $\boldsymbol{\tau}$. More precisely, let $\mathcal{U}_2(\boldsymbol{\tau})$ be the set of

all decorated trees which may be obtained by identifying a non-empty subset of elements (i.e. of vertices, edges and arrows) of τ and τ'. The embedding associated with ϕ is given by the images of τ and τ' in T_n via ϕ (in particular σ_ϕ is the union of the images of τ and τ'), see Figure 5.

Note that the function $\phi \mapsto (\sigma_\phi, \phi')$ defined above is not one to one, since an element of $\mathcal{U}_2(\tau)$ may be obtained in several ways by overlapping τ and τ'. However, it is easy to see that there exists a constant $C(\tau) > 0$ such that any decorated tree $\sigma \in \mathcal{U}_2(\tau)$ and any embedding ϕ' of σ in T_n is associated with at most $C(\tau)$ decorated maps ϕ. We may therefore conclude that

$$D_\tau(T_n)^2 = \#\mathcal{E}_\tau^1(T_n) + \#\mathcal{E}_\tau^2(T_n) \;\; \leqslant \;\; D_{\tau,\tau}(T_n) + C(\tau) \cdot \sum_{\sigma \in \mathcal{U}_2(\tau)} D_\sigma(T_n).$$

Observe that $w(\sigma) \leqslant 2 \cdot w(\tau)$ for every $\sigma \in \mathcal{U}_2(\tau)$. Lemma 9, Corollary 7 and the fact that $\mathcal{U}_2(\tau)$ is a finite set imply the desired bound.
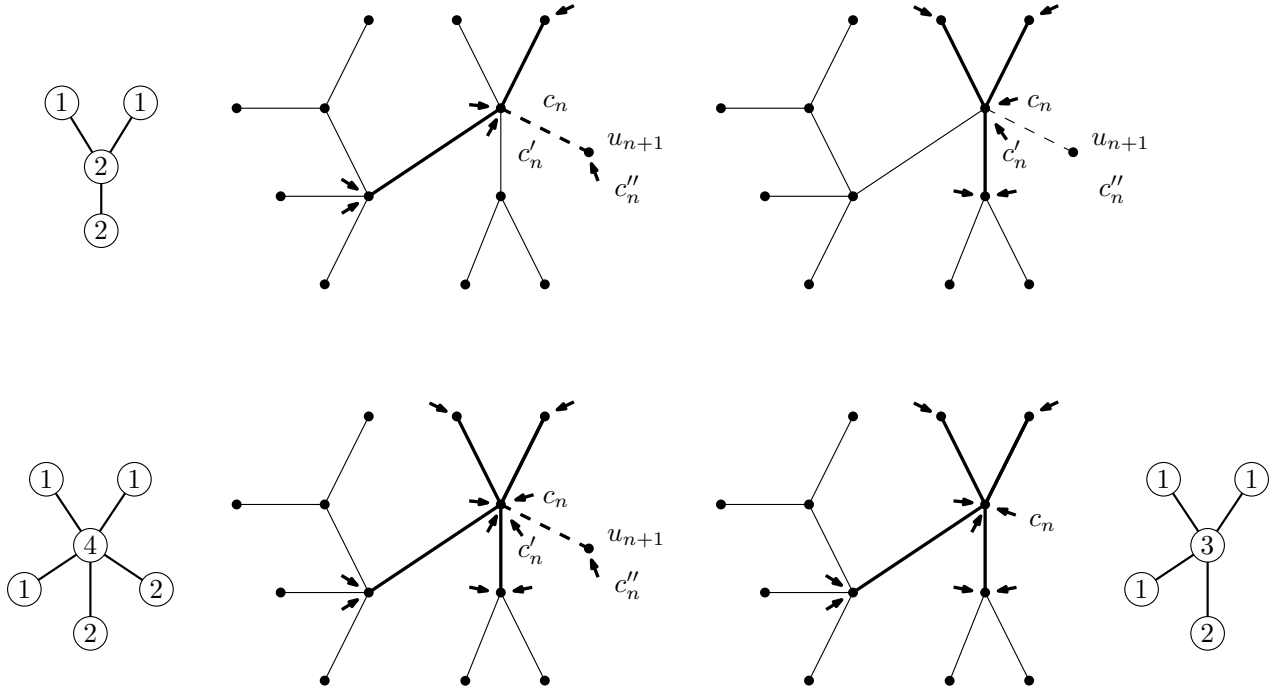
For the second assertion, we work conditionally on $\mathcal{F}_n$. As in the discussion after the proof of Lemma 6, let u_{n+1} be the vertex added to T_n in the transition from T_n to T_{n+1}, and c_n'' be the corner adjacent to u_{n+1} in T_{n+1} (since u_{n+1} is a leaf of T_{n+1}, there is only one corner adjacent to it). Also let c_n be the corner of T_n in which the additional edge of T_{n+1} is grafted, and denote by c_n and c_n' the two corners of T_{n+1} resulting from splitting c_n. Finally let v_n, be the vertex adjacent to c_n. We refer to Fig. 5 for an example.

Note that $D_\tau(T_{n+1}) - D_\tau(T_n)$ is the number of decorated embeddings of τ in T_{n+1} that use at least one of the corners c_n', c_n''. Similarly, $(D_\tau(T_{n+1}) - D_\tau(T_n))^2$ is the number of decorated maps $\psi : \tau \cup \tau' \to T_{n+1}$ such that $\psi|_\tau$ and $\psi|_{\tau'}$ both use at least one of the corners c_n', c_n''. To simplify notation, denote by $\mathcal{E}_\tau(T_{n+1})$ the set of all such decorated maps. Fix $\psi \in \mathcal{E}_\tau(T_{n+1})$. Since $w(\tau) > 1$, we have $v_n \in \psi(\tau)$ and $v_n \in \psi(\tau')$. As in the proof of (i), we may associate with the decorated map ψ a decorated embedding $\phi : \sigma \to T_{n+1}$ of a decorated tree $\sigma \in \mathcal{U}_2(\tau)$. We shall now furthermore associate with ϕ a decorating embedding $\phi' : \sigma' \to T_n$ of a modified decorated tree σ' in T_n.

Let w be the vertex of σ such that $\phi(w) = v_n$. Define the modified decorated tree σ' by altering σ as follows. Remove from σ the vertex mapped by ϕ to u_{n+1} as well as the arrow pointing to it, if such a vertex exists. Remove in addition the arrows of σ mapped by ϕ to c_n or c_n'. Finally add an arrow $\vec{a}$ to σ' pointing to w. The decorated embedding $\phi' : \sigma' \to T_n$ is defined to be equal to ϕ on $\sigma \cap \sigma'$, and maps the arrow $\vec{a}$ of σ' on the corner c_n (we stay loose on the order of the distinguished corners).

To sum up, with every decorated map $\psi \in \mathcal{E}_\tau(T_{n+1})$, we have associated a decorated tree σ' with a distinguished arrow $\vec{a}$ and a decorated embedding ϕ of σ' in T_n. Moreover, we have done this in such a way that $\phi(\vec{a}) = c_n$. In addition, σ' satisfies $w(\sigma') \leqslant 2w(\tau) - 1$ and $|\sigma'| \leqslant 2|\tau| - 1$. As in (i), this association is not injective but the number of pre-images of any given image may be bounded by a constant $C(\tau) > 0$. Hence

$$\left(D_\tau(T_{n+1}) - D_\tau(T_n)\right)^2 \leqslant \sum_{\substack{\sigma';|\sigma'|<2|\tau|, \\ w(\sigma')<2w(\tau)}} \sum_{\vec{a}\in\sigma'} \sum_{\phi:\sigma'\to T_n} C(\tau) \cdot \mathbb{1}_{\{\phi(\vec{a})=c_n\}},$$

Figure 5: *Top diagram:* Two decorated embeddings ϕ_1 and ϕ_2 of the decorated tree τ in T_{n+1}. Both embeddings use corners of T_{n+1} not present in T_n, hence contribute to $D_\tau(T_{n+1}) - D_\tau(T_n)$. *Bottom diagram:* The overlapping of the two decorated embeddings ϕ_1 and ϕ_2 of τ induces a decorated embedding of the decorated tree σ depicted on the left. To σ we associate the embedding of σ' in T_n (right diagrams).

where the first sum is taken on decorated trees σ' with the displayed constraints. Since the corner c_n is uniform in T_n, we obtain

$$\mathbb{E}\left[\left(D_\tau(T_{n+1}) - D_\tau(T_n)\right)^2 \middle| \mathcal{F}_n\right] \leqslant \sum_{\substack{\sigma';|\sigma'|<2|\tau|, \\ w(\sigma')<2w(\tau)}} \sum_{\vec{a}\in\sigma'} \frac{C(\tau)}{2n-2} \cdot D_{\sigma'}(T_n) \ll n^{w(\tau)-3/2}.$$

For the last estimate, we have used the fact that $D_{\sigma'}(T_n) \ll n^{w(\tau)-1/2}$ for all values of σ' (see Corollary 7) and that the number of terms in the sum is bounded in terms of τ only. Taking the expectation of the expression above leads to the desired inequality. $\qquad\square$

2.4 Constructing martingales

We now use the recurrence relation (3) in order to construct the martingales of Proposition 5. It may be instructive for the reader to compute the martingale $M_\tau^{(S)}(n)$ using (3) in some simple cases, for instance when τ is formed of a single vertex with label 2 or 3, or two vertices of label 1 linked by an edge.

Proof of Proposition 5. Fix a seed tree S with $|S| = n_0 \geqslant 2$. For a decorated tree τ and $n \geqslant 2$, set

$$\alpha_n^\tau = \prod_{j=2}^{n-1} \left(1 + \frac{w(\tau)}{2j-2}\right)^{-1}, \text{ when } w(\tau) > 1; \qquad \alpha_n^\tau = \frac{1}{2n-2}, \text{ when } w(\tau) = 1.$$

In particular, if $w(\tau) > 1$, observe that $n^{-w(\tau)/2} \ll \alpha_n^\tau \ll n^{-w(\tau)/2}$. For a sequence of real numbers $(a_n)_{n \geqslant 0}$ set $\Delta_n a = a_{n+1} - a_n$ for $n \geqslant 0$.

We start by constructing by induction (on the order $\prec$ on decorated trees) coefficients $\{a_n(\tau, \tau') : \tau' \prec \tau, n \geqslant n_0\}$ such that

$$a_n(\tau, \tau') \ll 1, \qquad \Delta_n a(\tau, \tau') \ll 1/n \tag{7}$$

$$\text{and} \qquad M_\tau^{(S)}(n) = \alpha_n^\tau \left(D_\tau(T_n^{(S)}) - \sum_{\tau' \prec \tau} a_n(\tau, \tau') \cdot D_{\tau'}(T_n^{(S)})\right) \qquad \text{is a martingale.} \tag{8}$$

We emphasize that by construction, the coefficients $a_n(\tau, \tau')$ will not depend on S (this is essential). To simplify notation, we write T_n and $M_\tau(n)$ for respectively $T_n^{(S)}$ and $M_\tau^{(S)}(n)$. First, $M_{①}(n) = \alpha_n^{①} \cdot D_{①}(T_n) = 1$, which is clearly a martingale.

Next, fix a decorated tree τ with $w(\tau) \geqslant 2$ and assume that the coefficients $a_n(\sigma, \sigma')$ have been constructed for every $\sigma' \prec \sigma \prec \tau$ and $n \geqslant n_0$, and that they have the desired properties. Then we claim that there exist constants $(b_n(\sigma, \sigma') : \sigma' \prec \sigma \prec \tau, n \geqslant n_0)$ such that $b_n(\sigma, \sigma') \ll 1$ and

$$D_\sigma(T_n) = \frac{1}{\alpha_n^\sigma} M_\sigma(n) + \sum_{\sigma' \prec \sigma} \frac{b_n(\sigma, \sigma')}{\alpha_n^{\sigma'}} M_{\sigma'}(n), \qquad n \geqslant n_0. \tag{9}$$

Indeed, define the matrix $(\mathcal{A}_n(\sigma, \rho))_{\sigma, \rho \prec \tau}$ taking value $-a_n(\sigma, \rho)$ if $\rho \prec \sigma$, 1 if $\sigma = \rho$ and 0 otherwise. Then, by (8), for every $n \geqslant n_0$, we have the following equality of vectors indexed by $\sigma \prec \tau$:

$$\left(\frac{1}{\alpha_n^\sigma} \cdot M_\sigma(n)\right)_{\sigma \prec \tau} = \mathcal{A}_n \cdot (D_\sigma(T_n))_{\sigma \prec \tau}.$$

We may write $\{\sigma : \sigma \prec \tau\} = \{\sigma_1, \ldots, \sigma_K\}$ in such a way that $\sigma_i \prec \sigma_j \Rightarrow i < j$. In this setting, $\mathcal{A}_n$ is a triangular matrix with values 1 on the diagonal and all coefficients $\ll 1$. It follows that $\mathcal{A}_n$ is invertible, and that its inverse shares this same property. If we denote by $(b_n(\sigma, \sigma'))_{\sigma' \prec \sigma}$ the above-diagonal entries of the inverse of $\mathcal{A}_n$, we obtain (9).

Then Lemma 6 and (9) yield, for $n \geqslant n_0$,

$$\mathbb{E}\left[\alpha_{n+1}^\tau \cdot D_\tau(T_{n+1}) \middle| \mathcal{F}_n\right] = \alpha_n^\tau \left(D_\tau(T_n) + \frac{1}{2n - 2 + w(\tau)} \sum_{\tau' \prec \tau} c(\tau, \tau') D_{\tau'}(T_n)\right)$$

$$= \alpha_n^\tau \cdot D_\tau(T_n) + \sum_{\sigma \prec \tau} \frac{1}{2n - 2 + w(\tau)} \cdot \left(c(\tau, \sigma) + \sum_{\sigma \prec \tau' \prec \tau} c(\tau, \tau') b_n(\sigma, \tau')\right) \cdot \frac{\alpha_n^\tau}{\alpha_n^\sigma} \cdot M_\sigma(n).$$

Now set

$$\bar{a}_n(\tau, \sigma) = \sum_{j=2}^{n-1} \frac{1}{2j - 2 + w(\tau)} \cdot \left(c(\tau, \sigma) + \sum_{\sigma \prec \tau' \prec \tau} c(\tau, \tau') b_j(\tau', \sigma)\right) \cdot \frac{\alpha_j^\tau}{\alpha_j^\sigma}, \tag{10}$$

for $n \geqslant n_0$, so that

$$\mathbb{E}\left[\alpha_{n+1}^{\tau} D_{\tau}(T_{n+1}) \big| \mathcal{F}_n\right] = \alpha_n^{\tau} D_{\tau}(T_n) + \sum_{\sigma \prec \tau} \left(\overline{a}_{n+1}(\tau, \sigma) - \overline{a}_n(\tau, \sigma)\right) \cdot M_{\sigma}(n).$$

Since $(M_{\tau'}(n))_{n \geqslant n_0}$ is a $(\mathcal{F}_n)$-martingale for every $\tau' \prec \tau$ by our induction hypothesis, the above implies that

$$\begin{aligned}
M_{\tau}(n) \; &:= \; \alpha_n^{\tau} D_{\tau}(T_n) - \sum_{\tau' \prec \tau} \overline{a}_n(\tau, \tau') M_{\tau'}(n) \\[2mm]
&= \; \alpha_n^{\tau}\left(D_{\tau}(T_n) - \sum_{\tau' \prec \tau} \overline{a}_n(\tau, \tau') \cdot \frac{\alpha_n^{\tau'}}{\alpha_n^{\tau}} \cdot \left(D_{\tau'}(T_n) + \sum_{\sigma \prec \tau'} a_n(\tau', \sigma) \cdot D_{\sigma}(T_n)\right)\right) \\[2mm]
&= \; \alpha_n^{\tau}\left(D_{\tau}(T_n) - \sum_{\sigma \prec \tau}\left(\overline{a}_n(\tau, \sigma) \cdot \frac{\alpha_n^{\sigma}}{\alpha_n^{\tau}} + \sum_{\sigma \prec \tau' \prec \tau} \overline{a}_n(\tau, \tau') \cdot \frac{\alpha_n^{\tau'}}{\alpha_n^{\tau}} \cdot a_n(\tau', \sigma)\right) \cdot D_{\sigma}(T_n)\right).
\end{aligned}$$

is a $(\mathcal{F}_n)$ martingale. Finally, for $\sigma \prec \tau$ and $n \geqslant n_0$, set

$$a_n(\tau, \sigma) \quad := \quad \overline{a}_n(\tau, \sigma) \cdot \frac{\alpha_n^{\sigma}}{\alpha_n^{\tau}} + \sum_{\sigma \prec \tau' \prec \tau} \overline{a}_n(\tau, \tau') \cdot \frac{\alpha_n^{\tau'}}{\alpha_n^{\tau}} \cdot a_n(\tau', \sigma). \tag{11}$$

With this notation, it is now clear that the martingale $(M_{\tau}(n))_{n \geqslant n_0}$ defined as above satisfies (8).

Let us now analyse the orders of magnitude of the quantities $a_n(\tau, \sigma)$ and $\Delta_n a(\tau, \sigma)$ in order to establish (7). We have $\frac{1}{2n-2+w(\tau)} \ll 1/n$, and for two decorated trees σ, σ' with $w(\sigma), w(\sigma') > 1$ a straightforward computation yields

$$\frac{\alpha_n^{\sigma}}{\alpha_n^{\sigma'}} \ll n^{\frac{w(\sigma')-w(\sigma)}{2}} \qquad \text{and} \qquad \Delta_n \frac{\alpha^{\sigma}}{\alpha^{\sigma'}} \ll n^{\frac{w(\sigma')-w(\sigma)}{2}-1}.$$

In addition, by our induction hypothesis, we have $b_n(\sigma, \tau') \ll 1$ for every $\sigma \prec \tau' \prec \tau$. From (10) we get that

$$\Delta_n \overline{a}(\tau, \sigma) \ll n^{\frac{w(\sigma)-w(\tau)}{2}-1} \qquad \text{and} \qquad \overline{a}_n(\tau, \sigma) \ll n^{\frac{w(\sigma)-w(\tau)}{2}}$$

for every $\sigma \prec \tau$ such that $w(\sigma) \geqslant 2$. Hence, for $\sigma \neq ①$,

$$\overline{a}_n(\tau, \sigma) \cdot \frac{\alpha_n^{\sigma}}{\alpha_n^{\tau}} \ll 1 \quad \text{and} \quad \Delta_n\left(\overline{a}(\tau, \sigma) \cdot \frac{\alpha^{\sigma}}{\alpha^{\tau}}\right) \ll \frac{1}{n}. \tag{12}$$

A separate analysis shows that (12) also holds when $\sigma = ①$. By the induction hypothesis, we have that $a_n(\tau', \sigma) \ll 1$ and $\Delta_n a(\tau', \sigma) \ll 1/n$ for every $\sigma \prec \tau' \prec \tau$. By combining the previous estimates with Eq. (11) which defines $a_n(\tau, \sigma)$, we obtain

$$a_n(\tau, \sigma) \ll 1 \quad \text{and} \quad \Delta_n a(\tau, \sigma) \ll 1/n, \quad \text{for every } \tau' \prec \tau.$$

This completes the induction.

Finally, let us now prove that the martingales M_τ defined by (8) are indeed bounded in $\mathbb{L}^2$. To this end, since the increments of a martingale are orthogonal in $\mathbb{L}^2$, write

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

It is clear that $\mathbb{E}[M_\tau(n_0)^2] < \infty$, so it is enough to check that

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

By (8) and the Cauchy-Schwarz inequality, there exists a constant $c > 0$, depending only on τ, such that for $n \geqslant n_0$, the quantity $c \cdot \mathbb{E}\left[\left(M_\tau(n+1) - M_\tau(n)\right)^2\right]$ is bounded from above by

$$\mathbb{E}\left[\left(\Delta_n\left(\alpha^\tau \cdot D_\tau(T)\right)\right)^2\right]$$
$$+ \sum_{\tau' \prec \tau} \left(\left(a_{n+1}(\tau', \tau)\right)^2 \cdot \mathbb{E}\left[\left(\Delta_n\left(\alpha^{\tau'} \cdot D_{\tau'}(T)\right)\right)^2\right] + \left(\Delta_n a(\tau', \tau)\right)^2 \cdot \mathbb{E}\left[\left(\alpha_n^{\tau'} \cdot D_{\tau'}(T_n)\right)^2\right]\right).$$

To bound this quantity, it will be useful to note that, for every σ with $w(\sigma) > 1$, by Lemma 8 and a straightforward computation,

$$\mathbb{E}\left[\left(\Delta_n\left(\alpha^\sigma \cdot D_\sigma(T)\right)\right)^2\right] \leqslant 2\left(\Delta_n \alpha^\sigma\right)^2 \cdot \mathbb{E}\left[D_\sigma(T_n)^2\right] + 2\left(\alpha_{n+1}^\sigma\right)^2 \cdot \mathbb{E}\left[\left(\Delta_n D_\sigma(T)\right)^2\right] \ll n^{-3/2}.$$

In addition, when $\sigma = \textcircled{1}$, we have $\Delta_n\left(\alpha^\sigma \cdot D_\sigma(T)\right) = 0$. By combining the previous estimates with (7), we finally get that

$$\mathbb{E}\left[(M_\tau(n+1) - M_\tau(n))^2\right] \ll n^{-3/2}.$$

This implies $\sum_{n \geqslant n_0} \mathbb{E}\left[\left(M_\tau(n+1) - M_\tau(n)\right)^2\right] < \infty$, and the proof is complete. $\qquad\square$

3 Scaling limits of looptrees

In this section, we prove Theorem 2. The main tool is a coupling between the looptrees of the plane LPAM and a certain modification of binary trees obtained by Rémy's algorithm [27]. The coupling between the LPAM and Rémy's algorithm has already been noticed in the literature [25], but will be recalled here and extended to looptrees.

3.1 Coupling with Rémy's algorithm

We start by introducing some useful notation. In this section, unless stated otherwise, trees are not considered as embedded in the plane. A tree is binary when all its vertices have degree at

most 3. If $x, y \in \tau$ are two vertices of a tree τ, we let $[\![x, y]\!]$ be the geodesic in τ between x and y. If $x_0, x_1, \ldots, x_k \in \tau$ are distinct vertices, we let

$$\mathsf{Span}(\tau; x_0, \ldots, x_k) = \bigcup_{0 \leqslant i,j \leqslant k} [\![x_i, x_j]\!]$$

be the tree spanned by these vertices. A labeled tree $\boldsymbol{\tau} = (\tau; x_0, x_1, \ldots, x_n)$ is a pair formed of a tree τ and a collection of leaves $(x_0, \ldots, x_n)$ of τ.

For a labeled tree $\boldsymbol{\tau} = (\tau; x_0, x_1, \ldots, x_n)$ the gluing $\mathsf{Glu}(\boldsymbol{\tau})$ of $\boldsymbol{\tau}$ is the graph constructed as follows. Set $p_1 = x_0$, and for $2 \leqslant i \leqslant n$, let p_i be the vertex of $\mathsf{Span}(\tau; x_0, x_1, \ldots x_{i-1})$ which is the closest to x_i in τ. Then $\mathsf{Glu}(\boldsymbol{\tau})$ is by definition the graph obtained from τ by identifying the vertices x_i and p_i for every $1 \leqslant i \leqslant n$. See the second line of Fig. 6 for an illustration. Formally, the vertices of the graph $\mathsf{Glu}(\boldsymbol{\tau})$ are the equivalence classes of the vertices of τ for the equivalence relation generated by $x_i \sim p_i$ for every $1 \leqslant i \leqslant n$. We view $\mathsf{Glu}(\boldsymbol{\tau})$ as a compact metric space by endowing its vertex set with the graph distance.

Next we present Rémy's algorithm; it is a recursive procedure for building labeled binary trees. Start with the tree $\mathbf{B}_1 = (B_1; A_0, A_1)$ consisting of a single edge with two leaves labeled A_0 and A_1. At every step $n \geqslant 1$, build $\mathbf{B}_{n+1}$ from $\mathbf{B}_n$ by picking an edge e of $\mathbf{B}_n$ uniformly at random, adding a vertex v on e (thus splitting e into two edges) and attaching a new edge to v linking it to a new leaf denoted A_{n+1}. Rémy [27] showed that for every fixed $n \geqslant 1$, the labeled tree $\mathbf{B}_n$ is uniformly distributed over the set of all binary trees with $n + 1$ labeled leaves.

Let $(T_n^{-\circ})_{n \geqslant 1}$ be the plane LPAM with seed $-\circ$, as defined in the Introduction. Recall that $\mathsf{Loop}(T_n^{-\circ})$ is the looptree associated with $T_n^{-\circ}$. An important element of the proof of Theorem 2 is the following.

Proposition 10. *We have the following joint equality in distribution*

$$(\mathsf{Loop}(T_n^{-\circ}); n \geqslant 1) \quad \overset{(d)}{=} \quad (\mathsf{Glu}(\mathbf{B}_n); n \geqslant 1).$$

Proof. The growth mechanism of $\mathsf{Loop}(T_n^{-\circ})$ is the following: at each step, an edge is selected uniformly at random, split in its middle by adding a new vertex, with attached to it a new loop made of single edge. Let us now turn to the growth mechanism of $\mathsf{Glu}(\mathbf{B}_n)$: This graph is the collection of n loops made by the geodesics starting from $\mathsf{Span}(B_n; A_0, \ldots, A_{i-1})$ and going to A_i for $i \in \{1, \ldots, n\}$ which are turned into cycles by identifying their endpoints. Then an edge of B_n is selected uniformly at random, and split in its middle by adding a new edge carrying A_{n+1}. Then observe that the impact of this splitting on $\mathsf{Glu}(B_n)$ is equivalent to the growth procedure of $\mathsf{Loop}(T_n^{-\circ})$ we have described (see Fig. 6 for an illustration). □

3.2   Definition of the Brownian looptree

In this section, we define the Brownian looptree $\mathcal{L}$. We first introduce some notation concerning continuous trees. A metric space $(\mathcal{T}, d)$ is an $\mathbb{R}$-tree if it contains no cycle and if for any points
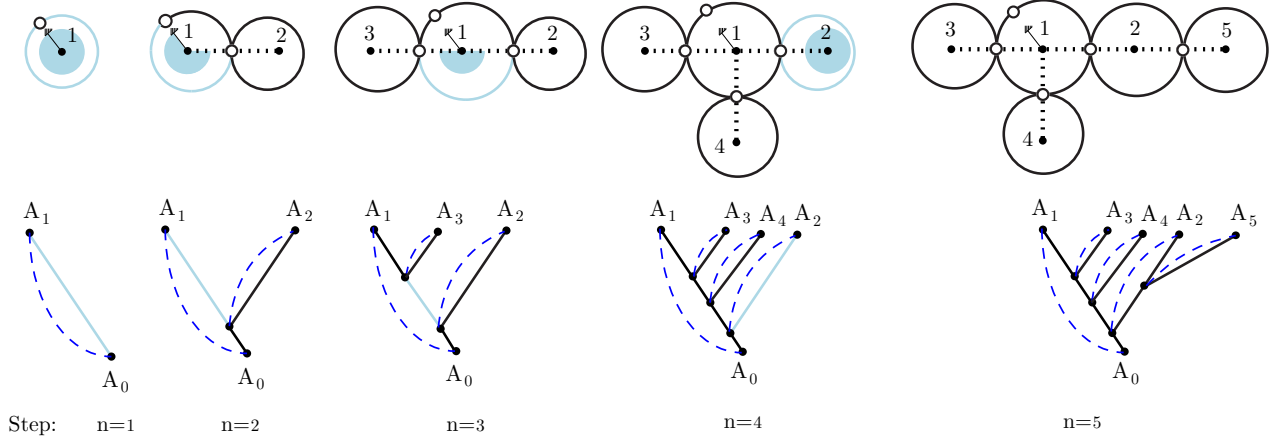
Figure 6: Illustration of the proof of Proposition 10, where in addition the edges that are split have been coupled in such a way that $\mathsf{Loop}(T_n^{-\circ}) = \mathsf{Glu}(\mathbf{B}_n)$. The first line represents the evolution of $(T_n^{-\circ}, \mathsf{Loop}(T_n^{-\circ}))_{1 \leqslant n \leqslant 5}$ and the second the evolution of $(\mathbf{B}_n, \mathsf{Glu}(\mathbf{B}_n))_{1 \leqslant n \leqslant 5}$. In the second line, the dashed lines represent the identifications that are made.

$x, y \in \mathcal{T}$ there exists a unique geodesic $[\![x, y]\!]$ between x and y which isometric to a segment of $\mathbb{R}$ (see [19, Sec. 3] for a more detailed definition). Moreover we impose that $(\mathcal{T}, d)$ be compact. To mark the difference between $\mathbb{R}$-trees and the regular trees used up to now, we will sometimes call the latter discrete trees.

We extend the notation introduced for discrete trees in Sec. 3.1 to continuous trees. If $\mathcal{T}$ is an $\mathbb{R}$-tree and $x_0, x_1, \ldots, x_k \in \mathcal{T}$ are distinct points, we let

$$\mathsf{Span}(\mathcal{T}; x_0, \ldots, x_k) = \bigcup_{0 \leqslant i, j \leqslant k} [\![x_i, x_j]\!]$$

be the $\mathbb{R}$-tree spanned by these vertices. The degree of a point $x \in \mathcal{T}$ is the number of connected components of $\mathcal{T} \backslash \{x\}$. A leaf is a point of degree 1. A labeled $\mathbb{R}$-tree is a pair consisting of an $\mathbb{R}$-tree $\mathcal{T}$ and a (finite or infinite) collection of leaves of $\mathcal{T}$.

Consider a labeled compact $\mathbb{R}$-tree $\mathscr{T} = (\mathcal{T}; (x_i)_{0 \leqslant i < N})$, where $N \in \overline{\mathbb{N}}_+ := \{1, 2, \ldots\} \cup \{+\infty\}$, and assume that $\mathcal{T}$ is binary (this assumption is not necessary, but it holds in our case and simplifies the exposition). The gluing $\mathsf{Glu}(\mathscr{T})$ of $\mathscr{T}$ is the quotient compact metric space constructed as follows. Set $p_1 = x_0$, and for $2 \leqslant i < N$, let p_i be the point of $\mathsf{Span}(\mathcal{T}; x_0, x_1, \ldots x_{i-1})$ which is the closest to x_i in τ. Write $\sim$ for the equivalence relation on $\mathcal{T}$ generated by $p_i \sim x_i$ for $1 \leqslant i < N$. If d denotes the graph distance on $\mathcal{T}$, we define a pseudo-distance Δ on $\mathcal{T}$ by

$$\Delta(a, b) = \inf \left\{ \sum_{i=0}^{k} d(p_i, q_i) : p_0 = a; q_k = b \right\},$$

where the infimum runs over all choices of $k \in \mathbb{N}$ and points $(p_i)_{0 \leqslant i \leqslant k}$ and $(q_i)_{0 \leqslant i \leqslant k}$ so that $q_i \sim p_{i+1}$ for $0 \leqslant i \leqslant k-1$.

In the case of an generic metric space, defining a "gluing" metric could yield to more identifications that those prescribed by $\sim$. This is not the case in our setup, as explained next.

Lemma 11. *For every* $a, b \in \mathcal{T}$, $\Delta(a, b) = 0$ *if and only if* $a \sim b$.

Proof. We first check that $\sim$ is closed. For this, consider a sequence $(a_i, b_i)_{i \geqslant 0}$ converging to (a, b) as $i \to \infty$. We can suppose without loss of generality that all the points $\{a_i, b_i\}_{i \geqslant 0}$ are distinct. Since $\mathcal{T}$ is binary, it is simple to see that we have $u \sim v$ if and only if $u = v$ or $\{u, v\} = \{p_i, x_i\}$ for a certain $i \geqslant 1$. In particular, if $\{a, b\} \neq \{c, d\}$ with $a \sim b$ and $c \sim d$ we must have $]\!]a, b[\![\cap]\!]c, d[\![= \varnothing$. By compactness this implies that $d(a_i, b_i) \to 0$, and hence $a = b$. The relation $\sim$ is thus closed.

Now let $a, b \in \mathcal{T}$ be such that $\Delta(a, b) = 0$. For every $i \geqslant 1$, we denote by a_i and b_i the projections (i.e. closest point) of respectively a and b on $\mathsf{Span}(\mathcal{T}; x_0, \dots, x_i)$. A moment's thought shows that a_i must be equal to b_i inside $\mathsf{Glu}(\mathsf{Span}(\mathcal{T}; x_0, \dots, x_i); x_0, \dots, x_i)$, since otherwise we would have $\Delta(a, b) > 0$. In particular, $a_i \sim b_i$. As $i \to \infty$, we have $a_i \to \tilde{a}$ where $\tilde{a}$ is the projection of a on the closure of $\bigcup_i \mathsf{Span}(\mathcal{T}; x_0, \dots, x_i)$, and similarly $b_i \to \tilde{b}$. If $a \neq \tilde{a}$ (or $b \neq \tilde{b}$), we would have $\Delta(a, b) > 0$ since there would exist a small ball around a unaffected by the gluings. Hence $(a, b) = (\tilde{a}, \tilde{b})$, and $a_i \to a$ and $b_i \to b$. Since $\sim$ is closed, we have $a \sim b$ as desired. $\qquad\square$

Using the above we may deduce (see for instance [10, Exercise 3.1.14]) that

$$\mathsf{Glu}(\mathcal{T}; (x_i)_{0 \leqslant i < N}) \quad := \quad (\mathcal{T}/\sim, \Delta). \tag{13}$$

is a *compact* metric space, which we call the (continuous) gluing of $(\mathcal{T}; (x_i)_{0 \leqslant i < N})$. We shall denote by $\pi : \mathcal{T} \to \mathsf{Glu}(\mathcal{T}; (x_i)_{0 \leqslant i < N})$ the canonical projection.

In the case where $\mathcal{T} = \mathcal{T}_e$ is the Brownian CRT and $x_i = X_i$ for $i \geqslant 0$ is a sequence of i.i.d. random variables sampled according to the mass measure of $\mathcal{T}_e$, the random compact metric space $\mathcal{L} = \mathsf{Glu}(\mathcal{T}_e, (X_i)_{i \geqslant 0})$ is called the Brownian looptree.

Remark 12. *The Brownian looptree may also be constructed through a line breaking procedure, very similar to the one designed by Aldous to construct the Brownian CRT (see [26, Theorem 7.6]). Consider $0 < \theta_1 < \theta_2 < \cdots$ to be the points of a Poisson point process on $\mathbb{R}_+$ with intensity $t/2 \cdot dt$. Break the line $[0, \infty)$ at points θ_k to create segments of length $\theta_1, \theta_2 - \theta_1, \dots$. Glue the two end-points of each such segment together to create metric circles $\mathcal{C}_1, \mathcal{C}_2, \dots$. Construct recursively metric spaces $\mathcal{G}_1, \mathcal{G}_2, \dots$ by setting $\mathcal{G}_1 = \mathcal{C}_1$ and, for each $k \geqslant 1$, glueing $\mathcal{C}_{k+1}$ to a point chosen uniformly at random on $\mathcal{G}_k$. The Brownian looptree is then the completion of $\cup_{k \geqslant 1} \mathcal{G}_k$.*

3.3 Convergence towards the Brownian looptree

We briefly describe the k-pointed Gromov–Hausdorff topology (we refer to [10, 16, 22] for additional details). A k-pointed compact metric space is a triple $(E, d, (x_1, \dots, x_k))$, where (E, d) is a compact metric space and $x_1, \dots, x_k \in E$. Two k-pointed compact metric spaces are said to be isometric if there exists an isometry between them mapping the k distinguished points of one of them to the

distinguished points of the other (preserving the order). The set of isometry classes of k-pointed compact metric spaces is endowed with the k-pointed Gromov–Hausdorff distance defined next. If $(E, d, (x_1, \ldots, x_k))$ and $(E', d', (x'_1, \ldots, x'_k))$ are two k-pointed compact metric spaces,

$$d_{GH}\big((E, d, (x_1, \ldots, x_k)), (E', d', (x'_1, \ldots, x'_k))\big) \;=\; \inf\left\{d^F_H(\phi(E), \phi'(E')) \vee \max_{1 \leqslant i \leqslant k} \delta(\phi(x_i), \phi'(x'_i))\right\},$$

where the infimum is taken over all choices of metric spaces (F, δ) and isometric embeddings $\phi : E \to F$ and $\phi' : E' \to F$ of E and E' into F, and where d^F_H denotes the Hausdorff distance between compacts sets in F. The k-pointed Gromov–Hausdorff distance is indeed a metric on the space of isometry classes of k-pointed compact metric spaces. It renders this space separable and complete. For $k = 0$, d_{GH} is the usual Gromov–Hausdorff distance on (isometry classes of) compact metric spaces.

We now state a continuity proposition inspired from [12, Proposition 12]. If (E, d) is a metric space and $x_0, \ldots, x_n \in E$, we say that $x_0, \ldots, x_n$ is an ε-net in E if $E = \bigcup_{0 \leqslant i \leqslant n} \{y \in E; d(y, x_i) < \varepsilon\}$.

Proposition 13. *Let* $(\tau; (x_i)_{0 \leqslant i < N})$ *be either a labeled discrete tree, or a labeled $\mathbb{R}$-tree, with $N \in \overline{\mathbb{N}}_+$. Then, for every integer $0 \leqslant k < N$,*

$$d_{GH}\big(\mathsf{Glu}(\tau; (x_i)_{0 \leqslant i \leqslant k}), \mathsf{Glu}(\tau; (x_i)_{0 \leqslant i < N})\big) \leqslant 2 \inf\{\varepsilon > 0; x_0, \ldots, x_k \text{ is an } \varepsilon\text{-net in } \tau\},$$

where τ is equipped with its graph distance in the discrete case or with its metric in the continuous case.

Proof. For $k \geqslant 0$, set $\tau_k = \mathsf{Span}(\tau; x_0, \ldots, x_k)$. We clearly have

$$d_H(\tau, \tau_k) \leqslant \inf\{\varepsilon > 0; x_0, \ldots, x_k \text{ is an } \varepsilon\text{-net in } \tau\}. \tag{14}$$

We thus can bound $d_{GH}\big(\mathsf{Glu}(\tau; (x_i)_{0 \leqslant i \leqslant k}), \mathsf{Glu}(\tau; (x_i)_{0 \leqslant i < N})\big)$ above by

$$d_{GH}\big(\mathsf{Glu}(\tau; (x_i)_{0 \leqslant i \leqslant k}), \mathsf{Glu}(\tau_k; (x_i)_{0 \leqslant i \leqslant k})\big) + d_{GH}\big(\mathsf{Glu}(\tau_k; (x_i)_{0 \leqslant i \leqslant k}), \mathsf{Glu}(\tau; (x_i)_{0 \leqslant i < N})\big),$$

which is less than or equal to $d_H(\tau, \tau_k) + d_H(\tau, \tau_k)$ since $\mathsf{Glu}(\tau_k; x_0, \ldots, x_k) \to \mathsf{Glu}(\tau; x_0, \ldots, x_k)$ and similarly $\mathsf{Glu}(\tau_k; x_0, \ldots, x_k) \to \mathsf{Glu}(\tau; (x_i)_{0 \leqslant i < N})$ are isometric embeddings and Glu is a contraction. Combining this with (14) finishes the proof. $\qquad\square$

Before proceeding to the proof of of Theorem 2, we state a final simple property that we will not prove.

Lemma 14. *Fix an integer $k \geqslant 0$. Let $(\mathcal{T}^{(n)})_{n \geqslant 1} = \big(\mathcal{T}^{(n)}; x_0^{(n)}, \ldots, x_k^{(n)}\big)_{n \geqslant 1}$ be a sequence of labeled $\mathbb{R}$-trees and $\mathcal{T} = (\mathcal{T}; x_0, \ldots, x_k)$ be a labeled $\mathbb{R}$-tree. Suppose that $\mathcal{T}^{(n)} \to \mathcal{T}$ holds almost surely for the k-pointed Gromov–Hausdorff topology. Then $\mathsf{Glu}(\mathcal{T}^{(n)}) \to \mathsf{Glu}(\mathcal{T})$ also holds almost surely for the Gromov–Hausdorff topology.*

Proof of Theorem 2. Recall from Section 3.1 the notation $\mathbf{B}_n = (B_n; A_0, \ldots, A_n)$ for the sequence of trees grown by Rémy's algorithm. By [12, Theorem 5 (ii)], there exists a pair $(\mathcal{T}_e, (X_i; i \geqslant 0))$, where $\mathcal{T}_e$ is a Brownian CRT and $(X_i; i \geqslant 0)$ is a collection of i.i.d. vertices sampled according to its mass measure, such that for every $k \geqslant 1$ we have the following convergence for the $k+1$-pointed Gromov-Hausdorff topology

$$\left(n^{-1/2} \cdot B_n; A_0, \ldots, A_k\right) \xrightarrow[n\to\infty]{a.s.} \left(2\sqrt{2} \cdot \mathcal{T}_e; X_0, \ldots, X_k\right). \tag{15}$$

Hence, by Lemma 14, the following holds in the regular Gromov-Hausdorff topology

$$n^{-1/2} \cdot \mathsf{Glu}(B_n; A_0, \ldots, A_k) \xrightarrow[n\to\infty]{a.s.} 2\sqrt{2} \cdot \mathsf{Glu}(\mathcal{T}_e; X_0, \ldots, X_k). \tag{16}$$

For $0 \leqslant k \leqslant n$, set $L_n^{(k)} = \mathsf{Glu}(B_n; A_0, \ldots, A_k)$ so that $\mathsf{Glu}(\mathbf{B}_n) = L_n^{(n)}$. Also set $\mathcal{L}_k = \mathsf{Glu}(\mathcal{T}_e; X_0, \ldots, X_k)$. Now, for $n \geqslant k \geqslant 1$,

$$d_{\mathrm{GH}}\left(\frac{L_n^{(n)}}{\sqrt{n}}, 2\sqrt{2} \cdot \mathcal{L}\right) \leqslant d_{\mathrm{H}}\left(\frac{L_n^{(n)}}{\sqrt{n}}, \frac{L_n^{(k)}}{\sqrt{n}}\right) + d_{\mathrm{GH}}\left(\frac{L_n^{(k)}}{\sqrt{n}}, 2\sqrt{2} \cdot \mathcal{L}_k\right) + d_{\mathrm{H}}(2\sqrt{2} \cdot \mathcal{L}_k, 2\sqrt{2} \cdot \mathcal{L}),$$

Denote by respectively $U_{n,k}, V_{n,k}$ and W_k the three terms appearing in the previous sum. In order to prove that the right-hand side above converges to 0 as $n \to \infty$, we will first take the $\limsup$ of the above as $n \to \infty$, then make k tend to ∞.

By (16), $\lim_{n\to\infty} V_{n,k} = 0$ for every fixed $k \geqslant 1$. Also, since $(X_i)_{i\geqslant 0}$ is a.s. dense in $\mathcal{T}_e$, by Proposition 13 we get that $\lim_{k\to\infty} W_k = 0$. By Proposition 13,

$$U_{n,k} = d_{\mathrm{H}}\left(\frac{L_n^{(n)}}{\sqrt{n}}, \frac{L_n^{(k)}}{\sqrt{n}}\right) \leqslant 2 \inf\left\{\varepsilon > 0; A_0, \ldots, A_k \text{ is an } \varepsilon\text{-net in } n^{-1/2} \cdot B_n\right\}.$$

But by (15) we have

$$\inf\left\{\varepsilon > 0; A_0, \ldots, A_k \text{ is an } \varepsilon\text{-net in } n^{-1/2} \cdot B_n\right\} \xrightarrow[n\to\infty]{a.s.} \inf\left\{\varepsilon > 0; X_0, \ldots, X_k \text{ is an } \varepsilon\text{-net in } 2\sqrt{2} \cdot \mathcal{T}_e\right\}.$$

We deduce that

$$\limsup_{k\to\infty} \limsup_{n\to\infty} d_{\mathrm{H}}\left(\frac{\mathcal{L}_n^{(n)}}{\sqrt{n}}, \frac{\mathcal{L}_n^{(k)}}{\sqrt{n}}\right) \leqslant 4\sqrt{2} \cdot \limsup_{k\to\infty} \inf\{\varepsilon > 0; X_0, \ldots, X_k \text{ is an } \varepsilon\text{-net in } \mathcal{T}_e\} = 0,$$

since the collection $(X_i; i \geqslant 0)$ is almost surely dense in $\mathcal{T}_e$. The proof is complete. $\qquad\square$

3.4 Convergence towards Brownian looptrees for general seeds

In this section we prove the Corollary 3. In order to describe the construction of $\mathcal{L}^{(S)}$ and prove this result, a preliminary discussion is required on how $T_n^{(S)}$ may be constructed from independent copies of the processes $(T_n^{-\circ})_{n\geqslant 1}$.

For $n \geqslant 0$ and $N \geqslant 1$, denote by $\mathrm{Pól}(n, N)$ the law after n draws of the state of a Pólya urn with N colors, starting with one ball of each color and diagonal replacement matrix $\mathrm{Diag}(2, 2, \ldots, 2)$. In other words, consider an urn with N balls of different colors. At each step a ball is taken out uniformly at random, inspected, and then put back in the urn along with two additional balls of the same color. Then $\mathrm{Pól}(n, N)$ is the law of $(X_1^n, \ldots, X_N^n)$, where X_i^n represents the numbers of balls of the i-th color after n draws. If $\underline{P}$ is a plane planted tree and c is a corner of some plane tree S, then a new plane tree may be obtained by gluing $\underline{P}$ inside c, as depicted in Fig. 7.
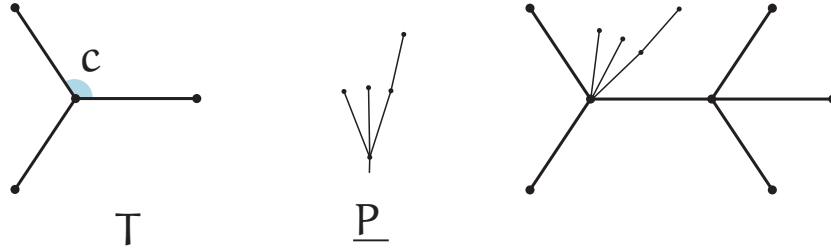

Figure 7: A plane tree T with a distinguished corner c, a plane planted tree, and the plane tree obtained by gluing $\underline{P}$ inside c.

Proposition 15. *Fix a plane tree S and let $(c_1, \ldots, c_N)$ be an exhaustive enumeration of its corners with $N := 2|S| - 2$. If $n \geqslant |S|$ is an integer, let $(\alpha_1^n, \ldots, \alpha_N^n)$ be a random variable sampled according to $\mathrm{Pól}(n - |S|, N)$. Then, conditionally on $(\alpha_1^n, \ldots, \alpha_N^n)$, let $\underline{P}_1^n, \ldots, \underline{P}_N^n$ be independent random variables distributed as respectively $\mathsf{T}_{(\alpha_1^n+1)/2}^{\to\circ}, \ldots, \mathsf{T}_{(\alpha_N^n+1)/2}^{\to\circ}$. Finally, let S_n be the tree obtained by gluing, for every $1 \leqslant i \leqslant N$, the planted tree $\underline{P}_i^n$ in each each corner c_i of S. Then S_n has the same law as $\mathsf{T}_n^{(S)}$.*

Rather than a formal proof, we give a brief explanation of this fact. Combined with Figure 8, it should be enough to convince the reader. As $\mathsf{T}_n^{(S)}$ grows from S, vertices are added sequentially. For every $1 \leqslant i \leqslant N$, there are $(\alpha_i^n - 1)/2$ vertices added to the corner c_i of S (that is either direct neighbours of c_i, or linked to c_i by edges not belonging S). In particular, the subtree of $\mathsf{T}_n^{(S)}$ emanating from c_i is a planted tree with α_i^n corners (including the corners at its base). Thus, in order to construct $\mathsf{T}_{n+1}^{(S)}$ from $\mathsf{T}_n^{(S)}$, in order to construct $\mathsf{T}_{n+1}^{(S)}$, the new vertex is added in the tree emanating from c_i with probability $\alpha_i^n / \sum_{j=1}^N \alpha_j^n$. This shows that $(\alpha_1^n, \ldots, \alpha_N^n)$ indeed follows the law $\mathrm{Pól}(n, N)$. Moreover, conditionally on the number $\frac{\alpha_i^n - 1}{2}$ of vertices added to c_i, these vertices are added following the rules of the LPAM starting with $-\!\circ$ as the seed. Hence the tree emanating from c_i in $\mathsf{T}_n^{(S)}$ has the law of $\mathsf{T}_{\frac{\alpha_i^n+1}{2}}^{\to\circ}$. Finally, the trees growing inside the different corners of T are independent conditionally on their size.

We are now ready to describe the construction of the limit space $\mathcal{L}^{(S)}$ of Corollary 3. For this we need to introduce notation.

If $\underline{P}$ is a planted tree, define a modified looptree $\widetilde{\mathrm{Loop}}(\underline{P})$ by "cutting" $\mathrm{Loop}(\underline{P})$ at the vertex associated with the root half-edge of $\underline{P}$. More precisely delete this vertex and add two distinct vertices as
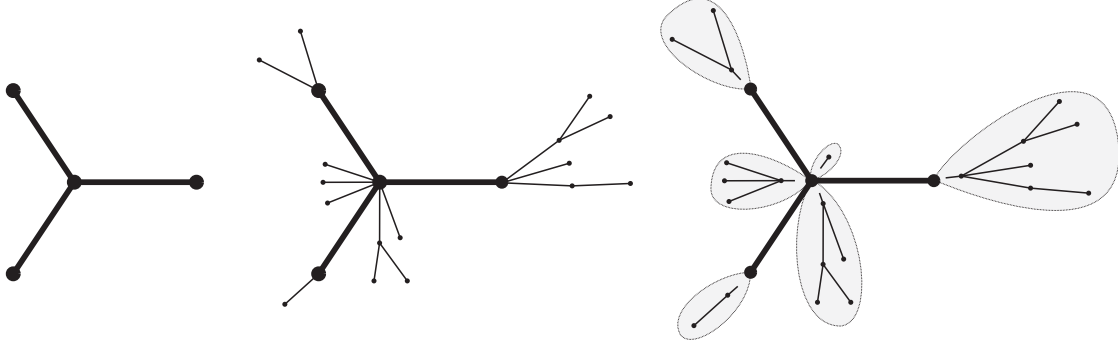
Figure 8: The tree $T_n^{(S)}$ is obtained by gluing planted trees in the corners of the seed S.

endpoints of the two edges of $\mathsf{Loop}(\underline{P})$ incident to the removed vertex. Let $g(\underline{P})$, resp. $d(\underline{P})$, denote the endpoints of the edge to the left, resp. right, of the root half-edge of $\underline{P}$, when the latter is oriented towards its only endpoint.
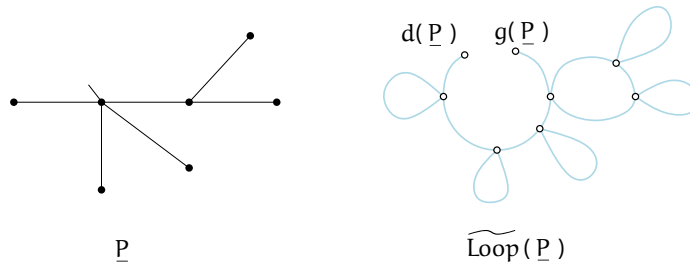
Figure 9: A planted tree $\underline{P}$ and its associated modified looptree $\widetilde{\mathsf{Loop}}(\underline{P})$.

A simple extension of Theorem 2 then shows that we have the following almost sure 2-pointed Gromov–Hausdorff convergence

$$\left(n^{-1/2} \cdot \widetilde{\mathsf{Loop}}(T_n^{-\circ}); \; g(T_n^{-\circ}), d(T_n^{-\circ})\right) \xrightarrow[n\to\infty]{a.s.} \left(2\sqrt{2} \cdot \widetilde{\mathcal{L}} \; ; \; \pi(X_0), \pi(X_1)\right),$$

where $\widetilde{\mathcal{L}}$ is constructed exactly as $\mathcal{L}$ except that we do not make the identification $X_0 \sim X_1$. Equivalently, $\widetilde{\mathcal{L}}$ is obtained from the Brownian looptree by "cutting" it at the vertex $\pi(X_0)$ and distinguishing the two newly obtained points. To simplify notation, write $\widetilde{\mathcal{L}} = (\widetilde{\mathcal{L}} \; ; \; \pi(X_0), \pi(X_1))$.

Assume that $T_n^{(S)}$ is constructed as in Proposition 15. By standard results concerning Pólya urns (see e.g [4] or [11, Prop. 3]), we have

$$\left(\frac{\alpha_1^n}{n}, \ldots, \frac{\alpha_N^n}{n}\right) \xrightarrow[n\to\infty]{a.s.} 2 \cdot (\alpha_1, \ldots, \alpha_N),$$

where $(\alpha_1, \ldots, \alpha_N)$ follows the Dirichlet distribution $\mathrm{Dir}\left(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}\right)$. It follows there exists a collection $(\widetilde{\mathcal{L}}_i \; ; \; \pi(X_i^0), \pi(X_i^1))_{1 \leqslant i \leqslant N}$ of independent pointed modified Brownian looptrees such that the

convergence

$$\left(\frac{\widetilde{\mathrm{Loop}}(\underline{P}_i^n)}{\sqrt{n}} \; ; \; g(\underline{P}_i^n), d(\underline{P}_i^n) \right) \quad \xrightarrow[n\to\infty]{a.s.} \quad \left(2\sqrt{2} \cdot \sqrt{\alpha_i} \cdot \widetilde{\mathcal{L}}_i \; ; \; \pi(X_i^0), \pi(X_i^1) \right) \tag{17}$$

holds in the 2-pointed Gromov–Hausdorff topology for every $1 \leqslant i \leqslant N$.

Then $\mathcal{L}^{(S)}$ is obtained by gluing these metric spaces along the structure given by the seed, as described next. Let $\{m(e) : e \in E(S)\}$ denote the collection of midpoints of edges of S. For each corner c_i of S, let $e_g(i)$ and $e_d(i)$ be the edges to its left and right, respectively (note that they are not necessarily distinct). For every $i \geqslant 1$, identify the points $\pi(X_i^0), \pi(X_i^1)$ of $2\sqrt{2} \cdot \sqrt{\alpha_i} \cdot \widetilde{\mathcal{L}}_i$ to $m(e_g(i))$ and $m(e_d(i))$, respectively. This creates a compact metric space which we denote by $\mathcal{L}^{(S)}$. The same construction may be performed in the discrete setting, see Fig. 10 for an illustration.

Corollary 3 follows readily from (17) and from the fact that $\mathrm{Loop}(T_n^{(S)})$ may be obtained from the modified looptrees of $\underline{P}_1^n, \ldots, \underline{P}_N^n$ in the same way as $\mathcal{L}^{(S)}$ is obtained from $\widetilde{\mathcal{L}}_1, \ldots, \widetilde{\mathcal{L}}_N$.
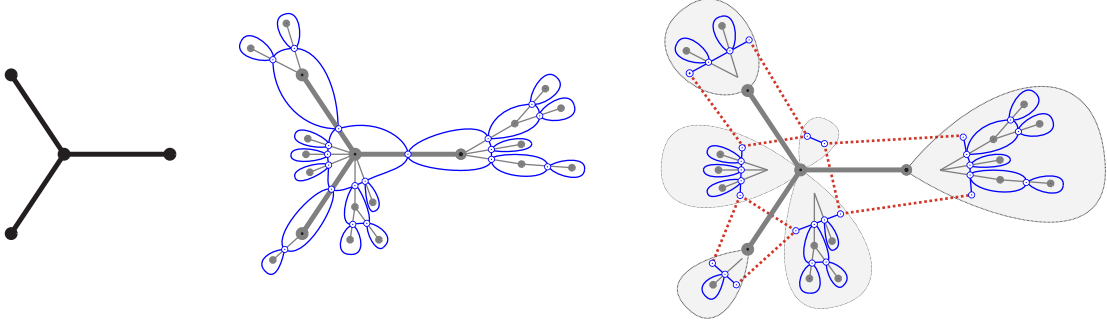
Figure 10: The discrete looptree $\mathrm{Loop}(T_n^{(S)})$ can be constructed from the collection of modified looptrees $(\widetilde{\mathrm{Loop}}(\underline{P}_i^n))_i$ by identifying vertices connected by dashed edges.

3.5 Dimension of the Brownian looptree

In this section we establish Proposition 4.

Write $\mathcal{L} = \mathrm{Glu}(\mathcal{T}_e, (X_i)_{i \geqslant 0})$, where $\mathcal{T}_e$ is a Brownian CRT and $(X_i)_{i \geqslant 0}$ is a collection of independent leaves sampled according to its mass measure μ. Recall that $\pi : \mathcal{T}_e \to \mathcal{L}$ is the canonical projection. The upper bound on the Hausdorff dimension is a consequence of the fact that π is a contraction. Since $\dim(\mathcal{T}_e) = 2$, it follows that $\dim(\mathcal{L}) \leqslant 2$ (see e.g. [21, Theorem 7.5]). To establish the lower bound, we will use the probability mass measure ν on $\mathcal{L}$, which is defined as push-forward of μ by the canonical projection. We shall show the following result:

Lemma 16. *For every $\varepsilon > 0$, almost surely, for ν-every x we have*

$$\limsup_{r \to 0} \frac{\nu(B_r(x))}{r^{2-\varepsilon}} = 0,$$

where $B_r(x)$ denotes the open ball of radius r around the point x in $\mathcal{L}$.

By standard density theorems for Hausdorff measures [21, Theorem 8.8] (this reference covers the case of measures on $\mathbb{R}^n$, but the proof remains valid here), this implies that the Hausdorff dimension of $\mathcal{L}$ is greater than or equal to $2 - \varepsilon$, almost surely. The lower bound will thus follow.

The rest of this section is devoted to the proof Lemma 16. To simplify, we say that a point is chosen uniformly in $\mathcal{T}_e$ if it is sampled according to its mass measure μ. Consider an additional uniform random leaf $Y \in \mathcal{T}_e$, independent of $(X_i)_{i \geqslant 0}$. Note that almost surely, $Y \neq X_i$ for every $i \geqslant 0$. We shall prove that for every $\epsilon \in (0, 2)$, almost surely,

$$\limsup_{r \to 0} \frac{\nu(B_r(\pi(Y)))}{r^{2-\varepsilon}} = 0,$$

By Fubini's theorem, this indeed implies Lemma 16. To this end, define a nested sequence of rooted subtrees $\mathcal{T}_e = \mathcal{T}^1 \supset \mathcal{T}^2 \supset \mathcal{T}^3 \supset \cdots$ all containing the point Y and defined recursively as follows. First, set $\mathcal{T}_e = \mathcal{T}^1$ which is rooted at $\tilde{P}_1 := X_0$. For every $j \geqslant 1$, if $\mathcal{T}^1, \ldots, \mathcal{T}^j$ have been constructed, set $k_j = \min\{i \geqslant 1; X_i \in \mathcal{T}^j\}$. Next, consider $\tilde{P}_{j+1}$ the branching point between $\tilde{P}_j, X_{k_j}$ and Y (if a, b, c are different leaves of $\mathcal{T}_e$, the branching point between a, b and c is defined to be the unique element of $[\![a, b]\!] \cap [\![b, c]\!] \cap [\![a, c]\!]$). The tree $\mathcal{T}^{j+1}$ is finally defined to be the subtree of $\mathcal{T}^j \backslash \{\tilde{P}_{j+1}\}$ containing Y to which we add the vertex $\tilde{P}_{j+1}$. Moreover $\tilde{P}_{j+1}$ is declared to be the root of $\mathcal{T}^{j+1}$. We refer to Fig. 11 for an illustration.
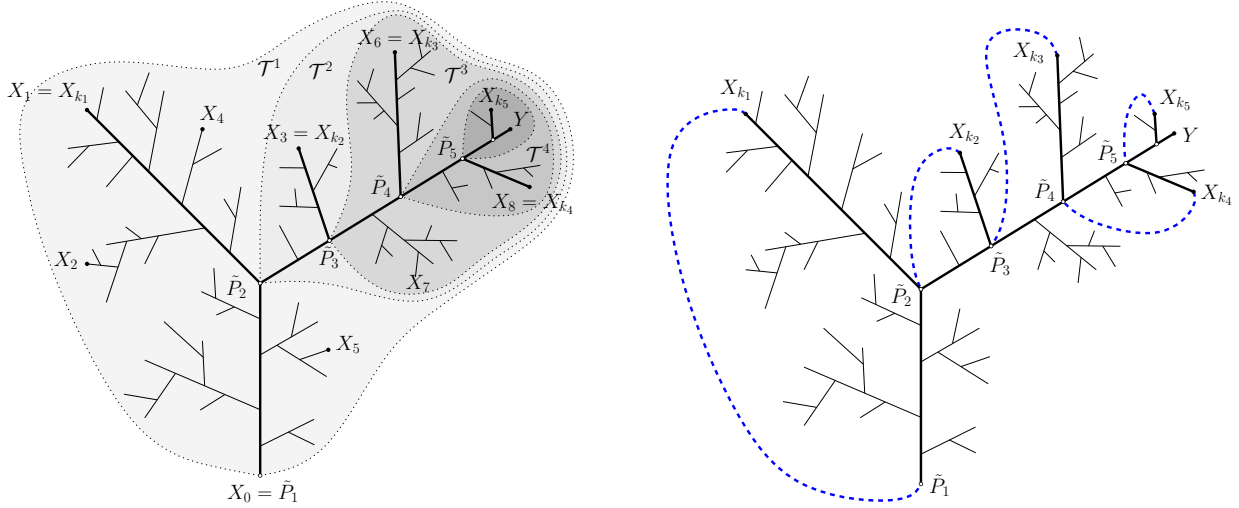
Figure 11: Illustration of the construction of $\mathcal{T}^i$ for $i \geqslant 0$ and of the proof of Proposition 18.

Proposition 17. *The following assertions hold.*

(i) The process $(-\log \mu(\mathcal{T}^i))_{i \geqslant 1}$ is a random walk, and its step distribution is an exponential random variable of parameter $1/2$.

(ii) For every $i \geqslant 1$, the random tree $\frac{1}{\sqrt{\mu(\mathcal{T}^i)}} \cdot \mathcal{T}^i$ has the law of a Brownian CRT. In addition Y and $\tilde{P}_i$ are two independent uniform leaves of $\mathcal{T}^i$.

Proof. We prove the statement by induction on $i \geqslant 1$. For $i = 1$, this is simply because X_0 and Y are two independent leaves in $\mathcal{T}_e$. By induction, at step $i \geqslant 1$, we assume that $\mathcal{T}^i$ is a random multiple of a Brownian CRT and that $Z_0 := \tilde{P}_i$ and $Z_1 := Y$ are two independent uniform leaves of $\mathcal{T}^i$. Observe that by construction, $Z_2 := X_{k_i}$ is a uniform leaf of $\mathcal{T}^i$, independent of (Z_0, Z_1). In addition, if B denotes the branching point between Z_0, Z_1 and Z_2, note that $\mathcal{T}^i$ is the union of three subtrees containing respectively Z_0, Z_1, Z_2 and having B as the only common element, and that the subtree $\mathcal{T}^{i+1}$ is the one containing Z_1, rooted at B. It follows from Aldous' decomposition in three parts of the CRT [3, Theorem 2] that $\mu(\mathcal{T}^{i+1}) = \alpha \cdot \mu(\mathcal{T}^i)$, where α is the first coordinate of a Dirichlet $\mathrm{Dir}(1/2, 1/2, 1/2)$ random variable independent of $\mathcal{T}^i$, that $\mathcal{T}^{i+1}$ has the same distribution as $\sqrt{\alpha}$ times $\mathcal{T}^i$, and that B and Z_1 are independent uniform leaves in $\mathcal{T}^{i+1}$. This implies the second assertion. It is a simple matter to check that α has density $(2\sqrt{x})^{-1}$ on $[0, 1]$, so that $-\log(\alpha)$ is distributed according to an exponential random variable of parameter $1/2$. This completes the proof. $\qquad \square$

Now, for every tree $\mathcal{T}^i$ we introduce the quantity

$$\mathcal{X}_i = \min\left(d_{\mathcal{T}_e}(\tilde{P}_i, \tilde{P}_{i+1}), d_{\mathcal{T}_e}(X_{k_i}, \tilde{P}_{i+1})\right).$$

The reason for considering this random variable lies in the following geometric proposition:

Proposition 18. *For every $i \geqslant 2$ and every $x \in \mathcal{T}_e$ such that $x \notin \mathcal{T}^i$, we have*

$$\Delta\left(\pi(x), \pi(Y)\right) \geqslant \mathcal{X}_i.$$

Proof (Sketch). In the construction of the Brownian looptree from $\mathcal{T}_e$ and $(X_i)_{i \geqslant 0}$, the points X_{k_i} are glued to $\tilde{P}_i$ for every $i \geqslant 0$. Specifically, each segment $[\![\tilde{P}_i, X_{k_i}]\!]$ becomes a loop denoted by L_i in $\mathcal{R}$, and the loops L_i and L_{i+1} share the common point $\pi(\tilde{P}_{i+1}) = \pi(X_{k_i})$. It should then be clear that in $\mathcal{R}$, the region $\pi(\mathcal{T}^1 \backslash \mathcal{T}^i)$ is separated from $\pi(Y)$ and that the only way to go from $\pi(Y)$ to this region is to travel along $\mathcal{R}$ and cross the loop L_i from $\pi(\tilde{P}_{i+1})$ to $\pi(\tilde{P}_i)$ but this requires at least a length $\mathcal{X}_i$. We leave the details to the reader. $\qquad \square$

Proof of Lemma 16. By the first assertion of Proposition 17 and the strong law of large numbers we have

$$\frac{\log\left(\mu(\mathcal{T}^i)\right)}{i} \xrightarrow[i \to \infty]{\text{a.s.}} -2. \tag{18}$$

By the second assertion of the last proposition we have $\mathcal{X}_i / \sqrt{\mu(\mathcal{T}^i)} = \mathcal{X}_1$ in distribution. In addition, by [19, Theorem 2.11], $\mathcal{X}_1$ has the same law as $\min(\ell_1, \ell_2)$ where (ℓ_1, ℓ_2, ℓ_3) has density on $\mathbb{R}_+^3$ given by $16(\ell_1 + \ell_2 + \ell_3)e^{-2(\ell_1 + \ell_2 + \ell_3)^2}\, d\ell_1 d\ell_2 d\ell_3$. From this expression, it is a simple matter to establish the existence of a constant $c > 0$ such that for every $i \geqslant 1$, $\mathbb{P}(\mathcal{X}_1 \geqslant i \text{ or } \mathcal{X}_1 \leqslant i^{-2}) \leqslant ci^{-2}$. An application of Borel–Cantelli's yields that almost surely, for every i sufficiently large, $\mathcal{X}_i / \sqrt{\mu(\mathcal{T}^i)} \geqslant i^{-2}$ and $\mathcal{X}_i / \sqrt{\mu(\mathcal{T}^i)} \leqslant i^2$. Combining this with (18), we get that

$$\frac{\log(\mathcal{X}_i)}{i} \xrightarrow[i \to \infty]{\text{a.s.}} -1. \tag{19}$$

Now, by Proposition 18, we have $B_{\mathcal{X}_i}(\pi(Y))) \subset \pi(\mathcal{T}^i)$. Noting that $\mu(\pi^{-1}(\pi(A))) = \mu(A)$ for every $A \subset \mathcal{T}_e$, this implies that $\nu(B_{\mathcal{X}_i}(\pi(Y))) \leqslant \mu(\mathcal{T}^i)$. By combining (19) and (18), we finally obtain that

$$\limsup_{i \to \infty} \frac{\nu(B_{\mathcal{X}_i}(\pi(Y)))}{\mathcal{X}_i^{2-\varepsilon}} = 0.$$

Since $\mathcal{X}_i \to 0$ a.s. as $i \to \infty$, this completes the proof. $\qquad\qquad\square$

4 Comments, extensions, conjectures and open questions

4.1 Affine reinforcement

We first investigate the extension of our results to the more general LPAM$^\delta$ model, in which vertices are chosen proportionally to an affine function of their degree. To describe this model, first fix a parameter

$$\delta > -1.$$

Let S be a finite tree with n_0 vertices. Define the random sequence of trees $(T_n^{(S),\delta})_{n \geqslant n_0}$ by $T_{n_0}^{(S),\delta} = S$ and, for $n \geqslant n_0$, conditionally on $T_n^{(S),\delta}$, the tree $T_{n+1}^{(S),\delta}$ is obtained from $T_n^{(S),\delta}$ by choosing a vertex $u \in T_n^{(S),\delta}$ with probability proportional to $\deg(u) + \delta$, and connecting it via an edge to a new vertex. We call this the LPAM$^\delta$ model. It was first introduced in [23]. For $\delta = 0$ we recover LPAM studied in the previous sections. The parameter δ has a dramatic impact on the geometry of $T_n^{(S),\delta}$ as $n \to \infty$. For instance, it is known that in this context the maximal degree in $T_n^{-\circ,\delta}$ is of order $n^{1/(2+\delta)}$, see e.g. [24, 29]. Still, we conjecture that the analogs of our results hold in this setting with the appropriate modifications.

Conjecture 1 (Influence of the seed). *For two trees S_1, S_2 set $d_\delta(S_1, S_2) = \lim_{n \to \infty} d_{TV}(T_n^{(S_1),\delta}, T_n^{(S_2),\delta})$. Then the function d_δ is a metric on trees with at least 3 vertices.*

We believe that a way to prove this conjecture is to use the same observables (namely the number of embeddings of a certain structure in the tree at step n) as those used to prove Theorem 1. However we will not pursue this goal in this paper.

A plane version of the above algorithm may also be considered. Assume that $T_n^{(S),\delta}$ is a plane tree and let us describe how to construct $T_{n+1}^{(S),\delta}$. Choose a vertex $u \in T_n^{(S),\delta}$ at random as before, and then choose uniformly at random a corner c among all the corners adjacent to u. Now graft the edge leading to the new vertex of $T_{n+1}^{(S),\delta}$ in c. Using this construction, $T_n^{(S),\delta}$ has indeed the tree structure of the LPAM$^\delta$, and its embedding is a uniform embedding of such a tree (assuming that this is also true for S). Other planar versions may be considered, but we choose this one for its symmetry.

Conjecture 2 (Scaling limit). *There exists a random compact metric space $\mathcal{L}_\delta^{(S)}$ such that the convergence*

$$n^{-\frac{1}{2+\delta}} \cdot \mathsf{Loop}(T_n^{(S),\delta}) \xrightarrow[n \to \infty]{a.s.} \mathcal{L}_\delta^{(S)}$$

holds almost surely for the Gromov–Hausdorff convergence. In addition, almost surely, the Hausdorff dimension of $\mathcal{L}_\delta^{(S)}$ is $2 + \delta$.

As seen previously, it is natural to scale $\mathsf{Loop}(T_n^{(S),\delta})$ by a factor $n^{-1/(2+\delta)}$, since the large degrees of $T_n^{(S),\delta}$ are of order $n^{1/(2+\delta)}$. We now give some arguments to support this Conjecture 2. To simplify, as in the introduction, treat only the case $S = -\circ$.

It may be shown that the plane LPAM^δ is closely related to a modification of Ford's algorithm with parameter $\alpha = 1/(2 + \delta)$. Ford's algorithm is a means to grow recursively a sequence of binary trees that generalizes Rémy's algorithm. For references see [17].

FORD'S ALGORITHM: Fix a parameter $\alpha \in [0, 1]$. We will construct a random sequence of labeled binary trees $(\mathbf{F}_n)_{n \geqslant 1}$. Start with $\mathbf{F}_1$ being a binary tree with two leaves labeled A_0 and A_1. For $n \geqslant 1$, given $\mathbf{F}_n$, to obtain $\mathbf{F}_{n+1}$ we assign a weight $1 - \alpha$ to each of the n edges of $\mathbf{F}_n$ adjacent to a leaf and a weight α to each of the $n - 1$ other edges; then we select at random an edge e proportionally to its weight and split it as in Rémy's algorithm. That is we place a middle vertex on e, to which we attach a new edge carrying a new leaf denoted by A_{n+1}.

FORD'S MODIFIED ALGORITHM: We consider now the following modification of Ford's algorithm, which we denote $(\tilde{\mathbf{F}}_n)_{n \geqslant 1}$. We proceed exactly as in Ford's algorithm except that, once the edge e has been selected at step n, we first find the unique $i \in \{1, 2, \ldots, n\}$ such that e belongs to the geodesic joining the leaf A_i to the set $\mathsf{Span}(\tilde{\mathbf{F}}_n; A_0, A_1, \ldots, A_{i-1})$, then we choose a *new* edge f *uniformly at random* on this geodesic, split it as in Rémy's algorithm and attach the new leaf A_{n+1} to it.

Observe that in the case $\alpha = 1/2$ both Ford's algorithm and its modified version have the same distribution as Rémy's algorithm. The analog of Proposition 10 is this case is the following: For $\alpha = 1/(2 + \delta)$, we have the following joint equality in distribution

$$(\mathsf{Loop}(T_n^\delta); n \geqslant 1) \quad \overset{(d)}{=} \quad (\mathsf{Glu}(\tilde{\mathbf{F}}_n); n \geqslant 1).$$

This follows from the fact that choosing the first edge in Ford's modified algorithm amounts to choosing a vertex of T_n^δ according to the LPAM^δ rule, and choosing the second edge amounts to choosing a corner of this vertex uniformly at random. We leave details to the reader. We also mention that the original Ford algorithm also corresponds to a plane LPAM^δ, but in which corners do not play exchangeable roles (the first corner around each vertex has weight $1 - \alpha$ and all others weight α).

An analog of (15) is known for Ford's algorithm. The sequence of random rescaled label trees $n^{-\alpha} \cdot \mathbf{F}_n$ converges almost surely towards a random compact labeled self-similar $\mathbb{R}$-tree of Hausdorff dimension $1/\alpha$ (belonging to the family of so-called fragmentation trees). See [18] for details. A way to prove Conjecture 2 would be to first prove analog convergences for the trees $\tilde{\mathbf{F}}_n$ arising from Ford's modified algorithm. We hope to exploit these connections in a future work.

4.2 Connections with the Poisson boundary

Finally, we connect the concept of the influence of the seed with the notion of the Poisson boundary of a transient Markov chain, which captures the information contained in its tail σ-field. Consider a Markov chain X on a countable state space V. Assume that we may write $V = V_0 \sqcup V_1 \sqcup V_2 \sqcup \cdots$ in such a way that the transitions from V_i always belong to V_{i+1} for $i \geqslant 0$. We call this the "layer" condition, and call V_i a layer. In our case, V_n is just the set of all looptrees associated with trees with n vertices. In particular, this Markov chain is transient. For $x \in V_i$, we denote by $(X_n^{(x)})_{n \geqslant i}$ the Markov chain started from x. In particular, $X_n^{(x)} \in V_n$ for every $n \geqslant i$. For every starting points $x, y \in V$, we define the asymptotic total variation:

$$d(x, y) := \lim_{n \to \infty} d_{\mathrm{TV}}(X_n^{(x)}, X_n^{(y)}).$$

We shall give an alternative expression for the pseudo-distance d by using the Poisson boundary of the chain. The Poisson boundary of X is a measurable space $\mathcal{P} = (E, \mathcal{A})$, which is also endowed with a family of probability measures $(\nu_x)_{x \in V}$ such that any bounded harmonic function h on V can be represented as

$$h(x) = \int d\nu_x(\xi) \mathbf{h}(\xi) \tag{20}$$

where $\mathbf{h} : E \to \mathbb{R}$ is a bounded measurable function on E. The measures ν_x can be interpreted as the harmonic measures on E seen from x. The most classical way to construct the Poisson boundary is via the construction of the Martin boundary of the chain, we refer to [30, Chap. 4] for details. We also mention that the Poisson boundary captures the information contained in the tail σ-field of X. Indeed, there is a one-to-one correspondence between bounded harmonic functions h and equivalence classes of bounded random variables Z measurable with respect to the tail σ-field of X which is given by the formula $h(x) = \mathbb{E}_x[Z]$ for $x \in V$. In our setting, we have

Proposition 19. *For every* $x, y \in V$, *we have* $d(x, y) = d_{\mathrm{TV}}(\nu_x, \nu_y)$.

Proof. We first express $d(x, y)$ in terms of harmonic functions. If $x, y \in V_1 \cup V_2 \cup \ldots \cup V_i$ we claim that for $n \geqslant i$ we have

$$d_{\mathrm{TV}}(X_n^{(x)}, X_n^{(y)}) = \frac{1}{2} \sup_{h \in \mathcal{H}_n} |h(x) - h(y)|, \tag{21}$$

where $\mathcal{H}_n = \{h : \bigsqcup_{i=1}^n V_i \to \mathbb{R},$ harmonic on $V_0 \sqcup \cdots \sqcup V_{n-1}$ and $\|h\|_\infty \leqslant 1\}$. To establish this equality, remark that if we denote by $\nu_x^{(n)}$ the law of the first hitting point of V_n by the chain starting from x (which is also unique visited point in V_n by our layer condition). Then observe that by classical potential theory, for every set $A \subset V_n$ we have

$$\left| \mathbb{P}(X_n^{(x)} \in A) - \mathbb{P}(X_n^{(y)} \in A) \right| = \left| \int_{V_n} d\nu_x^{(n)}(\xi) \mathbf{1}_A(\xi) - \int_{V_n} d\nu_y^{(n)}(\xi) \mathbf{1}_A(\xi) \right|$$

$$= \frac{1}{2} \left| \int_{V_n} d\nu_x^{(n)}(\xi) \left(\mathbf{1}_A(\xi) - \mathbf{1}_{A^c}(\xi) \right) - \int_{V_n} d\nu_y^{(n)}(\xi) \left(\mathbf{1}_A(\xi) - \mathbf{1}_{A^c}(\xi) \right) \right|.$$

It is plain to see that functions $h_A : x \mapsto \int_{V_n} d\nu_x^{(n)}(\xi)(\mathbf{1}_A(\xi) - \mathbf{1}_{A^c}(\xi))$ are the extreme points of the convex set $\mathcal{H}_n$. Hence, by convexity of $h \mapsto |h(x) - h(y)|$, we get that

$$d_{TV}(X_n^{(x)}, X_n^{(y)}) = \frac{1}{2} \sup_{A \subset V_n} |h_A(x) - h_A(x)| = \frac{1}{2} \sup_{h \in \mathcal{H}_n} |h(x) - h(y)|.$$

This establishes (21).

By taking the limit $n \to \infty$, we get then get that

$$d(x, y) = \frac{1}{2} \sup \{|h(x) - h(y)|\}, \tag{22}$$

where the supremum runs over all harmonic functions h on V whose $\| \cdot \|_\infty$ norm is bounded by one. Using the Poisson representation of bounded harmonic functions (20), it is a simple matter to check that the supremum on the right-hand side of (22) is actually equal to

$$\sup_{\substack{h : E \to \mathbb{R} \\ \text{Borel, } \|h\|_\infty \leqslant 1}} \left| \int d\nu_x(\xi) h(\xi) - \int d\nu_y(\xi) h(\xi) \right| = 2 \sup_{A \subset E, \text{ Borel}} \left| \int d\nu_x(\xi) \mathbf{1}_A(\xi) - \int d\nu_y(\xi) \mathbf{1}_A(\xi) \right|$$

$$= 2 \sup_{A \subset E, \text{ Borel}} |\nu_x(A) - \nu_y(A)| = 2 d_{TV}(\nu_x, \nu_y).$$

This completes the proof. □

In view of Proposition 19, a natural open question raised by our work is the following.

Open Question. *Is the measured space of scaling limits of discrete looptrees isomorphic to the Poisson boundary of the chain of planar preferential attachment trees thus implying (1)? Or, equivalently, are all asymptotic events of the chain measurable with respect to the scaling limit $\mathcal{L}$?*

In particular, we believe that for every decorated tree τ, the limiting value of the martingale $M_\tau^{(S)}(n)$ used to prove Theorem 1 is a measurable function of $\mathcal{L}^{(S)}$.

References

[1] L. ADDARIO-BERRY, N. BROUTIN, AND C. GOLDSCHMIDT, *The continuum limit of critical random graphs*, Probab. Theory Related Fields, 152 (2012), pp. 367–406.

[2] D. ALDOUS, *The continuum random tree. I*, Ann. Probab., 19 (1991), pp. 1–28.

[3] D. ALDOUS, *Recursive self-similarity for random trees, random triangulations and Brownian excursion.*, Ann. Probab., 22 (1994), pp. 527–545.

[4] K. B. ATHREYA, *On a characteristic property of Polya's urn*, Studia Sci. Math. Hungar., 4 (1969), pp. 31–35.

[5] A.-L. BARABÁSI AND R. ALBERT, *Emergence of scaling in random networks*, Science, 286 (1999), pp. 509–512.

[6] N. BERGER, C. BORGS, J. T. CHAYES, AND A. SABERI, *Asymptotic behavior and distributional limits of preferential attachment graphs*, Ann. Probab., 42 (2014), pp. 1–40.

[7] B. BOLLOBÁS, O. RIORDAN, J. SPENCER, AND G. TUSNÁDY, *The degree sequence of a scale-free random graph process*, Random Structures Algorithms, 18 (2001), pp. 279–290.

[8] S. BUBECK, E. MOSSEL, AND M. Z. RÁCZ, *On the influence of the seed graph in the preferential attachment model*, Preprint available on arxiv, http://arxiv.org/abs/1401.4849v3, (2014).

[9] ——, *On the influence of the seed graph in the preferential attachment model*, Preprint available on arxiv, http://arxiv.org/abs/1401.4849v2, (2014).

[10] D. BURAGO, Y. BURAGO, AND S. IVANOV, *A course in metric geometry*, vol. 33 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2001.

[11] B. CHAUVIN, C. MAILLER, AND N. POUYANNE, *Smoothing equations for large Pólya urns.*, Journal of Theoretical Probability, (To appear).

[12] N. CURIEN AND B. HAAS, *The stable trees are nested*, Probab. Theory Related Fields, 157 (2013), pp. 847–883.

[13] N. CURIEN AND I. KORTCHEMSKI, *Percolation on random triangulations and stable looptrees*, Preprint available on arxiv, http://arxiv.org/abs/1307.6818, (Submitted).

[14] ——, *Random stable looptrees*, arXiv:1304.1044, (submitted).

[15] S. DEREICH AND P. MÖRTERS, *Random networks with sublinear preferential attachment: the giant component*, Ann. Probab., 41 (2013), pp. 329–384.

[16] S. N. EVANS, *Probability and real trees*, vol. 1920 of Lecture Notes in Mathematics, Springer, Berlin, 2008. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005.

[17] D. J. FORD, *Probabilities on cladograms: Introduction to the alpha model*, Preprint. Available at arXiv:math/0511246v1.

[18] B. HAAS AND G. MIERMONT, *Scaling limits of Markov branching trees, with applications to Galton-Watson and random unordered trees*, Ann. of Probab., 40 (2012), pp. 2589–2666.

[19] J.-F. LE GALL, *Random trees and applications*, Probability Surveys, (2005).

[20] ——, *Random geometry on the sphere*, To appear in the Proceedings of ICM 2014, Seoul, available on arXiv, (2014).

REFERENCES

[21] P. MATTILA, *Geometry of sets and measures in Euclidean spaces*, vol. 44 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1995. Fractals and rectifiability.

[22] G. MIERMONT, *Tessellations of random maps of arbitrary genus*, Ann. Sci. Éc. Norm. Supér. (4), 42 (2009), pp. 725–781.

[23] T. F. MÓRI, *On random trees*, Studia Sci. Math. Hungar., 39 (2002), pp. 143–155.

[24] T. F. MÓRI, *The maximum degree of the Barabási-Albert random tree*, Combin. Probab. Comput., 14 (2005), pp. 339–348.

[25] E. A. PEKÖZ, A. RÖLLIN, AND N. ROSS, *Joint degree distributions of preferential attachment random graphs*, Preprint available on arxiv, http://arxiv.org/abs/1402.4686, (2014).

[26] J. PITMAN, *Combinatorial stochastic processes*, vol. 1875 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.

[27] J.-L. RÉMY, *Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire*, RAIRO Inform. Théor., 19 (1985), pp. 179–195.

[28] J. SZYMAŃSKI, *On a nonuniform random recursive tree*, in Random graphs '85 (Poznań, 1985), vol. 144 of North-Holland Math. Stud., North-Holland, Amsterdam, 1987, pp. 297–306.

[29] R. VAN DER HOFSTAD, *Lecture notes random graphs and complex networks*, In preparation, R. van der Hofstad (May 2013). Random graphs and complex networks. http://www.win.tue.nl/ rhofstad/NotesRGCN.pdf, (2013).

[30] W. WOESS, *Random walks on infinite graphs and groups*, vol. 138 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2000.