Combinatorics, Probability and Computing (2021), **30**, pp. 849–893
doi:10.1017/S0963548321000055

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Tree limits and limits of random trees

Svante Janson†

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden
Email: svante.janson@math.uu.se

(Received 28 May 2020; revised 12 February 2021; first published online 31 March 2021)

Abstract
We explore the tree limits recently defined by Elek and Tardos. In particular, we find tree limits for many classes of random trees. We give general theorems for three classes of conditional Galton–Watson trees and simply generated trees, for split trees and generalized split trees (as defined here), and for trees defined by a continuous-time branching process. These general results include, for example, random labelled trees, ordered trees, random recursive trees, preferential attachment trees, and binary search trees.

2020 MSC Codes: 60C05, 05C05; 05C12, 60J80.

1. Introduction

Elek and Tardos [20] have recently introduced a theory of tree limits, in analogy with the theory of graph limits [39] and other similar limits of various combinatorial objects (e.g. hypergraphs, permutations, …). Their idea is to regard a tree as a metric space with a probability measure; the metric is the usual graph distance, suitably rescaled, and the probability measure is the uniform measure on the vertices. Then, for each integer r, consider the random matrix $(d(\xi_i, \xi_j))_{i,j=1}^r$ of distances between r random vertices $\xi_1, \dots, \xi_r$. A sequence of trees is said to converge if, for each $r \geqslant 1$, the resulting random $r \times r$ matrices converge in distribution. (This type of convergence for metric spaces with a measure goes back to Gromov [22, Chapter $3\frac{1}{2}$].) See Section 3 for details of this and of other topics mentioned below.

Elek and Tardos [20] choose to normalize the metrics of the trees by dividing the graph distance by the diameter; hence the trees become metric spaces with diameter 1. One reason for this normalization is that this embeds the trees in a compact space, and thus every sequence of trees has a convergent subsequence. However, the theory developed in [20] treats also more general real trees, and include trees with a different normalization. We will in general not use the Elek–Tardos normalization, since other scalings often seem more natural, in particular for random trees, see e.g. Examples 7.2 and 7.3, and Sections 9 and 10.

The main results by Elek and Tardos [20] are that there exists a set of limit objects called *dendrons* such that each convergent sequence of finite trees (with their normalization) converges to a unique dendron. The dendrons can be regarded as real trees equipped with probability measures, but the precise definition is slightly different. The dendrons are defined as special cases of *long dendrons*; dendrons are long dendrons with diameter at most 1. This is tied to the Elek–Tardos normalization, and in the present paper, the main limit objects are the long dendrons.

†Supported by the Knut and Alice Wallenberg Foundation

In some cases, the (long) dendrons can be identified with real trees, and the tree limits then coincide with limits in the Gromov–Prohorov metric (see Remark 3.6). Such limits have been studied earlier (also in the stronger Gromov–Hausdorff–Prohorov metric); one much studied example going back to Aldous [5, 6, 7] is provided by conditioned Galton–Watson trees, see Section 9. However the Elek–Tardos limits are more general, and include also other types of limits; one example is provided by a different class of conditioned Galton–Watson trees (where condensation appears in the limit), see Section 10.

The tree limits by Elek and Tardos [20] thus seem to be very interesting, and promising for future research. The purpose of the present paper is to further develop the theory of tree limits. We give some general results in Sections 3–6. In particular, we show how the set of all tree limits, or equivalently the set of all long dendrons, can be regarded as a metric (and Polish) space; this makes it possible to define and study random tree limits and limits of random trees in a convenient way. Moreover, we characterize relative compactness of a sequence (or set) of rescaled trees (Theorem 6.1 and Corollary 6.3).

The second, and perhaps main, part of the paper applies the general theory to several classes of random trees and finds tree limits for them. As a preparation, we give in Section 7 some simple examples of limits of deterministic trees. A few general results on limits of random trees are given in Section 8.

The following sections study first different classes of conditioned Galton–Watson tree and simply generated trees (Sections 9–11), and then different classes of random trees with logarithmic height (Sections 12–14), in particular split trees (Section 13) and trees defined by continuous time branching processes (Section 14). We find tree limits in all these cases, as the size $n \to \infty$; in some cases with convergence in distribution to a random tree limit, and in others with convergence in probability to a fixed tree limit.

The found limits are of different types. In particular, for a class of conditioned Galton–Watson trees including many standard classes of random trees with height of order $\sqrt{n}$ (Section 9), the well-known limit theorem by Aldous [7] gives convergence to a random real tree known as the the Brownian continuum random tree; this tree can be regarded as a (random) long dendron, and Aldous's result holds in the present sense too.

On the other hand, many standard classes of random trees with height of order $\log n$ are covered by the general results in Sections 12–14 and have tree limits of a quite different type; these limits are long dendrons of a very simple type (but distinct from real trees), which is equivalent to the fact that in these trees, almost all pairs of vertices have almost the same distance. (This is shown more generally in Theorem 8.2.)

Remark 1.1. The tree limits by Elek and Tardos [20] studied in the present paper are global limits, in general quite different from local limits studied in e.g. [29]. Nevertheless, there are cases (see for example Section 10) where the trees are such that there is a strong relation between the tree limits and local limits.

2. Some notation

Probability measures

Recall that a Polish space is a separable completely metrizable topological space. In other words, it can be regarded as a complete separable metric space, but we ignore the metric. (When necessary or convenient, we can choose a metric, but there is no distinguished one.) A Polish space is often regarded as a measurable space, equipped with its Borel σ-field.

If $X = (X, \mathcal{F})$ is a measurable space, then $\mathcal{P}(X)$ is the space of probability measures on X. In particular, if X is a metric space, then $\mathcal{P}(X)$ is the space of Borel probability measures on X, and in

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

this case we equip $\mathcal{P}(X)$ with the standard weak topology, see e.g. [11]. If X is a Polish space, then so is $\mathcal{P}(X)$, see [11, Appendix III] or [12, Theorem 8.9.5].

The Dirac measure (unit point mass) at a point x is denoted δ_x.

If μ is a probability measure on a space X, then $\xi \sim \mu$ and $\mu = \mathcal{L}(\xi)$ both denote that ξ is a random element of X with distribution μ.

If $X = (X, \mathcal{F})$ and $Y = (Y, \mathcal{G})$ are measurable spaces, $\varphi: X \to Y$ is a measurable map, and $\mu \in \mathcal{P}(X)$, then the *push-forward* $\varphi(\mu) \in \mathcal{P}(Y)$ of μ is defined by

$$\varphi(\mu)(A) := \mu\big(\varphi^{-1}(A)\big), \qquad A \in \mathcal{G}. \tag{2.1}$$

(This is often denoted $\mu \circ \varphi^{-1}$ or $\varphi_*(\mu)$.) Equivalently, if ξ is a random element of X then

$$\xi \sim \mu \implies \varphi(\xi) \sim \varphi(\mu). \tag{2.2}$$

Limits

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

As usual, w.h.p. (*with high probability*) means with probability tending to 1 as a parameter (here always n) tends to ∞.

If Z, Z_n are random elements of a metric space X, then $Z_n \xrightarrow{\mathrm{d}} Z$, $Z_n \xrightarrow{\mathrm{p}} Z$, and $Z_n \xrightarrow{\mathrm{a.s.}} Z$ denote convergence in distribution, in probability and almost surely (a.s.), respectively. Note that $Z_n \xrightarrow{\mathrm{d}} Z$ is the same as convergence in $\mathcal{P}(X)$ of the distributions, i.e., $\mathcal{L}(Z_n) \to \mathcal{L}(Z)$.

If $(a_n)_n$ is a sequence of positive numbers, then $o_{\mathrm{p}}(a_n)$ denotes a sequence of random variables Z_n such that $Z_n/a_n \xrightarrow{\mathrm{p}} 0$; this is equivalent to $|Z_n|/a_n < \varepsilon$ w.h.p. for every $\varepsilon > 0$.

Miscellaneous

If T is a tree, we abuse notation and write T for its vertex set $V(T)$. The number of vertices is denoted by $|T|$. If T is a rooted tree, then the root is denoted by o.

If $x, y \in \mathbb{R}$, then $x \wedge y := \min\{x, y\}$. On the other hand, if v and w are vertices in a rooted tree, then $v \wedge w$ denotes their last common ancestor.

For a sequence of random variables, i.i.d. means independent and identically distributed.

$\mathbb{N} := \{1, 2, \dots\}$ and $\mathbb{N}_0 := \{0, 1, 2, \dots\}$.

C and c denote positive constants; these may vary from one occurrence to another. (We sometimes distinguish them by subscripts.)

3. Convergence of trees and long dendrons

We give here a summary of the main definitions and results of [20], together with some further notation.

3.1 Convergence of trees

For $r \geqslant 1$, let M_r be the space of real $r \times r$ matrices; note that $M_r = \mathbb{R}^{r^2}$ is a Polish space, and thus $\mathcal{P}(M_r)$ is a Polish space.

For a set X with a given function $d: X^2 \to \mathbb{R}$, and $r \geqslant 1$, let $\rho_r: X^r \to M_r$ be the map given by the entries

$$\rho_r(x_1, \dots, x_r)_{ij} = \rho_r(x_1, \dots, x_r; X, d)_{ij} := \begin{cases} d(x_i, x_j), & i \neq j, \\ 0, & i = j. \end{cases} \tag{3.1}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

We often consider ρ_r when d is a metric on X; then the special definition in (3.1) when $i=j$ is redundant. However, for the long dendrons defined below, we typically have $d(x,x)>0$, and then the definition (3.1) is important. See also Remark 3.10.

A *metric measure space* is a triple (X,d,μ), where X is a measurable space (so $X=(X,\mathcal{F})$ with $\mathcal{F}$ hidden in the notation), $\mu\in\mathcal{P}(X)$, and $d:X^2\to\mathbb{R}$ is a measurable metric on X.

Suppose, more generally, that $X=(X,\mathcal{F},\mu)$ is a probability space and that $d:X^2\to\mathbb{R}$ is a measurable function. For $r\geqslant 1$, define the sampling measure

$$\tau_r(X)=\tau_r(X,d,\mu):=\rho_r(\mu^r)\in\mathcal{P}(M_r), \tag{3.2}$$

the push-forward of the measure $\mu^r\in\mathcal{P}(X^r)$ along ρ_r. In other words, if $\xi_1,\dots,\xi_r$ are i.i.d. random points in X with $\xi_i\sim\mu$, then

$$\tau_r(X):=\mathcal{L}\big(\rho_r(\xi_1,\dots,\xi_r;X)\big), \tag{3.3}$$

the distribution of the random matrix $\rho_r(\xi_1,\dots,\xi_r)\in M_r$.

A finite tree T is regarded as a metric space (T,d_T), where d_T is the graph distance. Furthermore, if $c>0$, we let cT denote the metric space (T,cd_T), where all distances are rescaled by c. We regard cT as a metric probability space by equipping it with the uniform measure μ_T defined by $\mu_T\{x\}=1/|T|$ for $x\in T$. Then $\tau_r(cT)\in\mathcal{P}(M_r)$ is defined by (3.2).

Definition 3.1. Let $(T_n)_1^\infty$ be a sequence of finite trees and $(c_n)_1^\infty$ a sequence of positive numbers. Then the sequence $(c_nT_n)_1^\infty$ converges if the sampling measures converge for every fixed r, i.e., if there exist $\lambda_r\in\mathcal{P}(M_r)$ such that, as $n\to\infty$,

$$\tau_r(c_nT_n)=\tau_r\big(T_n,c_nd_{T_n},\mu_{T_n}\big)\to\lambda_r \qquad \text{in } \mathcal{P}(M_r),\ r\geqslant 1. \tag{3.4}$$

By (3.3), the condition (3.4) is equivalent to convergence in distribution of the random matrices $\rho_r(\xi_1^{(n)},\dots,\xi_r^{(n)};c_nT_n)$, where for each n, $(\xi_i^{(n)})_i$ are i.i.d. uniform random vertices of T_n.

Remark 3.2. As said in the introduction, Elek and Tardos [20] consider only the normalization $c_n=1/\operatorname{diam}(T_n)$, but we will not assume this.

A *real tree* is a complete non-empty metric space (T,d) such that for any pair of distinct points $x,y\in T$, there exists a unique isometric map $\alpha:[0,d(x,y)]\to T$ with $\alpha(0)=x$ and $\alpha(d(x,y))=y$, and furthermore, for every $s\in(0,d(x,y))$, x and y are in different components of $T\setminus\{\alpha(s)\}$. (There are several different but equivalent versions of the definition; see e.g. [17, 18, 36, 37].)

Remark 3.3. Note that we define the trees as complete (as do [20]); this is often not required. For our purposes completeness is convenient and no real loss of generality; if T is an incomplete real tree (defined as above without completeness), then the completion $\overline{T}$ is also a real tree (see e.g. [17, Theorem 8]), and in the limit theory below we can use $\overline{T}$ instead of T.

If $T=(T,d)$ is a real tree and $c>0$, let $cT:=(T,cd)$. Then cT is also a real tree.

A *measured real tree* is a real tree $T=(T,d)$ equipped with a probability measure μ. We will only consider separable trees T and Borel measures μ, and then (T,d,μ) is always a metric measure space. (For non-separable measured real trees, see [20], where they e.g. are used in the proofs; then μ might be defined on a smaller σ-field than the Borel one, and the condition that d has to be measurable is added. See also Remark 3.10.)

Example 3.4. *If T is any finite tree (in the usual combinatorial sense), let $\hat{T}$ denote the real tree obtained by regarding each edge in T as an interval of length 1. Then $\hat{T}$ is a compact real tree, and T is isometrically embedded as a subset of $\hat{T}$. Hence, we can regard μ_T as a probability measure on*

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

$\hat{T}$, and $(\hat{T}, \mu_T) = (\hat{T}, d, \mu_T)$ is a measured real tree. Obviously, $\tau_r(T) = \tau_r(\hat{T}, \mu_T)$. More generally, cT is isometrically embedded in $c\hat{T}$ for any $c > 0$, and

$$\tau_r(cT) = \tau_r(c\hat{T}, \mu_T). \tag{3.5}$$

We can therefore sometimes identify $\hat{T}$ and T; see Section 5.

Consequently, we can regard Definition 3.1 as a special case of the following definition.

Definition 3.5. Let $(T_n)_1^\infty = (T_n, d_n, \mu_n)_1^\infty$ be a sequence of measured real trees. Then the sequence $(T_n)_1^\infty$ converges if the sampling measures converge for every fixed r, i.e., if there exist $\lambda_r \in \mathcal{P}(M_r)$ such that, as $n \to \infty$,

$$\tau_r(T_n) \to \lambda_r \qquad \text{in } \mathcal{P}(M_r), r \geqslant 1. \tag{3.6}$$

Again, (3.6) is equivalent to convergence in distribution of the random matrices $\rho_r(\xi_1^{(n)}, \dots, \xi_r^{(n)}; T_n)$, where, for each n, $(\xi_i^{(n)})_i$ are i.i.d. random points in T_n with $\xi_i^{(n)} \sim \mu_n$.

Remark 3.6. Gromov [22, Chapter $3\frac{1}{2}$] studied general complete separable metric measure spaces (with a finite Borel measure, which we may normalize to be a probability measure as above). He defined the Gromov–Prohorov metric (see Villani [50, p. 762] for another version, and Löhr [38] for the equivalence), and he also considered convergence in the sense above, i.e., $\tau_r(X_n) \to \tau_r(X)$ for every r, where X_n and X are metric measure spaces; it turns out that this is equivalent to convergence in the Gromov–Prohorov metric, see Greven, Pfaffelhuber and Winter [21]; see also [31]. Gromov [22, $3\frac{1}{2}$.14 and $3\frac{1}{2}$.18] noted also that it is possible that $\tau_r(X_n)$ converges for every r to some limit measure, but that there is no metric measure space X that is the limit. (One of Gromov's examples is the sequence of unit spheres S^n with uniform measure, which behave as in Theorem 3.13 below with almost all distances being almost equal; a metric measure space limit would have to have almost all distances equal to some positive constant, which is impossible for separable spaces.) The new idea by Elek and Tardos [20] is to define another type of limit object (long dendrons) that works in general when X_n are trees.

3.2 Long dendrons

Elek and Tardos [20] defined limit objects as follows. Note that a real tree T is locally connected (and locally pathwise connected); thus, if $p \in T$, then $T \setminus \{p\}$ is the disjoint union of one or several (possibly infinitely many) open connected components; these are called *p-branches*. A *branch* of T is a p-branch for some $p \in T$.

Definition 3.7. A *long dendron* $D = (T, d, \nu)$ is a real tree (T, d) together with a (Borel) probability measure ν on $A_D := T \times [0, \infty)$ satisfying $\nu(B \times [0, \infty)) > 0$ for every branch B of T. We define $d_D: A_D^2 \to [0, \infty)$ by

$$d_D\big((x, a), (y, b)\big) := d(x, y) + a + b. \tag{3.7}$$

An isomorphism between two long dendrons $D = (T, d, \nu)$ and $D' = (T', d', \nu')$ is an isometry f from (T, d) onto (T', d') such that the mapping $\bar{f} := (p, a) \mapsto (f(p), a)$ is measure-preserving $(A_D, \nu) \to (A_{D'}, \nu')$.

We call the real tree T the *base* of the long dendron D; we may identify T with $T \times \{0\} \subset A_D$. It is shown in [20, Lemma 6.2] that the base T of a dendron necessarily is separable; thus T and A_D are Polish spaces.

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Remark 3.8. Elek and Tardos [20] also define a *dendron* as a long dendron such that if $\xi_1, \xi_2 \in A_D$ are i.i.d. random points with distribution ν, then $d_D(\xi_1, \xi_2) \leqslant 1$ a.s. These are the limit objects for real trees with diameter $\leqslant 1$, and thus for trees with the Elek–Tardos normalization in Remark 3.2, but they have no special importance in the present paper. We may call them *short dendrons*. (For consistency with [20], we keep the name long dendron, although for our purposes it would be more natural to change notation and call them dendrons.)

For a long dendron $D = (T, d, \nu)$, we use again (3.2)–(3.3) and define the sampling measure

$$\tau_r(D) := \tau_r(A_D, d_D, \nu) \in \mathcal{P}(M_r), \tag{3.8}$$

i.e., the distribution of the random matrix $\rho_r(\xi_1, \dots, \xi_r; A_D, d_D) \in M_r$ if $\xi_1, \dots, \xi_r$ are i.i.d. random points in A_D with $\xi_i \sim \nu$.

Convergence of finite or real trees to a long dendron is defined by adding to Definitions 3.1 and 3.5 that the limits of the sampling measures are the sampling measures for the limit.

Definition 3.9. Let $(T_n)_1^\infty$ be a sequence of finite trees and $(c_n)_1^\infty$ a sequence of positive numbers, and let D be a long dendron. Then the sequence $(c_n T_n)_1^\infty$ converges to D if, as $n \to \infty$,

$$\tau_r(c_n T_n) \to \tau_r(D) \qquad \text{in } \mathcal{P}(M_r),\ r \geqslant 1. \tag{3.9}$$

Similarly, if $(T_n)_1^\infty$ is a sequence of real trees and D a long dendron, then T_n converges to D if, as $n \to \infty$,

$$\tau_r(T_n) \to \tau_r(D) \qquad \text{in } \mathcal{P}(M_r),\ r \geqslant 1. \tag{3.10}$$

Again, (3.9) and (3.10) are equivalent to convergence in distribution of the random matrices $\rho_r(\xi_1^{(n)}, \dots, \xi_r^{(n)})$, where $(\xi_i^{(n)})_i$ are i.i.d. as above.

Remark 3.10. As discussed in [20, Remark 4], the long dendrons could be replaced by real trees as follows. (We might think of long dendrons as proxies for some measured real trees.) Let $D = (T, d, \nu)$ be a long dendron. First, if $\nu(\{t\} \times [0, \infty)) = 0$ for every $t \in T$, consider $A_D = T \times [0, \infty)$ as a real tree T' consisting of $T = T \times \{0\}$ with a half-line $\{t\} \times [0, \infty)$ attached at each $t \in T$. In general, we have to attach a continuum of half-lines at each point t (so that each half-line has measure 0), for example by defining $T' := T \times \mathbb{C}$ regarded as T with the half-lines $\{(t, re^{i\theta}) : r \geqslant 0\}$ attached, for every $t \in T$ and $\theta \in [0, 2\pi)$, and with the measure ν' on T' equal to the push-forward by the map $(t, r, \theta) \mapsto (t, re^{i\theta})$ of the measure $\nu \times d\theta/2\pi$. Note that then $\tau_k(T') = \tau_k(D)$ for every $k \geqslant 1$. (Note how the special definition for $i = j$ in (3.1) interacts with (3.7) to give the desired result.)

However, we agree with Elek and Tardos [20] that it is more convenient to use long dendrons as limit objects. One reason is that the trees just constructed are nonseparable, and that the measures are not Borel measures on T'. (There are plenty of nonmeasurable open sets.) Another reason is that long dendrons provide uniqueness of the limits in a simple way.

3.3 Two examples

The following examples of long dendrons are rather simple, and extreme in the sense that the measure ν on $A_D = T \times [0, \infty)$ is supported on a 'one-dimensional' set with one of the coordinates fixed. Nevertheless, these two examples will play the main role in our limit theorems for random trees.

Example 3.11. *Let $T = (T, d, \mu)$ be a measured real tree such that every branch has positive measure. Identify $T \times \{0\}$ with T, and define ν as μ regarded as a measure on $T \times \{0\} \subset$*

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

$A_D := T \times [0,\infty)$. (More formally, ν is the push-forward of μ under $x \mapsto (x,0) \in A_D$.) Then (T,d,ν) is a long dendron.

Note that if $\xi \sim \mu$, then $(\xi,0) \sim \nu$. Since (3.7) implies that d_D equals d on $T \times \{0\} = T$, it follows that $\tau_r(D) = \tau_r(T)$ for every $r \geqslant 1$. We may thus identify the long dendron D with the measured real tree T.

By Remark 3.6, convergence of a sequence of (real) trees to D as in Definition 3.9 is equivalent to convergence to T in the Gromov–Prohorov metric.

Example 3.12. *Let $T = \{\bullet\}$ be the real tree consisting of a single point. Then the metric $d = 0$, and we let $\mu = \delta_\bullet$ (the only probability measure on T, so there is no choice).*

We may identify $A_D = \{\bullet\} \times [0,\infty)$ with $[0,\infty)$. Thus every probability measure ν on $[0,\infty)$ defines a long dendron $\Upsilon_\nu := (T,d,\nu)$.

By (3.1) and (3.7),

$$\rho_r(\xi_1,\ldots,\xi_r;\Upsilon_\nu) = \big((\xi_i+\xi_j)\mathbf{1}\{i\neq j\}\big)_{i,j=1}^r, \tag{3.11}$$

and thus $\tau_r(\Upsilon_\nu)$ is the distribution of the matrix (3.11) when $(\xi_i)_i$ are i.i.d. with $\xi_i \sim \nu$.

A particularly simple, and important, case is when $\nu = \delta_a$ for some $a \geqslant 0$. In this case we denote the long dendron by Υ_a, and note that $\xi_i = a$ is non-random, and thus (3.11) shows that $\rho_r(\xi_1,\ldots\xi_r)$ is the constant matrix

$$\rho_r(\xi_1,\ldots,\xi_r;\Upsilon_a) = \big(2a\mathbf{1}\{i\neq j\}\big)_{i,j=1}^r. \tag{3.12}$$

This leads to the following simple characterization of convergence to the long dendron Υ_a.

Theorem 3.13. *Let $(c_nT_n)_n$ be a sequence of rescaled trees, and $a \geqslant 0$. Then $c_nT_n \to \Upsilon_a$ if and only if*

$$c_nd_n(\xi_1^{(n)},\xi_2^{(n)}) \overset{\mathrm{p}}{\longrightarrow} 2a, \tag{3.13}$$

where d_n is the graph distance in T_n and $(\xi_i^{(n)})_i$ are i.i.d. uniformly random vertices in T_n.

The same holds, mutatis mutandis, for a sequence (T_n,d_n,μ_n) of measured real trees.

In the terminology of Gromov [22, p. 142], (3.13) says that c_nT_n have *(asymptotic) characteristic size* $2a$.

Proof. Convergence in distribution to a constant is the same as convergence in probability. Thus, (3.3), (3.1) and (3.12) show that Definition 3.9 now yields

$$\begin{aligned}
&c_nT_n \to \Upsilon_a \\
&\iff \big(c_nd_n(\xi_i^{(n)},\xi_j^{(n)})\mathbf{1}\{i\neq j\}\big)_{i,j=1}^r \overset{\mathrm{p}}{\longrightarrow} \big(2a\mathbf{1}\{i\neq j\}\big)_{i,j=1}^r, \qquad r\geqslant 1,\\
&\iff c_nd_n(\xi_i^{(n)},\xi_j^{(n)})\mathbf{1}\{i\neq j\} \overset{\mathrm{p}}{\longrightarrow} 2a\mathbf{1}\{i\neq j\}, \qquad i,j\geqslant 1.
\end{aligned} \tag{3.14}$$

By symmetry, it suffices to consider the case $i = 1, j = 2$. □

Remark 3.14. The (long) dendron Υ_0 is trivial, with $d_D(\xi_1,\xi_2) = 0$ a.s. if $\xi_i \sim \nu = \delta_0$. Note that Υ_0 equals the equally trivial real tree $T = \{\bullet\}$ consisting of a single point, regarded as a long dendron as in Example 3.11.

The trivial long dendron Υ_0 is by Theorem 3.13 the limit of c_nT_n when

$$c_nd_n(\xi_1^{(n)},\xi_2^{(n)}) \overset{\mathrm{p}}{\longrightarrow} 0, \tag{3.15}$$

which typically means that we have chosen the wrong rescaling.

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

3.4 Limit theorems

Some of the main results of Elek and Tardos [20] are the following, here somewhat reformulated.

Theorem 3.15. **([20, partly Theorems 1 and 4]).** *Any convergent sequence of rescaled finite trees converges to some long dendron. The same holds for any convergent sequence of measured real trees.*

This is not stated in quite this generality in [20]; we show in Section 15 how it follows from other results in [20]. (We postpone this proof until the end of the paper because it uses arguments from [20] quite different from the other arguments in the present paper.)

Theorem 3.16. **([20, Theorem 2, Lemmas 7.1 and 7.2]).** *Any long dendron is the limit of a convergent sequence $(c_nT_n)_1^\infty$ of rescaled finite trees.*

Again, this is not stated in quite this form in [20], but it is a simple consequence of [20, Lemmas 7.1 and 7.2]; we omit the details.

Theorem 3.17. **([20, Theorem 3]).** *Two long dendrons D and D' are isomorphic if and only if $\tau_r(D)=\tau_r(D')$ for every $r\geqslant 1$. Consequently, the limit of a sequence of real trees (or rescaled finite trees) is unique (up to isomorphism) if it exists.*

4. Infinite matrices

We extend the definitions in Section 3 to the case $r=\infty$, i.e. to infinite matrices. Let M_∞ be the space of infinite real matrices $(a_{ij})_{i,j=1}^\infty$. Define ρ_r and τ_r by (3.1) and (3.2)–(3.3) also for $r=\infty$; thus $\tau_\infty(X,d,\mu)$ is the distribution of the infinite random matrix $\big(d(\xi_i,\xi_j)\mathbf{1}\{i\neq j\}\big)_{i,j\geqslant 1}$ where ξ_i are i.i.d. with $\xi_i\sim\mu$.

Given any $A=(a_{ij})_{i,j=1}^s\in M_s$, with $r\leqslant s\leqslant\infty$, define the restriction

$$\Pi_r(A)=(a_{ij})_{i,j=1}^r\in M_r, \tag{4.1}$$

i.e., the $r\times r$ top left corner of A. Furthermore, if $A\in M_s$ is a random matrix with distribution $\nu\in\mathcal{P}(M_s)$, we denote the distribution of $\Pi_r(A)$ by $\Pi_r(\nu)\in\mathcal{P}(M_r)$. (This is the push-forward of ν, see (2.1).) In other words, $\Pi_r(\nu)$ is the marginal distribution of the $r\times r$ top left corner.

Say that a sequence $\lambda_r\in\mathcal{P}(M_r)$, $1\leqslant r<\infty$, is *consistent* if $\Pi_r(\lambda_s)=\lambda_r$ when $r\leqslant s$.

If $\lambda\in\mathcal{P}(M_\infty)$, then the sequence $\lambda_r:=\Pi_r(\lambda)$ is obviously consistent. Conversely, every consistent sequence arises in this way for a unique $\lambda\in\mathcal{P}(M_\infty)$; the corresponding statement for distributions of random vectors in $\mathbb{R}^\infty$ is well-known [34, Theorem 6.14], and the result for M_∞ follows immediately by reading the entries of the matrices in a suitable fixed order. Furthermore, if $\lambda,\lambda_n\in M_\infty$, then

$$\lambda_n\to\lambda \text{ in } \mathcal{P}(M_\infty) \iff \Pi_r(\lambda_n)\to\Pi_r(\lambda) \text{ in } \mathcal{P}(M_r) \text{ for each } r\geqslant 1. \tag{4.2}$$

Again, this follows immediately from the corresponding well-known fact for $\mathbb{R}^\infty$ [11, p. 19].

A sequence $\tau_r(X)$, $r\geqslant 1$, given by (3.3) is obviously consistent; furthermore, $\tau_r(X)=\Pi_r(\tau_\infty(X))$ for every r. Consequently, (4.2) implies the following.

Theorem 4.1. *Let $(T_n)_1^\infty=(T_n,d_n,\mu_n)_1^\infty$ be a sequence of measured real trees, and let $D=(T,d,\nu)$ be a long dendron.*

(i) *The sequence $(T_n)_1^\infty$ converges if and only if there exists $\lambda\in\mathcal{P}(M_\infty)$ such that, as $n\to\infty$,*

$$\tau_\infty(T_n)\to\lambda \qquad \textit{in } \mathcal{P}(M_\infty), \tag{4.3}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

i.e., if and only if the infinite random matrices $\rho_\infty(\xi_1^{(n)}, \xi_2^{(n)}, \ldots; T_n)$ converge in distribution, where $\xi_i^{(n)}$ are i.i.d. random points in T_n with $\xi_i^{(n)} \sim \mu_n$.

(ii) *The sequence $(T_n)_1^\infty$ converges to D if and only if as $n \to \infty$,*

$$\tau_\infty(T_n) \to \tau_\infty(D) \qquad \text{in } \mathcal{P}(M_\infty), \tag{4.4}$$

i.e., if and only if the infinite random matrices $\rho_\infty(\xi_1^{(n)}, \xi_2^{(n)}, \ldots; T_n)$ converge in distribution to $\rho_\infty(\xi_1, \xi_2, \ldots; D)$, where $\xi_i^{(n)}$ are as in (i) and ξ_i are i.i.d. with $\xi_i \sim \nu$.

In particular, the same results holds for a sequence $(c_n T_n)_1^\infty$ of rescaled finite trees.

Proof. This follows from the remarks before the theorem. Note that if (3.6) holds for every $r \geqslant 1$, then $(\lambda_r)_r$ is a consistent sequence, since $(\tau_r(T_n))_r$ is for every n. □

5. Abstract tree limits

Based on the preceding section, we can define tree limits in an abstract way as follows, using only (part of) the definitions and elementary considerations above and none of the deep results of [20]. (Cf. [15] for graph limits.)

Let $\mathcal{T}_\mathsf{f}$ be the set of all rescaled finite trees cT (with arbitrary $c > 0$). Then $\tau_\infty : \mathcal{T}_\mathsf{f} \to \mathcal{P}(M_\infty)$. Let $\mathfrak{T}_\mathsf{f} := \tau_\infty(\mathcal{T}_\mathsf{f}) \subseteq \mathcal{P}(M_\infty)$ and

$$\mathfrak{T} := \overline{\mathfrak{T}_\mathsf{f}} = \overline{\tau_\infty(\mathcal{T})} \subseteq \mathcal{P}(M_\infty). \tag{5.1}$$

This defines $\mathfrak{T}$ as a closed subset of the Polish space $\mathcal{P}(M_\infty)$; thus $\mathfrak{T}$ is a Polish space. Hence, we can regard $\mathfrak{T}$ as a (complete and separable) metric space whenever convenient; if necessary we can define a metric of $\mathfrak{T}$ e.g. as the Prohorov metric on $\mathcal{P}(M_\infty)$ [11, Appendix III], [12, Theorem 8.3.2], but we have in the present paper no need for a specific choice of metric.

We can identify a rescaled finite tree cT with its image $\tau_\infty(cT) \in \mathfrak{T}$ (temporarily ignoring the question whether this is a one-to-one map). Then convergence as in Definition 3.1 is, by Theorem 4.1, the same as convergence in the metric space $\mathfrak{T}$. Furthermore, $\mathfrak{T}$ is the set of all possible limits of convergent sequences; thus it is natural to say that $\mathfrak{T}$ is the set of *tree limits*.

We have thus defined a set of tree limits; moreover, this set has turned out to be a Polish space.

Similarly, a measured real tree T defines an element $\tau_\infty(T) \in \mathcal{P}(M_\infty)$. We define $\mathcal{T}_\mathsf{r}$ as the set of all measured real trees and $\mathfrak{T}_\mathsf{r} := \tau_\infty(\mathcal{T}_\mathsf{r}) \subset \mathcal{P}(M_\infty)$. (We ignore the set-theoretic difficulty of defining the "set of all measurable real trees"; formally we either consider trees that are subsets of some huge universe, or suitable equivalence classes under isomorphisms.) Then the following holds.

Theorem 5.1. *With notations as above,*

$$\mathfrak{T}_\mathsf{f} \subseteq \mathfrak{T}_\mathsf{r} \subseteq \mathfrak{T} = \overline{\mathfrak{T}_\mathsf{f}} = \overline{\mathfrak{T}_\mathsf{r}}. \tag{5.2}$$

We postpone the proof. It follows that convergence of measured real trees as in Definition 3.5 also is the same as convergence in $\mathfrak{T}$. From now on, whenever convenient, we identify finite trees and measured real trees with their images in $\mathfrak{T}$.

Returning to the deep results by Elek and Tardos [20] in Theorems 3.15–3.17, we first note that, similarly, each long dendron D defines an element $\tau_\infty(D) \in \mathcal{P}(M_\infty)$. Theorem 3.17 and the remarks in Section 4 show that $\tau_\infty(D) = \tau_\infty(D')$ if and only if D and D' are isomorphic. Thus, letting $\mathfrak{D}$ be the set of all equivalence classes of long dendrons modulo isomorphism, $\tau_\infty : \mathfrak{D} \to \mathcal{P}(M_\infty)$ is injective.

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Theorem 5.2. *$\tau_\infty(\mathfrak{D}) = \mathfrak{T}$, and the mapping $\tau_\infty : \mathfrak{D} \to \mathfrak{T}$ is a bijection.*

Proof. If D is a long dendron, then by Theorem 3.16, there exists a convergent sequence of rescaled finite trees $(c_n T_n)_n$ that converges to D. In other words, $\tau_\infty(c_n T_n) \to \tau_\infty(D)$. Thus $\tau_\infty(D) \in \overline{\tau_\infty(\mathcal{T}_\mathsf{f})} = \mathfrak{T}$.

Conversely, if $\mu \in \mathfrak{T}$, then there exists a sequence $c_n T_n \in \mathcal{T}_\mathsf{f}$ such that $\tau_\infty(c_n T_n) \to \mu$. Thus the sequence $c_n T_n$ is convergent, and by Theorem 3.15, there exists a long dendron D such that $c_n T_n \to D$, which means $\tau_\infty(c_n T_n) \to \tau_\infty(D)$. Consequently, $\mu = \tau_\infty(D)$.

Hence, $\tau_\infty(\mathfrak{D}) = \mathfrak{T}$, and we have already remarked that τ_∞ is injective on $\mathfrak{D}$ by Theorem 3.17. □

Consequently, we can identify $\mathfrak{D}$ and $\mathfrak{T}$, and regard also $\mathfrak{D}$ as the set of all tree limits. (As done by Elek and Tardos [20].) Note that this defines a topology on $\mathfrak{D}$, making $\mathfrak{D}$ into a Polish space.

We ignore the taking of equivalence classes, and regard $\mathfrak{D}$ as the set of all long dendrons. Thus, the topology on $\mathfrak{D}$ gives a notion of convergence for long dendrons.

Theorem 5.3. *Let $D = (T, d, \nu)$ and $D_n = (T_n, d_n, \nu_n)$, $n \geqslant 1$ be long dendrons. Then the following are equivalent.*

(i) *$D_n \to D$ in $\mathfrak{D}$.*
(ii) *$\tau_\infty(D_n) \to \tau_\infty(D)$ in $\mathcal{P}(M_\infty)$.*
(iii) *The infinite random matrices $\rho_\infty(\xi_1^{(n)}, \xi_2^{(n)}, \dots; D_n)$ converge in distribution to $\rho_\infty(\xi_1, \xi_2, \dots; D)$, where $\xi_i^{(n)}$ are i.i.d. random points in A_{D_n} with $\xi_i^{(n)} \sim \mu_n$ and ξ_i are i.i.d. random points in A_D with $\xi_i \sim \nu$.*
(iv) *$\tau_r(D_n) \to \tau_r(D)$ in $\mathcal{P}(M_r)$, for every $r \geqslant 1$.*
(v) *The finite random matrices $\rho_r(\xi_1^{(n)}, \xi_2^{(n)}, \dots; D_n)$ converge in distribution to $\rho_r(\xi_1, \xi_2, \dots; D)$ for every $r \geqslant 1$, where $\xi_i^{(n)}$ and ξ_i are as in (iii).*

Proof. Immediate by the definitions and comments before the theorem together with (4.2). □

Summarizing, we may thus regard finite trees, real trees, and long dendrons as elements of the Polish space $\mathfrak{T} \subset \mathcal{P}(M_\infty)$. This gives a unified meaning to convergence of trees and real trees to a long dendron, and also a notion of convergence of long dendrons.

We turn to the question whether τ_∞ is injective (up to obvious isomorphisms) on the sets $\mathcal{T}_\mathsf{f}$ of finite trees and $\mathcal{T}_\mathsf{r}$ of measured real trees; recall that for long dendrons, this is answered (positively) by Theorem 3.17. Gromov [22, $3\frac{1}{2}$.5 and $3\frac{1}{2}$.7] studied a more general setting and proved that if $X_1 = (X_1, d_1, \mu_1)$ and $X_2 = (X_2, d_2, \mu_2)$ are two separable and complete metric measure spaces such that the measures have full support, and $\tau_\infty(X_1) = \tau_\infty(X_2)$, then X_1 and X_2 are isomorphic. This applies immediately to rescaled finite trees, and it follows that if $c_1 T_1$ and $c_2 T_2$ are rescaled trees with $\tau_\infty(c_1 T_1) = \tau_\infty(c_2 T_2)$, then $T_1 \cong T_2$ as metric spaces, and thus as trees, and $c_1 = c_2$. (Except in the trivial case $|T_1| = |T_2| = 1$, when c_1 and c_2 are arbitrary.) In other words, $\tau_\infty : \mathcal{T}_\mathsf{f} \to \mathfrak{T}$ is injective up to isomorphism.

For measured real trees (T, d, μ), this is not quite true, since it may happen that μ is concentrated on a subtree $T' \subset T$, and then $\tau_\infty(T, d, \mu) = \tau_\infty(T', d, \mu)$. However, if $\mathcal{T}_\mathsf{c}$ is the set of measured real trees such that every branch has positive measure, then τ_∞ is injective on $\mathcal{T}_\mathsf{c}$ (up to isomorphism). One way to see that is to note that every $T = (T, d, \mu) \in \mathcal{T}_\mathsf{c}$ may be regarded as a long dendron as in Example 3.12, and then use Theorem 3.17.

In general, given a measured real tree T, we may prune branches of measure 0 and obtain a subtree $T' \in \mathcal{T}_\mathsf{c}$; this is called the *core* of T in [20], where a detailed definition is given. We see that the mapping τ_∞ does not distinguish between a measured real tree T and its core T'.

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

In other words, our identification of measured real trees with tree limits in $\mathfrak{T}$ means that we ignore branches of measure 0, and thus identify a tree with its core, but trees with different cores are distinguished. With some care, we may thus also regard measured real trees as elements of $\mathfrak{T}$.

One important consequence of regarding trees, measured real trees and long dendrons as elements of the Polish space $\mathfrak{T}$ is that then standard theory (e.g. [11]) defines for us random trees, random measured real trees and random long dendrons, as well as convergence in probability or distribution of such random objects. This will be a central topic in the remainder of the paper.

First, however, it remains to prove Theorem 5.1.

Proof of Theorem 5.1. First, as explained in Example 3.4, a rescaled finite tree $cT \in \mathcal{T}_\mathsf{f}$ can be embedded in a measured real tree $\hat{cT} \in \mathcal{T}_\mathsf{r}$ such that (3.5) holds for all finite r, and thus also for $r = \infty$. This proves $\mathfrak{T}_\mathsf{f} \subseteq \mathfrak{T}_\mathsf{r}$.

Recalling (5.1), it remains only to show $\mathfrak{T}_\mathsf{r} \subseteq \mathfrak{T}$.

We give first a short proof using the results of [20]. If T is a measured real tree, then the constant sequence $T, T, \ldots$ trivially is convergent, and thus Theorem 3.15 shows that there exists a long dendron D such that $T \to D$, which by Theorem 5.3 means $\tau_\infty(T) = \tau_\infty(D)$. Hence, by Theorem 5.2,

$$\tau_\infty(T) = \tau_\infty(D) \in \tau_\infty(\mathfrak{D}) = \mathfrak{T}. \tag{5.3}$$

We give also an alternative, elementary proof. We do this in several steps. We consider for simplicity, as said earlier, only separable trees.

Step 1. As in [20], say that a measured real tree is a *finite real tree* if it can be obtained from a finite tree by regarding each edge as an interval of some positive length (not necessarily the same for all edges), and adding a probability measure on the (finite) set of vertices. Note that a finite real tree has finite diameter. [20, Lemma 7.2] shows that every finite real tree T of diameter $\leqslant 1$ is a limit of rescaled finite trees, i.e., $T \in \mathfrak{T}$. By rescaling, the same holds for every finite real tree.

Step 2. Suppose that $T = (T, d, \mu)$ is a measured real tree such that μ is concentrated on a finite set of points $\{x_1, \ldots, x_m\}$. Let $T' := \bigcup_{i=1}^m [x_1, x_i]$ be the subtree spanned by $\{x_1, \ldots, x_m\}$. Then (T', d, μ) is a finite real tree. Furthermore, $\tau_\infty(T) = \tau_\infty(T', d, \mu) \in \mathfrak{T}$, using Step 1.

Step 3. Suppose that $T = (T, d, \mu)$ is a measured real tree such that μ is concentrated on a countable set of points $\{x_1, x_2, \ldots\}$. Let

$$\mu_n := \sum_{k=1}^{n} \mu\{x_k\}\delta_{x_k} + \Bigl(\sum_{k=n+1}^{\infty} \mu\{x_k\}\Bigr)\delta_{x_1}, \tag{5.4}$$

where δ_x is the point mass at x. Let ξ_i be i.i.d. with $\xi_i \sim \mu$, and let

$$\xi_i^{(n)} := \begin{cases} \xi_i & \text{if } \xi_i \in \{x_1, \ldots, x_n\}, \\ x_1 & \text{otherwise.} \end{cases} \tag{5.5}$$

Then $(\xi_i^{(n)})_i$ are i.i.d. random points in T with $\xi_i^{(n)} \sim \mu_n$. Furthermore, $\mathbb{P}(\xi_i^{(n)} \neq \xi_i) \to 0$ as $n \to \infty$, and thus $\rho_r(\xi_1^{(n)}, \ldots, \xi_r^{(n)}) \xrightarrow{\text{p}} \rho_r(\xi_1, \ldots, \xi_r)$ for each $r \geqslant 1$. Hence, $\tau_r(T, d, \mu_n) \to \tau_r(T, d, \mu)$ for every finite r, and thus also $\tau_\infty(T, d, \mu_n) \to \tau_\infty(T, d, \mu)$. Since $\tau_\infty(T, d, \mu_n) \in \mathfrak{T}$ by Step 2, it follows that $\tau_\infty(T, d, \mu) \in \mathfrak{T}$.

Step 4. Let $T = (T, d, \mu)$ be any separable tree. There exists a countable dense subset $A := \{x_1, x_2, \ldots\}$.

For each $n \geqslant 1$, define a measurable function $f_n : T \to A$ such that $d(x, f_n(x)) < 1/n$ for all x. (For example, let $f_n(x) := x_i$ for the smallest i such that $d(x, x_i) < 1/n$.) Let ξ_i be i.i.d. random

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

points in T with $\xi_i \sim \mu$, and let $\xi_i^{(n)} := f_n(\xi_i)$. Then $(\xi_i^{(n)})_i$ are i.i.d. with $\xi_i^{(n)} \sim \mu_n := f_n(\mu)$, which is concentrated on the countable set A. By Step 3, $\tau_\infty(T, d, \mu_n) \in \mathfrak{T}$ for every n. Furthermore,

$$\left|d(\xi_i^{(n)}, \xi_j^{(n)}) - d(\xi_i, \xi_j)\right| \leqslant d(\xi_i^{(n)}, \xi_i) + d(\xi_j^{(n)}, \xi_j) < 2/n \tag{5.6}$$

for every i and j, and and thus $\rho_r(\xi_1^{(n)}, \dots, \xi_r^{(n)}) \xrightarrow{\text{a.s.}} \rho_r(\xi_1, \dots, \xi_r)$ as $n \to \infty$ for each $r \geqslant 1$. Hence, $\tau_r(T, d, \mu_n) \to \tau_r(T, d, \mu)$ for every finite r, and thus also $\tau_\infty(T, d, \mu_n) \to \tau_\infty(T, d, \mu)$. Consequently, $\tau_\infty(T, d, \mu) \in \mathfrak{T}$. □

6. Compactness

Recall that a set S in a metric space X is *relatively compact* if every sequence in S has a convergent subsequence. (This is equivalent to $\overline{S}$ being compact.)

Recall also that a family $\{Z_\alpha : \alpha \in \mathcal{A}\}$ of random variables in a metric space X is *tight* if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subseteq X$ such that $\mathbb{P}(Z_\alpha \notin K_\varepsilon) < \varepsilon$ for every $\alpha \in \mathcal{A}$. In this case we also say that the family of distributions $\{\mathcal{L}(Z_\alpha)\}$ is tight.

Prohorov's theorem [11, Section 6] says that for a Polish space X, the set of distributions $\{\mathcal{L}(Z_\alpha)\}$ is relatively compact in $\mathcal{P}(X)$ if and only if $\{Z_\alpha\}$ is tight. In particular, this leads to the following characterization of relative compactness in $\mathfrak{T}$.

Theorem 6.1. *Let $A = \{c_\alpha T_\alpha : \alpha \in \mathcal{A}\}$ be a set of rescaled trees. Then the following are equivalent, where $(\xi_i^{(\alpha)})_i$ are i.i.d. uniformly random vertices in T_α and d_α is the graph distance in T_α.*

(i) *A is relatively compact.*
(ii) *The set of measures $\{\tau_\infty(c_\alpha T_\alpha) : \alpha \in \mathcal{A}\} \subseteq \mathcal{P}(M_\infty)$ is tight.*
(iii) *The set of random variables $\{\rho_\infty(\xi_1^{(\alpha)}, \xi_2^{(\alpha)} \dots; c_\alpha T_\alpha) : \alpha \in \mathcal{A}\}$ in M_∞ is tight.*
(iv) *The set of random variables $\{c_\alpha d_\alpha(\xi_1^{(\alpha)}, \xi_2^{(\alpha)}) : \alpha \in \mathcal{A}\}$ is tight.*
(v) *There exists $x_\alpha \in T_\alpha$, $\alpha \in \mathcal{A}$, such that the set of random variables $\{c_\alpha d_\alpha(\xi_1^{(\alpha)}, x_\alpha) : \alpha \in \mathcal{A}\}$ is tight.*

The same holds, mutatis mutandis, for sets of measured real trees $\{(T_\alpha, d_\alpha, \mu_\alpha)\}$, and for sets of long dendrons $\{(T_\alpha, d_\alpha, \nu_\alpha)\}$; in these cases, $\xi_i^{(\alpha)} \sim \mu_\alpha$ and $\xi_i^{(\alpha)} \sim \nu_\alpha$, respectively, and for long dendrons we use d_D defined in (3.7).

Proof. (i) $\iff$ (ii) $\iff$ (iii): Prohorov's theorem, together with the definition of convergence and Theorem 4.1.

(iii) $\implies$ (iv): Immediate by (3.1), since the mapping $(a_{ij})_{i,j} \mapsto a_{1,2}$ is continuous $M_\infty \to \mathbb{R}$.

(iv) $\implies$ (iii): Follows by symmetry and the fact that M_∞ has the product topology. To be more precise, let $\varepsilon > 0$. By (iv), there exist constants C_k, $k \geqslant 0$, such that $\mathbb{P}\big(c_\alpha d_\alpha(\xi_1^{(\alpha)}, \xi_2^{(\alpha)}) > C_k\big) < 2^{-k}\varepsilon$ for every α. Then $K := \{(a_{ij})_{i,j} : |a_{ij}| \leqslant C_{i+j}\}$ is a compact subset of M_∞, and, by symmetry,

$$\begin{aligned}\mathbb{P}\big(\rho_\infty(\xi_1^{(\alpha)}, \xi_2^{(\alpha)} \dots; c_\alpha T_\alpha) \notin K\big) &\leqslant \sum_{i,j=1}^{\infty} \mathbb{P}\big(c_\alpha d_\alpha(\xi_i^{(\alpha)}, \xi_j^{(\alpha)}) > C_{i+j}\big) \\ &< \sum_{i,j=1}^{\infty} 2^{-i-j}\varepsilon = \varepsilon.\end{aligned} \tag{6.1}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

(iv) $\Longrightarrow$ (v): If C_ε is such that $\mathbb{P}\big(c_\alpha d_\alpha(\xi_1^{(\alpha)}, \xi_2^{(\alpha)}) > C_\varepsilon\big) < \varepsilon$, then (by Fubini's theorem), there exists x_α^ε such that $\mathbb{P}\big(c_\alpha d_\alpha(\xi_1^{(\alpha)}, x_\alpha^\varepsilon) > C_\varepsilon\big) < \varepsilon$. It suffices to consider $\varepsilon < 1/2$. It then follows that

$$\mathbb{P}\big(c_\alpha d_\alpha(\xi_1^{(\alpha)}, x_\alpha^\varepsilon) \leqslant C_\varepsilon \text{ and } c_\alpha d_\alpha(\xi_1^{(\alpha)}, x_\alpha^{1/2}) \leqslant C_{1/2}\big) > 1 - \varepsilon - 1/2 > 0 \tag{6.2}$$

and thus the events $\{c_\alpha d_\alpha(\xi_1^{(\alpha)}, x_\alpha^\varepsilon) \leqslant C_\varepsilon\}$ and $\{c_\alpha d_\alpha(\xi_1^{(\alpha)}, x_\alpha^{1/2}) \leqslant C_{1/2}\}$ are not disjoint. Hence, $c_\alpha d_\alpha(x_\alpha^\varepsilon, x_\alpha^{1/2}) \leqslant C_\varepsilon + C_{1/2}$. Consequently,

$$\mathbb{P}\big(c_\alpha d_\alpha(\xi_1^{(\alpha)}, x_\alpha^{1/2}) > 2C_\varepsilon + C_{1/2}\big) \leqslant \mathbb{P}\big(c_\alpha d_\alpha(\xi_1^{(\alpha)}, x_\alpha^\varepsilon) > C_\varepsilon\big) < \varepsilon, \tag{6.3}$$

and thus we may choose $x_\alpha := x_\alpha^{1/2}$.

(v) $\Longrightarrow$ (iv): If C_ε is such that $\mathbb{P}\big(c_\alpha d_\alpha(\xi_1^{(\alpha)}, x_\alpha) > C_\varepsilon\big) < \varepsilon/2$, then $\mathbb{P}\big(c_\alpha d_\alpha(\xi_1^{(\alpha)}, \xi_2^{(\alpha)}) > 2C_\varepsilon\big) < \varepsilon$. □

Definition 6.2. A set of rescaled trees, measured real trees, or long dendrons, is *tight* if (iv) (or, equivalently, (v)) in Theorem 6.1 holds.

With this definition, Theorem 6.1 simply says that a set of rescaled trees, measured real trees or long dendrons is relatively compact if and only if it is tight. Usually, we consider sequences rather than general sets, and then Theorem 6.1 has the following corollary.

Corollary 6.3. *If $(c_n T_n)_n$ is a tight sequence of rescaled trees, then some subsequence converges to some long dendron.*

The same holds for tight sequences of measured real trees and for tight sequences of long dendrons.

7. Simple examples

As a preparation for the study of limits of random trees in the following sections, we give here a few simple examples of limits of deterministic trees.

Example 7.1 (paths). *Let P_n be the path with n vertices. We may take the vertices to be $\{1, \ldots, n\}$, and then $d_{P_n}(x, y) = |x - y|$.*

Let $I = (I, d, \mu)$ be the unit interval $I := [0, 1]$ considered as a measured real tree with the usual metric d and Lebesgue measure μ. We regard I as a long dendron as in Example 3.11, and claim that $\frac{1}{n}P_n \to I$.

To see this, let $(\xi_i)_i$ be i.i.d. with $\xi_i \sim \mu$, and let $\xi_i^{(n)} := \lceil n\xi_i \rceil$. Then $(\xi_i^{(n)})_i$ are i.i.d. uniform vertices of P_n, and $\frac{1}{n}\xi_i^{(n)} \to \xi_i$ as $n \to \infty$. Hence, (3.1) shows that

$$\rho_r\big(\xi_1^{(n)}, \ldots, \xi_r^{(n)}; \tfrac{1}{n}P_n\big) \xrightarrow{\text{a.s.}} \rho_r\big(\xi_1, \ldots, \xi_r; I\big) \tag{7.1}$$

for every r. This implies convergence in distribution, and thus $\tau_r\big(\frac{1}{n}P_n\big) \to \tau_r(I)$, and thus

$$\frac{1}{n}P_n \to I. \tag{7.2}$$

The diameter of P_n is $n-1$. Obviously, we obtain the same limit I if we use the Elek–Tardos normalization $\frac{1}{n-1}P_n$. (Note that the limit I is a short dendron.)

Example 7.2 (stars). *Let $S_n = K_{n-1,1}$ be a star with n vertices. If $(\xi_i^{(n)})_i$ are i.i.d. random vertices in S_n, then with probability $1 - O(1/n)$, $\xi_1^{(n)}$ and $\xi_2^{(n)}$ are distinct peripheral vertices, and thus $d(\xi_1^{(n)}, \xi_2^{(n)}) = 2$. Hence, $d(\xi_1^{(n)}, \xi_2^{(n)}) \xrightarrow{\text{p}} 2$ as $n \to \infty$, and thus Theorem 3.13 shows that*

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

$$S_n \to \Upsilon_1. \tag{7.3}$$

Of course, we can use the Elek–Tardos normalization and consider $\frac{1}{2}S_n$*, which has diameter 1, and obtain the equivalent result* $\frac{1}{2}S_n \to \Upsilon_{1/2}$.

Example 7.3 (complete binary trees). *Let* B_n *be a complete binary tree with height* $n-1$ *and thus* $2^n - 1$ *vertices. Let* o *be the root and let* $h(x) := d(x, o)$ *(known as the depth of* x*) denote the distance from a vertex* x *to the root.*

If $(\xi_i^{(n)})_i$ *are i.i.d. random vertices in* B_n*, then for* $0 \leqslant k < n$,

$$\mathbb{P}\big(h(\xi_i^{(n)}) < n-k\big) = \frac{2^{n-k}-1}{2^n-1} \leqslant 2^{-k}. \tag{7.4}$$

Since also $h(\xi_i^{(n)}) < n$*, it follows that*

$$\frac{1}{n}h(\xi_i^{(n)}) \overset{\mathrm{p}}{\longrightarrow} 1. \tag{7.5}$$

Recall that $x \wedge y$ *denotes the last common ancestor of* $x, y \in B_n$*. If* $h(x \wedge y) \geqslant k$*, then* x *and* y *are both descendants of one of the* 2^k *vertices* z *with depth* k*. For each* z*, the number of such* x *(or* y*) is* $2^{n-k} - 1$*. Hence,*

$$\mathbb{P}\big(h(\xi_1^{(n)} \wedge \xi_2^{(n)}) \geqslant k\big) = 2^k\frac{(2^{n-k}-1)^2}{(2^n-1)^2} \leqslant 2^{-k}, \qquad k < n. \tag{7.6}$$

Consequently,

$$\frac{1}{n}h(\xi_1^{(n)} \wedge \xi_2^{(n)}) \overset{\mathrm{p}}{\longrightarrow} 0. \tag{7.7}$$

Since $d(x, y) = h(x) + h(y) - 2h(x \wedge y)$ *for* $x, y \in B_n$*, (7.5) and (7.7) imply*

$$\frac{1}{n}d(\xi_1^{(n)}, \xi_2^{(n)}) = \frac{1}{n}h\left(\xi_1^{(n)}\right) + \frac{1}{n}h\left(\xi_2^{(n)}\right) - \frac{2}{n}h\left(\xi_1^{(n)} \wedge \xi_2^{(n)}\right) \overset{\mathrm{p}}{\longrightarrow} 1+1-0 = 2. \tag{7.8}$$

Consequently, Theorem 3.13 yields

$$\frac{1}{n}B_n \to \Upsilon_1. \tag{7.9}$$

We see that (7.9) encapsulates (and formalizes) the fact that almost all pairs of vertices in B_n *have distance* $\approx 2n$.

Recall that B_n *has* $N := 2^n - 1$ *vertices. Thus, (7.9) can also be written*

$$\frac{1}{\log N}B_n \to \Upsilon_{1/\log 2}. \tag{7.10}$$

The results extend to complete b*-ary trees* T_n^b*, for any* $b \geqslant 2$*, with* $N = (b^n - 1)/(b-1)$ *nodes. In this case,*

$$\frac{1}{\log N}T_n^b \to \Upsilon_{1/\log b}. \tag{7.11}$$

Example 7.4 (superstars). *Let* T_n *consist of a central vertex* o *with* n *paths attached:* N_{kn} *paths with* k *edges for* $k \geqslant 1$*, all having* o *as one endpoint but otherwise disjoint, for some numbers* $N_{kn} \geqslant 0$ *with* $\sum_k N_{kn} = n$*. The number of vertices is thus* $|T_n| = 1 + \sum_k kN_{kn}$*. We assume that as* $n \to \infty$*, for some* $p_k \geqslant 0$ *with* $\sum_{k=1}^\infty p_k = 1$,

$$\frac{N_{kn}}{n} \to p_k, \qquad k \geqslant 1, \tag{7.12}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

and

$$\sum_k k\frac{N_{kn}}{n} \to \gamma := \sum_{k=1}^{\infty} kp_k < \infty. \tag{7.13}$$

Thus

$$|T_n| \sim \gamma n. \tag{7.14}$$

Suppose further (this actually follows from the other assumptions) that

$$\sum_k k^2\frac{N_{kn}}{n} = o(n). \tag{7.15}$$

Let $(\xi_i^{(n)})_i$ be i.i.d. uniformly random vertices of T_n. It follows from the assumptions above that, for $k \geqslant 1$,

$$\mathbb{P}\big(d(\xi_i^{(n)}, o) = k\big) = \frac{\sum_{j\geqslant k} N_{jn}}{|T_n|} \to \frac{\sum_{j\geqslant k} p_j}{\gamma} =: q_k. \tag{7.16}$$

Note that

$$\sum_{k=1}^{\infty} q_k = \frac{\sum_{k=1}^{\infty}\sum_{j\geqslant k} p_j}{\gamma} = \frac{\sum_{j=1}^{\infty} jp_j}{\gamma} = 1. \tag{7.17}$$

Let ν be the probability distribution on $\mathbb{N}$ given by $\nu\{k\} = q_k$, and let $(\xi_i)_i$ be i.i.d. with $\xi_i \sim \nu$. Then, (7.16) shows that

$$d(\xi_i^{(n)}, o) \xrightarrow{\text{d}} \xi_i, \qquad n \to \infty. \tag{7.18}$$

Furthermore, for any $i, j \geqslant 1$, by (7.15) and (7.14), as in the special case in Example 7.2,

$$\begin{aligned}\mathbb{P}\big(d(\xi_i^{(n)}, \xi_j^{(n)}) &\neq d(\xi_i^{(n)}, o) + d(\xi_i^{(n)}, o)\big)\\ &= \mathbb{P}\big(\xi_i^{(n)} \text{ and } \xi_j^{(n)} \text{ are in the same path}\big) = \frac{\sum_k k^2 N_{kn}}{|T_n|^2} \to 0.\end{aligned} \tag{7.19}$$

It follows from (7.18) and (7.19) that for any $r \geqslant 1$,

$$\rho_r\big(\xi_1^{(n)}, \dots \xi_r^{(n)}; T_n\big) \xrightarrow{\text{d}} \big((\xi_i + \xi_j)\mathbf{1}\{i \neq j\}\big)_{i,j=1}^{r} = \rho\big(\xi_1, \dots, \xi_r; \Upsilon_\nu\big). \tag{7.20}$$

Hence,

$$T_n \to \Upsilon_\nu. \tag{7.21}$$

8. Limits of random trees

In the rest of the paper we consider limits of random (finite) trees. Suppose that $\mathcal{T}_n$, $n \geqslant 1$, are random trees (with any distributions) and let, conditioned on $\mathcal{T}_n$, $(\xi_i^{(n)})_i$ be i.i.d. uniformly random vertices of $\mathcal{T}_n$. We are concerned with limits in distribution or probability of $c_n\mathcal{T}_n$ to some random or deterministic long dendron (tree limit). (Here, c_n are some given positive numbers.) By the definitions above, this is equivalent to convergence of the conditional distributions

$$\tau_\infty(c_n\mathcal{T}_n) = \mathcal{L}\big(\rho_\infty(\xi_1^{(n)}, \xi_2^{(n)}, \dots; c_n\mathcal{T}_n) \mid \mathcal{T}_n\big), \tag{8.1}$$

regarded as random elements of $\mathcal{P}(M_\infty)$; we thus want to show either that $\tau_\infty(c_n\mathcal{T}_n)$ converges in distribution to $\tau_\infty(D)$ for a random long dendron D, or (as a special case) that it converges in probability to $\tau_\infty(D)$ for a fixed D.

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Remark 8.1. It is important that we consider randomness in two steps: first $\mathcal{T}_n$ is a random tree and then $(\xi_i^{(n)})_i$ are random vertices in $\mathcal{T}_n$. As seen in (8.1), we are interested in the *quenched* version, where we first sample $\mathcal{T}_n$ and then condition on $\mathcal{T}_n$.

The alternative *annealed* version considers $\mathcal{T}_n$ and $(\xi_i^{(n)})_i$ as random together; the annealed distribution of $\rho_\infty(\xi_1^{(n)}, \xi_2^{(n)}, \dots; c_n\mathcal{T}_n)$ is the mean (or intensity) $\mathbb{E}\,\tau_\infty(c_n\mathcal{T}_n)$ of the random measure in (8.1), which in general is not what we want.

We note one simple case where the difference between quenched and annealed disappears.

Theorem 8.2. *Let $(\mathcal{T}_n)_n$ be a sequence of rescaled random trees and c_n some positive numbers. Let further $a \geqslant 0$. Then the following are equivalent, where d_n is the graph distance in $\mathcal{T}_n$ and $(\xi_i^{(n)})_i$ are i.i.d. uniformly random vertices in $\mathcal{T}_n$.*

(i) $c_nT_n \overset{\mathrm{p}}{\longrightarrow} \Upsilon_a$.

(ii)

$$c_nd_n(\xi_1^{(n)}, \xi_2^{(n)}) \overset{\mathrm{p}}{\longrightarrow} 2a. \tag{8.2}$$

(iii) *For every $\varepsilon > 0$,*

$$\mathbb{P}\big[|c_nd_n(\xi_1^{(n)}, \xi_2^{(n)}) - 2a| > \varepsilon \mid \mathcal{T}_n\big] \overset{\mathrm{p}}{\longrightarrow} 0. \tag{8.3}$$

Proof. Recall that for any random variables Z_n,

$$Z_n \overset{\mathrm{p}}{\longrightarrow} 0 \iff \mathbb{E}\big[|Z_n| \wedge 1\big] \to 0. \tag{8.4}$$

Thus, for deterministic trees T_n, the convergence in probability (3.13) is equivalent to

$$\mathbb{E}\big[|c_nd_n(\xi_1^{(n)}, \xi_2^{(n)}) - 2a| \wedge 1\big] \to 0. \tag{8.5}$$

Consequently, by Theorem 3.13, (i) is equivalent to

$$\mathbb{E}\big[|c_nd_n(\xi_1^{(n)}, \xi_2^{(n)}) - 2a| \wedge 1 \mid \mathcal{T}_n\big] \overset{\mathrm{p}}{\longrightarrow} 0. \tag{8.6}$$

A simple argument using Markov's inequality shows that (8.6) is equivalent to (iii). (This argument is a conditional version of (8.4).)

Furthermore, since the left-hand side of (8.6) is bounded by 1, (8.4) shows that (8.6) is equivalent to

$$\mathbb{E}\big[|c_nd_n(\xi_1^{(n)}, \xi_2^{(n)}) - 2a| \wedge 1\big] \overset{\mathrm{p}}{\longrightarrow} 0, \tag{8.7}$$

which by a final application of (8.4) is equivalent to (ii). □

We give also a version of the compactness criterion in Theorem 6.1 for random trees. We state the theorem for a sequence of random trees, although the statement and proof holds for an arbitrary set.

Theorem 8.3. *Let $(c_n\mathcal{T}_n)_n$ be a sequence of rescaled random trees. Then the following are equivalent, where $(\xi_i^{(n)})_i$ are i.i.d. uniformly random vertices in $\mathcal{T}_n$ and d_n is the graph distance in $\mathcal{T}_n$.*

(i) *The sequence $(c_n\mathcal{T}_n)_n$ of random elements of $\mathfrak{T}$ is relatively compact in $\mathcal{P}(\mathfrak{T})$.*
(ii) *The sequence $(c_n\mathcal{T}_n)_n$ of random elements of $\mathfrak{T}$ is tight.*
(iii) *The sequence of random variables $\big(c_nd_n(\xi_1^{(n)}, \xi_2^{(n)})\big)_n$ is tight.*

Proof. (i) $\iff$ (ii): Since $\mathfrak{T}$ is a Polish space, this is Prohorov's theorem.

(ii) $\implies$ (iii): By the definitions in Section 5, $\mathfrak{T}$ is a closed subspace of $\mathcal{P}(M_\infty)$ and it follows that (ii) means that for every $\varepsilon > 0$, there exists a compact set $\mathcal{K}_\varepsilon \subset \mathcal{P}(M_\infty)$ such that, for every $n \geqslant 1$,

$$\mathbb{P}\big(\tau_\infty(c_n\mathcal{T}_n) \notin \mathcal{K}_\varepsilon\big) < \varepsilon. \tag{8.8}$$

Furthermore, Prohorov's theorem (now applied to the Polish space M_∞) shows that for every $\delta > 0$, there exists a compact set $K_{\varepsilon,\delta} \subset M_\infty$ such that if $\lambda \in \mathcal{P}(M_\infty)$, then

$$\lambda \in \mathcal{K}_\varepsilon \implies \lambda(K_{\varepsilon,\delta}) > 1 - \delta. \tag{8.9}$$

Since the projection $(a_{ij})_{ij} \mapsto a_{12}$ is continuous $M_\infty \to \mathbb{R}$, there exists a constant $C_{\varepsilon,\delta}$ such that if $(a_{ij})_{ij} \in K_{\varepsilon,\delta}$, then $|a_{12}| \leqslant C_{\varepsilon,\delta}$.

Consequently, for every $\varepsilon, \delta > 0$, and all n,

$$|c_n d_n(\xi_1^{(n)}, \xi_2^{(n)})| > C_{\varepsilon,\delta} \implies \rho_\infty(\xi_1^{(n)}, \dots; c_n\mathcal{T}_n) \notin K_{\varepsilon,\delta} \tag{8.10}$$

and thus, using also (8.9),

$$\begin{aligned}&\mathbb{P}\big(|c_n d_n(\xi_1^{(n)}, \xi_2^{(n)})| > C_{\varepsilon,\delta} \mid \mathcal{T}_n\big) \geqslant \delta \\ &\quad\implies \mathbb{P}\big(\rho_\infty(\xi_1^{(n)}, \dots; c_n\mathcal{T}_n) \notin K_{\varepsilon,\delta} \mid \mathcal{T}_n\big) \geqslant \delta \\ &\quad\implies \tau_\infty(c_n\mathcal{T}_n) = \mathcal{L}\big(\rho_\infty(\xi_1^{(n)}, \dots; c_n\mathcal{T}_n)\big) \notin \mathcal{K}_\varepsilon.\end{aligned} \tag{8.11}$$

Hence, (8.8) implies

$$\mathbb{P}\Big(\mathbb{P}\big(|c_n d_n(\xi_1^{(n)}, \xi_2^{(n)})| > C_{\varepsilon,\delta} \mid \mathcal{T}_n\big) \geqslant \delta\Big) < \varepsilon \tag{8.12}$$

which yields

$$\mathbb{P}\big(|c_n d_n(\xi_1^{(n)}, \xi_2^{(n)})| > C_{\varepsilon,\delta}\big) = \mathbb{E}\,\mathbb{P}\big(|c_n d_n(\xi_1^{(n)}, \xi_2^{(n)})| > C_{\varepsilon,\delta} \mid \mathcal{T}_n\big) \leqslant \delta + \varepsilon. \tag{8.13}$$

By taking $\delta = \varepsilon$, this shows (iii).

(iii) $\implies$ (ii): By (iii), for every $\varepsilon > 0$, there exists C_ε such that

$$\mathbb{P}\big(|c_n d_n(\xi_1^{(n)}, \xi_2^{(n)})| > C_\varepsilon\big) < \varepsilon. \tag{8.14}$$

Define

$$K_\varepsilon := \big\{(a_{ij})_{ij} \in M_\infty : |a_{ij}| \leqslant C_{2^{-i-j}\varepsilon}\big\}. \tag{8.15}$$

This is a compact subset of M_∞, and (8.14) implies

$$\begin{aligned}\mathbb{P}\big(\rho_\infty(\xi_1^{(n)}, \dots; c_n\mathcal{T}_n) \notin K_\varepsilon\big) &\leqslant \sum_{i,j=1}^\infty \mathbb{P}\big(|c_n d_n(\xi_i^{(n)}, \xi_j^{(n)})| > C_{2^{-i-j}\varepsilon}\big) \\ &< \sum_{i,j=1}^\infty 2^{-i-j}\varepsilon = \varepsilon.\end{aligned} \tag{8.16}$$

Hence,

$$\mathbb{E}\big[\mathbb{P}\big(\rho_\infty(\xi_1^{(n)}, \dots; c_n\mathcal{T}_n) \notin K_\varepsilon \mid \mathcal{T}_n\big)\big] = \mathbb{P}\big(\rho_\infty(\xi_1^{(n)}, \dots; c_n\mathcal{T}_n) \notin K_\varepsilon\big) < \varepsilon, \tag{8.17}$$

and Markov's inequality shows that, for any $\ell \geqslant 1$,

$$\mathbb{P}\big[\mathbb{P}\big(\rho_\infty(\xi_1^{(n)}, \dots; c_n\mathcal{T}_n) \notin K_{4^{-\ell}\varepsilon} \mid \mathcal{T}_n\big) > 2^{-\ell}\big] < 2^{-\ell}\varepsilon. \tag{8.18}$$

By the definition of τ_∞, this is the same as

$$\mathbb{P}\big[\tau_\infty(c_n\mathcal{T}_n)\big(M_\infty \setminus K_{4^{-\ell}\varepsilon}\big) > 2^{-\ell}\big] < 2^{-\ell}\varepsilon. \tag{8.19}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Let

$$\mathcal{K}_\varepsilon := \{\lambda \in \mathcal{P}(M_\infty) : \lambda(K_{4^{-\ell}\varepsilon}) \geqslant 1 - 2^{-\ell}, \forall \ell \geqslant 1\} \tag{8.20}$$

and note that $\mathcal{K}_\varepsilon$ is compact by Prohorov's theorem. It follows by (8.19) that

$$\mathbb{P}\big(\tau_\infty(c_n\mathcal{T}_n) \notin \mathcal{K}_\varepsilon\big) < \sum_{\ell=1}^{\infty} 2^{-\ell}\varepsilon = \varepsilon. \tag{8.21}$$

Hence, the sequence $\tau_\infty(c_n\mathcal{T}_n)$ is tight in $\mathcal{P}(M_\infty)$, and thus in $\mathfrak{T}$. □

Remark 8.4. Again, the same holds, mutatis mutandis, for random measured real trees and for random long dendrons. In fact, the argument is quite general and holds for any measured metric spaces. We believe that this may be known, but we do not know a reference and have included a full proof for completeness.

9. Conditioned Galton–Watson trees, I

Consider a Galton–Watson process with some given offspring distribution ζ. (We let ζ denote both the distribution and a random variable with this distribution.) The family tree of the Galton–Watson process is a random tree $\mathcal{T}$, which in the subcritical and critical cases (i.e., when $\mathbb{E}\,\zeta \leqslant 1$) is a.s. finite. $\mathcal{T}$ is a *Galton–Watson tree*, and the random tree $\mathcal{T}_n := (\mathcal{T} \mid |\mathcal{T}| = n)$ obtained by conditioning $\mathcal{T}$ on a given size n is said to be a *conditioned Galton–Watson tree*. (We consider only n such that $\mathbb{P}\,(|\mathcal{T}| = n) > 0$.) For further details, see e.g. the survey [29].

In the standard case $\mathbb{E}\,\zeta = 1$ and $\operatorname{Var}\zeta < \infty$, Aldous [5, 6, 7] proved convergence in distribution of the conditioned Galton–Watson tree $\mathcal{T}_n$, after rescaling, to a limit object called the *Brownian continuum random tree*; this is a random measured real tree which we denote by $T_{2\mathbf{e}}$, for reasons given below. Aldous's original result was not in terms of the type of convergence discussed in the present paper, but it holds in the present context too. In fact, Aldous's result has been stated in several different forms, more or less equivalent; one version, stated e.g. in [24, Theorem 8] and [2, Theorem 5.2], is convergence in the Gromov–Hausdorff–Prohorov metric (defined in e.g. [50, Chapter 27] and [41, Section 6]), which is stronger than Gromov–Prohorov convergence and thus implies convergence in the tree limit sense used in the present paper (see Remark 3.6 and Example 3.11). We thus have the following.

Theorem 9.1. *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree with critical offspring distribution ζ with finite variance, i.e., we assume $\mathbb{E}\,\zeta = 1$ and $\sigma^2 := \operatorname{Var}\zeta \in (0,\infty)$. Then, as $n \to \infty$,*

$$\frac{1}{\sqrt{n}}\mathcal{T}_n \overset{\mathrm{d}}{\longrightarrow} \frac{1}{\sigma}T_{2\mathbf{e}}, \tag{9.1}$$

where $T_{2\mathbf{e}}$ is the Brownian continuum random tree.

Proof. As said before the theorem, this is known. For completeness, we sketch a proof in the present context; omitted details can be found e.g. in e.g. [7] and [36].

One standard version of Aldous's theorem uses the *contour function* $C_{\mathcal{T}_n}(t)$ of $\mathcal{T}_n$. In general, if T is a rooted tree with $|T| = n$, then C_T is a continuous function $[0, 2(n-1)] \to [0,\infty)$; informally, $C_T(t)$ is the distance, at time t, from the root to a particle that travels with unit speed along the "outside" of the tree, starting at the root at time 0 and returning at time $2(n-1)$, having traversed every edge once in each direction. Aldous [7] showed that

$$\frac{1}{\sqrt{n}}C_{\mathcal{T}_n}(2(n-1)t) \overset{\mathrm{d}}{\longrightarrow} \frac{2}{\sigma}\mathbf{e}(t) \qquad \text{in } C[0,1], \tag{9.2}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

where $\mathbf{e}(t)$ is the standard Brownian excursion, which is a random continuous function $[0, 1] \to [0, \infty)$ with $\mathbf{e}(0) = \mathbf{e}(1) = 0$.

Every continuous function $g:[0, 1] \to [0, \infty)$ with $g(0) = g(1) = 0$ defines a real tree T_g: define a pseudometric on $[0, 1]$ by

$$d(s, t) := g(s) + g(t) - 2 \min_{u \in [s,t]} g(u), \qquad 0 \leqslant s \leqslant t \leqslant 1, \tag{9.3}$$

and form the quotient of $[0, 1]$ modulo the equivalence relation $\{d(s, t) = 0\}$; see e.g. [36, Theorem 2.2]. The uniform (Lebesgue) measure on $[0, 1]$ induces a measure μ on T_g, making (T_g, μ) a measured real tree.

Taking $g(t) = C_T(2(n-1)t)$ for a rooted tree T with $|T| = n$ gives $T_g = \hat{T}$, the real tree obtained from T as in Example 3.4. The measure μ induced by g is the uniform measure on the edges of $\hat{T}$, and not the uniform measure μ' on the vertices of $T \subseteq \hat{T}$; however, it is easy to couple these measures and find $\xi \sim \mu$ and $\xi' \sim \mu'$ such that $\mathbb{P}(|\xi - \xi'| > 1) \leqslant 1/n$. It follows that (9.2) implies (9.1), both in the sense of the present paper and in the stronger Gromov–Hausdorff–Prohorov metric. □

Remark 9.2. Duquesne [19] considered the case when ζ has infinite variance and furthermore is in the domain of attraction of a stable distribution; he extended Aldous's result and showed convergence of the contour process of $c_n\mathcal{T}_n$ (for suitable c_n) to a certain stochastic process in this case too; this implies convergence of $c_n\mathcal{T}_n$ to a random real tree called the *stable tree* [37] in Gromov–Hausdorff–Prohorov sense, and thus in the weaker sense of tree limits, also in this case. (See [24, Theorem 8], with a somewhat stronger assumption on ζ.)

Remark 9.3. As is well-known, several important classes of random trees can be represented as conditioned Galton–Watson trees $\mathcal{T}_n$ satisfying the conditions above by choosing suitable offspring distributions ζ; thus Theorem 9.1 applies to them. This includes (uniformly) random labelled trees ($\sigma^2 = 1$), random ordered trees ($\sigma^2 = 2$) and random binary trees ($\sigma^2 = 1/2$); see e.g. [6] and [29].

Remark 9.4. Recall that random *simply generated trees* are defined by a weight sequence $(w_k)_k$; see, again, e.g. [6] or [29] for the definition and for the well-known fact that while simply generated trees are more general than conditioned Galton–Watson trees, they can in many cases be reduced to equivalent conditioned Galton–Watson trees. Thus Theorem 9.1 applies to simply generated trees under rather weak conditions. In other cases of simply generated trees, the results in Sections 10 and 11 may apply.

10. Conditioned Galton–Watson trees, II

Although a large class of conditioned Galton–Watson trees (and simply generated trees) are covered by Theorem 9.1, there are also other cases. One class of conditioned Galton–Watson trees with a different local limit behaviour showing condensation was found by Jonsson and Stefánsson [33]; this was generalized in [29], with further results in [1] and [49]. This class of conditioned Galton–Watson trees (called type II in [29] and [49]) has offspring distributions ζ satisfying

$$0 < \kappa := \mathbb{E}\,\zeta < 1, \tag{10.1}$$

$$\mathbb{E}\,R^{\zeta} = \infty, \qquad \forall R > 1. \tag{10.2}$$

In other words, the Galton–Watson trees are subcritical, and ζ has infinite moment generating function; see further [29, Section 8]. We will show that this class has tree limits that are very different from the ones in Section 9.

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

For a rooted tree T and a vertex $v \in T$, let $\delta(v)$ denote the outdegree of v. Furthermore, let $\Delta = \Delta(T) := \max_{v\in T} \delta(v)$ be the maximum outdegree, and let $v^\dagger$ be the vertex with maximum outdegree (chosen as e.g. the lexicographically first if there is a tie), so $\Delta = \delta(v^\dagger)$.

It is shown in [29, Section 19.6] that (10.1)–(10.2) imply the existence (asymptotically) of one or several vertices of very high (out)degree, with a total outdegree $\approx (1-\kappa)n$; typically, there is one single large vertex with degree $\approx (1-\kappa)n$, but this is not always the case; see [29] and Remark 10.8. We will assume that there is such a vertex; a case known as *complete condensation*. To be precise, we assume that ζ is such that

$$\Delta(\mathcal{T}_n) = (1-\kappa)n + o_p(n). \tag{10.3}$$

For example, this holds when the offspring distribution satisfies (10.1)–(10.2) and has a power law tail, as shown by Jonsson and Stefánsson [33, Theorem 5.5], see also [29, Theorem 19.34] and (more generally, with regularly varying tails) Kortchemski [35, Theorem 1].

We note that (10.3) implies that the second largest outdegree is $o_p(n)$; in particular the maximum degree vertex $v^\dagger$ is unique w.h.p., see [29, paragraph after Lemma 19.32].

Theorem 10.1. *Let $\mathcal{T}_n$ be a conditioned Galton–Watson tree with subcritical offspring distribution ζ satisfying (10.1)–(10.2) and (10.3). Then, as $n \to \infty$,*

$$\mathcal{T}_n \xrightarrow{p} \Upsilon_\nu, \tag{10.4}$$

where $\nu = \mathrm{Ge}\,(1-\kappa)$ is a geometric distribution on $\mathbb{N} := \{1, 2, \dots\}$.

Recall that Υ_ν is the deterministic long dendron defined in Example 3.12.

Remark 10.2. Note that there is no rescaling of $\mathcal{T}_n$ in (10.4); the situation is similar to Examples 7.2 and 7.4. Distances are typically small; formally, the distance $d(\xi_1^{(n)}, \xi_2^{(n)})$ between two random vertices is stochastically bounded (i.e., tight). Hence, the local limits studied in [29] and [49] are essentially global in this case.

Remark 10.3. The diameter $\mathrm{diam}\,(\mathcal{T}_n) \xrightarrow{p} \infty$, e.g. by Lemma 10.6 below. Hence, rescaling such that the diameter becomes 1 would only give the trivial limit Υ_0, see Remark 3.14.

The rest of this section contains the proof of Theorem 10.1. We begin with some further notation. In this proof, all trees are rooted and ordered. Trees that are equal up to order-preserving isomorphisms are regarded as equal. Let $\mathbf{T}$ be the countable set of all finite trees.

Let again $\mathcal{T}$ denote the (unconditioned) Galton–Watson tree with the chosen offspring distribution ζ. Since $\mathbb{E}\,\zeta < 1$, $\mathcal{T}$ is a.s. finite. If $\mathbf{t}$ is any fixed finite tree, let

$$\pi_{\mathbf{t}} := \mathbb{P}\,(\mathcal{T} = \mathbf{t}). \tag{10.5}$$

In other words, $(\pi_{\mathbf{t}})_{\mathbf{t}\in\mathbf{T}}$ is the probability distribution of $\mathcal{T} \in \mathbf{T}$.

The *fringe tree* [4] of a tree T at a vertex v, denoted T^v, is the subtree of T consisting of v and its descendants, rooted at v.

Let $\mathbf{t}$ denote a finite tree. For any tree T, let

$$N_{\mathbf{t}}(T) := \left|\{v \in T : T^v = \mathbf{t}\}\right|, \tag{10.6}$$

i.e., the number of fringe trees of T equal to $\mathbf{t}$.

It is shown in [29, Theorem 7.12] that for any fixed tree $\mathbf{t}$, assuming (10.1)–(10.2),

$$\frac{N_{\mathbf{t}}(\mathcal{T}_n)}{n} \xrightarrow{p} \pi_{\mathbf{t}}. \tag{10.7}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Both sides of (10.7) are, regarded as functions of $\mathbf{t}$, probability distributions on the countable set of finite trees. (Note that $\big(N_{\mathbf{t}}(\mathcal{T}_n)/n\big)_{\mathbf{t}}$ is the conditional distribution of $\mathcal{T}_n^v$ given $\mathcal{T}_n$, with v a random vertex, while $\big(\pi_{\mathbf{t}}\big)_{\mathbf{t}}$ is the distribution of $\mathcal{T}$.) Hence, (10.7) says that the random distribution $\big(N_{\mathbf{t}}(\mathcal{T}_n)/n\big)_{\mathbf{t}}$ converges in probability to $(\pi_{\mathbf{t}})_{\mathbf{t}}$ in the space $\mathcal{P}(\mathbf{T})$ with the usual weak topology. (Note that we here consider convergence of random probability distributions, regarded as elements of the space $\mathcal{P}(\mathbf{T})$ of probability distributions on finite trees.) We claim that, since $\mathbf{T}$ is countable, it follows that the random distribution $\big(N_{\mathbf{t}}(\mathcal{T}_n)/n\big)_{\mathbf{t}}$ converges in probability to $(\pi_{\mathbf{t}})_{\mathbf{t}}$ in total variation, and thus for any set $\mathbf{T}' \subseteq \mathbf{T}$ of finite trees,

$$\sum_{\mathbf{t}\in\mathbf{T}'} \frac{N_{\mathbf{t}}(\mathcal{T}_n)}{n} \overset{\mathrm{p}}{\longrightarrow} \sum_{\mathbf{t}\in\mathbf{T}'} \pi_{\mathbf{t}} = \mathbb{P}\,(\mathcal{T} \in \mathbf{T}'). \tag{10.8}$$

To see this, note that the corresponding result for sequences of probability distributions on a countable set is well known, see e.g. [23, Theorem 5.6.4]. The version used here with random distributions and convergence in probability follows by essentially the same proof, or by first using the Skorohod coupling theorem [34, Theorem 4.30], to see that we may assume that (10.7) holds a.s. for each $\mathbf{t}$, and then using the deterministic version.

We need an extension of (10.7). Let

$$N_{\mathbf{t},k}(T) := \big|\{v \in T : T^v = \mathbf{t} \text{ and } \delta(\hat{v}) = k\}\big|, \tag{10.9}$$

where $\hat{v}$ denotes the parent of v (undefined for the root). Also, let

$$p_k := \mathbb{P}\,(\zeta = k), \qquad k \geqslant 0 \tag{10.10}$$

and note that

$$\kappa := \mathbb{E}\,\zeta = \sum_{k=1}^{\infty} k p_k. \tag{10.11}$$

Lemma 10.4. *Assume* (10.1)–(10.2). *For every fixed* $\mathbf{t} \in \mathbf{T}$ *and* $k \in \mathbb{N}$,

$$\frac{N_{\mathbf{t},k}(\mathcal{T}_n)}{n} \overset{\mathrm{p}}{\longrightarrow} k p_k \pi_{\mathbf{t}}. \tag{10.12}$$

Proof. Let, for $j = 1, \dots, k$,

$$N_{\mathbf{t},k,j}(T) := \big|\{v \in T : T^v = \mathbf{t}, \delta(\hat{v}) = k, \text{ and } v \text{ is the } j\text{th child of } \hat{v}\}\big|, \tag{10.13}$$

Note that v is in the set in (10.13) if and only if $T^{\hat{v}} \in \mathbf{T}_j$, where $\mathbf{T}_j$ is the set of all trees $\hat{\mathbf{t}}$ such that the root has exactly k children, and if w is the jth of these, then the fringe tree $\hat{\mathbf{t}}^w = \mathbf{t}$. Hence, (10.8) shows that, using also the recursive property of the Galton–Watson tree $\mathcal{T}$,

$$\frac{N_{\mathbf{t},k,j}(\mathcal{T}_n)}{n} \overset{\mathrm{p}}{\longrightarrow} \mathbb{P}\,(\mathcal{T} \in \mathbf{T}_j) = p_k\,\mathbb{P}\,(\mathcal{T} = \mathbf{t}) = p_k \pi_{\mathbf{t}}. \tag{10.14}$$

The result follows, since $N_{\mathbf{t},k}(\mathcal{T}_n) = \sum_{j=1}^{k} N_{\mathbf{t},k,j}(\mathcal{T}_n)$. □

We have so far not used the assumption (10.3), but it is essential for the next lemma. Recall that $\Delta = \Delta(\mathcal{T}_n)$ is the maximum outdegree, and that w.h.p. $v^{\dagger}$ is the only vertex of outdegree Δ. Hence, w.h.p., $N_{\mathbf{t},\Delta}(\mathcal{T}_n)$ is the number of children v of $v^{\dagger}$ such that $\mathcal{T}_n^v = \mathbf{t}$.

Lemma 10.5. *Assume* (10.1)–(10.3). *For every fixed* $\mathbf{t} \in \mathbf{T}$,

$$\frac{N_{\mathbf{t},\Delta}(\mathcal{T}_n)}{n} \overset{\mathrm{p}}{\longrightarrow} (1-\kappa)\pi_{\mathbf{t}}. \tag{10.15}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Proof. Let $N_k := \left|\{v \in \mathcal{T}_n : \delta(v) = k\}\right|$. Then, as a consequence of (10.8) or as a simpler version of (10.7), see [29, Theorem 7.11],

$$N_k/n \xrightarrow{\text{p}} p_k, \qquad k \geqslant 0. \tag{10.16}$$

Let $\varepsilon > 0$, and choose K such that

$$\sum_{k>K} kp_k < \varepsilon. \tag{10.17}$$

The number of vertices having a parent of outdegree k is kN_k. Thus $\sum_k kN_k = n-1$. Hence, using (10.16), (10.3), (10.11) and (10.17),

$$\begin{aligned}\sum_{K+1}^{\Delta-1} kN_k &\leqslant \sum_{k=1}^{\infty} kN_k - \sum_{k=1}^{K} kN_k - \Delta \\ &= n-1-\sum_{k=1}^{K} kp_k n - (1-\kappa)n + o_{\text{p}}(n) \\ &= \sum_{k>K} kp_k n + o_{\text{p}}(n) \\ &< \varepsilon n + o_{\text{p}}(n) < 2\varepsilon n \qquad \text{w.h.p.}\end{aligned} \tag{10.18}$$

Now consider the $N_{\mathbf{t}}$ vertices v such that $\mathcal{T}_n^v = \mathbf{t}$. Assume for convenience $n > |\mathbf{t}|$, so that the root is not one of these vertices. Then, using (10.7) and (10.12),

$$\begin{aligned}N_{\mathbf{t},\Delta} &= N_{\mathbf{t}} - \sum_{k=1}^{K} N_{\mathbf{t},k} - \sum_{K+1}^{\Delta-1} N_{\mathbf{t},k} \\ &= \pi_{\mathbf{t}} n - \sum_{k=1}^{K} kp_k\pi_{\mathbf{t}} n - \sum_{K+1}^{\Delta-1} N_{\mathbf{t},k} + o_{\text{p}}(n).\end{aligned} \tag{10.19}$$

Thus, using (10.18) and $N_{\mathbf{t},k} \leqslant kN_k$, w.h.p.,

$$\pi_{\mathbf{t}} n - \sum_{k=1}^{K} kp_k\pi_{\mathbf{t}} n - 2\varepsilon n + o_{\text{p}}(n) \leqslant N_{\mathbf{t},\Delta} \leqslant \pi_{\mathbf{t}} n - \sum_{k=1}^{K} kp_k\pi_{\mathbf{t}} n + o_{\text{p}}(n). \tag{10.20}$$

Using also (10.17) and (10.11), we find that w.h.p.

$$(1-\kappa)\pi_{\mathbf{t}} n - 3\varepsilon n \leqslant N_{\mathbf{t},\Delta} \leqslant (1-\kappa)\pi_{\mathbf{t}} n + 2\varepsilon n. \tag{10.21}$$

The result (10.15) follows, since $\varepsilon > 0$ is arbitrary. □

Next, for a tree $\mathbf{t}$, and $\ell \geqslant 0$, let w_ℓ be the number of vertices at distance ℓ from the root. Furthermore, for $\ell \geqslant 1$, let $W_\ell := w_\ell(\mathcal{T}_n^{v^\dagger})$, the number of vertices in $\mathcal{T}_n$ that are descendants of $v^\dagger$ and are ℓ generations from it, and let $\overline{W} := n - \sum_{\ell\geqslant 1} W_\ell$ be the number of vertices that are not descendants of $v^\dagger$.

Lemma 10.6. *Assume (10.1)–(10.3). Then, for* $\ell \geqslant 1$,

$$\frac{W_\ell}{n} \xrightarrow{\text{p}} (1-\kappa)\kappa^{\ell-1} \tag{10.22}$$

and

$$\frac{\overline{W}}{n} \xrightarrow{\text{p}} 0. \tag{10.23}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Proof. We have, assuming $N_\Delta = 1$ which holds w.h.p.,

$$W_\ell = \sum_{\mathbf{t}\in\mathbf{T}} N_{\mathbf{t},\Delta}(\mathcal{T}_n) w_{\ell-1}(\mathbf{t}). \tag{10.24}$$

For any finite family $\mathbf{T}_0 \subset \mathbf{T}$, by Lemma 10.5,

$$\frac{1}{n}\sum_{\mathbf{t}\in\mathbf{T}_0} N_{\mathbf{t},\Delta}(\mathcal{T}_n) w_{\ell-1}(\mathbf{t}) \xrightarrow{\text{p}} (1-\kappa)\sum_{\mathbf{t}\in\mathbf{T}_0} \pi_{\mathbf{t}} w_{\ell-1}(\mathbf{t}). \tag{10.25}$$

Hence, by (10.24),

$$\frac{1}{n}W_\ell = \frac{1}{n}\sum_{\mathbf{t}\in\mathbf{T}} N_{\mathbf{t},\Delta}(\mathcal{T}_n) w_{\ell-1}(\mathbf{t}) \geqslant (1-\kappa)\sum_{\mathbf{t}\in\mathbf{T}_0} \pi_{\mathbf{t}} w_{\ell-1}(\mathbf{t}) + o_{\text{p}}(1) \tag{10.26}$$

for any finite $\mathbf{T}_0$. Furthermore, by elementary branching process theory,

$$\sum_{\mathbf{t}\in\mathbf{T}} \pi_{\mathbf{t}} w_{\ell-1}(\mathbf{t}) = \mathbb{E}\, w_{\ell-1}(\mathcal{T}) = (\mathbb{E}\,\zeta)^{\ell-1} = \kappa^{\ell-1}. \tag{10.27}$$

In particular, the sum converges, and it follows from (10.26) that

$$\frac{1}{n}W_\ell \geqslant (1-\kappa)\sum_{\mathbf{t}\in\mathbf{T}} \pi_{\mathbf{t}} w_{\ell-1}(\mathbf{t}) + o_{\text{p}}(1). \tag{10.28}$$

Thus, (10.28) yields

$$\frac{1}{n}W_\ell \geqslant (1-\kappa)\kappa^{\ell-1} + o_{\text{p}}(1), \qquad \ell \geqslant 1. \tag{10.29}$$

We can sum (10.29) over any set of ℓ, using the same argument as for (10.28) again. In particular, we obtain

$$\frac{1}{n}\sum_{j\neq\ell} W_j \geqslant (1-\kappa)\sum_{j\neq\ell}\kappa^{j-1} + o_{\text{p}}(1). \tag{10.30}$$

On the other hand, trivially,

$$\frac{1}{n}\sum_{j=1}^{\infty} W_j \leqslant 1 = (1-\kappa)\sum_{j=1}^{\infty}\kappa^{j-1}. \tag{10.31}$$

Subtracting (10.30) from (10.31) yields

$$\frac{1}{n}W_\ell \leqslant (1-\kappa)\kappa^{\ell-1} + o_{\text{p}}(1), \tag{10.32}$$

which together with (10.29) yields the result (10.22).

Furthermore, (10.29) and (10.30) yield

$$\frac{1}{n}\sum_{j=1}^{\infty} W_j \geqslant (1-\kappa)\sum_{j=1}^{\infty}\kappa^{j-1} + o_{\text{p}}(1) = 1 + o_{\text{p}}(1). \tag{10.33}$$

Thus,

$$\overline{W} = n - \sum_{j=1}^{\infty} W_j = o_{\text{p}}(n), \tag{10.34}$$

which yields (10.23) and completes the proof. □

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Lemma 10.7. *Assume (10.1)–(10.3). Let $(\xi_i^{(n)})_i$ be i.i.d. vertices in $\mathcal{T}_n$. Then*

$$\mathbb{P}\Big(d(\xi_1^{(n)},\xi_2^{(n)}) \neq d(\xi_1^{(n)},v^\dagger)+d(\xi_2^{(n)},v^\dagger) \Bigm| \mathcal{T}_n\Big) \overset{\mathrm{p}}{\longrightarrow} 0. \tag{10.35}$$

Proof. If $\xi_1^{(n)}$ and $\xi_2^{(n)}$ both are descendants of $v^\dagger$, then $d(\xi_1^{(n)},\xi_2^{(n)}) = d(\xi_1^{(n)},v^\dagger)+d(\xi_2^{(n)},v^\dagger)$ unless $\xi_1^{(n)}$ and $\xi_2^{(n)}$ are in the same fringe subtree rooted at a child of $v^\dagger$. Hence, the probability in (10.35) is at most

$$2\frac{\overline{W}}{n}+\frac{1}{n^2}\sum_{\mathbf{t}\in\mathbf{T}} N_{\mathbf{t},\Delta}|\mathbf{t}|^2. \tag{10.36}$$

By (10.23), it suffices to show that the sum in (10.36) is $o_{\mathrm{p}}(n^2)$. Fix $K>1$, and let $\mathbf{T}_K := \{\mathbf{t}\in\mathbf{T}: |\mathbf{t}| \leqslant K\}$ and $\mathbf{T}_{>K} := \{\mathbf{t}\in\mathbf{T}: |\mathbf{t}| > K\}$. First, deterministically,

$$\sum_{\mathbf{t}\in\mathbf{T}_K} N_{\mathbf{t},\Delta}|\mathbf{t}|^2 \leqslant K\sum_{\mathbf{t}\in\mathbf{T}_K} N_{\mathbf{t},\Delta}|\mathbf{t}| \leqslant Kn = o(n^2). \tag{10.37}$$

Secondly, since no subtree of $\mathcal{T}_n$ has more than n vertices,

$$\sum_{\mathbf{t}\in\mathbf{T}_{>K}} N_{\mathbf{t},\Delta}|\mathbf{t}|^2 \leqslant n\sum_{\mathbf{t}\in\mathbf{T}_{>K}} N_{\mathbf{t},\Delta}|\mathbf{t}|. \tag{10.38}$$

By (10.15), since the set $\mathbf{T}_K$ is finite,

$$\sum_{\mathbf{t}\in\mathbf{T}_{>K}} N_{\mathbf{t},\Delta}|\mathbf{t}| \leqslant n-\sum_{\mathbf{t}\in\mathbf{T}_K} N_{\mathbf{t},\Delta}|\mathbf{t}| = n-n\sum_{\mathbf{t}\in\mathbf{T}_K}(1-\kappa)\pi_{\mathbf{t}}|\mathbf{t}|+o_{\mathrm{p}}(n). \tag{10.39}$$

On the other hand

$$\sum_{\mathbf{t}\in\mathbf{T}}(1-\kappa)\pi_{\mathbf{t}}|\mathbf{t}| = (1-\kappa)\,\mathbb{E}\,|\mathcal{T}| = (1-\kappa)\frac{1}{1-\kappa} = 1. \tag{10.40}$$

Thus, for every $\varepsilon>0$, we may choose K such that $\sum_{\mathbf{t}\in\mathbf{T}_K}(1-\kappa)\pi_{\mathbf{t}}|\mathbf{t}| > 1-\varepsilon$, and then (10.38)–(10.39) yield, w.h.p.,

$$\sum_{\mathbf{t}\in\mathbf{T}_{>K}} N_{\mathbf{t},\Delta}|\mathbf{t}|^2 \leqslant n^2\varepsilon+o_{\mathrm{p}}(n^2) < 2\varepsilon n^2. \tag{10.41}$$

The result follows by (10.36), (10.37) and (10.41). □

Proof of Theorem 10.1. Let $(\xi_i^{(n)})$ be i.i.d. uniformly random vertices of $\mathcal{T}_n$, and let $Y_i^{(n)} := d(\xi_i^{(n)},v^\dagger)$. Then Lemma 10.6 yields

$$\mathbb{P}\big(Y_i^{(n)}=\ell \mid \mathcal{T}_n\big) \overset{\mathrm{p}}{\longrightarrow} (1-\kappa)\kappa^{\ell-1}, \qquad \ell\geqslant 1, \tag{10.42}$$

and Lemma 10.7 yields

$$\mathbb{P}\big(d(\xi_i^{(n)},\xi_j^{(n)}) \neq Y_i^{(n)}+Y_j^{(n)} \mid \mathcal{T}_n\big) \overset{\mathrm{p}}{\longrightarrow} 0. \tag{10.43}$$

We may for convenience, by the Skorohod coupling theorem [34, Theorem 4.30], or (more elementary) by considering suitable subsequences, assume that (10.42) and (10.43) hold with $\overset{\mathrm{a.s.}}{\longrightarrow}$. Then, (10.42) and the independence of $(Y_i^{(n)})_i$ shows that a.s. the sequence $\mathcal{T}_n$ is such that, conditioned on $\mathcal{T}_n$, we have $(Y_i^{(n)})_i \overset{\mathrm{d}}{\longrightarrow} (Y_i)_i$ with $Y_i \sim \mathrm{Ge}\,(1-\kappa)$ i.i.d. Consequently, for every $r\geqslant 1$, using also (10.43),

$$\rho_r(\xi_1^{(n)},\dots,\xi_r^{(n)};\mathcal{T}_n) \overset{\mathrm{d}}{\longrightarrow} \big((Y_i+Y_j)\mathbf{1}\{i\neq j\}\big)_{i,j=1}^r = \rho_r(\xi_1,\dots,\xi_r;\Upsilon_\nu), \tag{10.44}$$

where $\xi_i := (\bullet, Y_i) \in A_{\Upsilon_\nu}$ are i.i.d. with $\xi_i \sim \nu$. Hence, a.s., $\tau_r(\mathcal{T}_n)\to\tau_r(\Upsilon_\nu)$ and thus $\mathcal{T}_n\to\Upsilon_\nu$. □

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Remark 10.8. [29, Example 19.37] gives an example of an offspring distribution satisfying (10.1)–(10.2) but not (10.3). In this example, there exists a subsequence of n such that $\Delta(\mathcal{T}_n)/n \xrightarrow{\text{p}} 0$; there exists also another subsequences for which $\mathcal{T}_n$ w.h.p. contains two vertices of outdegree $n/3$.

It is an open problem to find tree limits in such cases. In the case just mentioned with two large vertices (but not more), we conjecture that the tree limit is similar to Υ_ν in Example 3.12, but has a base consisting of a unit interval with the marginal distribution of ν concentrated on the two endpoints.

Remark 10.9. The proof above is based on the result (10.7) for random fringe trees $\mathcal{T}_n^v$ of $\mathcal{T}_n$. However, we also consider the parent of the random node v, see e.g. (10.9); thus we really consider properties of (part of) the *extended fringe tree*, also defined by Aldous [4]. The asymptotic distribution of the entire extended fringe tree was found by Stufler [49]. However, his result is for the annealed version, where the tree $\mathcal{T}_n$ and the vertex v are chosen at random together, while we here need the quenched version, where we fix (i.e., condition on) $\mathcal{T}_n$ and then take a random vertex v. We have therefore used the (quenched) result (10.7) rather than the result of [49]. In fact, the argument above is easily extended to show that for the part of the extended fringe tree up to the first very large ancestor (i.e., w.h.p., $v^\dagger$), the infinite limit tree found by Stufler [49] is also the limit in the quenched sense. However, this does not hold for the remaining part of the extended fringe tree; this part is, for $n - o_{\text{p}}(n)$ choices of v, equal to the part of $\mathcal{T}_n$ between the root and $v^\dagger$, and conditioned on $\mathcal{T}_n$ it is thus w.h.p. equal to some random tree determined by $\mathcal{T}_n$. Consequently, there is no quenched limit of the entire extended fringe tree.

11. Simply generated trees, type III

As said in Remark 9.4, many simply generated trees are covered by the results for conditioned Galton–Watson trees in the preceding sections. However, there are also simply generated trees of a different type (called type III in [29]), where there is no equivalent conditioned Galton–Watson tree. These are defined by weight sequences $(w_k)_k$ such that the power series $\sum_k w_k z^k$ has radius of convergence 0, i.e.,

$$\sum_{k=0}^{\infty} w_k z^k = \infty, \qquad z > 0. \tag{11.1}$$

As shown in [29], such simply generated trees have many similarities with conditioned Galton–Watson trees satisfying (10.1)–(10.2), if we define $\kappa := 0$. In particular, there exists one or several vertices of high outdegree, with total outdegree $n - o_{\text{p}}(n)$. Again, cf. (10.3), we regard as typical the case of complete condensation now defined by

$$\Delta(\mathcal{T}_n) = n - o_{\text{p}}(n), \tag{11.2}$$

so that there is a single vertex $v^\dagger$ that has fathered almost all others. (In fact, then w.h.p. $v^\dagger$ is the root, see [29, (20.2)].) We then have an almost trivial result.

Theorem 11.1. *Let $\mathcal{T}_n$ be a simply generated tree defined by a weight sequence $(w_k)_1^\infty$ satisfying (11.1) and (11.2). Then*

$$\mathcal{T}_n \xrightarrow{\text{p}} \Upsilon_1. \tag{11.3}$$

Proof. If $(\xi_i^{(n)})$ are i.i.d. uniformly random vertices of $\mathcal{T}_n$, then (11.2) shows that w.h.p. $\xi_1^{(n)}$ and $\xi_2^{(n)}$ are children of the node $v^\dagger$ with highest degree. Furthermore, w.h.p., $\xi_1^{(n)} \neq \xi_2^{(n)}$. Hence, w.h.p. $d(\xi_1^{(n)}, \xi_2^{(n)}) = 2$, and the result follows by Theorem 8.2. □

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

The proof shows that (11.3) holds because $\mathcal{T}_n$ "almost" is a star S_n, see Example 7.2.

Example 11.2. *Consider the case $w_k = k!$, which satisfies (11.1). It was shown in [32] that then (11.2) holds. More precisely, if the fringe trees rooted at children of the root are called* branches, *w.h.p. $\mathcal{T}_n$ has a root of degree $n-1-Z_n$, where $Z_n \xrightarrow{\mathrm{d}} \mathrm{Po}(1)$, and of the $n-1-Z_n$ branches, Z_n have size 2 and all others are single vertices (i.e., leaves of $\mathcal{T}_n$).*

The case $w_k = (k!)^\alpha$ with $0 < \alpha < 1$ is similar [32]; there are more branches that have size $\geqslant 2$, and the largest may have size $\lceil 1/\alpha \rceil + 1$, but their number is still $o_{\mathrm{p}}(n)$ and (11.2) holds. If $w_k = (k!)^\alpha$ with $\alpha > 1$, then $\mathcal{T}_n = S_n$ w.h.p.

Thus, (11.3) holds for any $\alpha > 0$.

Remark 11.3. [29, Examples 19.18 and 19.39] give examples where (11.2) does not hold, and there are (at least for some subsequences) several large vertices. It is still true that w.h.p. almost all vertices are at distance 1 from one of the large vertices, so possible subsequence limits (in distribution) of $\mathcal{T}_n$ are determined by the structure of the subtree spanned by the large vertices. We leave further study of this case as an open problem.

12. Logarithmic trees

Many random trees $\mathcal{T}_n$ have heights that w.h.p. are of order $\log n$; we call such trees *logarithmic trees*. (Here, as usual, n measures the size of the tree in some sense. Note, however, that in some examples below, $|\mathcal{T}_n|$ is random and not always equal to n; nevertheless, it is always w.h.p. of order n.) Some examples are binary search trees, random recursive trees, m-ary search trees, digital search trees, preferential attachment trees and tries. Two general classes of such trees (overlapping, and together including the examples just mentioned) are studied in Sections 13 and 14.

In all these cases, it turns out that the random trees $\mathcal{T}_n$ after rescaling have a non-random tree limit (in probability) of the type Υ_a in Example 3.12. More precisely, for some $a \in (0, \infty)$,

$$\frac{1}{\log n}\mathcal{T}_n \xrightarrow{\mathrm{p}} \Upsilon_a. \tag{12.1}$$

We note that by Theorem 8.2, (12.1) is equivalent to

$$\frac{d(\xi_1^{(n)}, \xi_2^{(n)})}{\log n} \xrightarrow{\mathrm{p}} 2a, \tag{12.2}$$

where, as usual, d is the distance in $\mathcal{T}_n$ and $(\xi_i^{(n)})_i$ are i.i.d. vertices in $\mathcal{T}_n$. Equivalently, from our point of view, i.e. with regard to distances between random points, these classes of logarithmic trees behave just like the deterministic binary tree in Example 7.3. (We do not know any natural example of logarithmic random trees that do not satisfy (12.1)–(12.2).)

Remark 12.1. Note that Theorem 8.2 shows that in this case, with convergence to a constant, the annealed result (12.2) is sufficient. To prove (12.1) for some random trees $\mathcal{T}_n$, we therefore may work with annealed results and do not have to show quenched versions (which often are more difficult).

Before considering particular classes of random trees, we note the following simple result, which is used to prove (12.1) in many cases.

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Theorem 12.2. *Let $\mathcal{T}_n$ be random trees such that, as $n \to \infty$, for some $a \in [0, \infty)$,*

$$\frac{d(\xi_1^{(n)}, o)}{\log n} \xrightarrow{\text{p}} a \tag{12.3}$$

and

$$\frac{d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o)}{\log n} \xrightarrow{\text{p}} 0, \tag{12.4}$$

where $(\xi_i^{(n)})_i$ are i.i.d. random vertices in $\mathcal{T}_n$. Then (12.1) *and* (12.2) *hold.*

Proof. We have $d(v, w) = d(v, o) + d(w, o) - 2d(v \wedge w, o)$ for any $v, w \in \mathcal{T}_n$; thus (12.2) follows from (12.3) and (12.4). (Cf. Example 7.3, where the same argument was used.) □

The estimate (12.4) is usually easy, for example by arguments such as in Lemma 13.9 below, so the main task is to prove (12.3); this has been done for many logarithmic random trees.

The distance $d(\xi_1^{(n)}, \xi_2^{(n)})$ between two random vertices has previously been studied in a number of papers for various random trees. These results verify (12.2) and thus (12.1) for several random trees; we give some examples. (The references below show stronger results, which we ignore here.)

Example 12.3. Binary search trees *were studied by Mahmoud and Neininger [40] who showed (in particular)* (12.2) *with $a = 2$. (See also Panholzer and Prodinger [46].) Hence,*

$$\frac{1}{\log n}\mathcal{T}_n \xrightarrow{\text{p}} \Upsilon_2. \tag{12.5}$$

This also follows by any of Theorem 13.1, 13.4 or Theorem 14.3 below.

Example 12.4. Random recursive trees *were studied by Panholzer [45], who showed (in particular)* (12.2) *with $a = 1$. Hence,*

$$\frac{1}{\log n}\mathcal{T}_n \xrightarrow{\text{p}} \Upsilon_1. \tag{12.6}$$

This also follows by Theorem 14.3 below.

Example 12.5. Heap ordered trees *(also called* plane-oriented recursive trees *and* preferential attachment trees*) were studied by Morris, Panholzer and Prodinger [42] who showed (in particular)* (12.2) *with $a = 1/2$. Hence,*

$$\frac{1}{\log n}\mathcal{T}_n \xrightarrow{\text{p}} \Upsilon_{1/2}. \tag{12.7}$$

This also follows by Theorem 14.3 below.

Example 12.6. Random b-ary recursive trees *(*b-ary increasing trees*) were studied by Munsonius and Rüschendorf [43] who showed (in particular)* (12.2) *with $a = b/(b-1)$. Hence,*

$$\frac{1}{\log n}\mathcal{T}_n \xrightarrow{\text{p}} \Upsilon_{b/(b-1)}. \tag{12.8}$$

This also follows by Theorem 14.3 below, or (using [30, Theorem 6.1]) by Theorems 13.1 and 13.4. (The binary search tree in Example 12.3 is the case $b = 2$.)

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Example 12.7. *More generally, for a preferential attachment tree where, in each round, a node with outdegree k gets a child with probability proportional to $\chi k + \rho$, it follows from Theorem 14.3 and [26, Example 6.4] that*

$$\frac{1}{\log n}\mathcal{T}_n \xrightarrow{\text{p}} \Upsilon_{\rho/(\chi+\rho)}. \tag{12.9}$$

Examples 12.3–12.6 are the cases with $(\chi, \rho) = (-1, 2)$, $(0, 1)$, $(1, 1)$, $(-1, b)$, respectively.

We end with an example that, as far as we know, does not follow from the general results in Sections 13 and 14.

Example 12.8. Simple families of increasing trees (simply generated increasing trees) *were studied by Panholzer and Prodinger [47], who showed that if the generating function is a polynomial of degree $d \geqslant 2$, then* (12.2) *holds with $a = d/(d-1)$. Hence,*

$$\frac{1}{\log n}\mathcal{T}_n \xrightarrow{\text{p}} \Upsilon_{d/(d-1)}. \tag{12.10}$$

13. Split trees

Random *split trees* were introduced by Devroye [14] as a unified model that includes many important families of random trees (of logarithmic height), for example binary search trees, m-ary search trees, tries and digital search trees. Theorem 13.1 below shows that random split trees after rescaling have a non-random tree limit of the type Υ_a in Example 3.12. Equivalently, by Theorem 8.2, distances between random points satisfy (12.2).

The definition of split trees involves several parameters b, s, s_0, s_1 and a *split vector* $\mathcal{V} = (V_1, \dots, V_b)$ which is a random vector with $V_i \geqslant 0$ and $\sum_{i=1}^b V_i = 1$, i.e., a random probability distribution on $\{1, \dots, b\}$. The split tree is defined as a subtree of the infinite b-ary tree T_b. The tree is constructed by adding a sequence of n *balls* to the tree, which initially is empty. Each ball arrives at the root and then moves recursively as follows; see [14] for further details.

Each vertex is equipped with its own copy $\mathcal{V}^{(v)}$ of the random split vector $\mathcal{V}$; these copies are independent. Each vertex has maximum capacity $s \geqslant 1$; the first s balls that arrive at a vertex stay there (temporarily), but when the $(s+1)$th ball arrives at the vertex, some balls are sent to its children, leaving $s_0 \in [0, s]$ balls that remain in the vertex for ever. (The details of this step depend on s_1, see [14].) Any further ball that comes to the vertex is immediately passed along to one of its children, with probability $V_i^{(v)}$ for child i and independently of all previous events.

The split tree $\mathcal{T}_n$ is defined as the set of all vertices that have been visited by a ball; note that (if $s_0 = 0$) some vertices in $\mathcal{T}_n$ may be empty, but there is always at least one ball in some descendant of the vertex.

We exclude the trivial case $\max(V_1, \dots, V_b) = 1$ a.s., and then $\mathcal{T}_n$ is finite a.s. (Usually one assumes the slightly stronger $V_i < 1$ a.s. for every i [14].)

It is important to note that $\mathcal{T}_n$ is defined with a fixed number n of balls, while the number of vertices $|\mathcal{T}_n|$ in general is random. Nevertheless, it is easy to see that

$$\mathbb{E}\,|\mathcal{T}_n| = O(n). \tag{13.1}$$

Furthermore, since each node stores at most s balls, we have a deterministic lower bound

$$|\mathcal{T}_n| \geqslant n/s. \tag{13.2}$$

In fact, in most cases $\mathbb{E}\,|\mathcal{T}_n|/n$ converges to some constant, and, moreover, $|\mathcal{T}_n|/n$ converges in probability to the same constant, see [25, Theorem 1.1]. However, this is not always the case; for some tries, $\mathbb{E}\,|\mathcal{T}_n|/n$ oscillates.

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

We define

$$\chi := \sum_{i=1}^{b} \mathbb{E}\big[V_i \log(1/V_i)\big], \tag{13.3}$$

and note that $0 < \chi < \infty$.

Theorem 13.1. *Let $\mathcal{T}_n$ be a random split tree with a split vector $\mathcal{V} = (V_1, \ldots, V_b)$ and let χ be given by (13.3). Then,*

$$\frac{1}{\log n}\mathcal{T}_n \xrightarrow{\text{p}} \Upsilon_{1/\chi}. \tag{13.4}$$

Proof. As said in Section 12, by Theorem 8.2, (13.4) is equivalent to

$$\frac{d(\xi_1^{(n)}, \xi_2^{(n)})}{\log n} \xrightarrow{\text{p}} \frac{2}{\chi}. \tag{13.5}$$

where $(\xi_i^{(n)})_i$ are i.i.d. vertices in $\mathcal{T}_n$. Under a technical condition, (13.5) was proved by Berzunza, Cai and Holmgren [8, Corollary 1], as a corollary to some stronger estimates. (Actually, their d is slightly different, and includes the distance to the root, but the same proof yields (13.5).)

For completeness, we give a proof (by similar methods) in the following subsection, not requiring any further conditions; in fact, we consider there an even more general model. ☐

Without going into details, we note that the proof of (13.5) in [8], as well as our similar proof in Section 13.1, is based on showing the two results (12.3) and (12.4), and that (12.3) was shown by Holmgren [25].

Note also the related fact that if $\eta^{(n)}$ is a random *ball* in $\mathcal{T}_n$, then

$$\frac{d(\eta^{(n)}, o)}{\log n} \xrightarrow{\text{p}} \frac{1}{\chi}. \tag{13.6}$$

This was proved by Devroye [14, Theorem 2], see also the stronger results by Holmgren [25, Theorem 1.3] (under a weak technical assumption) and Berzunza, Cai and Holmgren [8, Lemma 13(ii)] (Actually, Devroye [14] considered the depth of the last added ball, and not a random one, but that easily implies the result (13.6) by Holmgren [25, Proposition 1.1], arguing as in Holmgren [25, proof of Corollary 1.1].)

Furthermore, Ryvkina [48] showed (in particular) the corresponding fact for the distance between two random balls:

$$\frac{d(\eta_1^{(n)}, \eta_2^{(n)})}{\log n} \xrightarrow{\text{p}} \frac{2}{\chi}. \tag{13.7}$$

See also Albert, Holmgren, Johansson and Skerman [3, Lemma 3.4], showing that $d(\eta_1^{(n)} \wedge \eta_2^{(n)}, o)$ is tight, which together with (13.6) implies (13.7),

Remark 13.2. In analogy with Theorem 8.2, we can interpret (13.7) as convergence

$$\frac{1}{\log n}(\mathcal{T}_n, \mu_n^*) \xrightarrow{\text{p}} \Upsilon_{1/\chi}, \tag{13.8}$$

where we equip $\mathcal{T}_n$ with the probability measure μ_n^* defined as the distribution of the balls on $\mathcal{T}_n$.

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

13.1 Generalized split trees

We define random *generalized split trees* as follows; this is a minor variation of the model in Broutin, Devroye, McLeish and de la Salle [13]. Let $2 \leqslant b < \infty$ be a fixed branching factor and suppose that for every integer $n \geqslant 1$ we have a random vector $\mathcal{N}^{(n)} = (N_i^{(n)})_{i=0}^b$ with $N_i^{(n)} \in \mathbb{N}_0$ and

$$\sum_{i=0}^{b} N_i^{(n)} = n. \tag{13.9}$$

Consider the infinite b-ary tree T_b. For a given number n of balls, all starting at the root, distribute the balls according to $\mathcal{N}^{(n)}$, with $N_0^{(n)}$ balls remaining in the root (for ever), and $N_i^{(n)}$ balls passed to the ith child. Continue recursively in each subtree that has received at least one ball, using an independent copy of $\mathcal{N}^{(m)}$ at each vertex that has received m balls.

It is convenient to begin by equipping each vertex v in the infinite tree T_b with a private copy $\mathcal{N}^{(n,v)}$ of $\mathcal{N}^{(n)}$ for each $n \geqslant 1$, with all these random vectors $\mathcal{N}^{(n,v)}$ independent. Then, at each vertex v that receives $m \geqslant 1$ balls, we apply $\mathcal{N}^{(m,v)}$.

The tree $\mathcal{T}_n$ is defined as the set of all vertices that have received at least one ball (whether or not any ball remains there). Equivalently, $\mathcal{T}_n$ is the set of all vertices $v \in \mathrm{T}_b$ such that the fringe tree T_b^v contains at least one ball. Note again that the size $|\mathcal{T}_n|$ is random.

We assume the following:

(ST1) There exists a constant C_0 such that for every n, a.s.,

$$0 \leqslant N_0^{(n)} \leqslant C_0. \tag{13.10}$$

(ST2) The random vector $n^{-1}\mathcal{N}^{(n)}$ converges in distribution as $n \to \infty$:

$$\frac{1}{n}\mathcal{N}^{(n)} = \Big(\frac{N_i^{(n)}}{n}\Big)_{i=0}^{b} \xrightarrow{\mathrm{d}} \mathcal{V} = \big(V_i\big)_{i=0}^{b}. \tag{13.11}$$

(ST3) For every $n \geqslant 1$,

$$\mathbb{P}\big(\max_{1\leqslant i\leqslant b} N_i^{(n)} = n\big) < 1 \tag{13.12}$$

and, similarly,

$$\mathbb{P}\big(\max_{1\leqslant i\leqslant b} V_i = 1\big) < 1. \tag{13.13}$$

We call the limit $\mathcal{V}$ in (13.11) the (asymptotic) *split vector*. Note that $V_0 = 0$ by (ST1) and (ST2); thus it suffices to consider $(V_i)_1^b$. Furthermore, (13.11) implies

$$\sum_{i=1}^{b} V_i = 1 \qquad \text{a.s.} \tag{13.14}$$

Thus, $\mathcal{V}$ is a random probability distribution on $\{1, \dots, b\}$.

It should be clear that the definition above includes the split trees defined by Devroye [14] and discussed above. (In particular, our model includes tries, unlike the version in [13].)

Remark 13.3. (13.12) only excludes the trivial case when, for some n, a.s. all n balls are passed to the same child, and therefore, by induction, continue along some infinite path so that $\mathcal{T}_n$ becomes infinite.

Conversely, it is easy to see by induction that (13.12) implies that $\mathcal{T}_n$ is finite a.s. for every $n \geqslant 1$.

Moreover, (13.13) implies uniformity in (13.12): it is easy to see that (13.12)–(13.13) is equivalent to the existence of $c, \delta > 0$ such that, for every $n \geqslant 1$,

$$\mathbb{P}\big(\max_{1\leqslant i\leqslant b} N_i^{(n)} > (1-\delta)n\big) \leqslant 1-c. \tag{13.15}$$

Theorem 13.4. *Let $\mathcal{T}_n$ be a random generalized split tree with a split vector $\mathcal{V} = (V_1, \ldots, V_b)$ and let χ be given by (13.3). Then,*

$$\frac{d(\xi_1^{(n)}, \xi_2^{(n)})}{\log n} \xrightarrow{\text{p}} \frac{2}{\chi}, \tag{13.16}$$

where d is the distance in $\mathcal{T}_n$ and $(\xi_i^{(n)})_i$ are i.i.d. vertices in $\mathcal{T}_n$. Consequently,

$$\frac{1}{\log n}\mathcal{T}_n \xrightarrow{\text{p}} \Upsilon_{1/\chi}. \tag{13.17}$$

To prove Theorem 13.4, we show a series of lemmas. We define random variables $W^{(n)}$ and W as size-biased selections from $N^{(n)}/n$ and $\mathcal{V}$. More precisely, conditionally on $N^{(n)}$, we select an index I with distribution $\mathbb{P}\,(I = i \mid N^{(n)}) = N_i^{(n)}/n$, and then define

$$W^{(n)} := \begin{cases} N_I^{(n)}/n, & I \geqslant 1, \\ 1, & I = 0. \end{cases} \tag{13.18}$$

(The special definition in the case $I = 0$, which has probability $O(1/n)$ only, will be convenient below.) Similarly, conditionally on $\mathcal{V}$ we select I with $\mathbb{P}\,(I = i \mid \mathcal{V}) = V_i$, and then take

$$W := V_I. \tag{13.19}$$

It follows from (ST2) that

$$W^{(n)} \xrightarrow{\text{d}} W. \tag{13.20}$$

Note that

$$\mathbb{E}\big(-\log(W^{(n)})\big) = \mathbb{E}\sum_{i=1}^{b} \frac{N_i^{(n)}}{n}\Big(-\log\frac{N_i^{(n)}}{n}\Big) \tag{13.21}$$

and, by (13.3),

$$\mathbb{E}\big(-\log W\big) = \mathbb{E}\sum_{i=1}^{b} V_i\big(-\log V_i\big) = \chi. \tag{13.22}$$

Lemma 13.5. *We may couple $-\log W^{(n)}$ with a copy $\zeta^{(n)}$ of $\zeta := -\log W$ such that*

$$\mathbb{E}\,|\zeta^{(n)} + \log W^{(n)}| \to 0 \qquad \text{as } n \to \infty. \tag{13.23}$$

Proof. By (13.20), we have

$$\log W^{(n)} \xrightarrow{\text{d}} \log W = -\zeta. \tag{13.24}$$

By the Skorohod coupling theorem [34, Theorem 4.30], we may assume that (13.24) holds a.s., and thus

$$\log W^{(n)} + \zeta \xrightarrow{\text{a.s.}} 0. \tag{13.25}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Furthermore,

$$\mathbb{E}\big[(\log W^{(n)})^2\big] = \mathbb{E}\sum_{i=1}^{b}\frac{N_i^{(n)}}{n}\Big(-\log\frac{N_i^{(n)}}{n}\Big)^2 \leqslant C, \tag{13.26}$$

since $x\log^2 x$ is bounded on $[0, 1]$, and similarly $\mathbb{E}\,\zeta^2 = \mathbb{E}\,[(\log W)^2] \leqslant C$. Hence the sequence $\mathbb{E}\,(\log W^{(n)} + \zeta)^2$ is uniformly bounded, and thus the sequence $\log W^{(n)} + \zeta$ is uniformly integrable [23, Theorem 5.4.2]. Consequently, (13.25) implies $\mathbb{E}\,|\log W^{(n)} + \zeta| \to 0$ [23, Theorem 5.5.2]. □

Let $\widehat{N}_v$ be the number of balls received by vertex $v \in \mathrm{T}_b$. Thus

$$\mathcal{T}_n = \{v \in \mathrm{T}_b : \widehat{N}_v \geqslant 1\}. \tag{13.27}$$

Lemma 13.6. (i) *There exists a constant C such that, for all n,*

$$\mathbb{E}\,|\mathcal{T}_n| \leqslant Cn \tag{13.28}$$

and, more generally, for any K,

$$\mathbb{E}\big|\{v \in \mathcal{T}_n : \widehat{N}_v \geqslant K\}\big| \leqslant Cn/K. \tag{13.29}$$

Furthermore,

$$|\mathcal{T}_n| \leqslant Cn \qquad w.h.p. \tag{13.30}$$

(ii) *Deterministically,*

$$|\mathcal{T}_n| \geqslant cn. \tag{13.31}$$

Proof. First, (13.31) follows immediately from the fact that by (ST1), no vertex contains more than C_0 balls when the construction is finished; hence there are at least n/C_0 vertices containing balls.

For (13.28), recall (13.15), and assume as we may that $\delta < 1/2$. Let $r := 1/(1-\delta) < 2$, and let X_k be the number of vertices v such that $\widehat{N}_v \in [r^k, r^{k+1})$. For a given $k \geqslant 0$, generate the tree as usual, but stop at every vertex v that receives $\widehat{N}_v < r^{k+1}$ balls, and colour these vertices pink. If a pink vertex v has $\widehat{N}_v \geqslant r^k$, recolour it red. Since the red vertices receive disjoint sets of balls, the number R_k of them is at most n/r^k. Condition on the set of red vertices and the numbers of balls in them, and continue the construction of the tree. Since $r < 2$, each red vertex has at most one child w with $\widehat{N}_w \geqslant r^k$, and by (13.15), with probability at least c it has none. Continuing, we see that for each red vertex, the number of descendants w with $\widehat{N}_w \geqslant r^k$ is dominated by a geometric distribution, and thus the expected number of such descendants is $O(1)$. Consequently, $\mathbb{E}\big(X_k \mid R_k\big) \leqslant CR_k$, and thus

$$\mathbb{E}\,X_k \leqslant C\,\mathbb{E}\,R_k \leqslant C\frac{n}{r^k}. \tag{13.32}$$

This yields

$$\mathbb{E}\,|\mathcal{T}_n| = \mathbb{E}\sum_{k=0}^{\infty} X_k \leqslant \sum_{k=0}^{\infty} C\frac{n}{r^k} = Cn, \tag{13.33}$$

which is (13.28).

We obtain (13.29) in the same way, summing only over k with $r^{k+1} > K$.

Finally, the argument above shows that X_k is stochastically dominated by a sum of $\lfloor n/r^k \rfloor$ independent copies of a geometric random variable ζ. Furthermore, we may choose these to be

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

independent also for different k. (The red vertices for different k are not independent, but the stochastic upper bound that we use holds also conditioned on events for larger k.) Hence,

$$|\mathcal{T}_n| \leqslant \sum_{i=1}^{m_n} \zeta_i, \tag{13.34}$$

where $\zeta_i \in \mathrm{Ge}\,(p)$ are i.i.d. with some fixed $0 < p < 1$, and $m_n := \sum_{k=0}^{\infty} \lfloor n/r^k \rfloor \leqslant Cn$. Hence, (13.30) follows by the law of large numbers. □

Lemma 13.7. *Let $\chi > 0$ be given by (13.3). If D_n is the depth of a random ball in $\mathcal{T}_n$, then*

$$\frac{D_n}{\log n} \xrightarrow{\mathrm{p}} \frac{1}{\chi}. \tag{13.35}$$

Proof. Consider a random ball, and suppose that it follows a path $v_0 = o, v_1, \dots, v_D$, ending up at a vertex v_D of depth $D = D_n$. For completeness, define $v_j := v_D$ for $j > D$. Let, for $k \geqslant 0$,

$$Y_k := \log \widehat{N}_{v_{k-1}} - \log \widehat{N}_{v_k} = -\log \frac{\widehat{N}_{v_k}}{\widehat{N}_{v_{k-1}}} \geqslant 0, \qquad k \geqslant 0. \tag{13.36}$$

For $k \geqslant 0$, let $\mathcal{F}_k$ be the σ-field generated by $\mathcal{N}^{(m,v)}$ with $m \geqslant 1$ and $d(v, o) < k$, together with v_j for $j \leqslant k$. Then $\widehat{N}_{v_k}$ and Y_k are $\mathcal{F}_k$-measurable, and so is the event $\{D \geqslant k\} = \{v_k \neq v_{k-1}\}$. Conditioned on $\mathcal{F}_k$, and assuming $\widehat{N}_{v_k} = m$ and $D \geqslant k$, Y_{k+1} has by the definitions (13.36) and (13.18) the same distribution as $-\log W^{(m)}$.

Let $\lambda > 0$ be fixed and let $\varepsilon > 0$. By Lemma 13.5, there exists $B = B_\varepsilon$ such that if $m \geqslant B$, then we can couple $-\log W^{(m)}$ with $\zeta := -\log W$ such that $\mathbb{E}\,|\zeta + \log W^{(m)}| < \varepsilon$.

Let $L = \lfloor \lambda \log n \rfloor$ and define the stopping time τ as the smallest k such that one of the following occurs.

(a) $k \geqslant L$,
(b) $\widehat{N}_{v_k} \leqslant B$,
(c) $k > D$, and thus the ball has come to rest.

By the comments just made, we can couple the sequence $(Y_k)_k$ with an i.i.d. sequence $(\zeta_k)_1^\infty$ with $\zeta_k \overset{\mathrm{d}}{=} -\log W$ such that on the event $\{\tau > k\} \in \mathcal{F}_k$,

$$\mathbb{E}\big(|Y_{k+1} - \zeta_{k+1}| \mid \mathcal{F}_k\big) < \varepsilon. \tag{13.37}$$

Let, recalling (13.36),

$$X := \sum_{k=1}^{\tau} \big(Y_k - \zeta_k\big) = \log n - \log \widehat{N}_{v_\tau} - \sum_{k=1}^{\tau} \zeta_k. \tag{13.38}$$

Then, (13.37) implies

$$\mathbb{E}\,|X| \leqslant \sum_{k=1}^{L} \mathbb{E}\big|\big(Y_k - \zeta_k\big)\mathbf{1}\{k \leqslant \tau\}\big| \leqslant L\varepsilon. \tag{13.39}$$

Let $\mathcal{E}_-$ be the event $\{D < L - 1\}$, and let $\mathcal{E}'_- := \mathcal{E}_- \cap \{\widehat{N}_{v_\tau} > B\}$ and $\mathcal{E}''_- := \mathcal{E}_- \cap \{\widehat{N}_{v_\tau} \leqslant B\}$. First, if $\mathcal{E}'_-$ occurs, then in the definition of τ, neither (a) nor (b) may occur. (If $\tau \geqslant L$, then already $\tau - 1 > D$, a contradiction.) Hence, (c) occurs, and thus the ball has v_τ as its final position. Let $\mathcal{S} := \{v \in \mathrm{T}_b \colon \widehat{N}_v > B\}$. By the definition of $\mathcal{E}'_-$, we have $v_\tau \in \mathcal{S}$, and thus the ball ends up in the set

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

$\mathcal{S}$. By (ST1), there are at most $C_0|\mathcal{S}|$ such balls, and thus the conditional probability given $\mathcal{T}_n$ that our random ball is one of them is $\leqslant C_0|\mathcal{S}|/n$. Hence, by (13.29),

$$P(\mathcal{E}'_-) \leqslant \frac{\mathbb{E}\,(C_0|\mathcal{S}|)}{n} \leqslant \frac{C_1}{B}. \tag{13.40}$$

We may increase B if necessary so that $B > C_1/\varepsilon$, and thus $\mathbb{P}\,(\mathcal{E}'_-) < \varepsilon$.

On the other hand, if $\mathcal{E}''_-$ holds, then, by (13.38),

$$X \geqslant \log n - \log B - \sum_{k=1}^{\tau} \zeta_k \geqslant \log n - \log B - \sum_{k=1}^{L} \zeta_k. \tag{13.41}$$

Thus, by the law of large numbers, recalling that $\mathbb{E}\,\zeta_k = \chi$ by (13.22), on the event $\mathcal{E}''_-$, w.h.p.

$$X \geqslant \log n - \log B - L(\chi + \varepsilon) \geqslant \bigl(1 - \lambda(\chi + \varepsilon) - \varepsilon\bigr) \log n. \tag{13.42}$$

If $\lambda < 1/\chi$, and ε is so small that $\lambda(\chi + \varepsilon) + \varepsilon < 1$, (13.42), (13.39) and Markov's inequality yield

$$\mathbb{P}\,(\mathcal{E}''_-) \leqslant \frac{L\varepsilon}{\bigl(1 - \lambda(\chi + \varepsilon) - \varepsilon\bigr) \log n} + o(1) \leqslant \frac{\lambda\varepsilon}{1 - \lambda(\chi + \varepsilon) - \varepsilon} + o(1). \tag{13.43}$$

Hence,

$$\mathbb{P}\,(D < L - 1) = \mathbb{P}\,(\mathcal{E}_-) = \mathbb{P}\,(\mathcal{E}'_-) + \mathbb{P}\,(\mathcal{E}''_-) \leqslant \varepsilon + \varepsilon \frac{\lambda}{1 - \lambda(\chi + \varepsilon) - \varepsilon} + o(1). \tag{13.44}$$

Letting $\varepsilon \to 0$, we see that $\mathbb{P}\,(D \leqslant \lambda \log n - 2) \leqslant \mathbb{P}\,(D < L - 1) \to 0$. In other words, for any $\lambda < 1/\chi$,

$$D > \lambda \log n - 2 \qquad \text{w.h.p.} \tag{13.45}$$

For the other side, assume $\lambda > 1/\chi$, and let $\mathcal{E}_+$ be the event $\{D \geqslant L\}$. Let $\mathcal{E}'_+ := \{\tau = L\}$ and $\mathcal{E}''_+ := \mathcal{E}_+ \cap \{\tau < L\}$.

The law of large numbers and (13.38) imply that on the event $\mathcal{E}'_+$, w.h.p.,

$$X \leqslant \log n - \sum_{k=1}^{L} \zeta_k \leqslant \log n - (\lambda\chi - \varepsilon) \log n = -(\lambda\chi - 1 - \varepsilon) \log n. \tag{13.46}$$

Hence, if ε is small enough, (13.39) and Markov's inequality yield

$$\mathbb{P}\,(\mathcal{E}'_+) \leqslant \frac{L\varepsilon}{(\lambda\chi - 1 - \varepsilon) \log n} + o(1) \leqslant \frac{\lambda\varepsilon}{\lambda\chi - 1 - \varepsilon} + o(1). \tag{13.47}$$

If $\mathcal{E}''_+$ holds, then (a) and (c) cannot hold, and thus $\widehat{N}_{v_\tau} \leqslant B$. Hence our chosen ball belongs to a subtree rooted at v_τ with at most B balls. Conditioned on $\widehat{N}_{v_\tau} = m$, this subtree is a copy of $\mathcal{T}_m$, and since the finitely many random trees $\mathcal{T}_m$, $1 \leqslant m \leqslant B$, all are a.s. finite and thus have finite (random) heights $H(\mathcal{T}_m)$, there exists a constant C_2 such that

$$\mathbb{P}\bigl(H(\mathcal{T}_m) > C_2\bigr) \leqslant \varepsilon, \qquad m = 1, \dots, B. \tag{13.48}$$

It follows that conditioned on $\mathcal{E}''$,

$$\mathbb{E}\bigl(D > L + C_2 \mid \mathcal{E}''_+\bigr) \leqslant \mathbb{E}\bigl(D > \tau + C_2 \mid \mathcal{E}''_+\bigr) \leqslant \varepsilon. \tag{13.49}$$

Finally, combining (13.47) and (13.49) we obtain

$$\begin{aligned} \mathbb{P}\bigl(D > L + C_2\bigr) &\leqslant \mathbb{P}\bigl(\mathcal{E}'\bigr) + \mathbb{P}\bigl(D > L + C_2 \text{ and } \mathcal{E}''\bigr) \\ &\leqslant \frac{\lambda\varepsilon}{\lambda\chi - 1 - \varepsilon} + \varepsilon + o(1). \end{aligned} \tag{13.50}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

The constant C_2 may depend on ε, but it follows that for large n,

$$\mathbb{P}\big(D > (\lambda+\varepsilon)\log n\big) \leqslant \mathbb{P}\big(D > L + C_2\big) \leqslant \frac{\lambda}{\lambda\chi - 1 - \varepsilon}\varepsilon + \varepsilon + o(1). \tag{13.51}$$

Since ε can be arbitrarily small, this shows that for any $\lambda > 1/\chi$ and $\delta > 0$,

$$D \leqslant (\lambda+\delta)\log n \qquad \text{w.h.p.}, \tag{13.52}$$

which together with (13.45) completes the proof. □

We transfer this result from balls to vertices.

Lemma 13.8. *Let $\mathcal{T}_n$ and $\xi_i^{(n)}$ be as above. Then*

$$\frac{d(\xi_1^{(n)}, o)}{\log n} \xrightarrow{\text{p}} \frac{1}{\chi}. \tag{13.53}$$

Proof. Again, let $L := \lfloor \lambda \log n \rfloor$ for a fixed $\lambda > 0$. Let $\mathcal{Z}_k$ be the set of vertices of $\mathcal{T}_n$ with depth k, and $\mathcal{Z}_k^{\text{b}}$ the set of balls with depth k; define $\mathcal{Z}_{\leqslant k}$, $\mathcal{Z}_{\geqslant k}$, $\mathcal{Z}^{\text{b}}_{\leqslant k}$, $\mathcal{Z}^{\text{b}}_{\geqslant k}$ analogously.

First, let $\lambda > 1/\chi$. Let U_L be the set of all vertices of T_b with depth L. For any $v \in \mathrm{T}_b$, conditioned on $\widehat{N}_v$, the fringe subtree $\mathcal{T}_n^v$ has the same distribution as $\mathcal{T}_m$ with $m = \widehat{N}_v$. Consequently, Lemma 13.6 shows that

$$\mathbb{E}\big(|\mathcal{T}_n^v| \mid \widehat{N}_v\big) \leqslant C\widehat{N}_v \tag{13.54}$$

and thus

$$\mathbb{E}\,|\mathcal{T}_n^v| \leqslant C\,\mathbb{E}\,\widehat{N}_v. \tag{13.55}$$

Since $\mathcal{Z}_{\geqslant L}$ is the union of the fringe trees $\mathcal{T}_n^v$ for $v \in U_L$, and $\mathcal{Z}^{\text{b}}_{\geqslant L}$ is the set of all balls that reach some vertex in U_L, it follows from (13.55) that

$$\mathbb{E}\,|\mathcal{Z}_{\geqslant L}| = \mathbb{E}\sum_{v\in U_L} |\mathcal{T}_n^v| \leqslant C \sum_{v\in U_L} \mathbb{E}\,\widehat{N}_v = C\,\mathbb{E}\sum_{v\in U_L} \widehat{N}_v = C\,\mathbb{E}\,|\mathcal{Z}^{\text{b}}_{\geqslant L}|. \tag{13.56}$$

However, we have by Lemma 13.7,

$$\mathbb{E}\,|\mathcal{Z}^{\text{b}}_{\geqslant L}| = n\,\mathbb{P}\,(D_n \geqslant L) = o(n). \tag{13.57}$$

Combining (13.56) and (13.57) yields, recalling (13.31),

$$\mathbb{P}\big(\xi_i^{(n)} \geqslant L\big) = \mathbb{E}\,\frac{|\mathcal{Z}_{\geqslant L}|}{|\mathcal{T}_n|} \leqslant C\,\mathbb{E}\,\frac{|\mathcal{Z}_{\geqslant L}|}{n} \leqslant C\frac{\mathbb{E}\,|\mathcal{Z}^{\text{b}}_{\geqslant L}|}{n} = o(1). \tag{13.58}$$

In the opposite direction, let $\lambda < 1/\chi$. Let $\varepsilon > 0$ and let B be a large number. We split $\mathcal{Z}_{\leqslant L}$ into the two sets $\mathcal{Z}'_{\leqslant L} := \{v \in \mathcal{Z}_{\leqslant L} : \widehat{N}_v > B\}$ and $\mathcal{Z}''_{\leqslant L} := \{v \in \mathcal{Z}_{\leqslant L} : \widehat{N}_v \leqslant B\}$. By (13.29), we may choose B so large that

$$\mathbb{E}\,|\mathcal{Z}'_{\leqslant L}| \leqslant \varepsilon n. \tag{13.59}$$

To treat $\mathcal{Z}''_{\leqslant L}$, we now stop the construction of $\mathcal{T}_n$ at each vertex v with $\widehat{N}_v \leqslant B$. If such a vertex also has depth $\leqslant L$, we colour it green. Let $\mathcal{G}$ be the set of all green vertices. Then the set $\mathcal{Z}''_{\leqslant L}$ is included in the union of the fringe trees $\mathcal{T}_n^v$ for $v \in \mathcal{G}$. Furthermore, conditioned on the set $\mathcal{G}$ and $(\widehat{N}_v)_{v\in\mathcal{G}}$, each $\mathcal{T}_n^v$ (for $v \in \mathcal{G}$) has the same distribution as $\mathcal{T}_m$ for $m = \widehat{N}_v$. Thus, by Lemma 13.6,

$$\begin{aligned}\mathbb{E}\big(|\mathcal{Z}''_{\leqslant L}| \mid \mathcal{G}|, (\widehat{N}_v)_{v\in\mathcal{G}}\big) &\leqslant \sum_{v\in\mathcal{G}} \mathbb{E}\big(|\mathcal{T}_n^v| \mid \mathcal{G}, (\widehat{N}_v)_{v\in\mathcal{G}}\big) \leqslant \sum_{v\in\mathcal{G}} C\widehat{N}_v \\ &\leqslant CB|\mathcal{G}| = C|\mathcal{G}|. \end{aligned} \tag{13.60}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Consequently,

$$\mathbb{E}\,|\mathcal{Z}''_{\leqslant L}| \leqslant C\,\mathbb{E}\,|\mathcal{G}|. \tag{13.61}$$

Next, let again C_2 be such that (13.48) holds, with ε replaced by $1/2$. Then, still conditioned on $\mathcal{G}$ and $(\widehat{N}_v)_{v\in\mathcal{G}}$, (13.48) shows that each fringe tree $\mathcal{T}_n^v$ (for $v\in\mathcal{G}$) with probability $\geqslant 1/2$ has height $\leqslant C_2$; if this happens, $\mathcal{T}_n^v$ has in particular at least one ball of depth $\leqslant C_2$ in the fringe tree, and thus depth $\leqslant L+C_2$ in $\mathcal{T}_n$. Hence,

$$\mathbb{E}\big(|\mathcal{Z}^{\mathsf{b}}_{\leqslant L+C_2}|\mid\mathcal{G}\big) \geqslant \tfrac{1}{2}|\mathcal{G}|. \tag{13.62}$$

Together with (13.61), this yields

$$\mathbb{E}\,|\mathcal{Z}''_{\leqslant L}| \leqslant C\,\mathbb{E}\,|\mathcal{G}| \leqslant C\,\mathbb{E}\,|\mathcal{Z}^{\mathsf{b}}_{\leqslant L+C_2}| \tag{13.63}$$

and then Lemma 13.7 implies

$$\mathbb{E}\,|\mathcal{Z}''_{\leqslant L}| \leqslant C\,\mathbb{E}\,|\mathcal{Z}^{\mathsf{b}}_{\leqslant L+C_2}| = Cn\,\mathbb{P}\big(D_n \leqslant L+C_2\big) = o(n). \tag{13.64}$$

By (13.59) and (13.64), we have for large n

$$\mathbb{E}\,|\mathcal{Z}_{\leqslant L}| = \mathbb{E}\,|\mathcal{Z}'_{\leqslant L}| + \mathbb{E}\,|\mathcal{Z}''_{\leqslant L}| \leqslant 2\varepsilon n. \tag{13.65}$$

Thus, $\mathbb{E}\,|\mathcal{Z}_{\leqslant L}| = o(n)$, and, similarly to (13.58),

$$\mathbb{P}\big(\xi_i^{(n)} \leqslant L\big) = \mathbb{E}\,\frac{|\mathcal{Z}_{\leqslant L}|}{|\mathcal{T}_n|} \leqslant C\,\mathbb{E}\,\frac{|\mathcal{Z}_{\leqslant L}|}{n} = o(1). \tag{13.66}$$

This completes the proof together with (13.58). □

Lemma 13.9. *With notations as above,*

$$\frac{d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o)}{\log n} \xrightarrow{\text{p}} 0. \tag{13.67}$$

Proof. There is a standard identification of the vertices of T_b with finite strings $i_1\cdots i_k$ with $k\geqslant 0$ and $i_j\in\{1,\dots,b\}$. If $v=i_1\cdots i_k\in\mathrm{T}_b$, let $v_j := i_1\cdots i_j$, $j\leqslant k$, and define

$$\widehat{V}_v := \prod_{j=0}^{k-1} V_{i_{j+1}}^{(v_j)}. \tag{13.68}$$

Then [13, Lemma 2], by (ST2) and induction over k, as $n\to\infty$,

$$\widehat{N}_v/n \xrightarrow{\text{p}} \widehat{V}_v, \qquad v\in\mathrm{T}_b. \tag{13.69}$$

Furthermore, it follows from (13.30) that for any $\varepsilon>0$ and any fixed $v\in\mathrm{T}_b$, w.h.p.

$$|\mathcal{T}_n^v| \leqslant C\widehat{N}_v + o_{\mathrm{p}}(n). \tag{13.70}$$

(The term $o_{\mathrm{p}}(n)$ takes care of the possibility that $\widehat{N}_v$ is small; we have not excluded the case $\widehat{V}_v=0$.) By (13.69) and (13.70),

$$|\mathcal{T}_n^v| \leqslant (C\widehat{V}_v + o_{\mathrm{p}}(1))n \tag{13.71}$$

and thus, for any fixed K,

$$\sum_{v\in U_K} |\mathcal{T}_n^v|^2 \leqslant \sum_{v\in U_K} (C\widehat{V}_v + o_{\mathrm{p}}(1))^2 n^2. \tag{13.72}$$

Since $d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o) \geqslant K)$ if and only if the two vertices $\xi_1^{(n)}$ and $\xi_2^{(n)}$ are in the same subtree $\mathcal{T}_n^v$ for some $v \in U_K$, it follows from (13.72) that, using also (13.31) and $\sum_{v\in U_K} \widehat{V}_v = 1$,

$$\begin{aligned}\mathbb{P}\big(d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o) \geqslant K \mid \mathcal{T}_n)\big) &= \frac{1}{|\mathcal{T}_n|^2} \sum_{v\in U_K} |\mathcal{T}_n^v|^2 \\ &\leqslant C \sum_{v\in U_K} (C\widehat{V}_v + o_{\mathrm{p}}(1))^2 \\ &= C \sum_{v\in U_K} \widehat{V}_v^2 + o_{\mathrm{p}}(1). \end{aligned} \tag{13.73}$$

Since the probability on the left-hand side is bounded by 1, we may assume that so is the term $o_{\mathrm{p}}(1)$ on the right-hand side, and thus we may take the expectation and use dominated convergence to conclude

$$\mathbb{P}\big(d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o) \geqslant K\big) \leqslant C\,\mathbb{E} \sum_{v\in U_K} \widehat{V}_v^2 + o(1). \tag{13.74}$$

Furthermore, by the definition (13.68) and independence,

$$\mathbb{E} \sum_{v\in U_K} \widehat{V}_v^2 = \sum_{i_1,\dots,i_K} \prod_{j=1}^{K} \mathbb{E}\, V_{i_j}^2 = \Big(\sum_{i=1}^{b} \mathbb{E}\, V_i^2\Big)^K. \tag{13.75}$$

Since $\sum_i V_i^2 \leqslant \sum_i V_i = 1$, and strict inequality holds with positive probability,

$$\sum_{i=1}^{b} \mathbb{E}\, V_i^2 = \mathbb{E} \sum_{i=1}^{b} V_i^2 < 1. \tag{13.76}$$

Hence, given any $\varepsilon > 0$, we can find K such that (13.74) yields

$$\mathbb{P}\big(d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o) \geqslant K)\big) \leqslant C\frac{\varepsilon}{2C} + o(1) < \varepsilon \tag{13.77}$$

for large n. In particular, (13.67) follows. (In fact, we have proved that the sequence $d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o)$ of random variables is tight.) □

Proof of Theorem 13.4. Theorem 13.4 follows from Lemmas 13.8 and 13.9 by Theorem 12.2. □

Remark 13.10. The random recursive tree and preferential attachment trees are not split trees in the sense above, since degrees are unbounded. Nevertheless, if the definition above is generalized to allow $b = \infty$, they too can be regarded as split trees, see [30]. We conjecture that under suitable conditions, Theorem 13.4 extends to the case $b = \infty$, but we have not pursued this. (Random recursive trees and preferential attachment trees can be handled by Theorem 14.3 below instead.)

14. Crump–Mode–Jagers branching process trees

A *Crump–Mode–Jagers* (CMJ) branching process (see e.g. [28]) is a continuous time process, where each individual gives birth to a (generally random) number of children at arbitrary random times; the times a single individual gets children are thus described by a point process Ξ on $[0, \infty)$. All individuals have independent and identically distributed such point processes. We start with a single individual, born at time 0; we also suppose that the CMJ process is supercritical and that it never dies out; hence its size a.s. grows to ∞.

The family tree of the CMJ process is a growing random tree $\widetilde{\mathcal{T}}_t$, $t \geqslant 0$, where the vertices are all individuals born up to time t. We stop the tree at the stopping time $\tau(n)$ where the tree first

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

reaches n vertices. Then (provided births a.s. occur at distinct times) $\mathcal{T}_n := \tilde{\mathcal{T}}_{\tau(n)}$ is a random tree with fixed size $|\mathcal{T}_n| = n$. More generally, $\tau(n)$ may be defined as the first time the total weight reaches n, where each individual has a weight given by some "characteristic" ψ; see [26] for details. (For example, for an m-ary search tree, ψ counts the balls, and we stop when there are n balls; cf. the split trees in Section 13.)

Many examples of such CMJ trees are discussed in the survey [26, Sections 6–8]; these include for example binary search trees and m-ary search trees (also covered by Section 13), and random recursive trees and preferential attachment trees. We give in Theorem 14.3 a general result for such trees. For example, this applies to Examples 12.3–12.6.

The point process Ξ can informally be regarded as the random set $\{\bar{\xi}_i\}_{i=1}^N$ of the times of births $\bar{\xi}_i$ of the children of the root, where the number of children $N \in \{0, 1, \dots, \infty\}$ in general is random. (We use the notation $\bar{\xi}_i$ to avoid confusion with the random vertices $\xi_i^{(n)}$.) Formally, Ξ is defined as the random measure $\sum_{i=1}^N \delta_{\bar{\xi}_i}$. Let $\mu := \mathbb{E}\,\Xi$ denote the intensity measure of Ξ.

We define the Laplace transform of the measure μ on $[0, \infty)$ by

$$\widehat{\mu}(\theta) = \int_0^\infty e^{-\theta t} \mu(\mathrm{d}t) = \mathbb{E} \int_0^\infty \sum_{i=1}^N e^{-\theta \bar{\xi}_i}, \qquad -\infty < \theta < \infty. \tag{14.1}$$

As in [26], we make the following assumptions; see further [26].

- (A1) $\mu\{0\} = \mathbb{E}\,\Xi\{0\} < 1$. (This rules out a rather trivial case with explosions already at the start. In all examples in [26], $\mu\{0\} = 0$.)
- (A2) μ is not concentrated on any lattice $h\mathbb{Z}$, $h > 0$. (This is for convenience only.)
- (A3) $\mathbb{E}\,N > 1$. (This is known as the *supercritical* case.) For simplicity, we further assume that $N \geqslant 1$ a.s. (In this case, every individual has at least one child, so the process never dies out and $|\tilde{\mathcal{T}}_\infty| = \infty$.)
- (A4) There exists a real number $\alpha > 0$ (the *Malthusian parameter*) such that $\widehat{\mu}(\alpha) = 1$, i.e.,
$$\int_0^\infty e^{-\alpha t} \mu(\mathrm{d}t) = 1. \tag{14.2}$$
- (A5) $\widehat{\mu}(\theta) < \infty$ for some $\theta < \alpha$.
- (A6ψ) (Only needed if the stopping time $\tau(n)$ is defined using a weight ψ. Thus void in the case that $\mathcal{T}_n$ always has n vertices.) The random variable $\sup_t \bigl(e^{-\theta t}\psi(t)\bigr)$ has finite expectation for some $\theta < \alpha$.

We assume also the following technical condition. (We conjecture that this is not necessary, but we use it in our proof.) Define the random variable

$$\widehat{\Xi}(\alpha) := \sum_{i=1}^N e^{-\alpha \bar{\xi}_i} \tag{14.3}$$

and note that (14.2) is equivalent to

$$\mathbb{E}\,\widehat{\Xi}(\alpha) = 1. \tag{14.4}$$

We assume a weak moment condition.

- (A7) We have $\mathbb{E}\bigl[\widehat{\Xi}(\alpha) \log \widehat{\Xi}(\alpha)\bigr] < \infty$.

Remark 14.1. Note that (A7) trivially holds if the outdegrees in $\mathcal{T}_n$ are bounded, so $N \leqslant C$ a.s. for some $C \leqslant \infty$. It is also easily seen that (A7) holds, as a consequence of the stronger $\mathbb{E}\bigl[\widehat{\Xi}(\alpha)^2\bigr] < \infty$, for random recursive trees and the linear preferential attachment trees in [26, Section 6].

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

We let for convenience $Z_t := |\tilde{\mathcal{T}}_t|$, and similarly $Z_t^v := |\tilde{\mathcal{T}}_t^v|$ for fringe trees. We also define

$$\beta := \int_0^\infty te^{-\alpha t}\mu(\mathrm{d}t) < \infty. \tag{14.5}$$

Remark 14.2. Nerman [44] showed that under the assumptions (A1)–(A6ψ) above, there exists a random variable W such that, as $t \to \infty$,

$$e^{-\alpha t} Z_t \xrightarrow{\text{a.s.}} W. \tag{14.6}$$

If furthermore (A7) holds, then $W > 0$ a.s. and

$$\mathbb{E}\, W = (\alpha\beta)^{-1}. \tag{14.7}$$

However, if (A7) fails, then $W = 0$ a.s. See also [16].

Theorem 14.3. *Assume (A1)–(A6ψ) and (A7). Then*

$$\frac{1}{\log n}\mathcal{T}_n \xrightarrow{\text{p}} \Upsilon_{1/(\alpha\beta)}. \tag{14.8}$$

Proof. It is shown in [27, Theorem 13.61], using results by Nerman [44] and Biggins [9, 10], that (12.3) holds with $a := 1/(\alpha\beta)$. Hence, by Theorem 12.2, it remains only to verify (12.4). We argue similarly as for Lemma 13.9.

We regard $\tilde{\mathcal{T}}_t$ as a subtree of the infinite tree T_∞, where the vertices are all finite strings $i_1 \cdots i_k$ of natural numbers $i_j \in \mathbb{N}$, with $0 \leqslant k < \infty$; thus the children of v are vi, for $i = 1, \dots,$ in this order. Note that the length of the string labelling $v \in \mathrm{T}_\infty$ equals $d(v, o)$; we denote this length by $|v|$.

For a vertex $v \in \mathrm{T}_\infty$, let b_v be the time that v is born in our CMJ branching process; if v never appears, then $b_v := \infty$.

The fringe tree $\tilde{\mathcal{T}}_t^v$ (defined as $\emptyset$ if $b_v > t$ so $v \notin \tilde{\mathcal{T}}_t$) is from the time b_v on a copy of the entire branching process tree, and thus (14.6) implies that for every v with $b_v < \infty$,

$$e^{-\alpha(t-b_v)} Z_t^v \xrightarrow{\text{a.s.}} W_v, \qquad t \to \infty, \tag{14.9}$$

where $W_v \overset{\text{d}}{=} W$ is independent of b_v. Thus

$$e^{-\alpha t} Z_t^v \xrightarrow{\text{a.s.}} e^{-\alpha b_v} W_v, \qquad t \to \infty, \tag{14.10}$$

which holds trivially also for $b_v = \infty$ (with $e^{-\infty} = 0$). Consequently,

$$\frac{Z_t^v}{Z_t} = \frac{e^{-\alpha t} Z_t^v}{e^{-\alpha t} Z_t} \xrightarrow{\text{a.s.}} \frac{e^{-\alpha b_v} W_v}{W} =: Y_v, \qquad t \to \infty. \tag{14.11}$$

Consider first the children of the root; these are labelled with $i \in \mathbb{N}$. Since $Z_t = 1 + \sum_i Z_t^i$, we have by (14.11) and (the elementary) Fatou's lemma for sums, a.s.,

$$\sum_i Y_i = \sum_i \liminf_{t\to\infty} \frac{Z_t^i}{Z_t} \leqslant \liminf_{t\to\infty} \sum_i \frac{Z_t^i}{Z_t} = 1. \tag{14.12}$$

Equivalently, by (14.11),

$$\sum_i e^{-\alpha b_i} W_i \leqslant W, \qquad \text{a.s.} \tag{14.13}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Furthermore, by (14.3)–(14.4), noting that $b_i = \bar{\xi}_i$,

$$\begin{aligned}\mathbb{E}\sum_i e^{-\alpha b_i} W_i &= \sum_i \mathbb{E}\big[e^{-\alpha b_i} W_i\big] = \sum_i \mathbb{E}\big[e^{-\alpha b_i}\big]\,\mathbb{E}\,[W_i] \\ &= \mathbb{E}\,[W] \sum_i \mathbb{E}\, e^{-\alpha b_i} = \mathbb{E}\,[W]\,\mathbb{E}\sum_i e^{-\alpha b_i} = \mathbb{E}\,[W]\,\mathbb{E}\,\widehat{\Xi}(\alpha) \\ &= \mathbb{E}\, W. \end{aligned} \tag{14.14}$$

By (14.7), $\mathbb{E}\, W < \infty$ and thus (14.13) and (14.14) imply

$$\sum_i e^{-\alpha b_i} W_i = W, \qquad \text{a.s.} \tag{14.15}$$

Equivalently, there is a.s. equality in (14.12).

Let $v \in \mathrm{T}_\infty$ and apply (14.15) to the fringe tree $\tilde{\mathcal{T}}_t^v$, again regarded as a copy of the original branching process; this shows that if $b_v < \infty$, then

$$\sum_i e^{-\alpha(b_{vi}-b_v)} W_{vi} = W_v, \qquad \text{a.s.} \tag{14.16}$$

and thus

$$\sum_i e^{-\alpha b_{vi}} W_{vi} = e^{-\alpha b_v} W_v, \qquad \text{a.s.,} \tag{14.17}$$

where (14.17) trivially holds also if $b_v = \infty$.

By (14.17) and induction we conclude that for every $k \geqslant 0$,

$$\sum_{|v|=k} e^{-\alpha b_v} W_v = W, \qquad \text{a.s.} \tag{14.18}$$

Equivalently, by the definition (14.11) again,

$$\sum_{|v|=k} Y_v = 1, \qquad \text{a.s.} \tag{14.19}$$

Next, fix an integer k. Two vertices v and w of $\tilde{\mathcal{T}}_t$ have $d(v \wedge w, o) \geqslant k$ if and only if they belong to the same subtree $\tilde{\mathcal{T}}_t^v$ for some v with $|v| = k$. Thus, if $\xi_j^{(t)}$ are i.i.d. uniformly random vertices in $\tilde{\mathcal{T}}_t$,

$$\mathbb{P}\big(d(\xi_1^{(t)} \wedge \xi_2^{(t)}, o) \geqslant k \mid \tilde{\mathcal{T}}_t\big) = \sum_{|v|=k} \Big(\frac{Z_t^v}{Z_t}\Big)^2. \tag{14.20}$$

By (14.19) and Fatou's lemma as in (14.12), a.s.,

$$1 = \sum_{|v|=k} Y_v \leqslant \liminf_{t\to\infty} \sum_{|v|=k} \frac{Z_t^v}{Z_t} \leqslant \limsup_{t\to\infty} \sum_{|v|=k} \frac{Z_t^v}{Z_t} \leqslant 1, \tag{14.21}$$

and thus

$$\sum_{|v|=k} \frac{Z_t^v}{Z_t} \overset{\text{a.s.}}{\longrightarrow} 1. \tag{14.22}$$

This together with (14.11) and (14.19) implies by a standard argument, cf. again [23, Theorem 5.6.4],

$$\sum_{|v|=k} \left| \frac{Z_t^v}{Z_t} - Y_v \right| \overset{\text{a.s.}}{\longrightarrow} 0. \tag{14.23}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

Hence,

$$\sum_{|v|=k}\left|\left(\frac{Z_t^v}{Z_t}\right)^2 - Y_v^2\right| \leqslant \sum_{|v|=k}\left|\frac{Z_t^v}{Z_t} - Y_v\right| \xrightarrow{\text{a.s.}} 0 \tag{14.24}$$

and thus (14.20) implies

$$\mathbb{P}\big(d(\xi_1^{(t)} \wedge \xi_2^{(t)}, o) \geqslant k \mid \tilde{\mathcal{T}}_t\big) \xrightarrow{\text{a.s.}} \sum_{|v|=k} Y_v^2. \tag{14.25}$$

By considering the sequence of times $\tau(n)$, this shows

$$\mathbb{P}\big(d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o) \geqslant k \mid \mathcal{T}_n\big) \xrightarrow{\text{a.s.}} \sum_{|v|=k} Y_v^2. \tag{14.26}$$

Taking the expectation yields, by dominated convergence,

$$\mathbb{P}\big(d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o) \geqslant k\big) \to \mathbb{E}\sum_{|v|=k} Y_v^2. \tag{14.27}$$

We want to show that the right-hand side of (14.27) tends to 0 as $k \to \infty$. Define, for $k \geqslant 0$,

$$Q_k := \sum_{|v|=k}\big(e^{-\alpha b_v} W_v\big)^2 = \sum_{|v|=k} e^{-2\alpha b_v} W_v^2 = W^2 \sum_{|v|=k} Y_v^2. \tag{14.28}$$

By (14.17), a.s.,

$$W^2 = Q_0 \geqslant Q_1 \geqslant Q_2 \geqslant \dots \tag{14.29}$$

Define

$$Q_\infty := \lim_{k\to\infty} Q_k. \tag{14.30}$$

Similarly, for each $i \in \mathbb{N}$ with $b_i < \infty$, consider the fringe tree $\tilde{\mathcal{T}}_t^i$, and define

$$Q_{k;i} := \sum_{|v|=k} e^{-2\alpha(b_{iv}-b_i)} W_{iv}^2, \tag{14.31}$$

$$Q_{\infty;i} := \lim_{k\to\infty} Q_{k;i} \overset{\text{d}}{=} Q_\infty. \tag{14.32}$$

For convenience, we define $Q_{\infty:i}$ also when $b_i = \infty$, as some copy of Q_∞ independent of everything else. Then, (14.31) and (14.28) yield, for any $k \geqslant 0$,

$$Q_{k+1} = \sum_{i=1}^{\infty} e^{-2\alpha b_i} Q_{k;i}. \tag{14.33}$$

Letting $k \to \infty$ in (14.33), we obtain by dominated convergence, since $Q_{k;i} \leqslant Q_{0;i}$ and $\sum_i e^{-2\alpha b_i} Q_{0;i} = Q_1 \leqslant Q_0 = W^2 < \infty$ a.s.,

$$Q_\infty = \sum_{i=1}^{\infty} e^{-2\alpha b_i} Q_{\infty;i} \qquad \text{a.s.} \tag{14.34}$$

We claim that $Q_\infty = 0$ a.s. To see this note first that (14.34) implies

$$Q_\infty^{1/2} \leqslant \sum_{i=1}^{\infty} e^{-\alpha b_i} Q_{\infty;i}^{1/2} \qquad \text{a.s.,} \tag{14.35}$$

with strict inequality as soon as there is more than one non-zero term in the sum. Moreover, since b_i and $Q_{\infty;i}$ are independent, using (14.3)–(14.4) again,

$$\begin{aligned}\mathbb{E}\sum_{i=1}^{\infty} e^{-\alpha b_i} Q_{\infty;i}^{1/2} &= \sum_{i=1}^{\infty} \mathbb{E}\big[e^{-\alpha b_i} Q_{\infty;i}^{1/2}\big] = \sum_{i=1}^{\infty} \mathbb{E}\big[e^{-\alpha b_i}\big]\, \mathbb{E}\big[Q_{\infty;i}^{1/2}\big] \\ &= \mathbb{E}\big[Q_{\infty}^{1/2}\big] \sum_{i=1}^{\infty} \mathbb{E}\big[e^{-\alpha b_i}\big] = \mathbb{E}\big[Q_{\infty}^{1/2}\big]\, \mathbb{E}\big[\widehat{\Xi}(\alpha)\big] \\ &= \mathbb{E}\, Q_{\infty}^{1/2}. \end{aligned} \tag{14.36}$$

Furthermore, $\mathbb{E}\, Q_{\infty}^{1/2} \leqslant \mathbb{E}\, W < \infty$. Hence, (14.36) implies

$$\mathbb{E}\Big(\sum_{i=1}^{\infty} e^{-\alpha b_i} Q_{\infty;i}^{1/2} - Q_{\infty}^{1/2}\Big) = \mathbb{E}\, Q_{\infty}^{1/2} - \mathbb{E}\, Q_{\infty}^{1/2} = 0, \tag{14.37}$$

and thus there is equality in (14.35) a.s.

Suppose that $\mathbb{P}\,(Q_\infty > 0) > 0$. Conditioned on the offspring Ξ of the root, the fringe trees $\tilde{\mathcal{T}}_t^i$, $i \leqslant N$, are independent copies of $\tilde{\mathcal{T}}_t$. Hence, the events $N \geqslant 2$, $Q_{\infty;1} > 0$ and $Q_{\infty;2} > 0$ are independent and thus with positive probability they occur together, and then there is strict inequality in (14.35). This contradiction shows that $Q_\infty = 0$ a.s.

Consequently, (14.28) shows that, since $W > 0$ a.s.,

$$\sum_{|v|=k} Y_v^2 = W^{-2} Q_k \xrightarrow{\text{a.s.}} W^{-2} Q_\infty = 0, \qquad k \to \infty. \tag{14.38}$$

Furthermore, $\sum_{|v|=k} Y_v^2 \leqslant 1$ by (14.19) or (14.26). Hence, by dominated convergence,

$$\mathbb{E} \sum_{|v|=k} Y_v^2 \to 0, \qquad k \to \infty. \tag{14.39}$$

Finally, (14.27) and (14.39) show that

$$\lim_{k\to\infty} \lim_{n\to\infty} \mathbb{P}\big(d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o) \geqslant k\big) = 0, \tag{14.40}$$

which shows that the sequence of random variables $d(\xi_1^{(n)} \wedge \xi_2^{(n)}, o)$ is tight, and in particular that (12.4) holds. □

15. Proof of Theorem 3.15

Theorem 3.15 is stated in [20, Theorem 1] for uniformly bounded rescaled finite trees. Furthermore, [20, Theorem 4] contains a related statement (for measured real trees); we show that it implies Theorem 3.15.

Proof of Theorem 3.15. This is the only place in the present paper where we use the machinery with ultraproducts used in [20] to prove the results there. We refer to [20] for definitions and basic properties, and will here only give the additional arguments needed. We fix, as in [20], an ultrafilter ω on $\mathbb{N}$. All ultralimits and ultraproducts are defined using ω.

Let $(T_n)_1^\infty = (T_n, d_n, \mu_n)_1^\infty$ be a convergent sequence of measured real trees. Thus (3.6) holds for some measures $\lambda_r \in \mathcal{P}(M_r)$.

Taking $r = 2$ in Definition 3.5, we see by (3.3) and (3.1) that, in particular,

$$d_n(\xi_1^{(n)}, \xi_2^{(n)}) \xrightarrow{\text{d}} \zeta, \tag{15.1}$$

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

for some random variable ζ. It follows from (15.1) that the sequence of random variables $d_n(\xi_1^{(n)}, \xi_2^{(n)})$ is tight, i.e., that for every $\varepsilon > 0$, there exists a constants C_ε such that for every n

$$\mathbb{P}\big(d_n(\xi_1^{(n)}, \xi_2^{(n)}) > C_\varepsilon\big) \leqslant \varepsilon. \tag{15.2}$$

Fix $\varepsilon > 0$. By (15.2) and Fubini's theorem, there exists $x_n \in T_n$ such that

$$\mathbb{P}\big(d_n(\xi_1^{(n)}, x_n) > C_\varepsilon\big) \leqslant \varepsilon. \tag{15.3}$$

Let $A_n := \{x \in T_n : d_n(x, x_n) \leqslant C_\varepsilon\}$. Then (15.3) says

$$\mu_n(A_n) \geqslant 1 - \varepsilon. \tag{15.4}$$

As in [20], form the ultraproduct $\mathbf{T} := \prod_\omega T_n$, and equip it with the pseudometric $\mathbf{d} := \lim_\omega d_n$ (which may take the value $+\infty$) and the probability measure $\mu := \prod_\omega \mu_n$. Let $\mathbf{x} := [(x_n)_n] \in \mathbf{T}$ and $\mathbf{A} := \prod_\omega A_n \subseteq \mathbf{T}$. For any $\mathbf{y} \in \mathbf{A}$, $\mathbf{y} = [(y_n)_n]$ for some $y_n \in T_n$ with $y_n \in A_n$ and thus $d_n(x_n, y_n) \leqslant C_\varepsilon$ for every n; hence

$$\mathbf{d}(\mathbf{x}, \mathbf{y}) = \lim_\omega d_n(x_n, y_n) \leqslant C_\varepsilon. \tag{15.5}$$

Furthermore, by (15.4),

$$\mu(\mathbf{A}) = \lim_\omega \mu_n(A_n) \geqslant 1 - \varepsilon. \tag{15.6}$$

Let $\mathbf{X} := B(\mathbf{x}, \infty) := \{\mathbf{y} : \mathbf{d}(\mathbf{y}, \mathbf{x}) < \infty\}$. Then (15.5) shows that $\mathbf{A} \subset \mathbf{X}$, and thus (15.6) shows

$$\mu(\mathbf{X}) \geqslant \mu(\mathbf{A}) \geqslant 1 - \varepsilon. \tag{15.7}$$

(It is shown in [20] that $\mathbf{X}$ is μ-measurable.) Here $\mathbf{x} = \mathbf{x}(\varepsilon)$ and $\mathbf{X} = \mathbf{X}(\varepsilon)$ may depend on ε. However, two infinite balls $B(\mathbf{x}_1, \infty)$ and $B(\mathbf{x}_2, \infty)$ in $\mathbf{T}$ either coincide or are disjoint. (Such infinite balls are called *clusters* in [20].) Hence, considering only $\varepsilon \leqslant \frac{1}{2}$, it follows from (15.7) that all $\mathbf{X}(\varepsilon)$ coincide, and consequently form a cluster $\mathbf{X}$ with, using (15.7) again, $\mu(\mathbf{X}) = 1$.

This means that the sequence $(T_n, d_n, \mu_n)_n$ is *essentially bounded*, in the terminology of [20]. Consequently, [20, Theorem 4] applies, and shows that

$$\lim_\omega \tau_r(T_n) = \lim_\omega \tau_r(T_n, d_n, \mu_n) = \tau_r(D), \tag{15.8}$$

for every $r \geqslant 1$ and some long dendron D (constructed from the ultraproduct $\mathbf{T}$ in a way that we do not have to consider further).

On the other hand, we have assumed (3.6), so the sequence $\tau_r(T_n) = \tau_r(T_n, d_n, \mu_n)$ converges. A convergent sequence has its limit as its ultralimit; hence (15.8) and (3.6) yield $\tau_r(D) = \lambda_r$. Consequently, (3.6) says

$$\tau_r(T_n) \to \tau_r(D), \qquad r \geqslant 1, \tag{15.9}$$

and thus $T_n \to D$, which completes the proof. □

Remark 15.1. The proof shows that a tight sequence $(T_n)_n$ of measured real trees is essentially bounded. The converse does not hold, since we may let T_n be arbitrary along some subsequences without affecting the ultraproduct and ultralimits, and thus the property of being essentially bounded. Nevertheless, a sequence $(T_n)_n$ such that every subsequence is essentially bounded is tight (as a consequence of [20, Theorem 4]). Similarly, a sequence is tight if and only if it is essentially bounded for every ultrafilter ω.

Acknowledgements

I thank Cecilia Holmgren and Gabriel Berzunza Ojeda for help with references, and an anonymous referee for a careful reading and finding a gap in a proof.

https://doi.org/10.1017/S0963548321000055 Published online by Cambridge University Press

References

[1] Romain Abraham & Jean-François Delmas: Local limits of conditioned Galton-Watson trees: the condensation case. *Electron. J. Probab.* 19 (2014), no. 56, 29 pp.

[2] Louigi Addario-Berry, Nicolas Broutin & Cecilia Holmgren: Cutting down trees with a Markov chainsaw. *Ann. Appl. Probab.* **24** (2014), no. 6, 2297–2339.

[3] Michael Albert, Cecilia Holmgren, Tony Johansson & Fiona Skerman: Embedding small digraphs and permutations in binary trees and split trees. *Algorithmica* **82** (2020), no. 3, 589–615.

[4] David Aldous: Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1** (1991), no. 2, 228–266.

[5] David Aldous: The continuum random tree I. *Ann. Probab.* **19** (1991), no. 1, 1–28.

[6] David Aldous: The continuum random tree II: an overview. *Stochastic Analysis (Durham, 1990)*, 23–70, London Math. Soc. Lecture Note Ser. 167, Cambridge Univ. Press, Cambridge, 1991.

[7] David Aldous: The continuum random tree III. *Ann. Probab.* **21** (1993), no. 1, 248–289.

[8] Gabriel Berzunza, Xing Shi Cai & Cecilia Holmgren: The asymptotic non-normality of the giant cluster for percolation on random split trees. Preprint, 2019. `arXiv`:1902.08109v4

[9] J. D. Biggins: The growth and spread of the general branching random walk. *Ann. Appl. Probab.* **5** (1995), no. 4, 1008–1024.

[10] J. D. Biggins: How fast does a general branching random walk spread? *Classical and Modern Branching Processes (Minneapolis, MN, 1994)*, 19–39, Springer, New York, 1997.

[11] Patrick Billingsley: *Convergence of Probability Measures*. Wiley, New York, 1968.

[12] Vladimir I. Bogachev: *Measure theory. Vol. I, II*. Springer-Verlag, Berlin, 2007.

[13] N. Broutin, L. Devroye, E. McLeish & M. de la Salle: The height of increasing trees. *Random Structures Algorithms* **32** (2008), no. 4, 494–518.

[14] Luc Devroye: Universal limit laws for depths in random trees. *SIAM J. Comput.* **28** (1999), no. 2, 409–432.

[15] Persi Diaconis & Svante Janson: Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)* **28** (2008), no. 1, 33–61.

[16] R. A. Doney: A limit theorem for a class of supercritical branching processes. *Journal of Applied Probability* **9** (1972), no. 4, 707–724.

[17] Andreas W. M. Dress: Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: A note on combinatorial properties of metric spaces. *Advances in Mathematics* **53**:3 (1984), 321–402.

[18] Andreas Dress, Vincent Moulton & Werner Terhalle: T-theory: an overview. *European J. Combin.* **17** (1996), no. 2-3, 161–175.

[19] Thomas Duquesne: A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab.* **31** (2003), no. 2, 996–1027.

[20] Gábor Elek & Gábor Tardos: Convergence and limits of finite trees. Preprint, 2020. `arXiv`:2001.00905

[21] Andreas Greven, Peter Pfaffelhuber & Anita Winter: Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Prob. Theo. Rel. Fields* **145** (2009), no. 1-2, 285–322.

[22] Misha Gromov: *Metric Structures for Riemannian and Non-Riemannian Spaces*. Birkhäuser, Boston, MA, 1999, 2001.

[23] Allan Gut. *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013.

[24] Bénédicte Haas & Grégory Miermont: Scaling limits of Markov branching trees with applications to Galton–Watson and random unordered trees. *Ann. Probab.* **40** (2012), no. 6, 2589–2666.

[25] Cecilia Holmgren: Novel characteristic of split trees by use of renewal theory. *Electron. J. Probab.* **17** (2012), no. 5, 27 pp.

[26] Cecilia Holmgren & Svante Janson: Fringe trees, Crump–Mode–Jagers branching processes and m-ary search trees. *Probab. Surv.* **14** (2017), 53–154.

[27] Cecilia Holmgren & Svante Janson: Fringe trees, Crump–Mode–Jagers branching processes and m-ary search trees. Preprint version of [26], 2016. arXiv:1601.03691

[28] Peter Jagers: *Branching Processes with Biological Applications*. John Wiley & Sons, London, 1975.

[29] Svante Janson: Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probab. Surv.* **9** (2012), 103–252.

[30] Svante Janson: Random recursive trees and preferential attachment trees are random split trees. *Combin. Probab. Comput.* **28** (2019), no. 1, 81–99.

[31] Svante Janson: On the Gromov–Prohorov distance. Preprint, 2020. arXiv:2005.13505

[32] Svante Janson, Thordur Jonsson & Sigurdur Örn Stefánsson: Random trees with superexponential branching weights. *J. Phys. A* **44** (2011), no. 48, 485002, 16 pp.

[33] Thordur Jonsson & Sigurdur Örn Stefánsson: Condensation in nongeneric trees. *J. Stat. Phys.* **142** (2011), no. 2, 277–313.

[34] Olav Kallenberg: *Foundations of Modern Probability*. 2nd ed., Springer, New York, 2002.

[35] Igor Kortchemski: Limit theorems for conditioned non-generic Galton–Watson trees. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** (2015), no. 2, 489–511.

[36] Jean-François Le Gall: Random trees and applications. *Probab. Surv.* **2** (2005), 245–311.

[37] Jean-François Le Gall: Random real trees. *Ann. Fac. Sci. Toulouse Math. (6)* **15** (2006), no. 1, 35–62.

[38] Wolfgang Löhr: Equivalence of Gromov–Prohorov– and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces. *Electron. Commun. Probab.* **18** (2013), no. 17, 10 pp.

[39] László Lovász: *Large Networks and Graph Limits*. American Mathematical Society, Providence, RI, 2012.

[40] Hosam M. Mahmoud & Ralph Neininger: Distribution of distances in random binary search trees. *Ann. Appl. Probab.* **13** (2003), no. 1, 253–276.

[41] Grégory Miermont: Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér. (4)* **42** (2009), no. 5, 725–781.

[42] Kate Morris, Alois Panholzer & Helmut Prodinger: On some parameters in heap ordered trees. *Combin. Probab. Comput.* **13** (2004), no. 4-5, 677–696.

[43] Götz Olaf Munsonius; Ludger Rüschendorf: Limit theorems for depths and distances in weighted random b-ary recursive trees. *J. Appl. Probab.* **48** (2011), no. 4, 1060–1080.

[44] Olle Nerman: On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrsch. Verw. Gebiete* **57** (1981), no. 3, 365–395.

[45] Alois Panholzer: The distribution of the size of the ancestor-tree and of the induced spanning subtree for random trees. *Random Structures Algorithms* **25** (2004), no. 2, 179–207.

[46] Alois Panholzer & Helmut Prodinger: Spanning tree size in random binary search trees. *Ann. Appl. Probab.* **14** (2004), no. 2, 718–733.

[47] Alois Panholzer & Helmut Prodinger: Analysis of some statistics for increasing tree families. *Discrete Math. Theor. Comput. Sci.* **6** (2004), no. 2, 437–460.

[48] Jelena Ryvkina: Ein universeller zentraler Grenzwertsatz für den Abstand zweier Kugeln in zufälligen Splitbäumen. Diploma Thesis, Johann Wolfgang Goethe-Universität, Frankfurt, 2008. urn:nbn:de:hebis:30-57448

[49] Benedikt Stufler: Local limits of large Galton-Watson trees rerooted at a random vertex. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** (2019), no. 1, 155–183.

[50] Cédric Villani: *Optimal transport*. Springer-Verlag, Berlin, 2009.

Cite this article: Janson S (2021). Tree limits and limits of random trees. *Combinatorics, Probability and Computing* **30**, 849–893. https://doi.org/10.1017/S0963548321000055