CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

The distance profile of rooted and unrooted simply generated trees

Gabriel Berzunza Ojeda[1,*] and Svante Janson[2]

[1]Department of Mathematical Sciences, University of Liverpool, Mathematical Sciences Building, Liverpool L69 7ZL, UK and [2]Department of Mathematics, Uppsala University, PO Box 480, Uppsala SE-751 06, Sweden
[*]Corresponding author. Email: gabriel.berzunza-ojeda@liverpool.ac.uk

Abstract

It is well known that the height profile of a critical conditioned Galton–Watson tree with finite offspring variance converges, after a suitable normalisation, to the local time of a standard Brownian excursion. In this work, we study the distance profile, defined as the profile of all distances between pairs of vertices. We show that after a proper rescaling the distance profile converges to a continuous random function that can be described as the density of distances between random points in the Brownian continuum random tree. We show that this limiting function a.s. is Hölder continuous of any order $\alpha < 1$, and that it is a.e. differentiable. We note that it cannot be differentiable at 0, but leave as open questions whether it is Lipschitz, and whether it is continuously differentiable on the half-line $(0, \infty)$. The distance profile is naturally defined also for unrooted trees contrary to the height profile that is designed for rooted trees. This is used in our proof, and we prove the corresponding convergence result for the distance profile of random unrooted simply generated trees. As a minor purpose of the present work, we also formalize the notion of unrooted simply generated trees and include some simple results relating them to rooted simply generated trees, which might be of independent interest.

1. Introduction

Consider a random simply generated tree. (For definitions of this and other concepts in the introduction, see Sections 2–3.) Under some technical conditions, amounting to the tree being equivalent to a critical conditioned Galton–Watson tree with finite offspring variance, the (height) profile of the tree converges in distribution, as a random function in $C[0, \infty)$. Moreover, the limiting random function can be identified with the local time of a standard Brownian excursion; this was conjectured by Aldous [3] and proved by Drmota and Gittenberger [18] (under a stronger assumption), see also Drmota [17, Section 4.2], and in general by Kersting [34] in a paper that unfortunately remains unpublished. See further Pitman [47] for related results and a proof in a special case. See also Kersting [34] for extensions when the offspring variance is infinite, a case not considered in the present paper.

Remark 1.1. To be precise, [17] and [18] assume that the offspring distribution for the conditioned Galton–Watson tree has a finite exponential moment. As said in [17, footnote on page 127], the analysis can be extended, but it seems that the proof of tightness in [17], which is based on estimating fourth moments, requires a finite fourth moment of the offspring distribution.

CrossMark

Note also that Drmota [17, 'a shortcut' pp. 123–125] besides the proof from [18] also gives an alternative proof that combines tightness (taken from the first proof) and the convergence of the contour process to a Brownian excursion shown by Aldous [4], and thus avoids some intricate calculations in the first proof. We will use this method of proof below.   □

Using notation introduced below, the result can be stated as follows.

Theorem 1.2. *(Drmota and Gittenberger [18], Kersting [34]). Let L_n be the (height) profile of a conditioned Galton–Watson tree of order n, with an offspring distribution that has mean 1 and finite variance σ^2. Then, as $n \to \infty$,*

$$n^{-1/2}L_n(xn^{1/2}) \xrightarrow{d} \frac{\sigma}{2}L_e\left(\frac{\sigma}{2}x\right), \tag{1}$$

in the space $C[0,\infty]$, where L_e is a random function that can be identified with the local time of a standard Brownian excursion e; this means that for every bounded measurable $f : [0,\infty) \to \mathbb{R}$,

$$\int_0^\infty f(x)L_e(x)\,dx = \int_0^1 f(e(t))\,dt. \tag{2}$$

Remark 1.3. This result is often stated with convergence (1) in the space $C[0,\infty]$; the version stated here with $C[0,\infty]$ is somewhat stronger but follows easily. (Note that the maximum is a continuous functional on $C[0,\infty]$ but not on $C[0,\infty)$.) See further Section 2.4.   □

The profile discussed above is the profile of the distances from the vertices to the root. Consider now instead the *distance profile*, defined as the profile of all distances between pairs of points. (Again, see Section 2 for details.) One of our main results is the following analogue of Theorem 1.2.

Theorem 1.4. *Let Λ_n be the distance profile of a conditioned Galton–Watson tree of order n, with an offspring distribution that has mean 1 and finite variance $\sigma^2 > 0$. Then, as $n \to \infty$,*

$$n^{-3/2}\Lambda_n(xn^{1/2}) \xrightarrow{d} \frac{\sigma}{2}\Lambda_e\left(\frac{\sigma}{2}x\right), \tag{3}$$

in the space $C[0,\infty]$, where $\Lambda_e(x)$ is a continuous random function that can be described as the density of distances between random points in the Brownian continuum random tree [2–4]; equivalently, for a standard Brownian excursion e, we have for every bounded measurable $f : [0,\infty) \to \mathbb{R}$,

$$\int_0^\infty f(x)\Lambda_e(x)\,dx = 2\iint_{0<s<t<1} f\left(e(s) + e(t) - 2\min_{u\in[s,t]}e(u)\right)ds\,dt. \tag{4}$$

The random distance profile Λ_n was earlier studied in [16], where the estimate (123) below was shown.

Remark 1.5. It is easy to see that the random function Λ_e really is random and not deterministic, e.g. as a consequence of Theorem 13.1. However, we do not know its distribution, although the expectation $\mathbb{E}\Lambda_e(x)$ is given in Lemma 15.11. In particular, the following problem is open. (See [17, Section 4.2.1] for such results, in several different forms, for L_e.)   □

Problem 1.6. Find a description of the (one-dimensional) distribution of $\Lambda_e(x)$ for fixed $x > 0$.

We have so far discussed rooted trees. However, the distance profile is defined also for unrooted trees, and we will find it convenient to use unrooted trees in parts of the proof. This leads us to consider random *unrooted simply generated trees*.

Simply generated families of rooted trees were introduced by Meir and Moon [42], leading to the notion of simply generated random rooted trees, see e.g. Drmota [17]. This class of random rooted trees is one of the most popular classes of random trees, and these trees have been frequently studied in many different contexts by many authors. Simply generated random *unrooted* trees have been much less studied, but they have occurred, e.g. in a work on non-crossing trees

by Kortchemski and Marzouk [38] (see also Marckert and Panholzer [41]). Nevertheless, we have not found a general treatment of them, so a minor purpose of the present paper is to do this in some detail, both for use in the paper and for future reference. We thus include (Sections 5–8) a general discussion of random unrooted simply generated trees, with some simple results relating them to rooted simply generated trees, allowing the transfer of many results for rooted simply generated trees to the unrooted case. Moreover, as part of the proof of Theorem 1.4, we prove the corresponding result (Theorem 11.2) for random unrooted simply generated trees.

As a preparation for the unrooted case, we also give (Section 4) some results (partly from Kortchemski and Marzouk [38]) on modified rooted simply generated trees (Galton–Watson trees), where the root has different weights (offspring distribution) than all other vertices.

The central parts of the proof of Theorem 1.4 are given in Sections 10–12, where we use both rooted and unrooted trees. As a preparation, in Section 9, we extend Theorem 1.2 to conditioned modified Galton–Watson trees. We later also extend Theorem 1.4 to conditioned modified Galton–Watson trees (Theorem 12.1).

We end the paper with some comments and further results related to our main results. In Section 13, we discuss a simple application to the Wiener index of unrooted simply generated trees. Section 14 contains some important moment estimations of the distance profile for conditioned Galton–Watson trees as well as for its continuous counterpart $\Lambda_{\mathbf{e}}$. In Section 15, we establish Hölder continuity properties of the continuous random function $L_{\mathbf{e}}$ and $\Lambda_{\mathbf{e}}$. It is known that $L_{\mathbf{e}}$ is a.s. Hölder continuous of order α (abbreviated to Hölder(α)) for $\alpha < \frac{1}{2}$, but not for $\alpha \geqslant \frac{1}{2}$. We show that $\Lambda_{\mathbf{e}}$ is smoother; it is a.s. Hölder(α) for $\alpha < 1$, and it is a.e. differentiable (Theorem 15.5). We do not know whether it is Lipschitz, or even continuously differentiable on $[0, \infty)$, but we show that it is does not a.s. have a two-sided derivative at 0 (Theorem 15.10), and we state some open problems.

Finally, some further remarks are given in Section 16.

2. Some notation

Trees are finite except when explicitly said to be infinite. Trees may be *rooted* or *unrooted*; in a rooted tree, the root is denoted o. The rooted trees may be *ordered* or not. The unrooted trees will always be *labelled*; we do not consider unrooted unlabelled trees in the present paper.

If T is a tree, then its number of vertices is denoted by $|T|$; this is called the *order* or the *size* of T. (Unlike some authors, we do not distinguish between order and size.) The notation $v \in T$ means that v is a vertex in T.

The degree of a vertex $v \in T$ is denoted $d(v)$. In a rooted tree, we also define the outdegree $d^+(v)$ as the number of children of v; thus,

$$d^+(v) = \begin{cases} d(v) - 1, & v \neq o, \\ d(v), & v = o. \end{cases} \tag{5}$$

A *leaf* in an unrooted tree is a vertex v with $d(v) = 1$. In a rooted tree, we instead require $d^+(v) = 0$; this may make a difference only for the root.

A *fringe subtree* in a rooted tree is a subtree consisting of some vertex v and all its descendants. We regard v as the root of the fringe tree. The *branches* of a rooted tree are the fringe trees rooted at the children of the root. The number of branches thus equals the degree $d(o)$ of the root.

Let $\mathfrak{T}_n$ be the set of all ordered rooted trees of order n, and let $\mathfrak{T} := \bigcup_1^\infty \mathfrak{T}_n$. Note that $\mathfrak{T}_n$ is a finite set; we may identify the vertices of an ordered rooted tree by finite strings of positive integers, such that the root is the empty string and the children of v are vi, $i = 1, \ldots, d(v)$. (Thus, an ordered rooted tree is regarded as a subtree of the infinite Ulam–Harris tree.) In fact, it is well known that $|\mathfrak{T}_n| = \frac{1}{n}\binom{2n-2}{n-1}$, the Catalan number C_{n-1}.

Let $\mathfrak{L}_n$ be the set of all unrooted trees of order n, with the labels $1, \ldots, n$; thus $\mathfrak{L}_n$ is the set of all trees on $[n] := \{1, \ldots, n\}$. $\mathfrak{L}_n$ is evidently finite and by Cayley's formula $|\mathfrak{L}_n| = n^{n-2}$. Let $\mathfrak{L} := \bigcup_1^\infty \mathfrak{L}_n$.

A *probability sequence* is the same as a probability distribution on $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$, i.e., a sequence $\mathbf{p} = (p_k)_0^\infty$ with $p_k \geqslant 0$ and $\sum_{k=0}^\infty p_k = 1$. The *mean* $\mu(\mathbf{p})$ and *variance* $\sigma^2(\mathbf{p})$ of a probability sequence are defined to be the mean and variance of a random variable with distribution $\mathbf{p}$, i.e.,

$$\mu(\mathbf{p}) := \sum_{k=0}^\infty k p_k, \qquad \sigma^2(\mathbf{p}) := \sum_{k=0}^\infty k^2 p_k - \mu(\mathbf{p})^2. \qquad (6)$$

We use $\xrightarrow{\text{d}}$ and $\xrightarrow{\text{p}}$ for convergence in distribution and in probability, respectively, for a sequence of random variables in some metric space; see e.g. [11]. Also, $\overset{\text{d}}{=}$ means convergence in distribution.

The total variation distance between two random variables X and Y in a metric space (or rather between their distributions) is defined by

$$d_{\text{TV}}(X, Y) := \sup_A \big|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)\big|, \qquad (7)$$

taking the supremum over all measurable subsets A. It is well known that in a complete separable metric space, there exists a coupling of X and Y (i.e., a joint distribution with the given marginal distributions) such that

$$\mathbb{P}(X \neq Y) = d_{\text{TV}}(X, Y), \qquad (8)$$

and this is best possible.

$O_p(1)$ denotes a sequence of real-valued random variables $(X_n)_n$ that is stochastically bounded, i.e., for every $\varepsilon > 0$, there exists C such that $\mathbb{P}(|X_n| > C) \leqslant \varepsilon$. This is equivalent to $(X_n)_n$ being tight. For tightness in more general metric spaces, see e.g. [11].

Let f be a real-valued function defined on an interval $I \subseteq \mathbb{R}$. The *modulus of continuity* of f is the function $[0, \infty) \to [0, \infty]$ defined by, for $\delta \geqslant 0$,

$$\omega(\delta; f) = \omega(\delta; f; I) := \sup\big(|f(s) - f(t)| : s, t \in I, |s - t| \leqslant \delta\big). \qquad (9)$$

If x and y are real numbers, $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$.

C denotes unspecified constants that may vary from one occurrence to the next. They may depend on parameters such as weight sequences or offspring distributions, but they never depend on the size of the trees. Sometimes we write, e.g., C_r to emphasize that the constant depends on the parameter r.

Unspecified limits are as $n \to \infty$.

2.1. Profiles

For two vertices v and w in a tree T, let $\mathsf{d}(x, y) = \mathsf{d}_T(x, y)$ denote the distance between v and w, i.e., the number of edges in the unique path joining v and w. In particular, in a rooted tree, $\mathsf{d}(v, o)$ is the distance to the root, often called the *depth* (or sometimes *height*) of v.[1]

For a rooted tree T, the *height* of T is $H(T) := \max_{v \in T} \mathsf{d}(v, o)$, i.e., the maximum depth. The *diameter* of a tree T, rooted or not, is $\text{diam}(T) := \max_{v,w \in T} \mathsf{d}(v, w)$.

The *profile* of a rooted tree is the function $L = L_T : \mathbb{R} \to [0, \infty)$ defined by

$$L(i) := \big|\{v \in T : \mathsf{d}(v, o) = i\}\big|, \qquad (10)$$

---

[1] We use different fonts to distinguish the distance d from the degree d; note also that the distance has two arguments and the degree only one.

for integers i, extended by linear interpolation to all real x. (We are mainly interested in $x \geqslant 0$, and trivially $L(x) = 0$ for $x \leqslant -1$, but it will be convenient to allow negative x.) The linear interpolation can be written as

$$L(x) := \sum_{i=0}^{\infty} L(i)\tau(x-i), \tag{11}$$

where τ is the triangular function $\tau(x) := (1 - |x|) \vee 0$.

Note that $L(0) = 1$, and that L is a continuous function with compact support $[-1, H(T) + 1]$. Furthermore, since $\int \tau(x)\,dx = 1$,

$$\int_{-1}^{\infty} L(x)\,dx = \sum_{i=0}^{\infty} L(i) = |T|, \tag{12}$$

where we integrate from -1 because of the linear interpolation; we have $\int_0^{\infty} L(x)\,dx = \sum_{i=0}^{\infty} L(i) - \frac{1}{2} = |T| - \frac{1}{2}$.

The *width* of T is defined as

$$W(T) := \max_{i \in \mathbb{N}_0} L(i) = \max_{x \in \mathbb{R}} L(x). \tag{13}$$

Similarly, in any tree T, rooted or unrooted, we define the *distance profile* as the function $\Lambda = \Lambda_T : [0, \infty) \to [0, \infty)$ defined by

$$\Lambda(i) := \left| \{(v, w) \in T : \mathsf{d}(v, w) = i\} \right| \tag{14}$$

for integers i, again extended by linear interpolaton to all real $x \geqslant 0$. For definiteness, we count ordered pairs in (14), and we include the case $v = w$, so $\Lambda(0) = |T|$. Λ is a continuous function on $[0, \infty)$ with support $[-1, \operatorname{diam}(T) + 1]$. We have, similarly to (12),

$$\int_{-1}^{\infty} \Lambda(t)\,dt = \sum_{i=0}^{\infty} \Lambda(i) = |T|^2. \tag{15}$$

If T is an unrooted tree, let $T(v)$ denote the rooted tree obtained by declaring v as the root, for $v \in T$. Then, as a consequence of (10) and (14),

$$\Lambda_T(x) = \sum_{v \in T} L_{T(v)}(x). \tag{16}$$

Hence, the distance profile can be regarded as the sum (or, after normalisation, average) of the profiles for all possible choices of a root.

Remark 2.1. Alternatively, one might extend L to a step function by $L(x) := L(\lfloor x \rfloor)$, and similarly for Λ. The asymptotic results are the same (and equivalent by simple arguments), with L and Λ elements of $D[0, \infty]$ instead of $C[0, \infty]$ and limit theorems taking place in that space. This has some advantages, but for technical reasons (e.g. simpler tightness criteria), we prefer to work in the space $C[0, \infty]$ of continuous functions. □

Remark 2.2. Another version of Λ would count unordered pairs of distinct vertices. The two versions are obviously equivalent and our results hold for the alternative version too, mutatis mutandis. □

2.2. Brownian excursion and its local time

The standard Brownian excursion $\mathbf{e}$ is a random continuous function $[0, 1] \to [0, \infty)$ such that $\mathbf{e}(0) = \mathbf{e}(1) = 0$ and $\mathbf{e}(t) > 0$ for $t \in (0, 1)$. Informally, $\mathbf{e}$ can be regarded as a Brownian motion conditioned on these properties; this can be formalised as an appropriate limit [23]. There are

several other, quite different but equivalent, definitions, see e.g. [49, XII.(2.13)], [13, Example II.1d)], and [17, Section 4.1.3].

The *local time* $L_{\mathbf{e}}$ of $\mathbf{e}$ is a continuous random function that is defined (almost surely) as a functional of $\mathbf{e}$ satisfying

$$\int_0^\infty f(x)L_{\mathbf{e}}(x)\,\mathrm{d}x = \int_0^1 f\big(\mathbf{e}(t)\big)\,\mathrm{d}t, \tag{17}$$

for every bounded (or non-negative) measurable function $f : [0, \infty) \to \mathbb{R}$. In particular, (17) yields, for any $x \geqslant 0$ and $\varepsilon > 0$,

$$\int_x^{x+\varepsilon} L_{\mathbf{e}}(y)\,\mathrm{d}y = \int_0^1 \mathbf{1}\{\mathbf{e}(t) \in [x, x+\varepsilon]\}\,\mathrm{d}t \tag{18}$$

and thus

$$L_{\mathbf{e}}(x) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^1 \mathbf{1}\{\mathbf{e}(t) \in [x, x+\varepsilon]\}\,\mathrm{d}t. \tag{19}$$

Hence, $L_{\mathbf{e}}(x)$ can be regarded as the occupation density of $\mathbf{e}$ at the value x.

Note that the existence (almost surely) of a function $L_{\mathbf{e}}(x)$ satisfying (17)–(19) is far from obvious; this is part of the general theory of local times for semimartingales, see e.g. [49, Chapter VI]. The existence also follows from (some of) the proofs of Theorem 1.2.

2.3. Brownian continuum random tree

Given a continuous function $g : [0, 1] \to [0, \infty)$ with $g(0) = g(1) = 0$, one can define a pseudo-metric d on $[0, 1]$ by

$$\mathsf{d}(s, t) = \mathsf{d}(s, t; g) := g(s) + g(t) - 2\min_{u \in [s,t]} g(u), \qquad 0 \leqslant s \leqslant t \leqslant 1. \tag{20}$$

By identifying points with distance 0, we obtain a metric space T_g, which is a compact real tree, see e.g. Le Gall [39, Theorem 2.2]. We denote the natural quotient map $[0, 1] \to T_g$ by ρ_g, and let T_g be rooted at $\rho_g(0)$. The *Brownian continuum random tree* defined by Aldous[2–4] can be defined as the random real tree $T_{\mathbf{e}}$ constructed in this way from the random Brownian excursion $\mathbf{e}$, see [39, Section 2.3]. (Aldous [2–4] used another definition, and another scaling corresponding to $T_{2\mathbf{e}}$.) Note that using (20), (4) can be written

$$\int_0^\infty f(x)\Lambda_{\mathbf{e}}(x)\,\mathrm{d}x = \iint_{s,t \in [0,1]} f\big(\mathsf{d}(s, t; \mathbf{e})\big)\,\mathrm{d}s\,\mathrm{d}t, \tag{21}$$

for any bounded (or non-negative) measurable function f. This means that $\Lambda_{\mathbf{e}}$ is the density of the distance in $T_{\mathbf{e}}$ between two random points, chosen independently with the probability measure on $T_{\mathbf{e}}$ induced by the uniform measure on $[0, 1]$. This justifies the equivalence of the two definitions of $\Lambda_{\mathbf{e}}$ stated in Theorem 1.4. As for the local time $L_{\mathbf{e}}$, the existence (almost surely) of a continuous function $\Lambda_{\mathbf{e}}$ satisfying (21) is far from trivial; this will be a consequence of our proof.

An important feature of the Brownian continuum random tree is its re-rooting invariance property. More precisely, fix $s \in [0, 1]$ and set

$$\mathbf{e}^{[s]}(t) = \begin{cases} \mathsf{d}(s, s+t; \mathbf{e}), & 0 \leqslant t < 1-s \\ \mathsf{d}(s, s+t-1; \mathbf{e}), & 1-s \leqslant t \leqslant 1. \end{cases} \tag{22}$$

Note that $\mathbf{e}^{[s]}$ is a random continuous function $[0, 1] \to [0, \infty)$ such that $\mathbf{e}^{[s]}(0) = \mathbf{e}^{[s]}(1) = 0$ and a.s. $\mathbf{e}^{[s]}(t) > 0$ for $t \in (0, 1)$; clearly, $\mathbf{e}^{[0]} = \mathbf{e}$. By Duquesne and Le Gall [21, Lemma 2.2], the compact real tree $T_{\mathbf{e}^{[s]}}$ is then canonically identified with the $T_{\mathbf{e}}$ tree re-rooted at the vertex $\rho_{\mathbf{e}}(s)$. Marckert and Mokkadem [40, Proposition 4.9] (see also Duquesne and Le Gall [22, Theorem 2.2]) have shown that for every fixed $s \in [0, 1]$,

$$\mathbf{e}^{[s]} \overset{\mathrm{d}}{=} \mathbf{e} \quad \text{and} \quad T_{\mathbf{e}^{[s]}} = T_{\mathbf{e}}, \tag{23}$$

in distribution. Thus, the re-rooted tree $T_{\mathbf{e}^{[s]}}$ is a version of the Brownian continuum random tree.

Remark 2.3. Indeed, Aldous [3, (20)] already observed that the Brownian continuum random tree is invariant under uniform re-rooting and that this property corresponds to the invariance of the law of the Brownian excursion under the path transformation (22) if $s = U$ is uniformly random on $[0, 1]$ and independent of $\mathbf{e}$. □

As a consequence of the previous re-rooting invariance property, we deduce the following explicit expression for the continuous function $\Lambda_{\mathbf{e}}$. For every fixed $s \in [0, 1]$, let $L_{\mathbf{e}^{[s]}}$ denote the local time of $\mathbf{e}^{[s]}$, which is perfectly defined thanks to (23). It follows from (20), (21) and (22) that

$$\int_0^\infty f(x) \Lambda_{\mathbf{e}}(x) \, \mathrm{d}x = \int_0^1 \int_0^1 f\big(\mathbf{e}^{[s]}(t)\big) \, \mathrm{d}s \, \mathrm{d}t = \int_0^1 \int_0^\infty f(x) L_{\mathbf{e}^{[s]}}(x) \, \mathrm{d}x \, \mathrm{d}s,$$

for any bounded (or non-negative) measurable function f, or equivalently,

$$\Lambda_{\mathbf{e}}(x) = \int_0^1 L_{\mathbf{e}^{[s]}}(x) \, \mathrm{d}s, \quad x \geqslant 0. \tag{24}$$

In accordance with the discrete analogue of $\Lambda_{\mathbf{e}}$ in (16), the identity (24) shows that $\Lambda_{\mathbf{e}}$ can be regarded as the average of the profiles for all possible choices of a root in $T_{\mathbf{e}}$.

2.4. The function spaces $C[0, \infty)$ and $C[0, \infty]$

Recall that $C[0, \infty)$ is the space of continuous functions on $[0, \infty)$ and that convergence in $C[0, \infty)$ means uniform convergence on each compact interval $[0, b]$. As said in Remark 1.3, we prefer to state our results in the space $C[0, \infty]$ of functions that are continuous on the extended half-line $[0, \infty]$. These are the functions f in $C[0, \infty)$ such that the limit $f(\infty) := \lim_{x \to \infty} f(x)$ exists; in our case, this is a triviality since all random functions on both sides of (1) and (3), and in similar later statements, have compact support, and thus trivially extend continuously to $[0, \infty]$ with $f(\infty) = 0$. The important difference between $C[0, \infty]$ and $C[0, \infty)$ is instead the topology: convergence in $C[0, \infty]$ means uniform convergence on $[0, \infty]$ (or, equivalently, on $[0, \infty)$).

In particular, the supremum is a continuous functional on $C[0, \infty]$, but not on $C[0, \infty)$ (where it also may be infinite). Thus, convergence of the width (after rescaling), follows immediately from Theorem 1.2 (see also the proof of Theorem 9.2); if this was stated with convergence in $C[0, \infty)$, a small extra argument would be needed (more or less equivalent to showing convergence in $C[0, \infty]$).

In the random setting, the difference between the two topologies can be expressed as in the following lemma. See also [27, Proposition 2.4], for the similar case of the spaces $D[0, \infty]$ and $D[0, \infty)$.

Lemma 2.4. *Let $X_n(t)$ and $X(t)$ be random functions in $C[0, \infty]$. Then $X_n(t) \overset{\mathrm{d}}{\longrightarrow} X(t)$ in $C[0, \infty]$ if and only if*

(i) $X_n(t) \overset{\mathrm{d}}{\longrightarrow} X(t)$ *in $C[0, \infty)$, and*

(ii) $X_n(t) \overset{\mathrm{p}}{\longrightarrow} X_n(\infty)$, *as $t \to \infty$, uniformly in n; i.e., for every $\varepsilon > 0$,*

$$\sup_{n \geqslant 1} \mathbb{P}\big(\sup_{u < t < \infty} |X_n(t) - X_n(\infty)| > \varepsilon \big) \to 0, \qquad \text{as } u \to \infty. \tag{25}$$

Proof. A straightforward exercise. □

In our cases, such as (1) and (3), the condition (25) is easily verified from convergence (or just tightness) of the normalised height $H_n/\sqrt{n}$, which can be used to bound the support of the left-hand sides. Hence, convergence in $C[0, \infty)$ and $C[0, \infty]$ is essentially equivalent.

Note that $C[0, \infty]$ is a separable Banach space, and that it is isomorphic to $C[0, 1]$ by a change of variable; thus, general results for $C[0, 1]$ may be transferred. Note also that all functions that we are interested in lie in the (Banach) subspace $C_0[0, \infty) := \{f \in C[0, \infty] : f(\infty) = 0\}$. Hence, the results may just as well be stated as convergence in distribution in $C_0[0, \infty)$.

3. Rooted simply generated trees

As a background, we recall first the definition of random rooted simply generated trees and the almost equivalent conditioned Galton–Watson trees, see e.g. [17] or [30] for further details, and [6] for more on Galton–Watson processes.

3.1. Simply generated trees

Let $\boldsymbol{\phi} = (\phi_k)_0^\infty$ be a given sequence of non-negative weights, with $\phi_0 > 0$ and $\phi_k > 0$ for at least one $k \geqslant 2$. (The latter conditions exclude only trivial cases when the random tree $\mathcal{T}_n^{\boldsymbol{\phi}}$ defined below either does not exist or is a deterministic path.)

For any rooted ordered tree $T \in \mathfrak{T}$, define the weight of T as

$$\phi(T) := \prod_{v \in T} \phi_{d^+(v)}. \tag{26}$$

For a given n, we define the random rooted *simply generated tree* $\mathcal{T}_n^{\boldsymbol{\phi}}$ of order n as a random tree in $\mathfrak{T}_n$ with probability proportional to its weight; i.e.,

$$\mathbb{P}(\mathcal{T}_n^{\boldsymbol{\phi}} = T) := \frac{\phi(T)}{\sum_{T' \in \mathfrak{T}_n} \phi(T')}, \qquad T \in \mathfrak{T}_n. \tag{27}$$

We consider only n such that at least one tree T with $\phi(T) > 0$ exists.

A weight sequence $\boldsymbol{\phi}' = (\phi_k')_0^\infty$ with

$$\phi_k' = ab^k \phi_k, \qquad k \geqslant 0, \tag{28}$$

for some $a, b > 0$ is said to be *equivalent* to $(\phi_k)_0^\infty$. It is easily seen that equivalent weight sequences define the same random tree $\mathcal{T}_n^{\boldsymbol{\phi}}$, i.e., $\mathcal{T}_n^{\boldsymbol{\phi}'} \overset{\mathrm{d}}{=} \mathcal{T}_n^{\boldsymbol{\phi}}$.

3.2. Galton–Watson trees

Given a probability sequence $\mathbf{p} = (p_k)_0^\infty$, the *Galton–Watson tree* $\mathcal{T}^{\mathbf{p}}$ is the family tree of a Galton–Watson process with offspring distribution $\mathbf{p}$. This means that $\mathcal{T}^{\mathbf{p}}$ is a random ordered rooted tree, which is constructed as follows: Start with a root and give it a random number of children with the distribution $\mathbf{p}$. Give each new vertex a random number of children with the same distribution and independent of previous choices, and continue as long as there are new vertices. In general, $\mathcal{T}^{\mathbf{p}}$ may be an infinite tree. We will mainly consider the *critical* case when the expectation $\mu(\mathbf{p}) = 1$, and then it is well known that $\mathcal{T}^{\mathbf{p}}$ is finite a.s. (We exclude the trivial case when $p_1 = 1$.)

The size $|\mathcal{T}^{\mathbf{p}}|$ of a Galton–Watson tree is random. Given $n \geqslant 1$, the *conditioned Galton–Watson tree* $\mathcal{T}_n^{\mathbf{p}}$ is defined as $\mathcal{T}^{\mathbf{p}}$ conditioned on $|\mathcal{T}^{\mathbf{p}}| = n$. (We consider only n such that this happens with positive probability.) Consequently, $\mathcal{T}_n^{\mathbf{p}}$ is a random ordered rooted tree of size n. It is easily seen that a conditioned Galton–Watson tree $\mathcal{T}_n^{\mathbf{p}}$ equals (in distribution) the simply generated tree with the weight sequence $\mathbf{p}$, and thus we use the same notation $\mathcal{T}_n^{\mathbf{p}}$ for both.

A (conditioned) Galton–Watson tree is *critical* if its offspring distribution **p** has mean $\mu(\mathbf{p}) = 1$. We will in the present paper mainly consider conditioned Galton–Watson trees that are critical and have a finite variance $\sigma^2(\mathbf{p})$; this condition is rather mild, as is seen in the following subsection.

3.3. Equivalence

A random simply generated tree $\mathcal{T}_n^{\boldsymbol{\phi}}$ with a weight sequence $(\phi_k)_0^\infty$ that is a probability sequence equals, as just said, the conditioned Galton–Watson tree $\mathcal{T}_n^{\boldsymbol{\phi}}$. Much more generally, any weight sequence $\boldsymbol{\phi}$ such that its generating function

$$\Phi(z) := \sum_{k=0}^\infty \phi_k z^k \tag{29}$$

has positive radius of convergence is equivalent to some probability weight sequence; hence, $\mathcal{T}_n^{\boldsymbol{\phi}}$ can be regarded as a conditioned Galton–Watson tree in this case too. Moreover, in many cases, we can choose an equivalent probability weight sequence that has mean 1 and finite variance; see e.g. [30, Section 4].

We will use this to switch between simply generated trees and conditioned Galton–Watson trees without comment in the sequel; we will use the name that seems best and most natural in different contexts.

3.4. Simply generated forests

The Galton–Watson process above starts with one individual. More generally, we may start with m individuals, which we may assume are numbered $1, \dots, m$; this yields a *Galton–Watson forest* consisting of m independent copies of $\mathcal{T}^{\mathbf{p}}$. Conditioning on the total size being $n \geqslant m$, we obtain a *conditioned Galton–Watson forest* $\mathcal{T}_{n,m}^{\mathbf{p}}$, which thus consists of m random trees $\mathcal{T}_{n,m;\,1}^{\mathbf{p}}, \dots, \mathcal{T}_{n,m;\,m}^{\mathbf{p}}$ with $|\mathcal{T}_{n,m;\,1}^{\mathbf{p}}| + \cdots + |\mathcal{T}_{n,m;\,m}^{\mathbf{p}}| = n$. Conditioned on the sizes $|\mathcal{T}_{n,m;\,1}^{\mathbf{p}}|, \dots, |\mathcal{T}_{n,m;\,m}^{\mathbf{p}}|$, the trees are independent conditioned Galton–Watson trees with the given sizes.

More generally, given any weight sequence $\boldsymbol{\phi}$, a random *simply generated forest* $\mathcal{T}_{n,m}^{\boldsymbol{\phi}}$ is a random forest with m rooted trees and total size n, chosen with probability proportional to its weight, defined as in (26). Again, conditioned on their sizes, the trees are independent simply generated trees.

Thus, the distribution of the sizes of the trees in the forest is of major importance. Consider the Galton–Watson case, and let $\mathcal{T}_{n,m;\,(1)}^{\mathbf{p}}, \dots, \mathcal{T}_{n,m;\,(m)}^{\mathbf{p}}$ denote the trees arranged in decreasing order: $|\mathcal{T}_{n,m;\,(1)}^{\mathbf{p}}| \geqslant \cdots \geqslant |\mathcal{T}_{n,m;\,(m)}^{\mathbf{p}}|$. (Ties are resolved randomly, say; this applies tacitly to all similar situations.) We have the following general result, which was proved by Marzouk [38, Lemma 5.7(iii)] under an additional regularity hypothesis.

Lemma 3.1. *Let $m \geqslant 1$ be fixed, and consider the conditioned Galton–Watson forest $\mathcal{T}_{n,m}^{\mathbf{p}}$ as $n \to \infty$. Then*

$$|\mathcal{T}_{n,m;\,(i)}^{\mathbf{p}}| = \begin{cases} n - O_p(1), & i = 1 \\ O_p(1), & i = 2, \dots, m. \end{cases} \tag{30}$$

Proof. Suppose first that $\mu(\mathbf{p}) = 1$. Suppose also, for simplicity, that $p_m > 0$. Consider the conditioned Galton–Watson tree $\mathcal{T}_{n+1}^{\mathbf{p}}$ and condition on the event $\mathcal{E}_m$ that the root degree is m. Conditioned on $\mathcal{E}_m$, there are m branches, which form a conditioned Galton–Watson forest $\mathcal{T}_{n,m}^{\mathbf{p}}$.

As $n \to \infty$, the random tree $\mathcal{T}_{n+1}^{\mathbf{p}}$ converges in distribution to an infinite random tree $\widehat{\mathcal{T}}$ (the size-biased Galton–Watson tree defined by Kesten [35]), see [30, Theorem 7.1]. Moreover, $\mathbb{P}(\mathcal{E}_m) \to m p_m > 0$ by [30, Theorem 7.10]. Hence, $\mathcal{T}_{n+1}^{\mathbf{p}}$ conditioned on $\mathcal{E}_m$ converges in distribution to $\widehat{\mathcal{T}}$ conditioned on $\mathcal{E}_m$. In other words, the forest $\mathcal{T}_{n,m}^{\mathbf{p}}$ converges in distribution to the branches of $\widehat{\mathcal{T}}$ conditioned on having exactly m branches; denote this random limit by $(\mathcal{T}_1, \ldots, \mathcal{T}_m)$. By the Skorohod coupling theorem [32, Theorem 4.30], we may (for simplicity) assume that this convergence is a.s. The convergence here is in the local topology used in [30], which means [30, Lemma 6.2] that for any fixed $\ell \geqslant 1$, if $T^{[\ell]}$ denotes the tree T truncated at height ℓ, then a.s., for sufficiently large n, $\mathcal{T}_{n,m;\,i}^{\mathbf{p},[\ell]} = \mathcal{T}_i^{[\ell]}$.

The infinite tree $\widehat{\mathcal{T}}$ has exactly one infinite branch; thus, there exists a (random) $j \leqslant m$ such that $\mathcal{T}_j$ is infinite but $\mathcal{T}_i$ is finite for $i \neq j$. Truncating the trees at an ℓ chosen larger than the heights $H(\mathcal{T}_i)$ for all $i \neq j$, we see that for large n, $\mathcal{T}_{n,m;\,i}^{\mathbf{p}} = \mathcal{T}_i$. Thus, $|\mathcal{T}_{n,m;\,i}^{\mathbf{p}}| = O(1)$ for $i \neq j$, and necessarily the remaining branch $\mathcal{T}_{n,m;\,j}^{\mathbf{p}}$ has size $n - O(1)$. Hence, for large enough n, $\mathcal{T}_{n,m;\,(1)}^{\mathbf{p}} = \mathcal{T}_{n,m;\,j}^{\mathbf{p}}$.

Consequently, $\mathcal{T}_{n,m;\,(2)}^{\mathbf{p}}, \ldots, \mathcal{T}_{n,m;\,(m)}^{\mathbf{p}}$ converge a.s., and thus in distribution, to the $m - 1$ finite branches of $\widehat{\mathcal{T}}$, arranged in decreasing order and conditioned on $\mathcal{E}_m$. In particular, their sizes converge in distribution and are thus $O_p(1)$.

We assumed for simplicity $p_m > 0$. In general, we may select a rooted tree T with $\geqslant m$ leaves, such that $p_{d^+(v)} > 0$ for every $v \in T$. Fix m leaves $v_1, \ldots, v_m$ in T, and consider the conditioned Galton–Watson tree $\mathcal{T}_{n+|T|-m}^{\mathbf{p}}$ conditioned on the event $\mathcal{E}_T$ that it consists of T with subtrees added at $v_1, \ldots, v_m$. Then these subtrees form a conditioned Galton–Watson forest $\mathcal{T}_{n,m}^{\mathbf{p}}$, and we can argue as above, conditioning $\widehat{\mathcal{T}}$ on $\mathcal{E}_T$.

This completes the proof when $\mu(\mathbf{p}) = 1$. If $\mu(\mathbf{p}) > 1$, there always exists an equivalent probability weight $\tilde{\mathbf{p}}$ with $\mu(\tilde{\mathbf{p}}) = 1$, and the result follows. If $\mu(\mathbf{p}) < 1$, the same may hold, and if it does not hold, then there is a similar infinite limit tree $\widehat{\mathcal{T}}$ [30, Theorem 7.1]; in this case, $\widehat{\mathcal{T}}$ has one vertex of infinite degree, but the proof above holds with minor modifications. $\square$

Remark 3.2. The proof shows that in the case $\mu(\mathbf{p}) = 1$, the small trees $\mathcal{T}_{n,m;\,(2)}^{\mathbf{p}}, \ldots, \mathcal{T}_{n,m;\,(m)}^{\mathbf{p}}$ in the forest converge in distribution to $m - 1$ independent copies of the unconditioned Galton–Watson tree $\mathcal{T}^{\mathbf{p}}$, arranged in decreasing order. More generally, the small trees converge in distribution to independent Galton–Watson trees for a probability distribution equivalent to $\mathbf{p}$. (This too was shown in [38, Lemma 5.7(iii)] under stronger assumptions.) $\square$

Remark 3.3. In the standard case $\mu(\mathbf{p}) = 1$, $\sigma^2(\mathbf{p}) < \infty$, it is also easy to show Lemma 3.1 using the fact $\mathbb{P}(|\mathcal{T}^{\mathbf{p}}| = n) \sim c n^{-3/2}$, for some $c > 0$, which is a well-known consequence of the local limit theorem, cf. (36)–(37). $\square$

Problem 3.4. A simply generated forest $\mathcal{T}_{n,m}^{\boldsymbol{\phi}}$ is covered by Lemma 3.1 when the generating function (29) has positive radius of convergence, since then it is equivalent to a conditioned Galton–Watson forest. We conjecture that Lemma 3.1 holds for simply generated forests also in the case when the generating function has radius of convergence 0, but we leave this as an open problem.

4. Modified simply generated trees

One frequently meets random trees where the root has a special distribution, see, for example [38, 41]. Thus, let $\boldsymbol{\phi}$ and $\boldsymbol{\phi}^0$ be two weight sequences, where $\boldsymbol{\phi}$ is as above, and $\boldsymbol{\phi}^0 = (\phi_k^0)_0^\infty$ satisfies $\phi_k^0 \geqslant 0$, with strict inequality for at least one k. We modify (26) and now define the weight of a tree $T \in \mathfrak{T}$ as

$$\phi^*(T) := \phi_{d^+(o)}^0 \prod_{v \neq o} \phi_{d^+(v)}. \tag{31}$$

The random *modified simply generated tree* $\mathcal{T}_n^{\phi,\phi^0}$ is defined as in (27), using the modified weight (31).

We say that a pair $(\phi', \phi^{0\prime})$ is equivalent to (ϕ, ϕ^0) if (28) holds and similarly

$$\phi_k^{0\prime} = a_0 b^k \phi_k^0, \qquad k \geqslant 0. \tag{32}$$

It is important that the same b is used in (28) and (32), while a and a_0 may be different. It is easy to see that equivalent pairs of weight sequences define the same modified simply generated tree.

Similarly, given two probability sequences $\mathbf{p} = (p_k)_0^\infty$ and $\mathbf{p}^0 = (p_k^0)_0^\infty$, we define the *modified Galton–Watson tree* $\mathcal{T}^{\mathbf{p},\mathbf{p}^0}$ and *conditioned modified Galton–Watson tree* $\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}$ as in Section 3.2, but now giving children to the root with distribution $\mathbf{p}^0$, and to everyone else with distribution $\mathbf{p}$.

Again, as indicated by our notation, we have an equality: the conditioned modified Galton–Watson tree $\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}$ equals the modified simply generated tree with weight sequences $\mathbf{p}$ and $\mathbf{p}^0$. Conversely, if two weight sequences ϕ and ϕ^0 both have positive radius of convergence, then it is possible (by taking b small enough) to find equivalent weight sequences ϕ' and $\phi^{0\prime}$ that are probability sequences, and thus $\mathcal{T}_n^{\phi,\phi^0} = \mathcal{T}_n^{\phi',\phi^{0\prime}}$ can be interpreted as a conditioned modified Galton–Watson tree.

Lemma 4.1. *Consider a modified simply generated tree $\mathcal{T}_n^{\phi,\phi^0}$ and denote its branches by $T_1, \ldots, T_{d(o)}$.*

  (i) *Conditioned on the root degree $d(o)$, the branches form a simply generated forest $\mathcal{T}_{n-1,d(o)}^{\phi}$.*

  (ii) *Conditioned on the root degree $d(o)$ and the sizes $|T_i|$ of the branches, the branches are independent simply generated trees $\mathcal{T}_{|T_i|}^{\phi}$.*

Proof. Exercise. □

Note that Lemma 4.1 applies also to the simply generated tree $\mathcal{T}_n^{\phi}$ (by taking $\phi^0 = \phi$). Thus, conditioned on the root degree the branches have the same distribution for $\mathcal{T}_n^{\phi}$ and $\mathcal{T}_n^{\phi,\phi^0}$. Hence, the distribution of the root degree is of central importance. The following lemma is partly shown by [38, Proposition 5.6] in greater generality (the stable case), although we add the estimate (35).

Lemma 4.2. *(mainly Kortchemski and Marzouk [38]) Suppose that $\mathbf{p}$ is a probability sequence with mean $\mu(\mathbf{p}) = 1$ and variance $\sigma^2(\mathbf{p}) \in (0, \infty)$ and that $\mathbf{p}^0$ is a probability sequence with finite mean $\mu(\mathbf{p}^0)$. Then the root degree $d(o)$ in the conditioned modified Galton–Watson tree $\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}$ converges in distribution to a random variable $\widetilde{D}$ with distribution*

$$\mathbb{P}(\widetilde{D} = k) = \frac{kp_k^0}{\sum_{j=1}^\infty jp_j^0} = \frac{kp_k^0}{\mu(\mathbf{p}^0)}. \tag{33}$$

In other words, for every fixed $k \geqslant 0$,

$$\mathbb{P}(d(o) = k) \to \mathbb{P}(\widetilde{D} = k), \qquad as\ n \to \infty. \tag{34}$$

Moreover, if n is large enough, we have uniformly

$$\mathbb{P}(d(o) = k) \leqslant 2\mathbb{P}(\widetilde{D} = k), \qquad k \geqslant 1. \tag{35}$$

As a consequence, $\mathbb{E}\widetilde{D} < \infty$ if and only if $\sigma^2(\boldsymbol{p}^0) < \infty$.

Proof. This uses well-known standard arguments, but we give a full proof for completeness; see also [38]. Let D be the root degree in the modified Galton–Watson tree $\mathcal{T}^{\mathbf{p},\mathbf{p}^0}$. If $D = k$, then

the rest of the tree consists of k independent copies of $\mathcal{T}^{\mathbf{p}}$. Thus, the conditional probability $\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n \mid D = k\big)$ equals the probability that a Galton–Watson process started with k individuals has in total $n - 1$ individuals; hence, by a formula by Dwass [24], see e.g. [30, Section 15] and the further references there,

$$\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n \mid D = k\big) = \frac{k}{n-1}\mathbb{P}\big(S_{n-1} = n - k - 1\big), \tag{36}$$

where S_{n-1} denotes the sum of $n - 1$ independent random variables with distribution $\mathbf{p}$.

Suppose for simplicity that the distribution $\mathbf{p}$ is aperiodic, i.e., not supported on any subgroup $d\mathbb{N}$. (The general case follows similarly using standard modifications.) It then follows by the local limit theorem, see e.g. [36, Theorem 1.4.2] or [46, Theorem VII.1], that, as $n \to \infty$,

$$\mathbb{P}\big(S_{n-1} = n - k - 1\big) = \frac{1}{\sqrt{2\pi\sigma^2 n}}\big(e^{-k^2/(2n\sigma^2)} + o(1)\big), \tag{37}$$

uniformly in k. Consequently, combining (36) and (37) with $\mathbb{P}(D = k) = p_k^0$,

$$\begin{aligned}
\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n \text{ and } D = k\big) &= \frac{kp_k^0}{n-1}\mathbb{P}\big(S_{n-1} = n - k - 1\big) \\
&= c\frac{kp_k^0}{n^{3/2}}\big(e^{-k^2/(2n\sigma^2)} + o(1)\big),
\end{aligned} \tag{38}$$

uniformly in k, where $c := (2\pi\sigma^2)^{-1/2}$.

Summing over k we find as $n \to \infty$, using $\sum kp_k^0 < \infty$ and monotone convergence,

$$\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n\big) = \frac{c}{n^{3/2}}\Big(\sum_{k=1}^{\infty}kp_k^0 e^{-k^2/(2n\sigma^2)} + o(1)\Big) \sim \frac{c}{n^{3/2}}\sum_{k=1}^{\infty}kp_k^0. \tag{39}$$

Thus, combining (38) and (39), for any fixed $k \geqslant 1$, as $n \to \infty$,

$$\begin{aligned}
\mathbb{P}\big(D = k \mid |\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n\big) &= \frac{\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n \text{ and } D = k\big)}{\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n\big)} \\
&\to \frac{kp_k^0}{\sum_{j=1}^{\infty}jp_j^0}.
\end{aligned} \tag{40}$$

The limits on the right-hand side sum to 1, and thus the result (33) follows.

Moreover, (38) and (39) also yield

$$\begin{aligned}
\mathbb{P}\big(D = k \mid |\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n\big) &= \frac{\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n \text{ and } D = k\big)}{\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n\big)} \\
&\leqslant \frac{kp_k^0}{\sum_{j=1}^{\infty}jp_j^0}\big(1 + o(1)\big),
\end{aligned} \tag{41}$$

uniformly in k. In particular, (35) holds for all k if n is large enough.

Finally, by (33), $\mathbb{E}\widetilde{D} = \sum k\mathbb{P}(\widetilde{D} = k) < \infty$ if and only if $\sum_k k^2 p_k^0 < \infty$. $\qquad\square$

It follows from Lemma 4.2 that the tree is overwhelmingly dominated by one branch. (Again, this was shown by Kortchemski and Marzouk [38] in greater generality.)

Lemma 4.3. *(essentially Kortchemski and Marzouk [38, Proposition 5.6]) Suppose that $\mathbf{p}$ is a probability sequence with mean $\mu(\mathbf{p}) = 1$ and variance $\sigma^2(\mathbf{p}) \in (0, \infty)$ and that $\mathbf{p}^0$ is a probability sequence with finite mean $\mu(\mathbf{p}^0)$. Let $\mathcal{T}_{n,(1)}, \ldots, \mathcal{T}_{n,(d(o))}$ be the branches of $\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}$ arranged in decreasing order. Then*

$$|\mathcal{T}_{n,(1)}| = n - O_p(1). \tag{42}$$

Proof. Let $D_n = d(o)$ and condition on $D_n = m$, for a fixed m. Then, by Lemmas 4.1 and 3.1, (42) holds. In other words, for every $\varepsilon > 0$, there exists $C_{m,\varepsilon}$ such that

$$\mathbb{P}\big(n - |\mathcal{T}_{n,(1)}| > C_{m,\varepsilon} \mid D_n = m\big) < \varepsilon. \tag{43}$$

By Lemma 4.2, $D_n \xrightarrow{d} \widetilde{D}$, and thus $(D_n)_n$ is tight, i.e., $O_p(1)$, so there exists M such that $\mathbb{P}(D_n > M) < \varepsilon$ for all n. Consequently, if $C_\varepsilon := \max_{m \leqslant M} C_{m,\varepsilon}$,

$$\mathbb{P}\big(n - |\mathcal{T}_{n,(1)}| > C_\varepsilon\big) = \mathbb{E}\mathbb{P}\big(n - |\mathcal{T}_{n,(1)}| > C_\varepsilon \mid D_n\big) \leqslant \varepsilon + \mathbb{P}(D_n > M)$$
$$\leqslant 2\varepsilon, \tag{44}$$

which completes the proof. □

Lemmas 4.1–4.3 make it possible to transfer many results that are known for simply generated trees (conditioned Galton–Watson trees) to the modified version. See Section 9 for a few examples.

Problem 4.4. Does Lemma 4.2 (and thus Lemma 4.3) hold without assuming finite variance $\sigma^2(\mathbf{p}) < \infty$. i.e., assuming only $\mu(\mathbf{p}) = 1$ and $\mu(\mathbf{p}^0) < \infty$? As said above, (33) was shown also when the variance is infinite by Kortchemski and Marzouk [38], but they then assume that $\mathbf{p}$ is in the domain of attraction of a stable distribution. What happens without this regularity assumption?

Remark 4.5. We assume in Lemma 4.2 that $\mu(\mathbf{p}^0) < \infty$. We claim that if $\mu(\mathbf{p}^0) = \infty$, then $d(o) \xrightarrow{P} \infty$; in other words, $\mathbb{P}(d(o) = k) \to 0$ for every fixed k, which can be seen as the natural interpretation of (33)–(34) in this case. □

We sketch a proof. First, from (38) and Fatou's lemma (for sums),

$$\liminf_{n \to \infty} n^{3/2}\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n\big) \geqslant \sum_{k=0}^{\infty} \liminf_{n \to \infty} n^{3/2}\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n \text{ and } D = k\big)$$
$$= \sum_{k=0}^{\infty} ckp_k^0 = \infty. \tag{45}$$

In other words, $n^{3/2}\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n\big) \to \infty$. Then, (38) and (45) yield, for any fixed $k \geqslant 0$,

$$\mathbb{P}\big(D = k \mid |\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n\big) = \frac{\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n \text{ and } D = k\big)}{\mathbb{P}\big(|\mathcal{T}^{\mathbf{p},\mathbf{p}^0}| = n\big)} \to 0. \tag{46}$$

This proves our claim. □

5. Unrooted simply generated trees

We make definitions corresponding to Section 3 for unrooted trees. In this case, we consider labelled trees, so that we can distinguish the vertices. (This is not needed for ordered trees, since their vertices can be labelled canonically as described in Section 2.) Of course, we may then ignore the labelling when we want.

Let $(w_k)_0^\infty$ be a given sequence of non-negative weights, with $w_1 > 0$ and $w_k > 0$ for some $k \geqslant 3$. (The weight w_0 is needed only for the trivial case $n = 1$ and might be ignored. We may take $w_0 = 0$ without essential loss of generality.)

For any labelled tree $T \in \mathfrak{L}_n$, now define the weight of T as

$$w(T) := \prod_{v \in T} w_{d(v)}. \tag{47}$$

Given $n \geqslant 1$, we define the random *unrooted simply generated tree* $\mathcal{T}_n^\circ = \mathcal{T}_n^{\mathbf{w},\circ}$ as a labelled tree in $\mathcal{L}_n$, chosen randomly with probability proportional to the weight (47). (We consider only n such that at least one tree of positive weight exists.)

Remark 5.1. Just as in the rooted case, replacing the weight sequence by an equivalent one (still defined as in (28)) gives the same random tree $\mathcal{T}_n^\circ$. $\qquad\square$

In the following sections, we give three (related but different) relations with the more well-known rooted simply generated trees.

6. Mark a vertex

Let $\mathcal{T}_n^{\mathbf{w},\circ}$ be a random unrooted simply generated tree as in Section 5 and mark one of its n vertices, chosen uniformly at random. Regard the marked vertex as a root and denote the resulting rooted tree by $\mathcal{T}_n^{\mathbf{w},\bullet}$.

Thus, $\mathcal{T}_n^{\mathbf{w},\bullet}$ is a random unordered rooted tree, where an unordered rooted tree T has probability proportional to its weight given by (47).

We make $\mathcal{T}_n^{\mathbf{w},\bullet}$ ordered by ordering the children of each vertex uniformly at random; denote the resulting random labelled ordered rooted tree by $\mathcal{T}_n^{\mathbf{w},*}$. Since each vertex v has $d^+(v)!$ possible orders, the probability that $\mathcal{T}_n^{\mathbf{w},*}$ equals a given ordered tree T is proportional to the weight

$$w^*(T) := \frac{w(T)}{\prod_{v \in T} d^+(v)!} = \frac{w_{d(o)}}{d(o)!} \prod_{v \neq o} \frac{w_{d(v)}}{d^+(v)!} = \frac{w_{d(o)}}{d(o)!} \prod_{v \neq o} \frac{w_{d^+(v)+1}}{d^+(v)!}. \tag{48}$$

The tree $\mathcal{T}_n^{\mathbf{w},*}$ is constructed as a labelled tree, but each ordered rooted tree $T \in \mathfrak{T}_n$ has the same number $n!$ of labellings, and they have the same weight (48) and thus appear with the same probability. Hence, we may forget the labelling and regard $\mathcal{T}_n^{\mathbf{w},*}$ as a random ordered tree in $\mathfrak{T}_n$, with probabilities proportional to the weight (48). This is the same as the weight (31) with

$$\phi_k := \frac{w_{k+1}}{k!}, \qquad k \geqslant 0, \tag{49}$$

$$\phi_k^0 := \frac{w_k}{k!}, \qquad k \geqslant 0. \tag{50}$$

Thus, $\mathcal{T}_n^{\mathbf{w},*} = \mathcal{T}_n^{\boldsymbol{\phi},\boldsymbol{\phi}^0}$, the modified simply generated tree defined in Section 4.

We recover $\mathcal{T}_n^{\mathbf{w},\circ}$ from $\mathcal{T}_n^* = \mathcal{T}_n^{\boldsymbol{\phi},\boldsymbol{\phi}^0}$ by ignoring the root (and adding a uniformly random labelling). This yields thus a method to construct $\mathcal{T}_n^{\mathbf{w},\circ}$.

Example 6.1. Marckert and Panholzer [41] studied uniformly random non-crossing trees of a given size n and found that if they are regarded as ordered rooted trees, then they have the same distribution as the conditioned modified Galton–Watson tree $\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}$, where

$$p_k = 4(k+1)3^{-k-2}, k \geqslant 0, \tag{51}$$

$$p_k^0 = 2 \cdot 3^{-k}, k \geqslant 1. \tag{52}$$

These weights are equivalent to $\phi_k = k + 1$ and $\phi_k^0 = 1$, which are given by (49)–(50) with $w_k = k!$. We may thus reformulate the result by Marckert and Panholzer [41] as: A uniformly random non-crossing tree is the same as a random unrooted simply generated tree with weights $w_k = k!$.

More generally, Kortchemski and Marzouk [38] studied *simply generated non-crossing trees*, which are random non-crossing trees with probability proportional to the weight (47) for some weight sequence $\mathbf{w} = (w_k)_k$, and showed that they (under a condition) are equivalent to conditioned modified Galton–Watson trees. In fact, for any weight sequence $\mathbf{w}$, the proofs in [41], in particular Lemma 2] and [38, in particular Proposition 2.1] show that the simply generated non-crossing tree, regarded as an ordered rooted tree, is the same as $\mathcal{T}_n^{\boldsymbol{\phi},\boldsymbol{\phi}^0}$ with

$$\phi_k := (k+1)w_{k+1}, \quad k \geqslant 0, \tag{53}$$

$$\phi_k^0 := w_k, \quad k \geqslant 0. \tag{54}$$

Thus, comparing with (49)–(50), it follows that the simply generated non-crossing tree is an unordered simply generated tree, with weight sequence $\overline{w}_k := w_k k!$.

Note that non-crossing trees are naturally defined as unrooted trees. A root is introduced in [38, 41] for the analysis, which as said above makes the trees conditioned modified Galton–Watson trees (or, more generally, modified simply generated trees). This is precisely the marking of an unrooted simply generated tree discussed in the present section.   □

Remark 6.2. The constructions in this and the next section lead to simple relations of generating functions (not used here); see [10, Appendix B].   □

7. Mark an edge

In the random unrooted tree $\mathcal{T}_n^{\mathbf{w},\circ}$, mark a (uniformly) random edge, and give it a direction; i.e., mark two adjacent vertices, say o_+ and o_-. Since each tree $\mathcal{T}_n^{\mathbf{w},\circ}$ has the same number $n - 1$ of edges, the resulting marked tree $\mathcal{T}_n^{\mathbf{w},\bullet\bullet}$ is distributed over all labelled trees on $[n]$ with a marked and directed edge with probabilities proportional to the weight (47).

Now ignore the marked edge, and regard the tree $\mathcal{T}_n^{\mathbf{w},\bullet\bullet}$ as two rooted trees $\mathcal{T}_{n,1}$ and $\mathcal{T}_{n,2}$ with roots o_+ and o_-, respectively. Furthermore, order randomly the children of each vertex in each of these rooted trees; this makes $\mathcal{T}_{n,1}$ and $\mathcal{T}_{n,2}$ a pair of ordered trees, and each pair (T_+, T_-) of labelled ordered rooted trees with $|T_+| + |T_-| = n$ and the labels $1, \ldots, n$ appears with probability proportional to

$$\widehat{w}(T_+, T_-) := \widehat{w}(T_+)\widehat{w}(T_-). \tag{55}$$

where, for a rooted tree T,

$$\widehat{w}(T) := \prod_{v \in T} \frac{w_{d^+(v)+1}}{d^+(v)!}. \tag{56}$$

Using again the definition (49), we have by (26),

$$\widehat{w}(T) = \prod_{v \in T} \phi_{d^+(v)} = \phi(T). \tag{57}$$

Moreover, since we now have ordered rooted trees, the vertices are distinguishable, and each pair (T_+, T_-) of ordered trees with $|T_+| + |T_-| = n$ has the same number $n!$ of labellings. Hence, we may ignore the labelling and regard the marked tree $\mathcal{T}_n^{\mathbf{w},\bullet\bullet}$ as a pair of ordered trees $(\mathcal{T}_{n,1}, \mathcal{T}_{n,2})$ with $|\mathcal{T}_{n,1}| + |\mathcal{T}_{n,2}| = n$ and probabilities proportional to the weight given by (55) and (57). This means that $\mathcal{T}_{n,1}$ and $\mathcal{T}_{n,2}$, conditioned on their sizes, are two independent random rooted simply

generated trees, with the weight sequence $\boldsymbol{\phi}$ given by (49); in other words, $(\mathcal{T}_{n,1}, \mathcal{T}_{n,2})$ is a simply generated forest $\mathcal{T}_{n,2}^{\boldsymbol{\phi}}$.

Consequently, we can construct the random unrooted simply generated tree $\mathcal{T}_n^{\mathbf{w},\circ}$ by taking two random rooted simply generated trees $\mathcal{T}_{n,1}$ and $\mathcal{T}_{n,2}$ constructed in this way, with the right distribution of their sizes, and joining their roots.

Note that $|\mathcal{T}_{n,1}| = n - |\mathcal{T}_{n,2}|$ is random, with a distribution given by the construction above. More precisely, if a_n is the total weight (57) summed over all ordered trees of order n, then

$$\mathbb{P}\bigl(|\mathcal{T}_{n,1}| = m\bigr) = \frac{a_m a_{n-m}}{\sum_{k=1}^{n-1} a_k a_{n-k}}. \tag{58}$$

Remark 7.1. If $\phi_2 > 0$, we can alternatively describe the result as follows: Use the weight sequence $(\phi_k)_0^\infty$ given by (49) and take a random rooted simply generated tree $\mathcal{T}_{n+1}^{\boldsymbol{\phi}}$ of order $n + 1$, conditioned on the root degree $= 2$; remove the root and join its two neighbours to each other (this is the marked edge). □

If $\phi_2 = 0$, we can instead take any $k > 2$ with $\phi_k > 0$, and take a random rooted simply generated tree $\mathcal{T}_{n+k-1}^{\boldsymbol{\phi}}$ of order $n + k - 1$, conditioned on the event that the root degree is k, and the $k - 2$ last children of the root are leaves; we remove the root and these children, and join the first two children.

Remark 7.2. Suppose that the weight sequence $(\phi_j)_0^\infty$ given by (49) satisfies $\sum_{j=1}^\infty \phi_j = 1$, so $(\phi_j)_0^\infty$ is a probability distribution. (Note that a large class of examples can be expressed with such weights, see Remark 5.1.) Then the construction above can be stated as follows:

Consider a Galton–Watson process with offspring distribution $(\phi_k)_0^\infty$, starting with two individuals, and conditioned on the total progeny being n. This creates a forest with two trees; join their roots to obtain $\mathcal{T}_n^{\mathbf{w},\circ}$.

Note that it follows from the arguments above that if we mark the edge joining the two roots, then the marked edge will be distributed uniformly over all edges in the tree $\mathcal{T}_n^{\mathbf{w},\circ}$. □

In the construction above, $\mathcal{T}_{n,1}$ and $\mathcal{T}_{n,2}$ have the same distribution by symmetry. Now define $\mathcal{T}_{n,+}$ as the largest and $\mathcal{T}_{n,-}$ as the smallest of $\mathcal{T}_{n,1}$ and $\mathcal{T}_{n,2}$. The next lemma shows that (at least under a weak condition), $\mathcal{T}_{n,-}$ is stochastically bounded, so $\mathcal{T}_n^{\mathbf{w},\circ}$ is dominated by the subtree $\mathcal{T}_{n,+}$.

Lemma 7.3. *Suppose that the generating function $\Phi(z)$ in (29) has a positive radius of convergence. Then, as $n \to \infty$, $\mathcal{T}_{n,-} \xrightarrow{\mathrm{d}} \mathcal{T}^{\mathbf{p}}$, an unconditioned Galton–Watson tree with offspring distribution $\mathbf{p}$ equivalent to $(\phi_k)_0^\infty$. In particular, $|\mathcal{T}_{n,-}| = O_{\mathrm{p}}(1)$, and thus $|\mathcal{T}_{n,+}| = n - O_{\mathrm{p}}(1)$.*

Proof. This is a special case of Lemma 3.1, see also Remark 3.2. □

As remarked in Problem 3.4, we conjecture that $|\mathcal{T}_{n,-}| = O_{\mathrm{p}}(1)$ also when the generating function has radius of convergence 0, but we leave this as an open problem.

8. Mark a leaf

This differs from the preceding two sections in that we do not recover the distribution of $\mathcal{T}_n^{\mathbf{w},\circ}$ exactly, but only asymptotically.

Let $N_0(T)$ be the number of leaves in an unrooted tree T. Let $\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}$ be a random unrooted labelled tree with probability proportional to $N_0(T)w(T)$; in other words, we bias the distribution of $\mathcal{T}_n^{\mathbf{w},\circ}$ by the factor $N_0(T)$.

Let $\widehat{\mathcal{T}}_n^{\mathbf{w},\bullet}$ be the random rooted tree obtained by marking a uniformly random leaf in $\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}$, regarding the marked leaf as the root. Then, any pair (T, o) with $T \in \mathfrak{L}_n$ and $o \in T$ with $d(o) = 1$

will be chosen as $\widehat{\mathcal{T}}_n^{\mathbf{w},\bullet}$ and its root with probability proportional to the weight (47). We order the children of each vertex at random as in Sections 6 and 7, and obtain an ordered rooted tree $\widehat{\mathcal{T}}_n^{\mathbf{w},*}$. Then each tree with root degree 1 appears with probability proportional to (48).

Consequently, if we ignore the labelling, $\widehat{\mathcal{T}}_n^{\mathbf{w},*} = \mathcal{T}_n^{\phi,\phi^0}$, where ϕ is given by (49), and $\phi_k^0 := \delta_{k1}$ (with a Kronecker delta). Equivalently, $\widehat{\mathcal{T}}_n^{\mathbf{w},*}$ has a root of degree 1, and its single branch is a $\mathcal{T}_{n-1}^{\phi}$.

Conversely, we may obtain $\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}$ from $\mathcal{T}_{n-1}^{\phi}$ by adding a new root under the old one and then adding a random labelling.

Remark 8.1. The construction above can also be regarded as a variant of the one in Section 7, where we mark an edge such that one endpoint is a leaf. Then, in the notation there, $|\mathcal{T}_{n,-}| = 1$ and $\mathcal{T}_{n,+} = \mathcal{T}_{n-1}^{\phi}$. □

As said above, $\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}$ does not have the distribution of $\mathcal{T}_n^{\mathbf{w},\circ}$, but it is not far from it.

Lemma 8.2. *Let $\mathbf{w}$ be any weight sequence. As $n \to \infty$, the total variation distance $d_{\mathrm{TV}}(\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}, \mathcal{T}_n^{\mathbf{w},\circ}) \to 0$. In other words, there exists a coupling such that $\mathbb{P}(\widehat{\mathcal{T}}_n^{\mathbf{w},\circ} \neq \mathcal{T}_n^{\mathbf{w},\circ}) \to 0$.*

Proof. We may construct $\mathcal{T}_n^{\mathbf{w},\circ}$ as in Section 7 from two random ordered trees $\mathcal{T}_{n,+}$ and $\mathcal{T}_{n,-}$, where $|\mathcal{T}_{n,-}| = O_{\mathrm{p}}(1)$. Conditioned on $|\mathcal{T}_{n,-}| = \ell$, for any fixed ℓ, we have $\mathcal{T}_{n,+} \overset{\mathrm{d}}{=} \mathcal{T}_{n-\ell}^{\phi}$, where ϕ is given by (49). Thus, by [30, Theorem 7.11] (see comments there for earlier references to special cases, and to further results), as $n \to \infty$, conditioned on $|\mathcal{T}_{n,-}| = \ell$ for any fixed ℓ,

$$\frac{N_0(\mathcal{T}_{n,+})}{n} \overset{\mathrm{d}}{=} \frac{N_0(\mathcal{T}_{n-\ell}^{\phi})}{n} \overset{\mathrm{P}}{\longrightarrow} \pi_0, \tag{59}$$

for some constant $\pi_0 > 0$. (If ϕ is a probability sequence, then $\pi_0 = \phi_0$.) Furthermore, $N_0(\mathcal{T}_n^{\mathbf{w},\circ}) = N_0(\mathcal{T}_{n,+}) + N_0(\mathcal{T}_{n,-}) = N_0(\mathcal{T}_{n,+}) + O(1)$, since $N_0(\mathcal{T}_{n,-}) \leqslant |\mathcal{T}_{n,-}| = \ell$. Consequently, still conditioned on $|\mathcal{T}_{n,-}| = \ell$ for any fixed ℓ,

$$\frac{N_0(\mathcal{T}_n^{\mathbf{w},\circ})}{n} \overset{\mathrm{P}}{\longrightarrow} \pi_0 > 0. \tag{60}$$

Since $|\mathcal{T}_{n,-}| = O_{\mathrm{p}}(1)$, it follows that (60) holds also unconditionally.

Since $N_0(\mathcal{T}_n^{\mathbf{w},\circ})/n \leqslant 1$, dominated convergence yields

$$\frac{\mathbb{E}N_0(\mathcal{T}_n^{\mathbf{w},\circ})}{n} = \mathbb{E}\frac{N_0(\mathcal{T}_n^{\mathbf{w},\circ})}{n} \to \pi_0. \tag{61}$$

By (60) and (61),

$$\frac{N_0(\mathcal{T}_n^{\mathbf{w},\circ})}{\mathbb{E}N_0(\mathcal{T}_n^{\mathbf{w},\circ})} \overset{\mathrm{P}}{\longrightarrow} 1, \tag{62}$$

and thus, by dominated convergence again,

$$\mathbb{E}\left| \frac{N_0(\mathcal{T}_n^{\mathbf{w},\circ})}{\mathbb{E}N_0(\mathcal{T}_n^{\mathbf{w},\circ})} - 1 \right| \to 0. \tag{63}$$

The definition of $\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}$ by biasing means that for any bounded (or non-negative) function $f : \mathfrak{L}_n \to \mathbb{R}$,

$$\mathbb{E}f(\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}) = \frac{\mathbb{E}[f(\mathcal{T}_n^{\mathbf{w},\circ})N_0(\mathcal{T}_n^{\mathbf{w},\circ})]}{\mathbb{E}N_0(\mathcal{T}_n^{\mathbf{w},\circ})}. \tag{64}$$

and thus, for any indicator function f,

$$\big|\mathbb{E}f(\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}) - \mathbb{E}f(\mathcal{T}_n^{\mathbf{w},\circ})\big| = \bigg|\mathbb{E}\Big[f(\mathcal{T}_n^{\mathbf{w},\circ})\Big(\frac{N_0(\mathcal{T}_n^{\mathbf{w},\circ})}{\mathbb{E}N_0(\mathcal{T}_n^{\mathbf{w},\circ})} - 1\Big)\Big]\bigg|$$
$$\leqslant \mathbb{E}\bigg|\frac{N_0(\mathcal{T}_n^{\mathbf{w},\circ})}{\mathbb{E}N_0(\mathcal{T}_n^{\mathbf{w},\circ})} - 1\bigg|. \tag{65}$$

Hence, taking the supremum over all $f = \mathbf{1}_A$,

$$d_{\mathrm{TV}}(\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}, \mathcal{T}_n^{\mathbf{w},\circ}) \leqslant \mathbb{E}\bigg|\frac{N_0(\mathcal{T}_n^{\mathbf{w},\circ})}{\mathbb{E}N_0(\mathcal{T}_n^{\mathbf{w},\circ})} - 1\bigg|, \tag{66}$$

and the result follows by (63). □

Lemma 8.2 implies that any result on convergence in probability or distribution for one of $\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}$ and $\mathcal{T}_n^{\mathbf{w},\circ}$ also hold for the other.

9. Profile of conditioned modified Galton–Watson trees

We will use the following extension of Theorem 1.2 to conditioned modified Galton–Watson trees.

Theorem 9.1. *Let L_n be the profile of a conditioned modified Galton–Watson tree $\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}$ of order n and assume that $\mu(\mathbf{p}) = 1$, $\sigma^2(\mathbf{p}) < \infty$ and $\mu(\mathbf{p}^0) < \infty$. Then, as $n \to \infty$,*

$$n^{-1/2}L_n(xn^{1/2}) \xrightarrow{\mathrm{d}} \frac{\sigma}{2}L_{\mathbf{e}}\Big(\frac{\sigma}{2}x\Big), \tag{67}$$

in the space $C[0,\infty]$, where $L_{\mathbf{e}}$ is, as in Theorem 1.2, the local time of a standard Brownian excursion $\mathbf{e}$.

Proof. Denote the branches of $\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}$ by $\mathcal{T}_1, \ldots, \mathcal{T}_{d(o)}$ and let $\mathcal{T}_0$ be a single root. Then, regarding the branches as rooted trees, which means that their vertices have their depths shifted by 1 from the original tree,

$$L_n(x) = \sum_{i=1}^{d(o)} L_{\mathcal{T}_i}(x-1) + L_{\mathcal{T}_0}(x). \tag{68}$$

Let $\mathcal{T}_{(1)}, \ldots, \mathcal{T}_{(d(o))}$ be the branches arranged in decreasing order. Lemma 4.3 shows that $|\mathcal{T}_{(1)}| = n - O_{\mathrm{p}}(1)$. Hence, (68) and the trivial estimate $0 \leqslant L_T(x) \leqslant |T|$ for any T and x yield

$$\big|L_n(x) - L_{\mathcal{T}_{(1)}}(x-1)\big| \leqslant \sum_{i=2}^{d(o)} |\mathcal{T}_{(i)}| + 1 = n - |\mathcal{T}_{(1)}| = O_{\mathrm{p}}(1). \tag{69}$$

Furthermore, conditioned on $|\mathcal{T}_{(1)}| = n - \ell$, for any fixed ℓ, $\mathcal{T}_{(1)}$ has the same distribution as $\mathcal{T}_{n-\ell}^{\mathbf{p}}$, and thus Theorem 1.2 shows that

$$(n-\ell)^{-1/2}L_{\mathcal{T}_{(1)}}(x(n-\ell)^{1/2}) \xrightarrow{\mathrm{d}} \frac{\sigma}{2}L_{\mathbf{e}}\Big(\frac{\sigma}{2}x\Big), \qquad \text{in } C[0,\infty], \tag{70}$$

and it follows easily that, still conditioned,

$$n^{-1/2}L_{\mathcal{T}_{(1)}}(xn^{1/2} - 1) \xrightarrow{\mathrm{d}} \frac{\sigma}{2}L_{\mathbf{e}}\Big(\frac{\sigma}{2}x\Big), \qquad \text{in } C[0,\infty]. \tag{71}$$

Together with (69), this shows that for every fixed ℓ,

$$\big(L_n(x) \mid |\mathcal{T}_{(1)}| = n - \ell\big) \xrightarrow{\mathrm{d}} \frac{\sigma}{2}L_{\mathbf{e}}\Big(\frac{\sigma}{2}x\Big), \qquad \text{in } C[0,\infty]. \tag{72}$$

It follows that (72) holds also if we condition on $n - |\mathcal{T}_{(1)}| \leqslant K$, for any fixed K, and then (67) follows easily from $n - |\mathcal{T}_{(1)}| = O_p(1)$.   □

Recall that for conditioned Galton–Watson trees $\mathcal{T}_n^{\mathbf{p}}$ with $\mu(\mathbf{p}) = 1$ and $\sigma^2(\mathbf{p}) < \infty$, the width divided by $\sqrt{n}$ converges in distribution: we have

$$n^{-1/2} W(\mathcal{T}_n^{\mathbf{p}}) \xrightarrow{\mathrm{d}} \sigma W, \tag{73}$$

for some random variable W (not depending on $\mathbf{p}$). In fact, as noted by Drmota and Gittenberger [18], this is an immediate consequence of (13) and (1), with

$$W := \tfrac{1}{2} \max_{x \geqslant 0} L_{\mathbf{e}}(x). \tag{74}$$

It is also known that all moments converge, see [19] (assuming an exponential moment) and [1] (in general).

The next theorem records that (73) extends to conditioned modified Galton–Watson trees, together with two partial results on moments.

Theorem 9.2. *Consider a conditioned modified Galton–Watson tree* $\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0}$ *where* $\mu(\mathbf{p}) = 1$, $\sigma^2(\mathbf{p}) < \infty$ *and* $\sigma^2(\mathbf{p}^0) < \infty$. *Then, as* $n \to \infty$,

$$n^{-1/2} W(\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0}) \xrightarrow{d} \sigma W, \tag{75}$$

$$n^{-1/2} \mathbb{E} W(\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0}) \to \sigma \mathbb{E} W = \sigma \sqrt{\pi/2}, \tag{76}$$

$$\mathbb{E}\big[W(\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0})^2 \big] = O(n). \tag{77}$$

Proof. First, (75) follows as in [18]: $f \to \sup f$ is a continuous functional on $C[0, \infty]$, and thus (75) follows from (67), (13) and (74).

We next prove (77). Denote the branches of $\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0}$ by $\mathcal{T}_1, \ldots, \mathcal{T}_{d(o)}$. Assume $n > 1$, then the width is attained above the root, and we have, for every $i \leqslant d(o)$,

$$W(\mathcal{T}_i) \leqslant W(\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0}) \leqslant \sum_{i=1}^{d(o)} W(\mathcal{T}_i). \tag{78}$$

Condition on $d(o)$ and $|\mathcal{T}_1|, \ldots, |\mathcal{T}_{d(o)}|$ as in Lemma 4.1. For a random variable X, denote its conditioned L^2 norm by

$$\|X\|_2' := \big(\mathbb{E}\big[X^2 \mid d(o), |\mathcal{T}_1|, \ldots, |\mathcal{T}_{d(o)}| \big] \big)^{1/2}. \tag{79}$$

By (78) and Minkowski's inequality, we have

$$\| W(\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0}) \|_2' \leqslant \sum_{i=1}^{d(o)} \| W(\mathcal{T}_i) \|_2'. \tag{80}$$

Furthermore, by Lemma 4.1 and [1, Corollary 1.3], if $|\mathcal{T}_i| = n_i$,

$$\mathbb{E}\big(W(\mathcal{T}_i)^2 \mid d(o), |\mathcal{T}_1|, \ldots, |\mathcal{T}_{d(o)}| \big) = \mathbb{E}\big[W(\mathcal{T}_{n_i}^{\mathbf{p}})^2 \big] \leqslant C n_i, \tag{81}$$

and thus $\| W(\mathcal{T}_i) \|_2' \leqslant C n_i^{1/2} = C |\mathcal{T}_i|^{1/2}$. Hence, by (80),

$$\| W(\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0}) \|_2' \leqslant \sum_{i=1}^{d(o)} C |\mathcal{T}_i|^{1/2} \tag{82}$$

and thus, by the Cauchy–Schwarz inequality,

$$\mathbb{E}\big[W(\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0})^2 \mid d(o), |\mathcal{T}_1|, \ldots, |\mathcal{T}_{d(o)}|\big] = \big(\|W(\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0})\|_2'\big)^2$$
$$\leqslant C\Big(\sum_{i=1}^{d(o)} |\mathcal{T}_i|^{1/2}\Big)^2 \leqslant Cd(o)\sum_{i=1}^{d(o)} |\mathcal{T}_i| \leqslant Cd(o)n. \tag{83}$$

Taking the expectation yields

$$\mathbb{E}\big[W(\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0})^2\big] \leqslant Cn\mathbb{E}\big[d(o)\big]. \tag{84}$$

Furthermore, (35) implies that for large n, $\mathbb{E}[d(o)] \leqslant 2\mathbb{E}\widetilde{D}$, where $\mathbb{E}\widetilde{D} < \infty$ by Lemma 4.2. Thus, $\mathbb{E}[d(o)] \leqslant C$, and (84) yields

$$\mathbb{E}\big[W(\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0})^2\big] \leqslant Cn, \tag{85}$$

showing (77).

Finally, (77) implies that the variables on the left-hand side of (75) are uniformly integrable [26, Theorem 5.4.2], and thus (76) follows from (75). $\mathbb{E}W = \sqrt{\pi/2}$ is well known, see e.g. [12]. $\square$

Problem 9.3. We conjecture that under the assumptions of Theorem 9.2, $\mathbb{E}\big[W(\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0})^r\big] = O(n^{r/2})$ for any $r > 0$, which implies convergence of all moments in (75), as shown for the case $\mathbf{p}^0 = \mathbf{p}$ in [1]. The proof above is easily generalized if $\mathbb{E}\widetilde{D}^{r/2} = O(1)$, which is equivalent to $\sum_k k^{1+r/2}p_k^0 < \infty$, but we leave the general case as an open problem.

Problem 9.4. Is $\sigma^2(\mathbf{p}^0) < \infty$ really needed in Theorem 9.2?

10. Distance profile, first step

We now turn to distance profiles. We begin with a weak version of Theorem 1.4; recall the pseudometric d defined in (20), and (21).

Lemma 10.1. *Consider a conditioned Galton–Watson tree $\mathcal{T}_n^{\mathbf{p}}$ where $\mu(\mathbf{p}) = 1$ and $\sigma^2 = \sigma^2(\mathbf{p}) < \infty$. Then, as $n \to \infty$, for any continuous function with compact support $f : [0, \infty) \to \mathbb{R}$,*

$$\int_0^\infty n^{-3/2}\Lambda_{\mathcal{T}_n^{\mathbf{p}}}(xn^{1/2})f(x)\,dx \overset{d}{\longrightarrow} \int_0^1\int_0^1 f\Big(\frac{2}{\sigma}\mathsf{d}(s,t;\mathbf{e})\Big)\,ds\,dt. \tag{86}$$

Proof. The function f is bounded, and also uniformly continuous, i.e., its modulus of continuity $\omega(\delta; f)$, defined in (9), satisfies $\omega(\delta; f) \to 0$ as $\delta \to 0$. Thus, for any rooted tree $T \in \mathfrak{T}_n$, noting that $\Lambda_T(x) \leqslant n$ on $[-1, 0]$ and using the analogue of (11) for Λ,

$$\int_0^\infty n^{-3/2}\Lambda_T(xn^{1/2})f(x)\,dx = n^{-2}\int_0^\infty \Lambda_T(x)f(n^{-1/2}x)\,dx$$
$$= n^{-2}\int_{-1}^\infty f(n^{-1/2}x)\Lambda_T(x)\,dx + O(n^{-1})$$
$$= n^{-2}\sum_{i=0}^\infty \int_{i-1}^{i+1} f(n^{-1/2}x)\Lambda_T(i)\tau(x-i)\,dx + O(n^{-1})$$
$$= n^{-2}\sum_{i=0}^\infty f(n^{-1/2}i)\Lambda_T(i) + O(\omega(n^{-1/2};f)) + O(n^{-1})$$
$$= n^{-2}\sum_{v,w\in T} f(n^{-1/2}\mathsf{d}(v,w)) + o(1), \tag{87}$$

where (as throughout the proof) $o(1)$ tends to 0 as $n \to \infty$, uniformly in $T \in \mathfrak{T}_n$. Recall that the *contour process* $C_T(x)$ of T is a continuous function $C_T : [0, 2n-2] \to [0, \infty)$ that describes the distance from the root to a particle that travels with speed 1 on the 'outside' of the tree. (Equivalently, it performs a depth first walk at integer times $0, 1, \ldots, 2n-2$.) For each vertex $v \neq o$, the particle travels through the edge leading from v towards the root during two time intervals of unit length (once in each direction). Thus, as is well known,

$$\int_0^{2n-2} f\big(n^{-1/2}C_T(x)\big)\, dx = 2\sum_{v \neq o} f\big(n^{-1/2}d(v, o)\big) + O\big(n\omega(n^{-1/2}; f)\big). \tag{88}$$

We will use a bivariate version of this. It is also well known that if $v(i)$ is the vertex visited by the particle at time i, then, for any integers $i, j \in [0, 2n-2]$,

$$d\big(v(i), v(j)\big) = d(i, j; C_T), \tag{89}$$

where the first d is the graph distance in T, and the second is the pseudometric defined by (20) (now on the interval $[0, 2n-2]$). Hence, the argument yielding (88) also yields

$$\int_0^{2n-2}\int_0^{2n-2} f\big(n^{-1/2}d(x, y; C_T)\big)\, dx\, dy = 4\sum_{v, w \neq o} f\big(n^{-1/2}d(v, w)\big) + O\big(n^2\omega(n^{-1/2}; f)\big). \tag{90}$$

We use the standard rescaling of the contour process

$$\widetilde{C}_T(t) := n^{-1/2}C_T\big((2n-2)t\big), \qquad t \in [0, 1], \tag{91}$$

and note that for any $g : [0, 1] \to [0, \infty)$ with $g(0) = g(1) = 0$ and $c > 0$,

$$d(s, t; cg) = cd(s, t; g), \qquad s, t \in [0, 1]. \tag{92}$$

Thus, by (90) and a change of variables,

$$\begin{aligned}
\int_0^1 \int_0^1 & f\big(d(s, t; \widetilde{C}_T)\big)\, ds\, dt \\
&= \frac{1}{(2n-2)^2}\int_0^{2n-2}\int_0^{2n-2} f\big(n^{-1/2}d(x, y; C_T)\big)\, dx\, dy \\
&= \frac{1}{(n-1)^2}\sum_{v, w \neq o} f\big(n^{-1/2}d(v, w)\big) + O\big(\omega(n^{-1/2}; f)\big) \\
&= \frac{1}{n^2}\sum_{v, w \neq o} f\big(n^{-1/2}d(v, w)\big) + o(1).
\end{aligned} \tag{93}$$

Combining (87) and (93), we find

$$\int_0^\infty n^{-3/2}\Lambda_T\big(xn^{1/2}\big)f(x)\, dx = \int_0^1 \int_0^1 f\big(d(s, t; \widetilde{C}_T)\big)\, ds\, dt + o(1). \tag{94}$$

We apply this to $T = \mathcal{T}_n^{\mathbf{p}}$ and use the result by Aldous [3, 4],

$$\widetilde{C}_{\mathcal{T}_n^{\mathbf{p}}}(t) \xrightarrow{d} \frac{2}{\sigma}\mathbf{e}(t), \qquad \text{in } C[0, 1]. \tag{95}$$

The functional $g \to \iint f\big(d(s, t; g)\big)\, ds\, dt$ is continuous on $C[0, 1]$, and the result (86) follows from (94) and (95) by the continuous mapping theorem, using also (92). $\qquad\square$

11. Distance profile of unrooted trees

We continue with the distance profile, now turning to unrooted simply generated trees for a while. Throughout this section, we assume that $\mathbf{w}$ is a weight sequence and that $\boldsymbol{\phi}$ and $\boldsymbol{\phi}^0$ are the weight sequences given by (49) and (50). We assume that the exponential generating function of $\mathbf{w}$ has positive radius of convergence; this means that the generating function $\Phi(z)$ in (29) has positive radius of convergence, which in turn implies that there exists a probability weight sequence $\mathbf{p}$ equivalent to $\boldsymbol{\phi}$. We assume furthermore that it is possible to choose $\mathbf{p}$ such that $\mu(\mathbf{p}) = 1$; $\mathbf{p}$ will denote this choice. (For algebraic conditions on Φ for such a $\mathbf{p}$ to exist, see e.g. [30].)

We note that by (49)–(50), $\phi_k^0 \leqslant \phi_{k-1}$, $k \geqslant 1$. Hence, if $p_k = ab^k \phi_k$, then $\sum_k b^k \phi_k^0 < \infty$, and it is possible to find $a_0 > 0$ such that $\mathbf{p}^0 := \boldsymbol{\phi}^{0\prime}$ given by (32) also is a probability sequence; hence $\mathcal{T}_n^{\boldsymbol{\phi},\boldsymbol{\phi}^0} = \mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}$ is a modified Galton–Watson tree. Furthermore, $p_k^0 \leqslant (a_0/a)p_k$, and thus if $\sigma^2(\mathbf{p}) < \infty$, then $\sigma^2(\mathbf{p}^0) < \infty$.

We begin with an unrooted version of Lemma 10.1.

Lemma 11.1. *Let $\mathbf{w}$, $\boldsymbol{\phi}$ and $\mathbf{p}$ be as above and assume $\sigma^2 := \sigma^2(\mathbf{p}) < \infty$. Let Λ_n be the distance profile of the unrooted simply generated tree $\mathcal{T}_n^{\mathbf{w},\circ}$. Then, as $n \to \infty$, for any continuous function with compact support $f : [0,\infty) \to \mathbb{R}$,*

$$\int_0^\infty n^{-3/2}\Lambda_n\big(xn^{1/2}\big)f(x)\,\mathrm{d}x \xrightarrow{\mathrm{d}} \int_0^1\int_0^1 f\Big(\frac{2}{\sigma}\mathrm{d}(s,t;\mathbf{e})\Big)\,\mathrm{d}s\,\mathrm{d}t. \tag{96}$$

Proof. Consider the leaf-biased random tree $\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}$ defined in Section 8. By Lemma 8.2, we may assume $\mathbb{P}(\widehat{\mathcal{T}}_n^{\mathbf{w},\circ} \neq \mathcal{T}_n^{\mathbf{w},\circ}) \to 0$ and thus it suffices to show (96) with $\Lambda_{\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}}$ instead of Λ_n. If $\mathcal{T}_{n,+}$ denotes the unique branch of $\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}$, then, trivially,

$$0 \leqslant \Lambda_{\widehat{\mathcal{T}}_n^{\mathbf{w},\circ}}(x) - \Lambda_{\mathcal{T}_{n,+}}(x) \leqslant 2n - 1, \qquad x \geqslant 0, \tag{97}$$

and thus we may further reduce and replace Λ_n in (96) by $\Lambda_{\mathcal{T}_{n,+}}$. As shown in Section 8, $\mathcal{T}_{n,+} \overset{\mathrm{d}}{=} \mathcal{T}_{n-1}^{\boldsymbol{\phi}} = \mathcal{T}_{n-1}^{\mathbf{p}}$, and the result now follows from Lemma 10.1, replacing n there by $n-1$ and x by $x = (n/(n-1))^{1/2}x$, noting that $\sup_x |f(x) - f((n/(n-1))^{1/2}x)| \to 0$ as $n \to \infty$. $\qquad\square$

Theorem 11.2. *Let $\mathbf{w}$, $\boldsymbol{\phi}$ and $\mathbf{p}$ be as above, and assume $\sigma^2 := \sigma^2(\mathbf{p}) < \infty$. Let Λ_n be the distance profile of the unrooted simply generated tree $\mathcal{T}_n^{\mathbf{w},\circ}$. Then, as $n \to \infty$,*

$$n^{-3/2}\Lambda_n\big(xn^{1/2}\big) \xrightarrow{\mathrm{d}} \frac{\sigma}{2}\Lambda_{\mathbf{e}}\Big(\frac{\sigma}{2}x\Big), \tag{98}$$

in the space $C[0,\infty]$, where $\Lambda_{\mathbf{e}}(x)$ is as in Theorem 1.4.

Proof. Let

$$Y_n(x) := n^{-3/2}\Lambda_n\big(xn^{1/2}\big) = n^{-3/2}\Lambda_{\mathcal{T}_n^{\mathbf{w},\circ}}\big(xn^{1/2}\big). \tag{99}$$

Regard Y_n as a random element of $C[0,\infty]$. Define also the mapping $\psi : C[0,\infty] \to \mathcal{M}([0,\infty))$, the space of all locally finite Borel measures on $[0,\infty)$, defined by $\psi(h) := h(x)\,\mathrm{d}x$; i.e., for $h \in C[0,\infty]$ and $f \in C[0,\infty)$ with compact support,

$$\int_0^\infty f(x)\,\mathrm{d}\psi(h) := \int_0^\infty f(x)h(x)\,\mathrm{d}x. \tag{100}$$

In other word, $\psi(h)$ has density h.

We give $\mathcal{M}([0,\infty))$ the vague topology, i.e., $\nu_n \to \nu$ in $\mathcal{M}([0,\infty))$ if $\int f\,\mathrm{d}\nu_n \to \int f\,\mathrm{d}\nu$ for every $f \in C[0,\infty)$ with compact support, and note that $\mathcal{M}([0,\infty))$ is a Polish space, see e.g. [32, Theorem A2.3]. Clearly, the separable Banach space $C[0,\infty]$ is also a Polish space. (Recall that a Polish space has a topology that can be defined by a complete separable metric.) It follows from the

definition (100) that ψ is continuous $C[0,\infty] \to \mathcal{M}([0,\infty))$. Furthermore, ψ is injective, since the density of a measure is a.e. uniquely determined.

We will use the alternative method of proof in [17, p. 123–125], and show the following two properties:

Claim 1. The sequence Y_n is tight in $C[0,\infty]$.

Claim 2. The sequence of random measures $\psi(Y_n)$ converges in distribution in $\mathcal{M}([0,\infty))$ to some random measure ζ.

It then follows from [14, Lemma 7.1] (see also [17, Theorem 4.17]) that

$$Y_n \xrightarrow{d} Z, \qquad \text{in } C[0,\infty], \tag{101}$$

for some random $Z \in C[0,\infty]$ such that

$$\psi(Z) \stackrel{d}{=} \zeta. \tag{102}$$

It will then be easy to complete the proof.

Proof of Claim 1: For $i = 1, \ldots, n$, let $\mathcal{T}(i)$ be $\mathcal{T}_n^{\mathbf{w},\circ}$ rooted at i. By symmetry, all $\mathcal{T}(i)$ have the same distribution; moreover, they equal in distribution $\widehat{\mathcal{T}}_n^{\mathbf{w},\bullet}$ defined in Section 6 (which has a random root). Hence, if we order each $\mathcal{T}(i)$ randomly, we have by Section 6

$$\mathcal{T}(i) \stackrel{d}{=} \mathcal{T}_n^{\mathbf{w},*} = \mathcal{T}_n^{\boldsymbol{\phi},\boldsymbol{\phi}^0} = \mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}. \tag{103}$$

By (16),

$$Y_n(x) = n^{-3/2} \Lambda_{\mathcal{T}_n^{\mathbf{w},\circ}}\left(xn^{1/2}\right) = \frac{1}{n} \sum_{i=1}^{n} n^{-1/2} L_{T(i)}\left(xn^{1/2}\right). \tag{104}$$

Since the sequence $n^{-1/2} L_{\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}}\left(xn^{1/2}\right)$ converges in $C[0,\infty]$ by Theorem 9.1, it is tight in $C[0,\infty]$. Furthermore,

$$\sup_x \left| n^{-1/2} L_{\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}}\left(xn^{1/2}\right) \right| = n^{-1/2} W(\mathcal{T}_n^{\mathbf{p},\mathbf{p}^0}), \tag{105}$$

which are uniformly integrable by (77) in Theorem 9.2. Hence, by Lemma A.2 (which we state and prove in Appendix A) and Remark A.3, (104) and (103) imply that the sequence Y_n is tight in $C[0,\infty]$, proving Claim 1.

Proof of Claim 2: Let $f \in C[0,\infty)$ have compact support. Then, Lemma 11.1 shows that

$$\int_0^\infty f(x)\,\mathrm{d}\psi(Y_n) = \int_0^\infty f(x) Y_n(x)\,\mathrm{d}x \xrightarrow{d} \int_0^1 \int_0^1 f\left(\frac{2}{\sigma}\mathrm{d}(s,t;\mathbf{e})\right)\mathrm{d}s\,\mathrm{d}t$$
$$= \int_0^\infty f(x)\,\mathrm{d}\zeta(x), \tag{106}$$

where ζ is the (random) probability measure on $[0,\infty)$ defined as the push-forward of Lebesgue measure on $[0,1] \times [0,1]$ by the map $(s,t) \to (2/\sigma)\mathrm{d}(s,t;\mathbf{e})$; in other words, ζ is the conditional distribution, given $\mathbf{e}$, of $(2/\sigma)\mathrm{d}(U_1, U_2; \mathbf{e})$ where U_1 and U_2 are independent uniform $U[0,1]$ random variables. This convergence in distribution for each f with compact support is equivalent to convergence in $\mathcal{M}([0,\infty))$, see [32, Theorem 16.16]. Thus, $\psi(Y_n) \xrightarrow{d} \zeta$ in $\mathcal{M}([0,\infty))$, proving Claim 2.

As said above, the claims imply (101)–(102). Thus, by the definition of ζ above, see (106), for any bounded measurable $f : [0, \infty) \to \mathbb{R}$,

$$
\int_0^\infty f(x)Z(x)\,\mathrm{d}x = \int_0^\infty f(x)\,\mathrm{d}\psi(Z) = \int_0^\infty f(x)\,\mathrm{d}\zeta(x)
$$

$$
= \int_0^1 \int_0^1 f\Big(\frac{2}{\sigma}\mathsf{d}(s, t; \mathbf{e})\Big)\,\mathrm{d}s\,\mathrm{d}t. \tag{107}
$$

Define

$$
\Lambda_{\mathbf{e}}(x) := \frac{2}{\sigma}Z\Big(\frac{2}{\sigma}x\Big). \tag{108}
$$

Then, (101) is the same as (98). Furthermore, replacing $f(x)$ by $f(\sigma x / 2)$ in (107) yields after a simple change of variables (21) and thus (4). (In particular, $\Lambda_{\mathbf{e}}$ does not depend on the weight sequence $\mathbf{w}$.) $\qquad\square$

12. Distance profile of rooted trees

Finally, we can prove Theorem 1.4 as a simple consequence of the corresponding result Theorem 11.2 for unrooted trees.

Proof of Theorem 1.4. Let $\boldsymbol{\phi} = \mathbf{p}$, and use (49) to define the weight sequence $\mathbf{w}$. Then, Theorem 11.2 applies to $\mathcal{T}_n^{\mathbf{w}, \circ}$. Consider, as in the proof of Lemma 11.1, the leaf-biased unrooted random tree $\widehat{\mathcal{T}}_n^{\mathbf{w}, \circ}$. By Lemma 8.2, Theorem 11.2 holds also for $\widehat{\mathcal{T}}_n^{\mathbf{w}, \circ}$.

Let again $\mathcal{T}_{n,+} \stackrel{\mathrm{d}}{=} \mathcal{T}_{n-1}^{\boldsymbol{\phi}} = \mathcal{T}_{n-1}^{\mathbf{p}}$ denote the unique branch of $\widehat{\mathcal{T}}_n^{\mathbf{w}, \circ}$. By (97), Theorem 11.2 holds also for $\mathcal{T}_{n,+}$ and thus for $\mathcal{T}_{n-1}^{\mathbf{p}}$. Replace n by $n + 1$, then a change of variables shows that Theorem 11.2 holds for $\mathcal{T}_n^{\mathbf{p}}$ too, which is Theorem 1.4. $\qquad\square$

As part of the proof, we have shown the corresponding result Theorem 11.2 for unrooted trees. The result also extends easily to conditioned modified Galton–Watson trees, using the method by Marckert and Panholzer [41] and Kortchemski and Marzouk [38].

Theorem 12.1. *Let Λ_n be the distance profile of a conditioned modified Galton–Watson tree $\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0}$ of order n and assume that $\mu(\mathbf{p}) = 1$, $\sigma^2(\mathbf{p}) < \infty$ and $\mu(\mathbf{p}^0) < \infty$. Then, (3) holds as $n \to \infty$.*

Proof. This is a simple consequence of Theorem 1.4 and Lemma 4.3. It follows from (42) that

$$
\sup_x \big|\Lambda_{\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0}}(x) - \Lambda_{\mathcal{T}_{n,(1)}}(x)\big| \leqslant 2n\big(n - |\mathcal{T}_{n,(1)}|\big) = O_{\mathrm{p}}(n), \tag{109}
$$

and thus (3) for $\mathcal{T}_n^{\mathbf{p}, \mathbf{p}^0}$ follows from the same result for $\mathcal{T}_{n,(1)}$, which follows from (42) and Theorem 1.4 by conditioning on $|\mathcal{T}_{n,(1)}|$. We omit the details. $\qquad\square$

13. Wiener index

Recall that the *Wiener index* of a tree T is defined as

$$
\mathrm{Wie}(T) := \frac{1}{2}\sum_{v, w \in T} \mathsf{d}(v, w), \tag{110}
$$

where d is the graph distance in T. Thus,

$$
\mathrm{Wie}(T) = \frac{1}{2}\sum_{i=1}^\infty i\Lambda_T(i) = \frac{1}{2}\int_{-1}^\infty x\Lambda_T(x)\,\mathrm{d}x. \tag{111}
$$

Since the integrand x in (111) is unbounded, convergence of the Wiener index does not follow immediately from convergence of the profile, but only a simple extra truncation argument is needed. For a conditioned Galton–Watson tree $\mathcal{T}_n^{\mathbf{p}}$, with $\mu(\mathbf{p}) = 1$ and $\sigma^2(\mathbf{p}) < \infty$, convergence is more easily proved directly from Aldous's result (95) [3, 4], see [28], but as an application of the results above, we show the corresponding result for unrooted trees.

Theorem 13.1. *Let* $\mathbf{w}$ *and* $\mathbf{p}$ *be as in Theorem 11.2. Then*

$$n^{-5/2}\mathrm{Wie}\big(\mathcal{T}_n^{\mathbf{w},\circ}\big) \xrightarrow{\mathrm{d}} \frac{1}{\sigma}\int_0^\infty x\Lambda_{\mathbf{e}}(x)\,\mathrm{d}x = \frac{1}{\sigma}\int_0^1\int_0^1 d(s,t;\mathbf{e})\,\mathrm{d}s\,\mathrm{d}t. \tag{112}$$

Proof. Let $T \in \mathfrak{T}_n$, and define a modified version by

$$\mathrm{Wie}'(T) := \frac{1}{2}\int_0^\infty x\Lambda_T(x)\,\mathrm{d}x = \mathrm{Wie}(T) + O(n), \tag{113}$$

using (111) and noting that $\Lambda_T(x) \leqslant n$ for $x \in [-1, 0]$. It suffices to prove (112) for $\mathrm{Wie}'(\mathcal{T}_n^{\mathbf{w},\circ})$. By (113),

$$n^{-5/2}\mathrm{Wie}'(T) = \frac{n^{-5/2}}{2}\int_0^\infty x\Lambda_T(x)\,\mathrm{d}x = \frac{1}{2}\int_0^\infty xn^{-3/2}\Lambda_T(n^{1/2}x)\,\mathrm{d}x. \tag{114}$$

Define a truncated version by, for $m \geq 0$,

$$\mathrm{Wie}_m(T) := \frac{n^{5/2}}{2}\int_0^\infty (x \wedge m)n^{-3/2}\Lambda_T(n^{1/2}x)\,\mathrm{d}x. \tag{115}$$

Then, since the support of Λ_T is $[-1, \mathrm{diam}(T) + 1]$,

$$\mathbb{P}\big(\mathrm{Wie}'(\mathcal{T}_n^{\mathbf{w},\circ}) \neq \mathrm{Wie}_m(\mathcal{T}_n^{\mathbf{w},\circ})\big) \leqslant \mathbb{P}\big(\mathrm{diam}(\mathcal{T}_n^{\mathbf{w},\circ}) > n^{1/2}m - 1\big). \tag{116}$$

Since $\mathrm{diam}(\mathcal{T}_n^{\mathbf{w},\circ}) = O_{\mathrm{p}}(n^{1/2})$, as an easy consequence of any of the constructions in Sections 6–8 and known results on the height of rooted Galton–Watson trees, see e.g. [3, 36] or [1, Theorem 1.2], it follows that

$$\sup_n \mathbb{P}\big(\mathrm{Wie}'(\mathcal{T}_n^{\mathbf{w},\circ}) \neq \mathrm{Wie}_m(\mathcal{T}_n^{\mathbf{w},\circ})\big) \to 0, \qquad \text{as } m \to \infty. \tag{117}$$

Furthermore, for each fixed m, (115) and Theorem 11.2 imply, as $n \to \infty$,

$$\begin{aligned}
n^{-5/2}\mathrm{Wie}_m(\mathcal{T}_n^{\mathbf{w},\circ}) &\xrightarrow{\mathrm{d}} \frac{1}{2}\int_0^\infty (x \wedge m)\frac{\sigma}{2}\Lambda_{\mathbf{e}}\Big(\frac{\sigma}{2}x\Big)\,\mathrm{d}x \\
&= \frac{1}{\sigma}\int_0^\infty \Big(x \wedge \frac{\sigma m}{2}\Big)\Lambda_{\mathbf{e}}(x)\,\mathrm{d}x.
\end{aligned} \tag{118}$$

The convergence in (112) follows by (117) and (118), see [11, Theorem 4.2]. The equality in (112) holds by (21).

Of course, the limit in (112) agrees with the limit in [28] for the rooted case.

14. Moments of the distance profile

In this section, we prove the following estimates on moments of the distance profile for a conditioned Galton–Watson tree; we use again the simplified notation L_n and Λ_n for the profile and distance profile as in Theorems 1.2 and 1.4. Throughout the section, C and c denote some positive constants that may depend on the offspring distribution $\mathbf{p}$ only; C_r denotes constants depending on $\mathbf{p}$ and the parameter r only. (As always, these may change from one occurrence to the next.)

Theorem 14.1. *Let* Λ_n *be the distance profile of a conditioned Galton–Watson tree of order* n, *with an offspring distribution* $\mathbf{p}$ *such that* $\mu(\mathbf{p}) = 1$ *and* $\sigma^2(\mathbf{p}) < \infty$.

(i) Let $r \geqslant 1$ be a real number. Then, for all $i, n \geqslant 1$,

$$\mathbb{E}[\Lambda_n(i)^r] \leqslant C_r n^{3r/2} e^{-ci^2/n}. \tag{119}$$

(ii) Let $r \geqslant 1$ be an integer, and suppose that $\mathbf{p}$ has a finite $(r+1)th$ moment:

$$\sum_k k^{1+r} p_k < \infty. \tag{120}$$

Then, for all $i, n \geqslant 1$,

$$\mathbb{E}[\Lambda_n(i)^r] \leqslant C_r i^r n^r. \tag{121}$$

Furthermore, we may in this case combine (119) and (121) to

$$\mathbb{E}[\Lambda_n(i)^r] \leqslant C_r i^r n^r e^{-ci^2/n}, \tag{122}$$

for all $i, n \geqslant 1$.

The proof is given later. Note that (122) trivially implies (121) and (changing C_r and c) (119); conversely (119) and (121) imply (122) by considering $i \leqslant n^{1/2}$ and $i > n^{1/2}$ separately.

The special case $r = 1$ of (121), i.e.,

$$\mathbb{E}\Lambda_n(i) \leqslant Cin, \tag{123}$$

was proved in [16, Theorem 1.3]; note that when $r = 1$, (120) holds automatically by our assumption $\sigma^2(\mathbf{p}) < \infty$.

Remark 14.2. The estimates above are natural analogues of estimates for the profile L_n. First, as proved in [1, Theorem 1.6], under the conditions in Theorem 14.1(i),

$$\mathbb{E}[L_n(i)^r] \leqslant C_r n^{r/2} e^{-ci^2/n}, \tag{124}$$

for all $i, n \geqslant 1$. Second, as proved in [29, Theorem 1.13], under the conditions in Theorem 14.1(ii),

$$\mathbb{E}[L_n(i)^r] \leqslant C_r i^r, \tag{125}$$

for all $i, n \geqslant 1$. The estimates (124)–(125) are used in our proof of Theorem 14.1.   □

Remark 14.3. We do not know whether the moment assumption (120) really is necessary for the result. This assumption is necessary for the corresponding estimate (125) for the profile $L_n(i)$, as noted in [1], but the argument there does not apply to the distance profile. We state this as an open problem.   □

Problem 14.4. Does (121) hold without the assumption (120)?

Remark 14.5. We also do not know whether the assumption that r is an integer is necessary in Theorem 14.1(i); we conjecture that it is not. (This assumption is used in the proof of (125) in [29].)   □

As an immediate consequence of Theorems 1.4 and 14.1, we obtain the corresponding results for the asymptotic profile $\Lambda_{\mathbf{e}}$.

Theorem 14.6. *For any $r \geqslant 1$ and all $x \geqslant 0$,*

$$\mathbb{E}[\Lambda_{\mathbf{e}}(x)^r] \leqslant C_r \min(x^r, e^{-cx^2}). \tag{126}$$

Proof. Fix an offspring distribution $\mathbf{p}$ with $\mu(\mathbf{p}) = 1$ and all moments finite. (For example, we may choose a well-known example such as $Po(1)$ or $Bi(2, \frac{1}{2})$.) Let $x \in (0, \infty)$. Define $i_n := \lceil 2\sigma^{-1} x n^{1/2} \rceil$. Then $i_n/n^{1/2} \to 2x/\sigma$ as $n \to \infty$, and (3) implies

$$n^{-3/2} \Lambda_n(i_n) \xrightarrow{\mathrm{d}} \frac{\sigma}{2} \Lambda_{\mathbf{e}}(x). \tag{127}$$

Hence, Fatou's lemma and (119) yield, for any $r \geqslant 1$,

$$\mathbb{E}[\Lambda_{\mathbf{e}}(x)^r] \leqslant (2/\sigma)^r \liminf_{n\to\infty} \mathbb{E}[n^{-3r/2}\Lambda_n(i_n)^r] \leqslant C_r e^{-cx^2}. \tag{128}$$

Similarly, Fatou's lemma and (121) yield

$$\mathbb{E}[\Lambda_{\mathbf{e}}(x)^r] \leqslant C_r x^r \tag{129}$$

for integer $r \geqslant 1$; equivalently, the L^r norm is estimated by $\|\Lambda_{\mathbf{e}}(x)\|_r \leqslant C_r x$. This estimate extends to all real $r \geqslant 1$ by Lyapounov's inequality. Hence, (128) and (129) both hold for all $r \geqslant 1$, which yields (126) for $x > 0$.

Finally, the result (126) is trivial for $x = 0$, since $\Lambda_{\mathbf{e}}(0) = 0$ a.s.   □

Remark 14.7. The same argument shows that Theorem 1.2 and (124)–(125) imply

$$\mathbb{E}[L_{\mathbf{e}}(x)^r] \leqslant C_r \min(x^r, e^{-cx^2}), \tag{130}$$

for any $r \geqslant 1$ and all $x \geqslant 0$.   □

The proof of Theorem 14.1 relies on an invariance property of the law of Galton–Watson trees under random re-rooting proved by Bertoin and Miermont [9, Proposition 2]. A pointed tree is a pair (T, v), where T is an ordered rooted tree (also called planar rooted tree) and v is a vertex of T. We endow the space of pointed trees with the σ-finite measure $\mathbb{P}^{\bullet}$ defined by

$$\mathbb{P}^{\bullet}((T, v)) = \mathbb{P}(\mathcal{T}^{\mathbf{p}} = T), \tag{131}$$

where $\mathcal{T}^{\mathbf{p}}$ is a Galton–Watson tree with offspring distribution $\mathbf{p}$ such that $\mu(\mathbf{p}) = 1$. We let $\mathbb{E}^{\bullet}$ denote the expectation under this 'law'. In particular, the conditional law $\mathbb{P}^{\bullet}(\cdot \mid |T| = n)$ on the space of pointed trees with n vertices is well defined and equals the distribution of $(\mathcal{T}_n^{\mathbf{p}}, v)$ where given the conditioned Galton–Watson tree $\mathcal{T}_n^{\mathbf{p}}$ of order n, v is a uniform random vertex of $\mathcal{T}_n^{\mathbf{p}}$.

Let us now describe the transformation of pointed trees of Bertoin and Miermont [9, Section 4]. They work with planted planar trees; the base in the planted tree is useful since it implicitly specifies the ordering of the transformed tree. However, we ignore this detail and formulate their transformation for rooted trees; our formulation is easily seen to be equivalent to theirs. For any rooted planar tree T and vertex v of T, let T_v be the fringe subtree of T rooted at v, and let T^v be the subtree of T obtained by deleting all the strict descendants of v in T. We define a new pointed tree $(\hat{T}, \hat{v})$ in the following way. If v is the root, we do nothing, and let $(\hat{T}, \hat{v}) = (T, v)$. Otherwise, first remove the edge $e(v)$ between v and its parent $pr(v)$ in T, and instead connect v to the root of T. We then re-root the resulting tree at $pr(v)$ and obtain the new rooted tree $\hat{T}$, which we point at $\hat{v} = v$. Note that $\hat{T}_{\hat{v}} = T_v$, and that $\hat{T}^{\hat{v}} \setminus \{\hat{v}\}$ equals $T^v \setminus \{v\} = T \setminus T_v$ rerooted at $pr(v)$.

Bertoin and Miermont [9, Proposition 2 and its proof] and its proof establish that this transformation preserves the measure $\mathbb{P}^{\bullet}$; this includes the following.

Proposition 14.8. (Bertoin and Miermont [9]). Under $\mathbb{P}^{\bullet}$, $(\hat{T}^{\hat{v}}, T_v)$ and (T^v, T_v) have the same 'law'. Furthermore, the trees T^v and T_v are independent, with T_v being a Galton–Watson tree with offspring distribution $\mathbf{p}$.   □

Proof of Theorem 14.1. For a rooted plane tree T and $i \geqslant 0$ an integer, it will be convenient to write $Z_i(T) := L_T(i)$ for the number of vertices at distance i from the root of T. For a vertex v in T, note that the number of vertices at distance $i \geqslant 1$ from v is less than or equal to $Z_{i-1}(\hat{T}^{\hat{v}}) + Z_i(T_v)$. (Strict inequality may occur because of the extra vertex $\hat{v}$ added in $\hat{T}^{\hat{v}}$.) From (16), we thus obtain that if Λ_T is the distance profile of T defined in (14), then

$$\Lambda_T(i) \leqslant \sum_{v \in T}(Z_{i-1}(\hat{T}^{\hat{v}}) + Z_i(T_v)), \qquad i \geqslant 1. \tag{132}$$

If $r \geqslant 1$ and $|T| = n$, then (132) and Jensen's inequality yield

$$\big(n^{-1}\Lambda_T(i)\big)^r \leqslant \frac{1}{n}\sum_{v\in T}\big(Z_{i-1}(\hat{T}^{\hat{v}}) + Z_i(T_v)\big)^r. \tag{133}$$

Consequently, using (131),

$$\begin{aligned}
\mathbb{E}[\Lambda_n(i)^r] &\leqslant n^r\mathbb{E}^\bullet\big[\big(Z_{i-1}(\hat{T}^{\hat{v}}) + Z_i(T_v)\big)^r \,\big|\, |T| = n\big] \\
&\leqslant 2^r n^r\mathbb{E}^\bullet\big[Z_{i-1}(\hat{T}^{\hat{v}})^r + Z_i(T_v)^r \,\big|\, |T| = n\big].
\end{aligned} \tag{134}$$

By Proposition 14.8, we have on the one hand that

$$\begin{aligned}
\mathbb{E}^\bullet\big[Z_{i-1}(\hat{T}^{\hat{v}})^r \,\big|\, |T| = n\big] &= \mathbb{E}^\bullet\big[Z_{i-1}(T^v)^r \,\big|\, |T| = n\big] \\
&\leqslant \mathbb{E}^\bullet\big[Z_{i-1}(T)^r \,\big|\, |T| = n\big] \\
&= \mathbb{E}[Z_{i-1}(\mathcal{T}_n^{\mathbf{p}})^r].
\end{aligned} \tag{135}$$

On the other hand, since $|T| = |T^v| + |T_v| - 1$, we see from Proposition 14.8 that

$$\begin{aligned}
\mathbb{E}^\bullet\big[Z_i(T_v)^r \,\big|\, |T| = n\big] &= \sum_{m=1}^n \mathbb{E}^\bullet\big[Z_i(T_v)^r \,\big|\, |T_v| = m, |T^v| = n - m + 1\big] \\
&\quad \times \mathbb{P}^\bullet\big[|T_v| = m \,\big|\, |T| = n\big] \\
&= \sum_{m=1}^n \mathbb{E}\big[Z_i(\mathcal{T}_m^{\mathbf{p}})^r\big]\mathbb{P}^\bullet\big[|T_v| = m \,\big|\, |T| = n\big] \\
&\leqslant \sup_{1\leqslant m\leqslant n}\mathbb{E}\big[Z_i(\mathcal{T}_m^{\mathbf{p}})^r\big].
\end{aligned} \tag{136}$$

Combining (134), (135) and (136), we have that

$$\begin{aligned}
\mathbb{E}[\Lambda_n(i)^r] &\leqslant 2^r n^r\big(\mathbb{E}\big[Z_{i-1}(\mathcal{T}_n^{\mathbf{p}})^r\big] + \sup_{1\leqslant m\leqslant n}\mathbb{E}\big[Z_i(\mathcal{T}_m^{\mathbf{p}})^r\big]\big) \\
&= 2^r n^r\big(\mathbb{E}\big[L_n(i-1)^r\big] + \sup_{1\leqslant m\leqslant n}\mathbb{E}\big[L_m(i)^r\big]\big).
\end{aligned} \tag{137}$$

Therefore, (119) and (121) follow from (124) and (125), proved in [1, Theorem 1.6] and [29, Theorem 1.13], respectively.

Finally (122) follows from (119) and (121) as noted above.

15. Hölder continuity

We now discuss Hölder continuity properties of the continuous random functions $L_{\mathbf{e}}$ and $\Lambda_{\mathbf{e}}$. We begin with $L_{\mathbf{e}}$, the local time of $\mathbf{e}$. The results are to a large extent known, although we do not know any reference to the form of them stated here; nevertheless, we treat also $L_{\mathbf{e}}$ in detail, as a background to and preparation for the discussion of $\Lambda_{\mathbf{e}}$ below.

It is well known that the local time of Brownian motion at some fixed time a.s. is Hölder continuous of order α for any $\alpha < 1/2$, see [49, VI.(1.8)]. It is also known that this Hölder continuity extends by standard arguments to the local times of Brownian bridge and Brownian excursion, and thus (see Theorem 1.2) to $L_{\mathbf{e}}$. We need a quantitative version of this.

For $\alpha > 0$ and an interval $I \subseteq \mathbb{R}$, let the Hölder space $\mathcal{H}_\alpha = \mathcal{H}_\alpha(I)$ be the space of functions $f : I \to \mathbb{R}$ such that

$$\|f\|_{\mathcal{H}_\alpha} := \sup\Big\{\frac{|f(x) - f(y)|}{|x - y|^\alpha} : x, y \in I, x \neq y\Big\} < \infty \tag{138}$$

This is a semi-norm. We may regard $\mathcal{H}_\alpha$ as a space of functions modulo constants, and then $\|\cdot\|_{\mathcal{H}_\alpha}$ is a norm and $\mathcal{H}_\alpha$ a Banach space.

Theorem 15.1. *Let $0 < \alpha < \frac{1}{2}$ and let $A < \infty$. Then*

$$\mathbb{E}\,\|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,A]} < \infty. \tag{139}$$

In particular, $L_{\mathbf{e}} \in \mathcal{H}_\alpha[0, \infty)$ a.s.

Proof. We let α and A be fixed. Constants C below may depend on α and A.

Recall that $L_{\mathbf{e}}(x) = L_1^x$, where L_t^x is the local time of a (standard) Brownian excursion $\mathbf{e}$. The proof will actually show the result for L_t^x for any fixed $t \in [0, 1]$.

We first note that if β is a 3-dimensional Bessel process started from 0 (see e.g. [49, Section VI.3]), then a standard Brownian excursion $\mathbf{e}$ is given by $\mathbf{e}(1) = 0$ and

$$\mathbf{e}(t) = (1-t)\beta\left(\frac{t}{1-t}\right), 0 \le t < 1; \tag{140}$$

see [13, II.(1.5)]. One then can deduce from [49, VI.(3.3)] and an application of the Itô integration by parts formula that $\mathbf{e}$ satisfies

$$\mathbf{e}(t) = \int_0^t \left(\frac{1}{\mathbf{e}(s)} - \frac{\mathbf{e}(s)}{1-s}\right) ds + B(t), \qquad 0 \le t \le 1, \tag{141}$$

where B is a standard Brownian motion. In particular, $\mathbf{e}$ is a continuous semi-martingale.

We now follow the proof of [49, VI.(1.7), see also VI.(1.32)], but avoid localising at the cost of further calculations. Let

$$\mathbf{v}(t) := \frac{1}{\mathbf{e}(t)} - \frac{\mathbf{e}(t)}{1-t}, \qquad 0 \le t < 1, \tag{142}$$

so (141) can be written

$$\mathbf{e}(t) = B(t) + \int_0^t \mathbf{v}(s)\, ds. \tag{143}$$

Then, by Tanaka's formula [49, VI.(1.2)], writing $x^+ := x \vee 0$,

$$L_t^x = 2\big(\mathbf{e}(t) - x\big)^+ - 2\big(-x\big)^+ - 2\int_0^t \mathbf{1}\{\mathbf{e}(s) > x\}\, dB(s)$$
$$- 2\int_0^t \mathbf{1}\{\mathbf{e}(t) > x\}\mathbf{v}(s)\, ds. \tag{144}$$

Fix $t \in [0, 1]$ and denote the four random functions of x on the right-hand side of (144) by $F_1(x), \ldots, F_4(x)$. Trivially, the first two are in $\mathcal{H}_1$ (= Lipschitz), with norm at most 2. Hence,

$$\|F_1\|_{\mathcal{H}_\alpha[0,A]} + \|F_2\|_{\mathcal{H}_\alpha[0,A]} \le C. \tag{145}$$

For each x, $F_3(x)$ is a continuous martingale in t. Thus, if $0 \le x < y \le A$, the Burkholder–Davis–Gundy inequality [49, IV.(4.1)] and the occupation times formula [49, VI.(1.9)] yield, for any $p \ge 2$,

$$\mathbb{E}|F_3(x) - F_3(y)|^p \leqslant C_p \mathbb{E}\Big[\Big(\int_0^t \mathbf{1}\{x < \mathbf{e}(s) \leqslant y\}\, ds\Big)^{p/2}\Big]$$

$$= C_p \mathbb{E}\Big[\Big(\int_x^y L_t^z\, dz\Big)^{p/2}\Big]$$

$$= C_p(y-x)^{p/2} \mathbb{E}\Big[\Big(\frac{1}{y-x}\int_x^y L_t^z\, dz\Big)^{p/2}\Big]$$

$$\leqslant C_p(y-x)^{p/2} \mathbb{E}\Big[\frac{1}{y-x}\int_x^y \big(L_t^z\big)^{p/2}\, dz\Big].$$

$$= C_p(y-x)^{p/2} \frac{1}{y-x}\int_x^y \mathbb{E}\big(L_t^z\big)^{p/2}\, dz. \tag{146}$$

We have $L_t^z \leqslant L_1^z = L_{\mathbf{e}}(z)$, and thus (130) implies $\mathbb{E}(L_t^z)^{p/2} \leqslant C_p$ for all z; hence (146) yields

$$\mathbb{E}|F_3(x) - F_3(y)|^p \leqslant C_p(y-x)^{p/2}. \tag{147}$$

Consequently, F_3 satisfies the Kolmogorov continuity criterion, and the version of it stated in [49, I.(2.1)] shows that if p is chosen so large that $(p/2 - 1)/p > \alpha$, then (147) implies

$$\mathbb{E}\, \|F_3\|_{\mathcal{H}_\alpha[0,A]} \leqslant \big(\mathbb{E}\, \|F_3\|^p_{\mathcal{H}_\alpha[0,A]}\big)^{1/p} < \infty. \tag{148}$$

Similarly, using the extension of the occupation times formula in [49, VI.(1.15)], if again $0 \leqslant x < y \leqslant A$,

$$F_4(y) - F_4(x) = 2\int_0^t \mathbf{1}\{x < \mathbf{e}(s) \leqslant y\}\mathbf{v}(s)\, ds$$

$$= 2\int_0^t \mathbf{1}\{x < \mathbf{e}(s) \leqslant y\}\Big(\frac{1}{\mathbf{e}(s)} - \frac{\mathbf{e}(s)}{1-s}\Big)\, ds$$

$$= 2\int_x^y dz \int_0^t \Big(\frac{1}{z} - \frac{z}{1-s}\Big)\, dL_s^z. \tag{149}$$

We now simplify and assume $0 \leqslant t \leqslant \frac{1}{2}$. Then (149) implies

$$\big|F_4(y) - F_4(x)\big| \leqslant 2\int_x^y \Big(\frac{1}{z} + 2z\Big)L_t^z\, dz \leqslant C\int_x^y \frac{1}{z}L_t^z\, dz \leqslant C\int_x^y \frac{L_{\mathbf{e}}(z)}{z}\, dz. \tag{150}$$

Let $p' := 1/\alpha$ and let $p := (1-\alpha)^{-1} > 1$ be the conjugate exponent. Then, by (150) and Hölder's inequality,

$$\big|F_4(y) - F_4(x)\big| \leqslant C(y-x)^\alpha \Big(\int_x^y \frac{L_{\mathbf{e}}(z)^p}{z^p}\, dz\Big)^{1/p}. \tag{151}$$

Consequently,

$$\|F_4\|_{\mathcal{H}_\alpha[0,\infty)} \leqslant C\Big(\int_0^\infty \frac{L_{\mathbf{e}}(z)^p}{z^p}\, dz\Big)^{1/p}. \tag{152}$$

Thus, using again (130),

$$\mathbb{E}\, \|F_4\|^p_{\mathcal{H}_\alpha[0,\infty)} \leqslant C\mathbb{E}\int_0^\infty \frac{L_{\mathbf{e}}(z)^p}{z^p}\, dz = C\int_0^\infty \frac{\mathbb{E}L_{\mathbf{e}}(z)^p}{z^p}\, dz < \infty \tag{153}$$

and thus

$$\mathbb{E}\, \|F_4\|_{\mathcal{H}_\alpha[0,A]} \leqslant \mathbb{E}\, \|F_4\|_{\mathcal{H}_\alpha[0,\infty)} < \infty. \tag{154}$$

Consequently, (144), (145), (148) and (154) yield

$$\mathbb{E} \, \|L_t^x\|_{\mathcal{H}_\alpha[0,A]} < \infty \tag{155}$$

for $0 \leqslant t \leqslant \frac{1}{2}$.

We have for simplicity assumed $t \leqslant \frac{1}{2}$. To complete the proof, we note that $\mathbf{e}$ is invariant under reflection: $\mathbf{e}(1-t) \overset{\text{d}}{=} \mathbf{e}(t)$ (as processes), and thus $L_1^x - L_{1/2}^x \overset{\text{d}}{=} L_{1/2}^x$ (as processes in x). Consequently,

$$\mathbb{E} \, \|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,A]} = \mathbb{E} \, \|L_1^x\|_{\mathcal{H}_\alpha[0,A]} \leqslant 2\mathbb{E} \, \|L_{1/2}^x\|_{\mathcal{H}_\alpha[0,A]} < \infty, \tag{156}$$

showing (139). (The case $\frac{1}{2} < t < 1$ follows similarly.)

Finally, (139) shows that a.s., $L_{\mathbf{e}} \in \mathcal{H}_\alpha[0,A]$ for every $A > 0$. Moreover, $L_{\mathbf{e}}$ has finite support $[0, \sup \mathbf{e}]$, and thus $L_{\mathbf{e}} \in \mathcal{H}_\alpha[0,\infty)$. □

We have for simplicity considered a finite interval $[0, A]$ in Theorem 15.1, but the result can easily be extended to $\mathcal{H}_\alpha[0,\infty)$.

Theorem 15.2. *Let* $0 < \alpha < \frac{1}{2}$. *Then*

$$\mathbb{E} \, \|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,\infty)} < \infty. \tag{157}$$

Proof. Note that the proof of Theorem 15.1 actually shows $\mathbb{E} \, \|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,A]}^p < \infty$ for some $p > 1$, see (148) and (153). Moreover, the same proof applied to the interval $[m, m+1]$ shows that for any $m \geqslant 1$,

$$\mathbb{E} \, \|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[m,m+1]}^p \leqslant Cm^p, \tag{158}$$

with C independent of m, where the factor m^p comes from (150). Furthermore, $L_{\mathbf{e}} = 0$ on $[m, m+1]$ unless $\sup \mathbf{e} > m$. The explicit formula for the distribution function of $\sup \mathbf{e}$, see e.g. [15, 33] or [17, p. 114], yields the well-known subgaussian decay

$$\mathbb{P}(\sup \mathbf{e} > x) \leqslant e^{-cx^2}, \qquad x \geqslant 1. \tag{159}$$

(See also [1] for the corresponding result for heights of conditioned Galton–Watson trees.) Combining (158) and (159) with Hölder's inequality, we obtain, with $1/q = 1 - 1/p$,

$$\mathbb{E} \, \|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[m,m+1]} \leqslant \left(\mathbb{E} \, \|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[m,m+1]}^p \right)^{1/p} \mathbb{P}\left(\|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[m,m+1]} \neq 0 \right)^{1/q}$$
$$\leqslant Cm\mathbb{P}(\sup \mathbf{e} > m)^{1/q} \leqslant Cme^{-cm^2}, \tag{160}$$

Finally, it is easy to see that

$$\|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,\infty)} \leqslant \sum_{m=0}^{\infty} \|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[m,m+1]} \tag{161}$$

and thus, by (139) and (160),

$$\mathbb{E} \, \|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,\infty)} \leqslant \sum_{m=0}^{\infty} \mathbb{E} \, \|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[m,m+1]} \leqslant C + \sum_{m=1}^{\infty} Cme^{-cm^2} < \infty. \tag{162}$$

□

Remark 15.3. The proofs above apply also to the local time of the Brownian bridge. The main differences are that we consider functions on $(-\infty, \infty)$ and that the term $1/\mathbf{e}(t)$ in (142) disappears. This shows that, writing $L_{\mathbf{b}}(x)$ for the local time of the Brownian bridge at time $t = 1$,

$$\mathbb{E} \, \|L_{\mathbf{b}}\|_{\mathcal{H}_\alpha(-\infty,\infty)} < \infty. \tag{163}$$

By Vervaat [51], the Brownian excursion can be constructed from a Brownian bridge by shifts in both time and space, and thus $L_{\mathbf{e}}$ is a (random) shift of $L_{\mathbf{b}}$. Consequently, with this coupling,

$$\|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,\infty)} = \|L_{\mathbf{b}}\|_{\mathcal{H}_\alpha(-\infty,\infty)}, \tag{164}$$

and thus (157) and (163) are equivalent. This yields an alternative proof of Theorem 15.2. □

We have so far considered $\alpha < 1/2$. For $\alpha \geqslant 1/2$, it is well known that Brownian motion is a.s. not locally Hölder(α); see e.g. [49, I.(2.7)] or [43, Section 1.2] for even more precise results. Moreover, the same holds for the local time $\bar{L}_t^x$ of Brownian motion, regarded as a function of the space variable x (for any fixed $t > 0$); we do not know an explicit reference but this follows easily, for example, from the first Ray–Knight theorem [49, XI.(2.2)], which says that if we stop at τ_1, the hitting time of 1, then the process $(\bar{L}_{\tau_1}^{1-x})$, $0 \leqslant x \leqslant 1$, is a 2-dimensional squared Bessel process, which (at least away from 0) has the same smoothness as a Brownian motion (since it can be written as the sum of squares of two independent Brownian motions); we omit the details.

Similarly, $L_{\mathbf{e}} = L_1^x$, which is the local time of a standard Brownian excursion is a.s. not Hölder($\frac{1}{2}$). One way to see this is that if we stop a Brownian motion when its local time at 0 reaches 1, i.e., at $\tau := \inf\{t : \bar{L}_t^0 = 1\}$, then the part of the Brownian motion before time τ and above any fixed $\delta > 0$ is a.s. included in a finite number of excursions, and these are independent, conditioned on the number of them and their lengths. Hence, if the local time of a Brownian excursion were Hölder($\frac{1}{2}$) with positive probability, then so would $\bar{L}_\tau^x$, restricted to $x \geqslant \delta$, be, and then $\bar{L}_t^x$, $x \geqslant \delta$, would be Hölder($\frac{1}{2}$) with positive probability for some rational t, and thus for all $t > 0$ by scaling, which contradicts the argument above.

15.1. Distance profile

We next consider the asymptotic distance profile $\Lambda_{\mathbf{e}}$. We first note that the nice re-rooting invariance property of the standard Brownian excursion $\mathbf{e}$ presented in Section 2.3 yields the following corollary of Theorem 15.2.

Corollary 15.4. *Let $0 < \alpha < \frac{1}{2}$. Then*

$$\mathbb{E}\, \|\Lambda_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,\infty)} < \infty. \tag{165}$$

In particular, the random function $\Lambda_{\mathbf{e}} \in \mathcal{H}_\alpha[0,\infty)$ a.s.

Proof. It follows from the identity (24) that

$$\|\Lambda_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,\infty)} \leqslant \int_0^1 \|L_{\mathbf{e}^{[s]}}\|_{\mathcal{H}_\alpha[0,\infty)}\, \mathrm{d}s, \tag{166}$$

where for every $s \in [0,1]$, $L_{\mathbf{e}^{[s]}}$ denotes the local time of the process $\mathbf{e}^{[s]}$ defined in (22), which is distributed as a standard Brownian excursion (23). Hence, Theorem 15.2 yields

$$\mathbb{E}\, \|\Lambda_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,\infty)} \leqslant \int_0^1 \mathbb{E}\, \|L_{\mathbf{e}^{[s]}}\|_{\mathcal{H}_\alpha[0,\infty)}\, \mathrm{d}s = \mathbb{E}\, \|L_{\mathbf{e}}\|_{\mathcal{H}_\alpha[0,\infty)} < \infty. \tag{167}$$

□

However, it turns out that the averaging in (24) actually makes $\Lambda_{\mathbf{e}}$ smoother than $L_{\mathbf{e}}$; we have the following stronger result, which improves Corollary 15.4.

Theorem 15.5. *The asymptotic distance profile $\Lambda_{\mathbf{e}} \in \mathcal{H}_\alpha$ for every $\alpha < 1$, a.s. Furthermore, $\Lambda_{\mathbf{e}}$ is a.s. absolutely continuous and with a derivative $\Lambda_{\mathbf{e}}'$ (defined a.e.) that belongs to $L^p(\mathbb{R})$ for every $p < \infty$.*

In order to prove Theorem 15.5, we first prove an estimate of the Fourier transform of $\Lambda_{\mathbf{e}}$. We choose to define Fourier transforms as, for $f \in L^1(\mathbb{R})$,

$$\widehat{f}(\xi) := \int_{-\infty}^{\infty} f(x) e^{i\xi x} \, dx, \qquad -\infty < \xi < \infty. \tag{168}$$

Note that by (21), the Fourier transform $\widehat{\Lambda_{\mathbf{e}}}$ can be written as

$$\widehat{\Lambda_{\mathbf{e}}}(\xi) := \int_0^{\infty} e^{i\xi x} \Lambda_{\mathbf{e}}(x) \, dx = \iint_{s,t \in [0,1]} e^{i\xi d(s,t;\mathbf{e})} \, ds \, dt. \tag{169}$$

Lemma 15.6. *There exists a constant C such that*

$$\mathbb{E}\, |\widehat{\Lambda_{\mathbf{e}}}(\xi)|^2 \leqslant C\big(\xi^{-4} \wedge 1\big), \qquad -\infty < \xi < \infty. \tag{170}$$

Proof. Note first that the profile $\Lambda_{\mathbf{e}}$ is a (random) non-negative function with integral 1, e.g., by (4). Hence, for every ξ,

$$\big|\widehat{\Lambda_{\mathbf{e}}}(\xi)\big| \leqslant \int_0^{\infty} \big|\Lambda_{\mathbf{e}}(x)\big| \, dx = 1. \tag{171}$$

Thus, (170) is trivial for $|\xi| \leqslant 1$. Assume in the remainder of the proof $|\xi| \geqslant 1$.

By (169),

$$\begin{aligned}
\mathbb{E}\, \big|\widehat{\Lambda_{\mathbf{e}}}(\xi)\big|^2 &= \mathbb{E} \int \cdots \int_{t_1,\dots t_4 \in [0,1]} e^{i\xi(d(t_1,t_2;\mathbf{e}) - d(t_3,t_4;\mathbf{e}))} \, dt_1 \cdots dt_4 \\
&= \mathbb{E} e^{i\xi(d(U_1,U_2;\mathbf{e}) - d(U_3,U_4;\mathbf{e}))},
\end{aligned} \tag{172}$$

where $U_1, \dots, U_4$ are i.i.d. uniform $U(0,1)$ random variables, independent of $\mathbf{e}$.

Recall from Section 2.3 that the Brownian excursion $\mathbf{e}$ defines the continuum random tree $T_{\mathbf{e}}$, with a quotient map $\rho_{\mathbf{e}} : [0,1] \to T_{\mathbf{e}}$. Let $\overline{U}_i := \rho_{\mathbf{e}}(U_i)$, $i = 1, \dots, 4$, be the points in $T_{\mathbf{e}}$ corresponding to U_i; these are i.i.d. uniformly random points in $T_{\mathbf{e}}$, and $d(U_i, U_j; \mathbf{e})$ equals the distance $d(\overline{U}_i, \overline{U}_j)$ between $\overline{U}_i$ and $\overline{U}_j$ in the real tree $T_{\mathbf{e}}$. Then (172) becomes

$$\mathbb{E}\, \big|\widehat{\Lambda_{\mathbf{e}}}(\xi)\big|^2 = \mathbb{E} e^{i\xi(d(\overline{U}_1,\overline{U}_2) - d(\overline{U}_3,\overline{U}_4))}. \tag{173}$$

Furthermore, we can simplify the calculations by rerooting $T_{\mathbf{e}}$ at $\overline{U}_4$; this preserves the distribution, see (23), and $\overline{U}_1, \dots, \overline{U}_3$ are still independent uniformly random points in the tree; hence, we also have, with $o = \rho_{\mathbf{e}}(0)$ the root of $T_{\mathbf{e}}$.

$$\mathbb{E}\, \big|\widehat{\Lambda_{\mathbf{e}}}(\xi)\big|^2 = \mathbb{E} e^{i\xi(d(\overline{U}_1,\overline{U}_2) - d(\overline{U}_3,o))}. \tag{174}$$

To calculate (174), it suffices to consider the (real) subtree of $T_{\mathbf{e}}$ spanned by the root o and $\overline{U}_1, \dots, \overline{U}_3$. [4, Corollary 22] showed that the distribution of the random real tree $\widetilde{\mathcal{T}}_k$ spanned in this way by $k \geqslant 1$ i.i.d. uniform random points $\overline{U}_1, \dots, \overline{U}_k \in T_{\mathbf{e}}$ can be described as follows. Let the *shape* τ of $\widetilde{\mathcal{T}}_k$ be the tree regarded as a combinatorial tree, i.e., ignoring the edge lengths. Then τ is a.s. a rooted binary tree with k leaves, labelled by $1, \dots, k$ (corresponding to $\overline{U}_1, \dots, \overline{U}_k$). Let $\bar{\mathfrak{T}}^k$ be the set of possible shapes, i.e., the set of binary trees with k labelled leaves. Then $|\bar{\mathfrak{T}}^k| = (2k - 3)!!$. Each shape τ has $k - 1$ internal vertices, and thus $2k$ vertices and $2k - 1$ edges. For each shape τ, label the edges $1, \dots, 2k - 1$ in some order, and for each tree $\widetilde{\mathcal{T}}_k$ with shape τ, let $\ell_1, \dots, \ell_{2k-1}$ be the corresponding edge lengths. Then $\widetilde{\mathcal{T}}_k$ is described by $(\tau, \ell_1, \dots, \ell_k)$, and the distribution of $(\tau, \ell_1, \dots, \ell_k)$ is given by the density, for $\tau \in \bar{\mathfrak{T}}^k$ and $\ell_i > 0$,

$$2^{2k} s e^{-2s^2} \, d\ell_1 \cdots d\ell_{2k-1}, \qquad \text{with } s := \ell_1 + \cdots \ell_{2k-1}. \tag{175}$$

Recall that our normalisation differs from [4], where $T_{2\mathbf{e}}$ is used instead of $T_{\mathbf{e}}$, and thus all edges are twice as long; hence (175) is obtained from [4, (33)] by a trivial change of variables.

For $k = 3$ we have $|\bar{\mathfrak{T}}^3| = 3!! = 3$ different shapes. It is easy to verify that in each of them, we may label the five edges such that $d(\overline{U}_1, \overline{U}_2) - d(\overline{U}_3, o) = \ell_1 + \ell_2 - \ell_3 - \ell_4$. Consequently, (174) and (175) yield, with s as in (175),

$$\mathbb{E}\left|\widehat{\Lambda_{\mathbf{e}}}(\xi)\right|^2 = 3 \int_{\mathbb{R}_+^5} e^{i\xi(\ell_1+\ell_2-\ell_3-\ell_4)} 2^6 s e^{-2s^2} \, d\ell_1 \cdots d\ell_5$$

$$= 192 \int_0^\infty \psi(s) s e^{-2s^2} \, ds, \tag{176}$$

where, with $\Sigma_5(s) := \{(\ell_1, \ldots, \ell_5) : \ell_i > 0 \text{ and } \ell_1 + \cdots + \ell_5 = s\}$,

$$\psi(s) := \int_{\Sigma_5(s)} e^{i\xi(\ell_1+\ell_2-\ell_3-\ell_4)} \, d\ell_1 \cdots d\ell_4. \tag{177}$$

We can easily find the Laplace transform of $\psi(s)$: for $\lambda > 0$,

$$\int_0^\infty e^{-\lambda s} \psi(s) \, ds = \int_{\mathbb{R}_+^5} e^{-\lambda(\ell_1+\cdots+\ell_5)} e^{i\xi(\ell_1+\ell_2-\ell_3-\ell_4)} \, d\ell_1 \cdots d\ell_5$$

$$= \frac{1}{(\lambda - i\xi)^2 (\lambda + i\xi)^2 \lambda}. \tag{178}$$

A partial fraction expansion of (178) yields,

$$\int_0^\infty e^{-\lambda s} \psi(s) \, ds = \frac{a(\xi)}{\lambda} + \frac{b_+(\xi)}{\lambda - i\xi} + \frac{b_-(\xi)}{\lambda + i\xi} + \frac{c_+(\xi)}{(\lambda - i\xi)^2} + \frac{c_-(\xi)}{(\lambda + i\xi)^2}, \tag{179}$$

for some coefficients $a(\xi), b_\pm(\xi), c_\pm(\xi)$. It is easy to calculate these, but it suffices to note that (178) is homogeneous of degree -5 in (λ, ξ), and thus $a(\xi) = a\xi^{-4}$, $b_\pm(\xi) = b_\pm \xi^{-4}$ and $c_\pm(\xi) = c_\pm \xi^{-3}$ for some complex $a, b_\pm, c_\pm$. By inverting the Laplace transform (179), we find

$$\psi(s) = a\xi^{-4} + b_+\xi^{-4}e^{i\xi s} + b_-\xi^{-4}e^{-i\xi s} + c_+\xi^{-3}se^{i\xi s} + c_-\xi^{-3}se^{-i\xi s}. \tag{180}$$

Finally, we substitute (180) in (176). We define

$$h_m(s) := s^m e^{-2s^2} \mathbf{1}\{s > 0\}, \tag{181}$$

and obtain, recalling (168),

$$\mathbb{E}\left|\widehat{\Lambda_{\mathbf{e}}}(s)\right|^2 = 192\xi^{-4}\left(a\hat{h}_1(0) + b_+\hat{h}_1(\xi) + b_-\hat{h}_1(-\xi) + c_+\xi\hat{h}_2(\xi) + c_-\xi\hat{h}_2(-\xi)\right). \tag{182}$$

We have $\hat{h}_1(\xi) = O(1)$ since h_1 is integrable, and $\xi\hat{h}_2(\xi) = i\widehat{h_2'}(\xi) = O(1)$ since h_2 is differentiable with integrable derivative. Consequently, the result (170) follows. $\qquad\square$

Remark 15.7. It is easy to see that in the proof above, $a = 1$ and $\hat{h}_1(0) = \int_0^\infty s e^{-2s^2} \, ds = 1/4$; moreover, $\hat{h}_1(\xi) = O(\xi^{-2})$ and $\hat{h}_2(\xi) = O(\xi^{-3})$. Hence, the proof yields

$$\mathbb{E}\left|\widehat{\Lambda_{\mathbf{e}}}(\xi)\right|^2 = 48\xi^{-4} + O(\xi^{-6}). \tag{183}$$

Thus, the estimate in Lemma 15.6 is sharp. $\qquad\square$

Remark 15.8. The result (170) (or (183)) can also be obtained in the same way from (172). However, for $k = 4$ we have $|\bar{\mathfrak{T}}^4| = 5!! = 15$ different shapes that we have to consider, and they yield several different terms in (176). The leading term in the partial fraction expansion (179) will now be $a\xi^{-4}/\lambda^3$. $\qquad\square$

Proof of Theorem 15.5. Let $0 < \alpha < 3/2$. Then, by Lemma 15.6,

$$
\mathbb{E} \int_{-\infty}^{\infty} \left| |\xi|^{\alpha} \widehat{\Lambda_{\mathbf{e}}}(\xi) \right|^2 d\xi = 2 \int_0^{\infty} |\xi|^{2\alpha} \mathbb{E} \left| \widehat{\Lambda_{\mathbf{e}}}(\xi) \right|^2 d\xi
$$
$$
\leqslant C \int_0^1 \xi^{2\alpha} d\xi + C \int_1^{\infty} \xi^{2\alpha - 4} d\xi < \infty. \tag{184}
$$

Hence, a.s., the function $\Lambda_{\mathbf{e}}$ belongs to the (generalised) Sobolev space $\mathcal{L}_\alpha^2$ (also called potential space; many different notations exist), which is defined as the space of all functions $f \in L^2(\mathbb{R})$ such that

$$
\|f\|_{\mathcal{L}_\alpha^2}^2 := \|f\|_{L^2}^2 + \int_{-\infty}^{\infty} \left| |\xi|^{\alpha} \widehat{\Lambda_{\mathbf{e}}}(\xi) \right|^2 d\xi < \infty, \tag{185}
$$

see, e.g. [50, Chapter V] or [8, Chapter 6]. Furthermore, this Sobolev space equals the Besov space $B_\alpha^{2,2}$, see again [50, Chapter V (there denoted $\Lambda_\alpha^{2,2}$)] or [8, Chapter 6.] $B_\alpha^{2,2}$ is an L^2 version of $\mathcal{H}_\alpha$, and for $0 < \alpha < 1$, $B_\alpha^{2,2}$ may be defined as the set of functions in L^2 such that

$$
\|f\|_{B_\alpha^{2,2}}^2 := \|f\|_{L^2}^2 + \int_0^{\infty} \left(\frac{\|f(\cdot + u) - f(\cdot)\|_{L^2(\mathbb{R})}}{u^\alpha} \right)^2 \frac{du}{u} < \infty; \tag{186}
$$

the equivalence (within constant factors) of (186) with (185) follows easily from the Plancherel formula. (For a much more general result, see [8, Theorem 6.2.5] or [50, (V.60)].) For $1 < \alpha < 2$, (177)) fails, but $B_\alpha^{2,2} = \{f \in L^2 : f' \in B_{\alpha-1}^{2,2}\}$, which again equals $\mathcal{L}_\alpha^2$.

If $1 \leqslant \alpha < 3/2$, it follows that the derivative (in distribution sense) $\Lambda_{\mathbf{e}}' \in \mathcal{L}_{\alpha-1}^2$. By the Sobolev (or Besov) embedding theorem [8, Theorem 6.5.1], see also [50, Theorem V.2] and, for a simplified version, [14, Lemma 6.2], this implies $\Lambda_{\mathbf{e}}' \in L^p(\mathbb{R})$ with $1/p = 1/2 - (\alpha - 1)$. Since α may be chosen arbitrarily close to 3/2, it follows that $\Lambda_{\mathbf{e}}' \in L^p$ for every $p < \infty$. (Recall that $\Lambda_{\mathbf{e}}$, and thus $\Lambda_{\mathbf{e}}'$, has compact support, so small p is not a problem.) The continuous function $\Lambda_{\mathbf{e}}$ thus has a distributional derivative $\Lambda_{\mathbf{e}}'$ that is integrable, which implies that $\Lambda_{\mathbf{e}}$ is the integral of this derivative, so $\Lambda_{\mathbf{e}}$ is absolutely continuous, with a derivative in the usual sense a.e, which equals the distributional derivative $\Lambda_{\mathbf{e}}'$.

Finally, $\Lambda_{\mathbf{e}} \in \mathcal{H}_\alpha$ for $\alpha < 1$ by $\Lambda_{\mathbf{e}}' \in L^p$ (with $p = 1/(1 - \alpha)$) and Hölder's inequality, or directly by $\Lambda_{\mathbf{e}} \in \mathcal{L}_{\alpha+1/2}^2 = B_{\alpha+1/2}^{2,2}$ and the Besov embedding theorem [8, Theorem 6.5.1], [50, V.6.7], noting $\mathcal{H}_\alpha = B_\alpha^{\infty,\infty}$. □

Remark 15.9. For the Fourier transform $\widehat{L_{\mathbf{e}}}$, we have instead of (170) and (183) the analogous estimate

$$
\mathbb{E} |\widehat{L_{\mathbf{e}}}(\xi)|^2 \leqslant C(\xi^{-2} \wedge 1), \qquad -\infty < \xi < \infty, \tag{187}
$$

and, more precisely,

$$
\mathbb{E} \left| \widehat{L_{\mathbf{e}}}(\xi) \right|^2 = 4\xi^{-2} + O(\xi^{-4}). \tag{188}
$$

These are proved by the same argument as above, now with $\left(d(\overline{U}_1, o) - d(\overline{U}_2, o) \right)$ in the exponent in (173) and using (175) with $k = 2$; we omit the details. □

Note that the exponent of ξ is -2 in (188), but -4 in (174). Hence, $\widehat{\Lambda_{\mathbf{e}}}(\xi)$ decays faster than $\widehat{L_{\mathbf{e}}}(\xi)$ (at least in an average sense), which intuitively means that $\Lambda_{\mathbf{e}}$ is smoother than $L_{\mathbf{e}}$, as seen in the results above.

Note also that the result of Vervaat [51] mentioned in Remark 15.3 implies that $|\widehat{L_{\mathbf{e}}}(\xi)| = |\widehat{L_{\mathbf{b}}}(\xi)|$. The expectation $\mathbb{E}|\widehat{L_{\mathbf{b}}}(\xi)|^2$ can easily be calculated and estimated directly, since $\mathbf{b}$ is a

Gaussian process, which leads to another proof of (187) and (188); we leave the details to the reader. □

The proof above shows that

$$\Lambda_{\mathbf{e}} \in B_\alpha^{2,2}, \qquad \alpha < 3/2, \tag{189}$$

and thus

$$\Lambda_{\mathbf{e}}' \in B_\alpha^{2,2}, \qquad \alpha < 1/2. \tag{190}$$

More precisely, it follows from the proof that

$$\mathbb{E}\,\|\Lambda_{\mathbf{e}}'\|^2_{B_{\alpha-1}^{2,2}} \leqslant C\mathbb{E}\,\|\Lambda_{\mathbf{e}}\|^2_{B_\alpha^{2,2}} < \infty \qquad \text{for } \alpha < 3/2, \tag{191}$$

but not for any $\alpha \geqslant 3/2$.

For comparison, (188) implies that

$$\mathbb{E}\,\|L_{\mathbf{e}}\|^2_{B_\alpha^{2,2}} < \infty \qquad \text{for } \alpha < 1/2, \tag{192}$$

but not for $\alpha \geqslant \frac{1}{2}$. In this case, the range of α in (192) is thus the same as the range of Hölder continuity, note that since $\mathcal{H}_\alpha = B_\alpha^{\infty,\infty}$ for $\alpha \in (0,1)$, this range equals the set of $\alpha \in (0,1)$ such that $L_{\mathbf{e}} \in B_\alpha^{\infty,\infty}$. Another, similar, example is the Brownian bridge $\mathbf{b}$; a simple calculation shows that $\mathbb{E}\,\|\mathbf{b}\|^2_{B_\alpha^{2,2}} < \infty$ for $\alpha < \frac{1}{2}$, but not for larger α.

This suggests (but does not prove) that $\Lambda_{\mathbf{e}}$ is even smoother than shown by Theorem 15.5. However, the derivative $\Lambda_{\mathbf{e}}'$ is *not* Hölder continuous on $(-\infty, \infty)$, as might be guessed from the analogy between (191) and (192). In fact, the (two-sided) derivative $\Lambda_{\mathbf{e}}'$ does not exist at 0, and thus $\Lambda_{\mathbf{e}}'$ is not even continuous, at least not at 0.

Theorem 15.10. $\Lambda_{\mathbf{e}}$ *does not a.s. have a (two-sided) derivative at 0.*

To see this, we note first the following.

Lemma 15.11. *For every $x \geqslant 0$,*

$$\mathbb{E}\Lambda_{\mathbf{e}}(x) = \mathbb{E}L_{\mathbf{e}}(x) = 4x e^{-2x^2}. \tag{193}$$

Proof. The result for $L_{\mathbf{e}}$ is well known; it was shown by [15, (6.2)], and it is equivalent to the case $k = 1$ of (175).

The result for $\Lambda_{\mathbf{e}}$ follows by (24) and (23). □

Proof of Theorem 15.10. Suppose that $\Lambda_{\mathbf{e}}$ is differentiable at 0. Since $\Lambda_{\mathbf{e}}(x) = 0$ for $x < 0$, the derivative has to be 0, and thus $\Lambda_{\mathbf{e}}(x)/x \to 0$ as $x \searrow 0$. Furthermore, Theorem 14.6 shows that $\mathbb{E}\big(\Lambda_{\mathbf{e}}(x)/x\big)^2 \leqslant C$, and thus $\Lambda_{\mathbf{e}}(x)/x$ is uniformly integrable for $x > 0$. Consequently, if $\Lambda_{\mathbf{e}}$ were a.s. differentiable at 0, then $\mathbb{E}\big(\Lambda_{\mathbf{e}}(x)/x\big) \to 0$ as $x \searrow 0$, which contradicts Lemma 15.11. □

Nevertheless, it is quite possible that $\Lambda_{\mathbf{e}}$ is continuously differentiable on $[0, \infty)$, with a one-sided derivative at 0. We end with some open problems suggested by the results above.

Problem 15.12.

   (i) Is $\Lambda_{\mathbf{e}}$ a.s. Lipschitz, i.e., is the derivative $\Lambda_{\mathbf{e}}'$ a.s. bounded?

  (ii) Does $\Lambda_{\mathbf{e}}'(0)$ exist a.s. as a right derivative?

 (iii) Is the derivative $\Lambda_{\mathbf{e}}'$ a.s. continuous on $[0, \infty)$?

 (iv) Is the derivative $\Lambda_{\mathbf{e}}'$ a.s. in $\mathcal{H}_\alpha[0, \infty)$ for some $\alpha > 0$? For every $\alpha < \frac{1}{2}$?

15.2. Finite n

We have in this section so far considered only the asymptotic profiles L_e and Λ_e. Consider now the profile L_n and distance profile Λ_n of a conditioned Galton–Watson tree $\mathcal{T}_n^{\mathbf{p}}$, for a given offspring distribution $\mathbf{p}$. We assume throughout the section that $\mu(\mathbf{p}) = 1$ and that $0 < \sigma^2(\mathbf{p}) < \infty$; we will often add the condition that $\mathbf{p}$ has a finite fourth moment $\mu_4(\mathbf{p}) := \sum_k k^4 p_k$.

It was shown by Drmota and Gittenberger [18], see also [17, Theorem 4.24], that if the offspring distribution has an exponential moment, then

$$\mathbb{E}\left|n^{-1/2}L_n(xn^{1/2}) - n^{-1/2}L_n(yn^{1/2})\right|^4 \leqslant C|x-y|^2, \tag{194}$$

which by the standard argument in (147)–(148) yields

$$\mathbb{E}\,\|n^{-1/2}L_n(n^{1/2}\cdot)\|_{\mathcal{H}_\alpha[0,A]} < \infty \tag{195}$$

for $\alpha < 1/4$. This and Theorem 1.2 then yield (139) for $\alpha < 1/4$. We conjecture that (194) can be extended to higher moments, yielding (195) and thus another proof of (15.1) for all $\alpha < \frac{1}{2}$, but this seems to require some non-trivial work.

Moreover, as said in [17, footnote on p. 127], the proof of (194) does not really require an exponential moment, but it seems to require at least a fourth moment for the offspring distribution. We do not know whether such a moment condition really is necessary for (194), and we state the following problem.

Problem 15.13. Let $p \geqslant 2$. Is it true for any offspring distribution $\mathbf{p}$ with $\mu(\mathbf{p}) = 1$ and $\sigma^2(\mathbf{p}) < \infty$ then

$$\mathbb{E}\left|n^{-1/2}L_n(xn^{1/2}) - n^{-1/2}L_n(yn^{1/2})\right|^p \leqslant C_p|x-y|^{p/2}\,? \tag{196}$$

Is this true assuming a pth moment for $\mathbf{p}$? Assuming an exponential moment?

Remark 15.14. A rerooting argument as in the proof of Theorem 14.1 shows that (196) would imply

$$\mathbb{E}\left|n^{-3/2}\Lambda_n(xn^{1/2}) - n^{-3/2}\Lambda_n(yn^{1/2})\right|^p \leqslant C_p|x-y|^{p/2}. \tag{197}$$

Again we can ask under which conditions this holds.   □

We can also use generating functions and singularity analysis to estimate the Fourier transform of L_n and Λ_n. Recall that we in Section 2.1 have defined L_n and Λ_n as functions on $\mathbb{R}$, using linear interpolation. We first consider their restrictions to the integers, and the corresponding Fourier transforms, for $-\infty < \xi < \infty$,

$$\widehat{L_n^{\mathbb{Z}}}(\xi) := \sum_{k=0}^{\infty} L_n(k)e^{i\xi k} \quad \text{and} \quad \widehat{\Lambda_n^{\mathbb{Z}}}(\xi) := \sum_{k=0}^{\infty} \Lambda_n(k)e^{i\xi k}. \tag{198}$$

Note that these are periodic functions with period 2π, and that, as a consequence of (12) and (15),

$$|\widehat{L_n^{\mathbb{Z}}}(\xi)| \leqslant \widehat{L_n^{\mathbb{Z}}}(0) = n \quad \text{and} \quad |\widehat{\Lambda_n^{\mathbb{Z}}}(\xi)| \leqslant \widehat{\Lambda_n^{\mathbb{Z}}}(0) = n^2. \tag{199}$$

Lemma 15.15. *Let $\mathbf{p}$ be an offspring distribution with $\mu(\mathbf{p}) = 1$.*

(i) *If $\mathbf{p}$ has a finite second moment, then*

$$\mathbb{E}\,|\widehat{L_n^{\mathbb{Z}}}(\xi n^{-1/2})|^2 \leqslant Cn^2(\xi^{-2} \wedge 1), \qquad |\xi| \leqslant \pi n^{1/2}. \tag{200}$$

(ii) *If $\mathbf{p}$ has a finite fourth moment, then*

$$\mathbb{E}\,|\widehat{\Lambda_n^{\mathbb{Z}}}(\xi n^{-1/2})|^2 \leqslant Cn^4(\xi^{-4} \wedge 1), \qquad |\xi| \leqslant \pi n^{1/2}. \tag{201}$$

The proofs of (200) and (201) are based on lengthy calculations of some generating functions; we omit these and refer to the preprint version of the present paper [10, Appendix C].

We return to the interpolated profiles L_n and Λ_n and their Fourier transforms $\widehat{L_n}$ and $\widehat{\Lambda_n}$ defined by (168). It follows from (11) that

$$\widehat{L_n}(\xi) = \widehat{L_n^{\mathbb{Z}}}(\xi)\widehat{\tau}(\xi) = \widehat{L_n^{\mathbb{Z}}}(\xi)\frac{\sin^2(\xi/2)}{(\xi/2)^2} \tag{202}$$

and similarly

$$\widehat{\Lambda_n}(\xi) = \widehat{\Lambda_n^{\mathbb{Z}}}(\xi)\widehat{\tau}(\xi) = \widehat{\Lambda_n^{\mathbb{Z}}}(\xi)\frac{\sin^2(\xi/2)}{(\xi/2)^2}. \tag{203}$$

In particular, $|\widehat{L_n}(\xi)| \leqslant |\widehat{L_n^{\mathbb{Z}}}(\xi)|$ and $|\widehat{\Lambda_n}(\xi)| \leqslant |\widehat{\Lambda_n^{\mathbb{Z}}}(\xi)|$; hence, the estimates in Lemma 15.15 hold also for $\widehat{L_n}$ and $\widehat{\Lambda_n}$.

We have stated Theorems 1.2 and 1.4 with convergence (in distribution) in $C[0, \infty]$; however, it is obvious that then the same statement holds with convergence in $C[-\infty, \infty]$. Note that both sides of (1) and (3) are non-negative functions on $(-\infty, \infty)$ with integral 1. It is easily seen that in the set of such functions, convergence a.e. (and *a fortiori* convergence in $C[-\infty, \infty]$, i.e., uniform convergence) implies convergence in $L^1(\mathbb{R})$, and thus (uniform) convergence of the Fourier transforms. Furthermore, the Fourier transforms of the left-hand sides of (1) and (3) are $n^{-1}\widehat{L_n}(\xi n^{-1/2})$ and $n^{-2}\widehat{\Lambda_n}(\xi n^{-1/2})$, respectively. Hence, Theorems 1.2 and 1.4 imply that for any fixed ξ

$$n^{-1}\widehat{L_n}(\xi n^{-1/2}) \xrightarrow{\mathrm{d}} \widehat{L_{\mathbf{e}}}\left(\tfrac{2}{\sigma}\xi\right), \tag{204}$$

$$n^{-2}\widehat{\Lambda_n}(\xi n^{-1/2}) \xrightarrow{\mathrm{d}} \widehat{\Lambda_{\mathbf{e}}}\left(\tfrac{2}{\sigma}\xi\right). \tag{205}$$

Consequently, by Fatou's lemma, the estimates (200) and (201) in Lemma 15.15 yield another proof of the estimates (187) and (170) above.

Moreover, we can argue similarly to the proof of Theorem 15.5 and obtain a corresponding result on the smoothness of Λ_n. Of course, this profile is by construction Lipschitz, but what is relevant is a smoothness estimate that is uniform in n.

Theorem 15.16. *If the offspring distribution* **p** *has a finite fourth moment, then*

$$\mathbb{E}\left\| n^{-3/2}\Lambda_n\left(xn^{1/2}\right) \right\|_{\mathcal{H}_\alpha}^2 \leqslant C \tag{206}$$

uniformly in n, for every fixed $\alpha < 1$.

For the proof, we need a discrete version of (a simple case of) the Sobolev embedding theorem. We do not know an explicit reference, so for completeness, we state this and give a proof.

Lemma 15.17. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function initially defined on $\mathbb{Z}$ and extended by linear interpolation to $\mathbb{R}$. Suppose that $\sum_{-\infty}^{\infty} |f(k)| < \infty$, and define, as in (198), $\widehat{f^{\mathbb{Z}}}(\xi) := \sum_k f(k)e^{i\xi k}$. Let $0 < \alpha < 1$. Then*

$$\|f\|_{\mathcal{H}_\alpha} \leqslant C_\alpha\left(\int_{-\pi}^{\pi} |\xi|^{2\alpha+1}|\widehat{f^{\mathbb{Z}}}(\xi)|^2 \, d\xi\right)^{1/2}. \tag{207}$$

Proof. Let $j, k \in \mathbb{Z}$. Then, by Fourier inversion,

$$|f(j) - f(k)| = \left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(e^{-ij\xi} - e^{-ik\xi}\right)\widehat{f^{\mathbb{Z}}}(\xi) \, d\xi\right|$$

$$\leqslant \int_{0}^{\pi}\left(|j - k|\xi \wedge 1\right)\left|\widehat{f^{\mathbb{Z}}}(\xi)\right| d\xi. \tag{208}$$

Hence, using the Cauchy–Schwarz inequality,

$$|f(j) - f(k)|^2 \leqslant \int_0^\pi \left(|j-k|^2 \xi^2 \wedge 1\right) \xi^{-2\alpha-1} \, \mathrm{d}\xi \int_0^\pi \left|\widehat{f^{\mathbb{Z}}}(\xi)\right|^2 \xi^{2\alpha+1} \, \mathrm{d}\xi. \tag{209}$$

Writing $M := \int_0^\pi \left|\widehat{f^{\mathbb{Z}}}(\xi)\right|^2 \xi^{2\alpha+1} \, \mathrm{d}\xi$, and changing variables to $t := |j-k|\xi$ in the first integral (assuming $j \neq k$), we obtain

$$|f(j) - f(k)|^2 \leqslant |j-k|^{2\alpha} \int_0^\infty \left(t^2 \wedge 1\right) t^{-2\alpha-1} \, \mathrm{d}t \cdot M = CM|j-k|^{2\alpha} \tag{210}$$

for some $C < \infty$ (depending on α). Hence,

$$|f(x) - f(y)| \leqslant CM^{1/2}|x-y|^\alpha, \tag{211}$$

whenever x and y are integers. Since f is linear between integers, it follows that (211) holds for all $x, y \in \mathbb{R}$ (with another C), i.e., $\|f\|_{\mathcal{H}_\alpha} \leqslant CM^{1/2}$, as claimed in (207). $\qquad\square$

Proof of Theorem 15.16. Lemma 15.17 and a change of variables together with (201) yield

$$\mathbb{E}\,\|\Lambda_n\|_{\mathcal{H}_\alpha}^2 \leqslant C\mathbb{E} \int_{-\pi}^\pi |\xi|^{2\alpha+1} \left|\widehat{\Lambda_n^{\mathbb{Z}}}(\xi)\right|^2 \, \mathrm{d}\xi$$

$$= Cn^{-\alpha-1} \int_{-\pi n^{1/2}}^{\pi n^{1/2}} |\xi|^{2\alpha+1} \mathbb{E}\left|\widehat{\Lambda_n^{\mathbb{Z}}}(\xi n^{-1/2})\right|^2 \, \mathrm{d}\xi$$

$$\leqslant Cn^{3-\alpha} \int_{-\infty}^\infty \left(|\xi|^{2\alpha-3} \wedge |\xi|^{2\alpha+1}\right) \, \mathrm{d}\xi = Cn^{3-\alpha}, \tag{212}$$

since the final integral converges for $0 < \alpha < 1$. A change of variables yields (206). $\qquad\square$

We end this subsection with a couple of open problems suggested by Theorem 15.16.

The fourth moment condition in Theorem 15.16 is used in our proof (see [10]) of (201), but it seems likely that it can be weakened. Hence, we ask the following.

Problem 15.18. Does (206) hold assuming only a finite second moment for **p**?

Just as for the limit $\Lambda_{\mathbf{e}}$, we do not know whether (206) holds for $\alpha = 1$. Since Hölder(1) equals Lipschitz, and Λ_n is defined by linear interpolation, it is easy to see that

$$\left\|n^{-3/2}\Lambda_n\left(xn^{1/2}\right)\right\|_{\mathcal{H}_1} = n^{-1} \max_k |\Lambda_n(k+1) - \Lambda_n(k)|. \tag{213}$$

Consequently, the question whether (206) holds for $\alpha = 1$ is equivalent to the following. (Cf. (123).)

Problem 15.19. Assume that $\mu(\mathbf{p}) = 1$ and that **p** has a finite second moment (or fourth moment, or an exponential moment). Is then

$$\mathbb{E} \max_k |\Lambda_n(k+1) - \Lambda_n(k)|^2 \leqslant Cn^2 \,? \tag{214}$$

A slightly weaker version is

$$\mathbb{E} \max_k |\Lambda_n(k+1) - \Lambda_n(k)| \leqslant Cn \,? \tag{215}$$

16. Further remarks

Remark 16.1. In analogy with Sections 3 and 5, we might, as a third alternative, also define random unlabelled simply generated trees by using the weights (47) on the set $\mathfrak{U}_n$ of all unlabelled unrooted trees of order n. (These are defined as equivalence classes of labelled trees in $\mathfrak{L}_n$ under

isomorphisms.) It is usually more challenging to consider unlabelled trees; we conjecture that results similar to those above hold for unlabelled simply generated trees too, but we leave this as an open problem.

There is some existing work [48, 52] on yet another related model for random trees: *simply generated unrooted plane trees* (i.e., unlabelled unrooted trees where each vertex is endowed with a cyclic ordering of its neighbourhood). In particular, the results in [48, Sections 2 and 3] seem to be reminiscent of some of our results in Sections 3–5. Informally, through a series of approximations (using rooted/labelled/differently weighted versions) one ends up comparing those tree models to a pair of Galton–Watson trees (with the roots joined by an edge), conditioned on the sum of vertices. One of the two trees being typically macroscopic while the other is microscopic.   □

Remark 16.2. Another class of 'simply generated' trees is simply generated increasing trees, see [7] and [17, Section 1.3.3]. These random trees are quite different: They have (under weak conditions) logarithmic height; moreover, both the profile and distance profile are degenerate, in the sense that the distribution of depths or distances divided by $\log n$ converges to a point mass at some positive constant. With a more refined scaling, the profile and distance profile are both Gaussian, see Panholzer and Prodinger [45].   □

Remark 16.3. The degenerate behaviour seen in Remark 16.2, with almost all distances close to $c \log n$ for some constant $c > 0$, seems to be typical for many classes of random trees with logarithmic height, see [31].   □

Problem 16.4. We have in this paper assumed that the offspring distribution is critical and has finite variance. It would be interesting to extend the results to the case of infinite variance, with the offspring distribution in the domain of attraction of a stable variable. In this case, the tree converges in a suitable sense to a random stable tree [20, 37]; moreover, there is a corresponding limit result for the profile [5, 34]. However, there seems to be several technical challenges to adapt the arguments above to this case, and we leave it as an open problem.

References

[1] Addario-Berry, L., Devroye, L. and Janson, S. (2013) Sub-Gaussian tail bounds for the width and height of conditioned Galton–Watson trees. *Ann. Probab.* **41**(2) 1072–1087.

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

[3] Aldous, D. (1991) The continuum random tree II: an overview. In *Stochastic Analysis (Durham, 1990)*, Vol. 167 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, Cambridge, pp. 23–70.

[4] Aldous, D. (1993) The continuum random tree III. *Ann. Probab.* **21**(1) 248–289.

[5] Angtuncio, O. and Bravo, G. U. (2020) On the profile of trees with a given degree sequence. Preprint. arXiv:2008.12242

[6] Athreya, K. B. and Ney, P. E. (1972) *Branching Processes*. Springer-Verlag, Berlin.

[7] Bergeron, F., Flajolet, P. and Salvy, B. (1992) Varieties of increasing trees. In *CAAP'92 (Rennes, 1992)*, Vol. 581 of *Lecture Notes in Computer Science*. Springer, Berlin, pp. 24–48.

[8] Bergh, J. and Löfström, J. (1976) *Interpolation Spaces. An Introduction*. Springer-Verlag, Berlin-New York.

[9] Bertoin, J. and Miermont, G. (2013) The cut-tree of large Galton-Watson trees and the Brownian CRT. *Ann. Appl. Probab.* **23**(4) 1469–1493.

[10] Berzunza Ojeda, G. and Janson, S. (2020) The distance profile of rooted and unrooted simply generated trees. (Preliminary version of the present paper.) Preprint. arXiv:2009.00572.

[11] Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley, New York.

[12] Biane, P., Pitman, J. and Yor, M. (2001) Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bull. Amer. Math. Soc. (N.S.)* **38**(4) 435–465.

[13] Blumenthal, R. M. (1992) *Excursions of Markov Processes*. BirkhÄuser Boston, Inc., Boston, MA.
[14] Bousquet-Mélou, M. and Janson, S. (2006) The density of the ISE and local limit laws for embedded trees. *Ann. Appl. Probab.* **16**(3), 1597–1632.
[15] Chung, K. L. (1976) Excursions in Brownian motion. *Ark. Mat.* **14**(2) 155–177.
[16] Devroye, L. and Janson, S. (2011) Distances between pairs of vertices and vertical profile in conditioned Galton–Watson trees. *Random Struct. Algorithms* **38**(4) 381–395.
[17] Drmota, M. (2009) *Random Trees*. Springer-Verlag, Vienna.
[18] Drmota, M. and Gittenberger, B. (1997) On the profile of random trees. *Random Struct. Algorithms* **10**(4) 421–451.
[19] Drmota, M. and Gittenberger, B. (2004) The width of Galton–Watson trees conditioned by the size. *Discrete Math. Theor. Comput. Sci.* **6**(2) 387–400.
[20] Duquesne, T. (2003) A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.* **31**(2) 996–1027.
[21] Duquesne, T. and Le Gall, J.-F. (2005) Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields* **131**(4) 553–603.
[22] Duquesne, T. and Le Gall, J.-F. (2009) On the re-rooting invariance property of Lévy trees. *Electron. Commun. Probab.* **14** 317–326.
[23] Durrett, R. T., Iglehart, D. L. and Miller, D. R. (1977) Weak convergence to Brownian meander and Brownian excursion. *Ann. Probab.* **5**(1) 117–129.
[24] Dwass, M. (1969) The total progeny in a branching process and a related random walk. *J. Appl. Probab.* **6** 682–686.
[25] Flajolet, P. and Sedgewick, R. (2009) *Analytic Combinatorics*. Cambridge University Press, Cambridge, UK.
[26] Gut, A. (2013) *Probability: A Graduate Course*, 2nd ed. Springer, New York.
[27] Janson, S. (1994) Orthogonal decompositions and functional limit theorems for random graph statistics. *Mem. Am. Math. Soc.* **111**(534).
[28] Janson, S. (2003) The Wiener index of simply generated random trees. *Random Struct. Algorithms* **22**(4) 337–358.
[29] Janson, S. (2006) Random cutting and records in deterministic and random trees. *Random Struct. Algorithms* **29**(2) 139–179.
[30] Janson, S. (2012) Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probab. Surv.* **9** 103–252.
[31] Janson, S. Tree limits and limits of random trees. *Combin. Probab. Comput.* (2021+), First View.
[32] Kallenberg, O. (2002) *Foundations of Modern Probability*, 2nd ed. Springer, New York.
[33] Kennedy, D. P. (1976) The distribution of the maximum Brownian excursion. *J. Appl. Probab.* **13**(2) 371–376.
[34] Kersting, G. (1998) On the height profile of a conditioned Galton–Watson tree. Unpublished paper. arXiv:1101.3656
[35] Kesten, H. (1986) Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.* **22**(4) 425–487.
[36] Kolchin, V. F. 1984 *Random Mappings*. Nauka, Moscow (Russian). English transl.: Optimization Software, New York, 1986.
[37] Kortchemski, I. (2013) A simple proof of Duquesne's theorem on contour processes of conditioned Galton-Watson trees. In *Séminaire de Probabilités XLV*, Vol. 2078 of *Lecture Notes in Mathematics*, Springer, Cham, 537–558.
[38] Kortchemski, I. and Marzouk, C. (2016) Triangulating stable laminations. *Electron. J. Probab.* 21, Paper No. 11, 31 p.
[39] Le Gall, J.-F. (2005) Random trees and applications. *Probab. Surv.* **2** 245–311.
[40] Marckert, J.-F. and Mokkadem, A. (2006) Limit of normalized quadrangulations: the Brownian map. *Ann. Probab.* **34**(6) 2144–2202.
[41] Marckert, J.-F. and Panholzer, A. (2002) Noncrossing trees are almost conditioned Galton–Watson trees. *Random Struct. Algorithms* **20**(1) 115–125.
[42] Meir, A. and Moon, J. W. (1978) On the altitude of nodes in random trees. *Canad. J. Math.* **30**(5) 997–1015.
[43] Mörters, P. and Peres, Y. (2010) *Brownian Motion*. Cambridge University Press, Cambridge.
[44] Otter, R. (1949) The multiplicative process. *Ann. Math. Stat.* **20** 206–224.
[45] Panholzer, A. and Prodinger, H. (2004) Analysis of some statistics for increasing tree families. *Discrete Math. Theor. Comput. Sci.* **6**(2) 437–460.
[46] Petrov, V. V. (1975) *Sums of Independent Random Variables*. Springer-Verlag, Berlin.
[47] Pitman, J. (1999) The SDE solved by local times of a Brownian excursion or bridge derived from the height profile of a random tree or forest. *Ann. Probab.* **27**(1) 261–283.
[48] Ramzews, L. and Stufler, B. (2019) Simply generated unrooted plane trees. *ALEA Lat. Am. J. Probab. Math. Stat.* **16**(1) 333–359.
[49] Revuz, D. and Yor, M. (1999) *Continuous Martingales and Brownian Motion*, 3rd ed. Springer, Berlin.
[50] Stein, E. M. (1970) *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, Princeton, NJ.
[51] Vervaat, W. (1979) A relation between Brownian bridge and Brownian excursion. *Ann. Probab.* **7**(1) 143–149.
[52] Wang, M. (2016) Scaling limits for a family of unrooted trees. *ALEA Lat. Am. J. Probab. Math. Stat.* **13**(2) 1039–1067.

Appendix A. Tightness

We show here a general technical result used in the proof of Theorem 11.2, and thus of our main theorem.

We recall the standard criterion for tightness in $C[0, 1]$ in, e.g., [11, Theorem 8.2], which we formulate as follows, using the modulus of continuity defined in (9). (The equivalence of our formulation and the one in [11] is a simple exercise.)

Lemma A.1. *A sequence of random functions $(X_n(t))_1^\infty$ in $C[0, 1]$ is tight if and only if the following two conditions hold:*

  (i) *The sequence $X_n(0)$ of random variables is tight.*

  (ii) $\omega(\delta; X_n) \xrightarrow{\mathrm{P}} 0$ *as $\delta \to 0$, uniformly in n.*

The uniform convergence in probability in (ii) means that for every $\varepsilon, \eta > 0$, there exists $\delta > 0$ such that

$$\sup_n \mathbb{P}\big(\omega(\delta; X_n) \geqslant \varepsilon\big) \leqslant \eta. \tag{A1}$$

Equivalently, for every sequence $\delta_k \to 0$ and every sequence $(n_k)_1^\infty$,

$$\omega(\delta_k; X_{n_k}) \xrightarrow{\mathrm{P}} 0 \qquad \text{as } k \to \infty. \tag{A2}$$

We use this to prove the following general lemma on tightness of averages.

Lemma A.2. *Suppose that $(X_n(t))_1^\infty$ is a sequence of random functions in $C[0, 1]$, that $(N_n)_1^\infty$ is an arbitrary sequence of positive integers, and that for each n, $(X_{ni}(t))_{i=1}^{N_n}$ is a, possibly dependent, family of identically distributed random functions with $X_{ni} \overset{\mathrm{d}}{=} X_n$. Assume that*

  (i) *The sequence (X_n) is tight in $C[0, 1]$.*

  (ii) *The sequence of random variables*

$$\|X_n\| := \sup_{t\in[0,1]} |X_n(t)| \tag{A3}$$

     is uniformly integrable.

Then the sequence of averages

$$Y_n(t) := \frac{1}{N_n} \sum_{i=1}^{N_n} X_{ni}(t) \tag{A4}$$

is tight in $C[0, 1]$.

Remark A.3. We state Lemma A.2 for $C[0, 1]$; it transfers immediately to, e.g., $C[0, \infty]$. It follows also that the lemma holds for $C[0, \infty)$, with (ii) replaced by the assumption that $\sup_{[0,b]} |X_n|$ over any finite interval [0,b] is uniformly integrable. □

Proof. Let $\delta_k \to 0$ and let n_k be arbitrary positive integers. Then (A2) holds by assumption (i) and Lemma A.1.

Furthermore, for any $\delta > 0$, we obviously have, by (9), $\omega(\delta; X_n) \leqslant 2\|X_n\|$. In particular, $\omega(\delta_k; X_{n_k}) \leqslant 2\|X_{n_k}\|$, and it follows from (ii) that the sequence $\omega(\delta_k; X_{n_k})$, $k \geqslant 1$, is uniformly integrable. Hence, (A2) implies

$$\mathbb{E}\omega(\delta_k; X_{n_k}) \to 0, \qquad \text{as } k \to \infty. \tag{A5}$$

Next, it follows from the definitions (A4) and (9) that, for any $\delta > 0$,

$$\omega(\delta; Y_n) = \frac{1}{N_n}\omega\Big(\delta; \sum_{i=1}^{N_n} X_{ni}\Big) \leqslant \frac{1}{N_n}\sum_{i=1}^{N_n}\omega\big(\delta; X_{ni}\big) \tag{A6}$$

and hence,

$$\mathbb{E}\omega(\delta; Y_n) \leqslant \frac{1}{N_n}\sum_{i=1}^{N_n}\mathbb{E}\omega\big(\delta; X_{ni}\big) = \mathbb{E}\omega\big(\delta; X_n\big). \tag{A7}$$

Consequently, by (A5) and (A7),

$$\mathbb{E}\omega(\delta_k; Y_{n_k}) \leqslant \mathbb{E}\omega(\delta_k; X_{n_k}) \to 0, \qquad \text{as } k \to \infty. \tag{A8}$$

Since the sequences $(\delta_k)_k$ and $(n_k)_k$ are arbitrary, this shows that the sequence $(Y_n)_n$ satisfies (A2) and thus condition Lemma A.1(ii).

Furthermore, condition Lemma A.1(i) holds too, because by (A4), (A3) and assumption (ii),

$$\mathbb{E}|Y_n(0)| \leqslant \frac{1}{N_n}\sum_{i=1}^{N_n}\mathbb{E}|X_{ni}(0)| = \mathbb{E}|X_n(0)| \leqslant \mathbb{E}\,\|X_n\| = O(1), \tag{A9}$$

which implies that the sequence $Y_n(0)$ is tight. Hence, the conclusion follows by Lemma A.1. $\square$

Remark A.4. The assumption on uniform integrability in Lemma A.2(ii) cannot be weakened to $\mathbb{E}\,\|X_n\| = O(1)$. For an example, let

$$X_n(t) := \xi_n n(1 \wedge nt), \qquad t \geqslant 0, \tag{A10}$$

where $\xi_n \sim Be(1/n)$, i.e., $\mathbb{P}(\xi_n = 1) = 1/n$ and otherwise $\xi_n = 0$. Then $\mathbb{E}\,\|X_n\| = 1$. On the other hand, take $N_n := n$ and let $X_{ni} \stackrel{d}{=} X_n$ be independent, $1 \leqslant i \leqslant N_n$. Then, for every n, with probability $e^{-1} + o(1)$ exactly one X_{ni} is not identically 0, and in this case $Y_n = (1 \wedge nt)$ and thus $\omega(1/n; Y_n) = 1$. Hence, $\mathbb{P}\big(\omega(1/n; Y_n) = 1\big) \geqslant e^{-1} + o(1)$ and thus (A2) and Lemma A.1(ii) do not hold for (Y_n); thus, the sequence Y_n is not tight. $\square$