Olivier Hénard & Pascal Maillard

On trees invariant under edge contraction

ON TREES INVARIANT UNDER EDGE CONTRACTION

by Olivier Hénard & Pascal Maillard

Abstract. — We study random trees which are invariant in law under the operation of contracting each edge independently with probability $p \in (0,1)$. We show that all such trees can be constructed through Poisson sampling from a certain class of random measured $\mathbb{R}$-trees satisfying a natural scale invariance property. This has connections to exchangeable partially ordered sets, real-valued self-similar increasing processes and quasi-stationary distributions of Galton–Watson processes.

Résumé (Au sujet des arbres invariants par contraction de leurs arêtes)
Nous étudions les arbres aléatoires dont la loi est invariante par la contraction indépendante de leurs arêtes avec probabilité $p \in (0,1)$. Nous montrons que ces arbres peuvent être construits par échantillonnage poissonnien à partir d'une classe de $\mathbb{R}$-arbres aléatoires mesurés qui satisfont à une propriété d'invariance naturelle. Cette étude est liée aux ordres partiels échangeables, aux processus autosimilaires croissants à valeurs réelles et aux distributions quasi-stationnaires de processus de Galton-Watson.

Contents

1. Introduction

Take a random rooted tree T (in the graph sense) and contract each edge independently with probability $p \in (0, 1)$. Are there (necessarily infinite) random trees which are invariant in law under this operation? Trivial examples are the (semi-)infinite ray, i.e., the one-dimensional half-lattice $\mathbb{N} := \{0, 1, \ldots\}$ and with root 0, or a (possibly random) number of copies of this graph whose roots are identified. Slightly less trivial examples are the previous graph with a bouquet of edges attached to each vertex, the number of edges in each bouquet being independent and identically distributed (iid) according to a geometric distribution (starting at 0). There are however many more examples of such trees. Indeed, the following is the main result of this article, which is an informal statement of Theorem 1:

THEOREM. — *There is a one-to-one correspondence between random trees invariant under the above contraction operation and a certain class of continuum random trees invariant under rescaling by the factor p.*

The proof of this result will involve quite an extensive machinery of tools, including, for instance, Gromov–Hausdorff–Prokhorov convergence of locally compact metric spaces and random exchangeable partial orders. We furthermore give in this article constructions of several examples of such trees and completely characterize the class of the trees which are also invariant (in law) under translations along the spine. This has connections with real-valued self-similar increasing processes and with quasi-stationary distributions of subcritical Galton–Watson processes.

The study of this problem originates in a geometrically motivated question asked to us by I. Benjamini: Is it possible to find a law on the space of trees and a suitable renormalization procedure, i.e., a suitable random coarse-graining operation that preserves the tree structure, such that the law of the tree is invariant under this operation? The contraction operation is an example of such a renormalization procedure, which has the pleasant feature of allowing to characterize all locally finite random trees it leaves invariant.

We are not aware of any similar results in the random tree literature. The operations that are usually considered on trees, for example random growth or pruning, always operate on leaves or whole subtrees instead of single interior vertices, see [Rém85, AP98]. There might be a good reason for that: The trees we obtain are indeed very different from usual trees in the sense that they are very elongated, with long chains of vertices of degree 2, to which might be attached some bouquets of edges as in the above example. In particular, neither exponentially growing trees nor critical Galton–Watson trees (conditioned on non-extinction) are amongst them.

DEFINITIONS AND STATEMENTS OF THE RESULTS. — The precise statements of our results require a fair amount of definitions which we give in this section, occasionally referring

to the appendix for details. Very impatient readers might directly jump to the statement of the main theorem (Theorem 1) and work their way backwards to understand all the definitions.

In this article, we consider rooted, locally finite trees $T = (V, E, \rho)$ in the graph-theoretic sense; in combinatorics these are also known as *unlabelled, unordered, non-plane* or *Pólya* trees [Sta97, FS09, Drm09]. We say that two such trees are *equivalent* if there exists a root-preserving graph isomorphism between them and denote by $\mathbb{T}$ the space of equivalence classes of trees. We usually identify an equivalence class with its representatives. A path in T is a finite or infinite sequence of adjacent, pairwise distinct vertices of T, and an end is an infinite path starting at the root. The *spine* of T, denoted by $\mathrm{Spine}(T)$, is the union of the ends of T, or $\{\rho\}$ in case T has no end. We denote by $\mathbb{T}_e \subset \mathbb{T}$ the subspace of trees with finitely many ends and by $\mathbb{T}_f \subset \mathbb{T}_e$ the subspace of the finite trees.

We endow the space $\mathbb{T}$ (and its subspaces) with the topology of local convergence defined as follows: For a tree $T \in \mathbb{T}$ and $k \in \mathbb{N}$, denote by $T^{\leqslant k}$ its restriction to the vertices at (graph) distance at most k from the root. A sequence $(T_n)_{n \geqslant 0}$ in $\mathbb{T}$ then is said to converge to $T \in \mathbb{T}$ if for every $k \in \mathbb{N}$, $(T_n^{\leqslant k})_{n \geqslant 0}$ converges to $T^{\leqslant k}$ in $\mathbb{T}_f$. This topology is metrizable in such a way that the space $\mathbb{T}$ is a complete separable metric space, but not the space $\mathbb{T}_e$ (see Section A.1). Note also that the induced topology on $\mathbb{T}_f$ is indeed the discrete topology.

In order to deal with *random* trees, we endow the space $\mathbb{T}$ with the Borel-σ-field induced by its topology. We then denote by $\mathscr{M}_1(\mathbb{T})$ the space of probability measures on $\mathbb{T}$, endowed with the topology of weak convergence. We will often denote the elements of $\mathscr{M}_1(\mathbb{T})$ by T as well, and will refer to them as *random* trees. We similarly define $\mathscr{M}_1(\mathbb{T}_e)$ and $\mathscr{M}_1(\mathbb{T}_f)$.

A rooted tree $T = (V, E, \rho) \in \mathbb{T}$ determines a partial order $\preceq_T$ on its vertex set V by $v \preceq_T w$ if and only if v lies on the (unique) path from ρ to w. In this case, we say that v is an *ancestor* of w, or that w is a *descendant* of v; accordingly, we call $\prec_T$ the *ancestral relation* of the tree T. We also write $v \prec_T w$ when $v \preceq_T w$ and $v \neq w$. Note that the tree T can in fact be completely recovered from $\preceq_T$ and $\prec_T$.[1] This allows us to formally define a generic contraction operation as follows:

Definition 1.1. — Let $T = (V, E, \rho) \in \mathbb{T}_e$ and let $V' \subset V$ be a subset of its vertices containing the root and containing an infinite number of vertices on each infinite path. The *contracted tree* $\mathscr{C}(T, V')$ is defined to be the rooted tree with vertex set V', root ρ and whose partial order $\preceq_{\mathscr{C}(T,V')}$ is the restriction of $\preceq_T$ to V'. It is easy to see that $\mathscr{C}(T, V')$ is indeed locally finite and has a finite number of ends only, i.e. $\mathscr{C}(T, V') \in \mathbb{T}_e$.

We then define the randomized contraction operation we will consider.

---

[1] In fact, the tree can be viewed as the transitive reduction of the (acyclic) directed graph $(V, \prec_T)$.

Definition 1.2. — Let $p, q \in (0, 1)$. Let $T = (V, E, \rho) \in \mathbb{T}_e$. Set $V_0 = V \setminus \mathrm{Spine}(T)$ and $V_1 = \mathrm{Spine}(T) \setminus \rho$. The random tree $\mathscr{C}_{p,q}(T) \in \mathscr{M}_1(\mathbb{T}_e)$ is defined to be equal to $\mathscr{C}(T, V')$, where V' is the random subset[2] of vertices containing ρ, every vertex in V_0 independently with probability p and every vertex in V_1 independently with probability q.

Note that if $p \neq q$, then the map $\mathscr{C}_{p,q} : \mathbb{T}_e \to \mathscr{M}_1(\mathbb{T}_e)$ is not continuous, because when a sequence of trees T_n converges to a tree T, non-spine vertices can become spine vertices in the limit. However, if we define for $M \in \mathbb{N}$ the map $\mathscr{C}_{p,q}^M$, in which a vertex is kept in V' with probability q if it has a descendant at distance M, and with probability p otherwise, then this map is easily seen to be continuous. In particular,[3] it extends to a continuous map $\mathscr{C}_{p,q}^M : \mathscr{M}_1(\mathbb{T}) \to \mathscr{M}_1(\mathbb{T})$. Since $\mathscr{C}_{p,q} = \lim_{M \to \infty} \mathscr{C}_{p,q}^M$, we conclude that the map $\mathscr{C}_{p,q}$ is measurable and extends to a (measurable) map $\mathscr{C}_{p,q} : \mathscr{M}_1(\mathbb{T}_e) \to \mathscr{M}_1(\mathbb{T}_e)$. This allows us to write $\mathscr{C}_{p,q}(T)$ for a *random* tree $T \in \mathscr{M}_1(\mathbb{T}_e)$.

Definition 1.3. — Let $p, q \in (0, 1)$. We say that a random tree $T \in \mathscr{M}_1(\mathbb{T}_e)$ is *(p, q)-self-similar*, if $T \stackrel{\mathrm{law}}{=} \mathscr{C}_{p,q}(T)$.

Remark 1.4. — In defining the contraction operation $\mathscr{C}_{p,q}$, we restricted our discussion to trees in $\mathbb{T}_e$. This brings no restriction when studying locally finite self-similar random trees, since there are no locally finite self-similar random trees with infinitely many ends: when applying the contraction operation repeatedly to such trees, the distance from the root to the branch points on the spine stochastically decreases, hence the degree of the root goes to infinity in law. Therefore, the degree of the root would have to be infinite in the first place.

We now extend the above definitions to $\mathbb{R}$-trees,[4] and point the reader to Section A.2 for precisions concerning the definitions below. We call $\mathfrak{T}$ the space of complete, locally compact, rooted, measured $\mathbb{R}$-trees $\mathscr{T} = (\mathscr{V}, d, \rho, \mu)$, with μ boundedly finite, modulo equivalence with respect to root- and measure-preserving isometries. This space is endowed with the GHP-topology and with its induced Borel σ-field. As above, we denote by $\mathfrak{T}_e \subset \mathfrak{T}$ the subspace of trees with finitely many ends (see before Lemma A.14 for the definition of an end in this setting), and by $\mathfrak{T}_f$ and $\mathfrak{T}_1$ the subspaces of $\mathfrak{T}_e$ for which the measure μ is finite or a probability measure, respectively.

Every $\mathbb{R}$-tree $\mathscr{T}$ defines a *length measure* $\ell_{\mathscr{T}}$ on its set of vertices, see (A.7). In many cases of interest in probability theory, the length measure is not locally finite,[5]

---

[2]It is easy to show that this subset contains an infinite number of vertices on each infinite path.

[3]See Section B.

[4]We have been notified by an anonymous referee that one can maybe streamline some arguments by working in the setting of 0-hyperbolic spaces which includes both discrete trees and $\mathbb{R}$-trees. Since we are not familiar with these notions, we did not pursue this direction.

[5]With respect to the topology on $\mathscr{V}$ induced by the metric d. One can define a finer topology generated by open segments with respect to which the length measure is always locally finite. Note that both topologies induce the same σ-algebra on the space of locally compact trees.

for example in the case of Aldous' Brownian continuum random tree. However, in this article, we will be interested in those trees $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_e$ for which the measure μ dominates the length measure $\ell_{\mathscr{T}}$, such that, in particular, the length measure is boundedly finite and μ has full support. We therefore define the spaces

$$\mathfrak{T}_e^\ell = \{\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_e : \mu \geqslant \ell_{\mathscr{T}}\}, \quad \mathfrak{T}_f^\ell = \mathfrak{T}_f \cap \mathfrak{T}_e^\ell, \quad \mathfrak{T}_1^\ell = \mathfrak{T}_1 \cap \mathfrak{T}_e^\ell,$$

The diameter of a tree $\mathscr{T} \in \mathfrak{T}_1^\ell$ is at most 1 since μ dominates $\ell_{\mathscr{T}}$ by definition. We show below that $\mathfrak{T}_e^\ell$, $\mathfrak{T}_f^\ell$ and $\mathfrak{T}_1^\ell$ are closed subspaces of $\mathfrak{T}_e$, $\mathfrak{T}_f$ and $\mathfrak{T}_1$, respectively (Lemma 3.4) and that the space $\mathfrak{T}_1^\ell$ is compact (Proposition 2.1).

We now define a (deterministic) rescaling operation $\mathscr{S}_{p,q}$ which will play the role of $\mathscr{C}_{p,q}$ for $\mathbb{R}$-trees. For a tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_e$ we denote by $\mathrm{Spine}(\mathscr{T})$ the subset of its vertices, called the *spine*, which lie on an end. If $x \in \mathscr{V}$, we denote by $\vec{x}$ the most recent ancestor of x on the spine, i.e., the vertex in $[\![\rho, x]\!] \cap \mathrm{Spine}(\mathscr{T})$ with maximal distance from the root. For two vertices $x, y \in \mathscr{V}$, we then have

$$d(x, y) = d(x, \vec{x}) + d(\vec{x}, \vec{y}) + d(\vec{y}, y), \quad \text{if } \vec{x} \neq \vec{y}.$$

DEFINITION 1.5. — If $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_e^\ell$ and $p, q \in (0, 1)$, then we define the tree $\mathscr{T}' = \mathscr{S}_{p,q}(\mathscr{T}) = (\mathscr{V}', d', \rho', \mu')$ by
- $\mathscr{V}' = \mathscr{V}$ and $\rho' = \rho$,
- $d'(x, y) = p\,d(x, y) + (q - p)d(\vec{x}, \vec{y})$ and
- $\mu' = p\mu + (q - p)\ell_{\mathscr{T}}(\cdot \cap \mathrm{Spine}(\mathscr{T}))$.

In words, we shrink distances *off* the spine by a factor p and *on* the spine by a factor q and scale the component $\mu - \ell_{\mathscr{T}}$ of the measure μ by a factor p.

As for $\mathscr{C}_{p,q}$, the map $\mathscr{S}_{p,q} : \mathfrak{T}_e^\ell \to \mathfrak{T}_e^\ell$ is not continuous when $p \neq q$, but is the limit as $R \to \infty$ of continuous maps $\mathscr{S}_{p,q}^R$ defined as follows: for $\mathscr{T} \in \mathfrak{T}_e^\ell$ and $R \geqslant 0$ consider the subset $\mathrm{Spine}_R(\mathscr{T}) \subset \mathscr{V}$ of the vertices of $\mathscr{T}$ that have a descendant at distance larger than R. Then define $\mathscr{S}_{p,q}^R$ analogously to $\mathscr{S}_{p,q}$, using Spine_R instead of Spine.[6] For every $R \geqslant 0$, the map $\mathscr{S}_{p,q}^R$ is continuous, as can be shown by straightforward but fairly technical arguments, see for example the proof of Lemma 2.6 (ii) in [EPW06] for a similar situation. The map $\mathscr{S}_{p,q} = \lim_{R \to \infty} \mathscr{S}_{p,q}^R$ is then measurable and extends to a map $\mathscr{S}_{p,q} : \mathscr{M}_1(\mathfrak{T}_e^\ell) \to \mathscr{M}_1(\mathfrak{T}_e^\ell)$, allowing us to write $\mathscr{S}_{p,q}(\mathscr{T})$ for a *random* tree $\mathscr{T} \in \mathscr{M}_1(\mathfrak{T}_e^\ell)$.

DEFINITION 1.6. — Let $p, q \in (0, 1)$. We say that a random rooted measured $\mathbb{R}$-tree $\mathscr{T} \in \mathscr{M}_1(\mathfrak{T}_e^\ell)$ is (p, q)-self-similar, if $\mathscr{T} \stackrel{\text{law}}{=} \mathscr{S}_{p,q}(\mathscr{T})$.

We now define a discretization operation on the space $\mathfrak{T}_e^\ell$, which will allow to turn a self-similar $\mathbb{R}$-tree into a self-similar discrete tree. For this, we recall that as for discrete trees, every rooted $\mathbb{R}$-tree $\mathscr{T} = (\mathscr{V}, d, \rho)$ induces a partial order $\preceq_{\mathscr{T}}$ on its

---

[6]Formally, in the tree $\mathscr{T}' = \mathscr{S}_{p,q}^R(\mathscr{T}) = (\mathscr{V}', d', \rho', \mu')$, one has $d'(x, y) = p\,d(x, y) + (q - p)d(\vec{x}_R, \vec{y}_R)$ and $\mu' = p\mu + (q - p)\ell_{\mathscr{T}}(\cdot \cap \mathrm{Spine}_R(\mathscr{T}))$ with $\vec{x}_R$ the most recent ancestor of x in $\mathrm{Spine}_R(\mathscr{T})$.

set of vertices $\mathscr{V}$ by $x \preceq_{\mathscr{T}} y$ if and only if $x \in [\![\rho, y]\!]$, where $[\![\rho, y]\!]$ is the range of the geodesic from ρ to y (however, it is not true anymore that $\mathscr{T}$ can be recovered from $\preceq_T$). Again, we write $x \prec_{\mathscr{T}} y$ when $x \preceq_{\mathscr{T}} y$ and $x \neq y$. We can now define the following discretization operation:
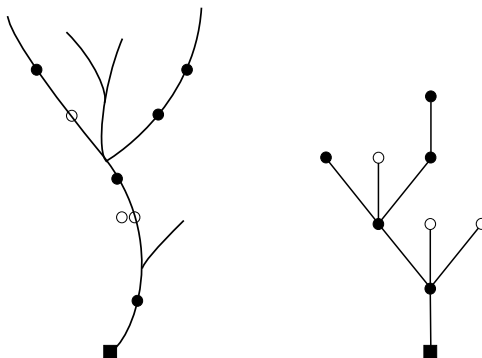

FIGURE 1.1. An $\mathbb{R}$-tree $\mathscr{T}$ (left) and a realization of its discretization $\mathscr{D}(\mathscr{T})$ (right). Black circles are vertices in V_0, and white circles are vertices in V_1. The square corresponds to the root. A vertex in V_1 is never an ancestor in $\mathscr{D}(\mathscr{T})$. The two neighbouring white circles on the left indicate that two points have been sampled at the same spot; since vertices in V_1 are sampled according to the measure $\mu - \ell_{\mathscr{T}}$, this is possible only if the measure μ has atoms.

DEFINITION 1.7. — Let $\mathscr{T} = (\mathscr{V}, d, \rho)$ be a rooted $\mathbb{R}$-tree. Let V_0 be a subset of $\mathscr{V}$ containing ρ and V_1 be a multiset of elements of $\mathscr{V}$, formally, this can be defined as a counting measure on $\mathscr{V}$. Suppose that V_0 and V_1 are boundedly finite in the sense that they contain a finite number of elements in each ball of finite radius. The *discretized tree* $\mathscr{D}(\mathscr{T}, V_0, V_1)$ is the rooted discrete tree T with vertex set $V_0 \cup V_1$ and whose ancestral relation $\prec_T$ is defined as follows:

$$\forall v, w \in V_0 \cup V_1 : v \prec_T w \iff v \prec_{\mathscr{T}} w \textbf{ and } v \in V_0.$$

See Figure 1 for a graphical illustration.

EXAMPLE 1.8. — If $\mathscr{T}$ is the tree reduced to the root element ρ, V_1 is the multi-set that contains n times the root ρ, V_0 is the empty set, and the discretized tree $\mathscr{D}(\mathscr{T}, V_0, V_1)$ has n edges adjacent to the root ρ.

DEFINITION 1.9. — For a tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_e^\ell$, define a random, rooted, discrete tree $\mathscr{D}(\mathscr{T}) \in \mathscr{M}_1(\mathbb{T}_e)$ by $\mathscr{D}(\mathscr{T}) = \mathscr{D}(\mathscr{T}, V_0 \cup \{\rho\}, V_1)$, where

   — V_0 is the set of atoms of a Poisson process on $\mathscr{V}$ with intensity measure $\ell_{\mathscr{T}}$, and
   — V_1 is the multiset of atoms of a Poisson process on $\mathscr{V}$ with intensity measure $\mu - \ell_{\mathscr{T}}$.

We will see below that the map $\mathscr{D}$ is actually continuous and thus can be extended to a map $\mathscr{D} : \mathscr{M}_1(\mathfrak{T}_e^\ell) \to \mathscr{M}_1(\mathbb{T}_e)$ (see Section B). Standard properties of Poisson processes now yield the following commutation relation between $\mathscr{S}_{p,q}$, $\mathscr{C}_{p,q}$ and $\mathscr{D}$.

LEMMA 1.10. — *We have the following equality of maps from $\mathscr{M}_1(\mathfrak{T}_e^\ell)$ to $\mathscr{M}_1(\mathbb{T}_e)$:*

$$\mathscr{D} \circ \mathscr{S}_{p,q} = \mathscr{C}_{p,q} \circ \mathscr{D}.$$

The (omitted) proof is based on the simple relation that exists between Poisson processes on $\mathscr{S}_{p,q}(\mathscr{T})$ and $\mathscr{T}$: a Poisson process on $\mathscr{S}_{p,q}(\mathscr{T})$ is distributed as a thinning of a Poisson process on $\mathscr{T}$, in which each atom is retained independently with probability q if the atom lies on $\mathrm{Spine}(\mathscr{T})$, and with probability p otherwise.

Lemma 1.10 gives a way of constructing self-similar discrete trees from self-similar $\mathbb{R}$-trees. The following theorem, the main result of this article, says that all self-similar discrete trees arise this way.

THEOREM 1. — *Let $p, q \in (0, 1)$. There exists a one-to-one correspondence between (p, q)-self-similar random rooted discrete trees $T \in \mathscr{M}_1(\mathbb{T}_e)$ and (p, q)-self-similar random rooted measured $\mathbb{R}$-trees $\mathscr{T} \in \mathscr{M}_1(\mathfrak{T}_e^\ell)$ given by*

$$T = \mathscr{D}(\mathscr{T}).$$

The methods used to prove Theorem 1 will also allow to characterize *compatible sequences* of random trees.

DEFINITION 1.11. — For a random rooted tree $T = (V, E, \rho) \in \mathscr{M}_1(\mathbb{T}_f)$ with $\#V = n + 1$ almost surely, define for $m \leqslant n$, $\mathscr{C}(T, m) = \mathscr{C}(T, V' \cup \{\rho\})$, where V' is a uniformly chosen subset of $V \smallsetminus \{\rho\}$ with m (distinct) elements. A family of random rooted trees $(T_n)_{n \in \mathbb{N}} \in \mathscr{M}_1(\mathbb{T}_f)$ with $\#V(T_n) = n + 1$ almost surely is called *compatible*, if for each $n \geqslant m \geqslant 1$, we have $T_m \overset{\mathrm{law}}{=} \mathscr{C}(T_n, m)$.

THEOREM 2. — *There exists a one-to-one correspondence between compatible families of random rooted trees $(T_n)_{n \in \mathbb{N}}$ and random rooted measured $\mathbb{R}$-trees $\mathscr{T} \in \mathscr{M}_1(\mathfrak{T}_1^\ell)$, given by $T_n = \mathscr{D}(\mathscr{T}, n)$ for $n \geqslant 1$, where $\mathscr{D}(\mathscr{T}, n)$ is the tree $\mathscr{D}(\mathscr{T})$ conditioned on having $n + 1$ vertices.*

The cornerstone in the proof of Theorems 1 and 2 will be the study of the continuity of the operator $\mathscr{D}$ and related topological issues. The following theorem summarizes the results thus obtained:

THEOREM 3. — *The map $\mathscr{D} : \mathfrak{T}_e^\ell \to \mathscr{M}_1(\mathbb{T}_e)$ as well as its extension $\mathscr{D} : \mathscr{M}_1(\mathfrak{T}_e^\ell) \to \mathscr{M}_1(\mathbb{T}_e)$ are continuous, closed, injective maps. In other words, they are homeomorphisms onto their images and these are closed subsets of $\mathscr{M}_1(\mathbb{T}_e)$.*

OVERVIEW OF THE PAPER. — We start by proving Theorem 3 in Sections 2 and 3. In Section 2 we consider the space $\mathfrak{T}_1^\ell$ only, the main result here is Proposition 2.1 which shows that $\mathfrak{T}_1^\ell$ is compact and that the restriction of $\mathscr{D}$ to $\mathfrak{T}_1^\ell$ is continuous, closed and injective. The proof relies on the use of random distance matrices and certain

exchangeable partial orders,[7] together with a somewhat intricate analysis of the continuity of a certain class of polynomial test functions.[8]

In Section 3, we prove Theorem 3 in full generality, i.e., on the space $\mathfrak{T}_e^\ell$ of infinite trees. This is probably the most technical section; we make use several different ways of truncating the trees and some technical arguments to bound the number of ends in the trees. Precompactness arguments play an important role.

In Section 4, Theorems 1 and 2 are proven. At the heart of the proofs is the following idea: First, we turn a discrete tree into an $\mathbb{R}$-tree by assigning length 1 to each edge. We then show that rescaling and discretizing that $\mathbb{R}$-tree yields with high probability the same result as contracting the original tree, on an arbitrarily large finite ball (Lemma 4.1). Together with Theorem 3, this allows to construct the real trees $\mathscr{T}$ in Theorems 1 and 2 as scaling limits of the discrete tree T, respectively, the sequence of discrete trees T_n.

In Section 5, motivated by the correspondence between self-similar discrete trees and self-similar $\mathbb{R}$-trees established in Theorem 1, we study examples of self-similar $\mathbb{R}$-trees. We give there an overview of the generality of examples that can be constructed. Like in the case of self-similar real-valued processes, see [OV85], it seems out of reach to completely characterize this family. We therefore consider in Section 6 a specific class of (p,q)-self-similar trees, namely those that are invariant with respect to translation along the spine (we suppose here for simplicity that the spine consists of a single infinite ray). In particular, in the case of self-similar trees consisting of a single spine to which iid subtrees are attached, we relate the construction of the corresponding $\mathbb{R}$-trees to the quasi-stationary distributions of linear-fractional subcritical Galton–Watson processes, see Proposition 6.4 and Remark 6.6.

In the short Section 7, we describe another attempt to prove Theorem 2 using exchangeability, which we initially pursued but dropped because of its drawbacks.

Finally, an appendix recalls some notions on the space of discrete trees and $\mathbb{R}$-trees that we consider in this work.

---

[7]We remark that exchangeable partial orders have been previously studied in generality by Janson [Jan11] who provided a limiting representation based on the dense graph limits introduced by Lovász and Szegedy [LS06]. It is however not clear to us how to make an efficient use of this representation for the questions studied here.

[8]A previous version of this article (published on the arXiv) contained a different, but incomplete proof.

2. The map $\mathscr{D}$: finite trees

In this section, we study the action of $\mathscr{D}$ on the space $\mathfrak{T}_1^\ell$, i.e., those measured trees in $\mathfrak{T}_e^\ell$ whose measure is a probability measure. The results are summarized in the following proposition:

PROPOSITION 2.1. — *The restriction of the map $\mathscr{D}$ to $\mathfrak{T}_1^\ell$ is a homeomorphism onto its image. Furthermore, the spaces $\mathfrak{T}_1^\ell$ and $\mathscr{D}(\mathfrak{T}_1^\ell)$ are compact.*
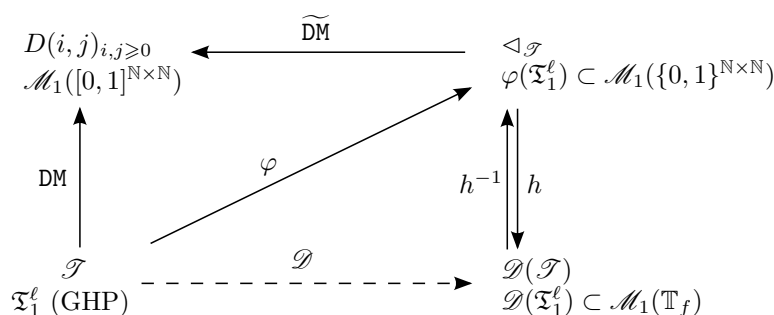

FIGURE 2.1. The different spaces used to prove Proposition 2.1 and the maps between them.

For this we rely on two different representations of random rooted measured trees, one in terms of its distance matrix distribution defined in Section A.2, the other in terms of an exchangeable partial order on $\mathbb{N}$ (the relationships between these representations is summarized in Figure 2.1). Set $\mathbb{N}^* := \{1, 2, \dots\}$. We recall that given a tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_1$, its *distance matrix distribution* $\mathrm{DM}(\mathscr{T})$ is defined as the push-forward of the probability measure $\delta_\rho \otimes \mu^{\otimes \mathbb{N}^*}$ by the map

$$(x_i)_{i \in \mathbb{N}} \longmapsto d(x_i, x_j)_{(i,j) \in \mathbb{N} \times \mathbb{N}}.$$

Since the distance between any two points of the tree $\mathscr{T} \in \mathfrak{T}_1^\ell$ is less than or equal to 1, its distance matrix distribution $\mathrm{DM}(\mathscr{T})$ is a probability measure on $[0,1]^{\mathbb{N} \times \mathbb{N}}$, endowed with the product topology.

The usefulness of the distance matrix distribution comes from the fact that it is convergence determining as recalled in the appendix. Here, we will use the following fact, which follows from the second part of Lemma A.8 together with Lemma A.9:

LEMMA 2.2. — *A sequence $\mathscr{T}_n \in \mathfrak{T}_1^\ell$ converges to a limit $\mathscr{T} \in \mathfrak{T}$ if and only if its distance matrix distributions $\mathrm{DM}(\mathscr{T}_n)$ converge. In this case, $\mathscr{T} \in \mathfrak{T}_1$ and $\mathrm{DM}(\mathscr{T}) = \lim_{n \to \infty} \mathrm{DM}(\mathscr{T}_n)$.*

Denote by $\mathscr{P}$ the space of random partial orders $\lhd$ on $\mathbb{N}$ which are invariant under finite permutations of $\mathbb{N}^*$ and such that $0 \lhd n$ for every $n \in \mathbb{N}^*$. $\mathscr{P}$ identifies with a closed subspace of the space of probability measures on $\{0,1\}^{\mathbb{N} \times \mathbb{N}}$ through the map $\lhd \mapsto (\mathbb{1}_{i \lhd j})_{(i,j) \in \mathbb{N} \times \mathbb{N}}$, therefore $\mathscr{P}$ is compact. Every tree $\mathscr{T} \in \mathfrak{T}_1^\ell$ defines an

element $\varphi(\mathscr{T})$ in $\mathscr{P}$ as follows: Fix a representative of the tree $\mathscr{T}$, also denoted by $\mathscr{T}$. Let $(X_1, S_1), (X_2, S_2), \ldots$ be an iid sequence of random variables in $\mathscr{V} \times \{0, 1\}$ with law $\ell_{\mathscr{T}} \otimes \delta_0 + (\mu - \ell_{\mathscr{T}}) \otimes \delta_1$. This means that $X_1, X_2, \ldots$ are iid according to μ and $S_i = 0$ if X_i was drawn according to $\ell_{\mathscr{T}}$ and $S_i = 1$ otherwise. Set $(X_0, S_0) = (\rho, 0)$. We then define the transitive relation $\lhd_{\mathscr{T}}$ on $\mathbb{N}$ by

$$(2.1) \qquad\qquad i \lhd_{\mathscr{T}} j \iff X_i \prec_{\mathscr{T}} X_j \text{ and } S_i = 0,$$

and denote its law by $\varphi(\mathscr{T})$. The relation $\lhd_{\mathscr{T}}$ is not reflexive, therefore is not a partial order.

Note that for a tree $\mathscr{T} \in \mathfrak{T}_1^\ell$, the binary relation $\lhd_{\mathscr{T}}$ is in direct relation to the discretization $\mathscr{D}(\mathscr{T})$. Indeed, if one denotes by $\lhd_{\mathscr{T}}^N$ the restriction of $\lhd_{\mathscr{T}}$ to the elements $\{0, \ldots, N\}$, with $N \sim \mathrm{Poi}(1)$, then the random rooted tree on $N + 1$ vertices with ancestral relation $\lhd_{\mathscr{T}}^N$ has the same distribution as $\mathscr{D}(\mathscr{T})$. Conversely, given $\mathscr{D}(\mathscr{T})$, one can define the restrictions $\lhd_{\mathscr{T}}^n$, $n = 1, 2, \ldots$ by conditioning the tree $\mathscr{D}(\mathscr{T})$ on having $n + 1$ vertices, assigning random labels $\{1, \ldots, n\}$ to the non-root vertices and then defining $\lhd_{\mathscr{T}}^n$ as the ancestral relation of this tree. This uniquely defines $\lhd_{\mathscr{T}}$.

We thus have constructed a bijection $h : \varphi(\mathfrak{T}_1^\ell) \to \mathscr{D}(\mathfrak{T}_1^\ell)$ such that $\mathscr{D} = h \circ \varphi$. Moreover, this bijection is a homeomorphism because for every n, the map assigning the law of $\mathscr{D}(\mathscr{T})$ conditioned on having $n + 1$ vertices to the law of $\lhd_{\mathscr{T}}^n$ is continuous (as well as its inverse) by finiteness of the corresponding spaces. We have thus proven the following lemma:

LEMMA 2.3. — *There exists a homeomorphism* $h : \varphi(\mathfrak{T}_1^\ell) \to \mathscr{D}(\mathfrak{T}_1^\ell)$ *such that* $\mathscr{D} = h \circ \varphi$.

As a consequence of the previous result, in order to show Proposition 2.1, it will be enough to prove the following lemma.

LEMMA 2.4. — φ *is a homeomorphism between* $\mathfrak{T}_1^\ell$ *and its image* $\varphi(\mathfrak{T}_1^\ell)$. *Moreover, both spaces are compact.*

In order to prove Lemma 2.4, we will make a detour by the space of random distance matrices. To wit, to every element $\lhd$ from $\mathscr{P}$, we associate a random distance matrix $\widetilde{D}(i, j)_{i,j \geqslant 0}$ by

$$(2.2) \qquad\qquad \widetilde{D}(i, j) = \lim_{n \to \infty} \frac{1}{n} \sum_{\substack{k=1 \\ k \notin \{i,j\}}}^{n} \mathbb{1}_{k \lhd i,\, k \ntriangleleft j \text{ or } k \lhd j,\, k \ntriangleleft i}.$$

The existence of the limit is provided by de Finetti's theorem applied to the exchangeable sequence $(\mathbb{1}_{k \lhd i,\, k \ntriangleleft j \text{ or } k \lhd j,\, k \ntriangleleft i})_{k \neq i,j}$. One easily verifies that $\widetilde{D}$ satisfies the triangle inequality almost surely, whence we call it a random distance matrix. We then denote the map associating the law of the random distance matrix $\widetilde{D}$ to (the law of) $\lhd$ by $\widetilde{\mathrm{DM}} : \mathscr{P} \to \mathscr{M}_1([0, 1]^{\mathbb{N} \times \mathbb{N}})$. We have the following lemma.

LEMMA 2.5. — $\mathrm{DM} = \widetilde{\mathrm{DM}} \circ \varphi$ *on* $\mathfrak{T}_1^\ell$.

Proof. — Fix a representative of a tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_1^\ell$. Let $((X_i, S_i))_{i \geqslant 0}$ be as above and define $D(i,j) = d(X_i, X_j)$, such that D follows the law $\mathtt{DM}(\mathscr{T})$. By the law of large numbers, this gives,

$$D(i,j) = \ell_{\mathscr{T}}(\llbracket X_i, X_j \rrbracket) = \lim_{n \to \infty} \frac{1}{n} \sum_{\substack{k=1 \\ k \notin \{i,j\}}}^{n} \mathbb{1}_{X_k \in \llbracket X_i, X_j \rrbracket, \, S_k = 0}$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{\substack{k=1 \\ k \notin \{i,j\}}}^{n} \mathbb{1}_{k \lhd_{\mathscr{T}} i, \, k \ntriangleleft_{\mathscr{T}} j \text{ or } k \lhd_{\mathscr{T}} j, \, k \ntriangleleft_{\mathscr{T}} i},$$

where the last equality follows from the definition of $\lhd_{\mathscr{T}}$. Equation (2.2) then shows that D follows the law $\widetilde{\mathtt{DM}}(\lhd_{\mathscr{T}})$, which proves the lemma. $\qquad \square$

Lemma 2.6. — *The space $\mathfrak{T}_1^\ell$ is compact.*

Proof. — By the first part of Lemma A.8, the space $\mathfrak{T}_1^\ell$ is precompact in $\mathfrak{T}$. It thus suffices to show that $\mathfrak{T}_1^\ell$ is closed.

We first introduce some more notation. Fix a representative of a tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_1$. Let $(X_i)_{i \geqslant 0}$ be as above and define $D(i,j) = d(X_i, X_j)$, such that D follows the law $\mathtt{DM}(\mathscr{T})$. Now set for $m \in \mathbb{N}^*$,

$$(2.3) \qquad M^{(m)}(1,2) = \frac{1}{m} \sum_{k=3}^{m} \mathbb{1}_{D(1,k)+D(k,2)=D(1,2)}.$$

By the law of large numbers, $M(1,2) = \lim_{m \to \infty} M^{(m)}(1,2)$ exists almost surely and equals $\mu(\llbracket X_1, X_2 \rrbracket)$. Moreover, conditioned on $M(1,2)$, the random variables $Y_k = \mathbb{1}_{D(i,k)+D(k,j)=D(i,j)}$, $k = 3, 4, \ldots$ are iid Bernoulli with parameter $M(1,2)$. By the conditional Chebychev inequality, we therefore have for every $\varepsilon > 0$,

$$(2.4) \qquad \mathbf{P}(|M^{(m)}(1,2) - M(1,2)| > \varepsilon|) \leqslant \frac{1}{m\varepsilon^2} \mathbf{E}[\mathrm{Var}(Y_3 \mid M(1,2))] \leqslant \frac{1}{4m\varepsilon^2}.$$

In other words, the convergence of $M^{(m)}(1,2)$ to $M(1,2)$ is uniform in $\mathscr{T}$.

Now, suppose there exist $\mathscr{T}_1, \mathscr{T}_2, \cdots \in \mathfrak{T}_1^\ell$ such that $\mathscr{T}_n$ converges to $\mathscr{T} \in \mathfrak{T}$ as $n \to \infty$. By Lemma 2.2, $\mathscr{T} \in \mathfrak{T}_1$ and $\mathtt{DM}(\mathscr{T}_n) \to \mathtt{DM}(\mathscr{T})$ as $n \to \infty$. Define D, $M(1,2)$ and $M^{(m)}(1,2)$ as above and define D_n, $M_n(1,2)$ and $M_n^{(m)}(1,2)$ analogously for every $n \in \mathbb{N}$. By Skorokhod's representation theorem we can and will assume that the distance matrices D_n converge (pointwise) almost surely to D. We then have almost surely, for every $m \in \mathbb{N}$,

$$M^{(m)}(1,2) \geqslant \lim_{n \to \infty} M_n^{(m)}(1,2),$$

since the RHS in (2.3) is an upper semi-continuous function in D for every m. The uniform convergence of $M^{(m)}(1,2)$ to $M(1,2)$ proven in (2.4) then shows that $M(1,2) \geqslant \lim_{n \to \infty} M_n(1,2)$. Moreover, since $\mathscr{T}_n \in \mathfrak{T}_1^\ell$ for every n, we have $M_n(1,2) \geqslant D_n(1,2)$ almost surely for every n, so that almost surely

$$M(1,2) \geqslant \lim_{n \to \infty} D_n(1,2) = D(1,2).$$

But since $M(1,2) = \mu([\![X_1, X_2]\!])$ and $D(1,2) = \ell_{\mathscr{T}}([\![X_1, X_2]\!])$ with X_1, X_2 iid according to μ, this implies that $\mu \geqslant \ell_{\mathscr{T}}$ on its support. By Lemma A.8, the measure μ has full support, whence $\mu \geqslant \ell_{\mathscr{T}}$. This shows that $\mathscr{T} \in \mathfrak{T}_1^\ell$ and hence $\mathfrak{T}_1^\ell$ is a closed subspace of the compact space $\mathfrak{T}_1$.   □

LEMMA 2.7. — *The map φ is continuous on $\mathfrak{T}_1^\ell$.*

Proof. — We will show more in fact: we show that the map which to a tree $\mathscr{T} \in \mathfrak{T}_1^\ell$ assigns the law of $(D_{\mathscr{T}}, \lhd_{\mathscr{T}})$ is continuous, where $D_{\mathscr{T}}$ is the distance matrix of the tree $\mathscr{T}$. For this, we will consider test functions of the form

$$(2.5) \qquad f(D, \lhd) = C \prod_{i,j=0}^{n} D(i,j)^{\beta_{ij}} \prod_{l=1}^{L} \mathbb{1}_{a_l \lhd b_l},$$

where $C \in \mathbb{R}$, $n \in \mathbb{N}$, $\beta_{ij} \in \mathbb{N}$, $L \geqslant 0$ and $a_l, b_l \in \{1, \ldots, n\}$. Note that $D_{\mathscr{T}}(i,j) \leqslant 1$ almost surely for all $\mathscr{T} \in \mathfrak{T}_1^\ell$ and $i, j \in \mathbb{N}$, whence we can formally see the pair $(D, \lhd)$ as a random variable taking values in the compact space $[0,1]^{\mathbb{N} \times \mathbb{N}} \times \{0,1\}^{\mathbb{N}^* \times \mathbb{N}^*}$. The vector space spanned by functions of the form (2.5) then forms an algebra of continuous functions on this space which separates points. By the Stone-Weierstrass theorem it is therefore enough to show that for a function f as in (2.5), the quantity $\mathbf{E}[f(D_{\mathscr{T}}, \lhd_{\mathscr{T}})]$ is continuous in $\mathscr{T}$. We will show this by induction on L. If $L = 0$, the assertion follows from the fact that the law of $D_{\mathscr{T}}$ is continuous in $\mathscr{T}$ by Lemma A.9 and Fact A.11. If $L > 0$, define the sets

$$A = \{a_l : l = 1, \ldots, L\} \quad \text{and} \quad B = \{b_l : l = 1, \ldots L\}.$$

We will distinguish two cases:

Case $A \subset B$. — In this case, there must exist a cycle[9] $l_1, \ldots, l_k, l_{k+1} = l_1$, such that $a_{l_i} = b_{l_{i+1}}$ for all $i = 1, \ldots, k$. In particular, $f(D, \lhd) \neq 0$ implies that

$$b_{l_1} = b_{l_{k+1}} = a_{l_k} \lhd b_{l_k} = a_{l_{k-1}} \lhd \cdots \lhd b_{l_1},$$

whence, by transitivity, $b_{l_1} \lhd b_{l_1}$. But by (2.1), we have $k \not\lhd_{\mathscr{T}} k$ for all $k \in \mathbb{N}$, whence $f(D_{\mathscr{T}}, \lhd_{\mathscr{T}}) \equiv 0$. In particular, $\mathbf{E}[f(D_{\mathscr{T}}, \lhd_{\mathscr{T}})]$ is continuous in $\mathscr{T}$.

Case $A \not\subset B$. — In this case, let $\alpha \in A \smallsetminus B$. Define the sets

$$\Lambda = \{l \in \{1, \ldots, L\} : a_l = \alpha\}, \quad \overline{\Lambda} = \{1, \ldots, L\} \smallsetminus \Lambda.$$

Note that $\Lambda \neq \varnothing$, such that $\#\overline{\Lambda} < L$. We will show that we can express $\mathbf{E}[f(D_{\mathscr{T}}, \lhd_{\mathscr{T}})]$ as a sum of expressions of the same type, with functions of the form (2.5) containing only the indicator functions corresponding to $l \in \overline{\Lambda}$. This will allow us to complete the induction step.

---

[9]To see this, start with $l_1 = 1$ and let l_2 be such that $a_{l_1} = b_{l_2}$. Then let l_3 be such that $a_{l_2} = b_{l_3}$ and so on. Since the l_i's take values in the finite set $\{1, \ldots, L\}$, a cycle has to appear at some point.

Recall the construction of $D_{\mathscr{T}}$ and $\lhd_{\mathscr{T}}$ from the sequence $(X_0, S_0), (X_1, S_1), \ldots$ defined above. Since $\Lambda \neq \varnothing$, we have by definition,

$$(2.6) \qquad \prod_{l \in \Lambda} \mathbb{1}_{a_l \lhd_{\mathscr{T}} b_l} = \mathbb{1}_{S_\alpha = 0} \prod_{l \in \Lambda} \mathbb{1}_{X_\alpha \prec_{\mathscr{T}} X_{b_l}}.$$

Now define $B_\Lambda = \{b_l : l \in \Lambda\}$. We can assume that $\alpha \notin B_\Lambda$, otherwise the function $\mathbb{1}_{\alpha \lhd \alpha}$ would be a factor of $f(D, \lhd)$ and therefore $f(D_{\mathscr{T}}, \lhd_{\mathscr{T}}) \equiv 0$. Let X' be the most recent common ancestor of the vertices X_b, $b \in B_\Lambda$, which is $\sigma(X_b; b \in B_\Lambda)$-measurable. By the definition of the sequence $(X_i, S_i)_{i \geqslant 0}$, conditioned on $(X_k, S_k)_{k \neq \alpha}$, the event $\{S_\alpha = 0, X_\alpha \prec_{\mathscr{T}} X'\}$ has probability $d_{\mathscr{T}}(\rho, X')$ and conditioned on this event, X_α is uniformly distributed on $[\![\rho, X']\!]$. Furthermore, if $X' \wedge X_i$ denotes the most recent common ancestor of X' and X_i, then if $X_\alpha \in [\![\rho, X']\!]$, we have for every $i \neq \alpha$,

$$D_{\mathscr{T}}(\alpha, i) = d_{\mathscr{T}}(X_\alpha, X_i) = d_{\mathscr{T}}(X_\alpha, X' \wedge X_i) + d_{\mathscr{T}}(X' \wedge X_i, X_i)$$
$$= |d_{\mathscr{T}}(\rho, X_\alpha) - d_{\mathscr{T}}(\rho, X' \wedge X_i)| + d_{\mathscr{T}}(\rho, X_i) - d_{\mathscr{T}}(\rho, X' \wedge X_i).$$

We now use this to calculate the expectation of the factors of $f(D_{\mathscr{T}}, \lhd_{\mathscr{T}})$ involving α, conditioned on the remaining ones. Since $D_{\mathscr{T}}$ is symmetric, we can assume that $\beta_{i\alpha} = 0$ for all $i \neq \alpha$, and furthermore, $\beta_{\alpha\alpha} = 0$, since $D_{\mathscr{T}}(\alpha, \alpha) = 0$ almost surely. From the above, we then have

$$(2.7) \quad \mathbf{E}\Big[\prod_{i \in \{0, \ldots, n\} \smallsetminus \alpha} D(\alpha, i)^{\beta_{\alpha i}} \mathbb{1}_{(S_\alpha = 0, \, X_\alpha \lhd_{\mathscr{T}} X')} \, \Big| \, (X_k, S_k)_{k \neq \alpha} \Big]$$
$$= \mathbf{E}_U\Big[\prod_{i \in \{0, \ldots, n\} \smallsetminus \alpha} \big(|U - x_i| + y_i\big)^{\beta_{\alpha i}} \mathbb{1}_{(U < z)} \Big],$$

where $U \sim \mathrm{Unif}(0, 1)$, $\mathbf{E}_U$ denotes expectation w.r.t. U and x_i, y_i, z are

$$x_i = d_{\mathscr{T}}(\rho, X' \wedge X_i), \quad y_i = d_{\mathscr{T}}(\rho, X_i) - d_{\mathscr{T}}(\rho, X' \wedge X_i), \quad z = d_{\mathscr{T}}(\rho, X').$$

Each factor in the product on the RHS of (2.7) can be written as

$$\big(|U - x_i| + y_i\big)^{\beta_{\alpha i}} \mathbb{1}_{(U < z)} = \big(U - x_i - 2(U - x_i)\mathbb{1}_{(U < x_i)} - y_i\big)^{\beta_{\alpha i}} \mathbb{1}_{(U < z)}.$$

Expanding the powers and the product then gives that

$$\prod_{i \in \{0, \ldots, n\} \smallsetminus \alpha} \big(|U - x_i| + y_i\big)^{\beta_{\alpha i}} \mathbb{1}_{(U < z)} = \sum_{\xi \in \{x_i, i \neq \alpha\} \cup \{z\}} P_\xi(U, (x_i)_{i \neq \alpha}, (y_i)_{i \neq \alpha}) \mathbb{1}_{(U < \xi)},$$

where each P_ξ is a polynomial in its arguments. Taking expectations and using the fact that $\mathbf{E}_U[U^{k-1} \mathbb{1}_{(U < \xi)}] = \xi^k / k$ for every $k \geqslant 1$ and $\xi \in [0, 1]$, we obtain that the RHS of (2.7) is a polynomial in $(x_i)_{i \neq \alpha}$, $(y_i)_{i \neq \alpha}$ and z.

Now note that for every $i \neq \alpha$, $d_{\mathscr{T}}(\rho, X')$ and $d_{\mathscr{T}}(\rho, X' \wedge X_i)$ are linear combinations[10] of $d_{\mathscr{T}}(X_i, X_j) = D_{\mathscr{T}}(i, j)$, $i, j \in \{0, \ldots, n\} \smallsetminus \{\alpha\}$. Together with the above, this implies that the RHS in (2.7) is in fact a polynomial in $D_{\mathscr{T}}(i, j)$, $i, j \in \{0, \ldots, n\} \smallsetminus \{\alpha\}$.

---

[10]For two vertices x, y, if z denotes their most recent common ancestor, $d_{\mathscr{T}}(\rho, z) = \frac{1}{2}(d_{\mathscr{T}}(\rho, x) + d_{\mathscr{T}}(\rho, y) - d_{\mathscr{T}}(x, y))$. The statement follows easily from this by induction.

Plugging this into (2.5) and using (2.6) yields that $\mathbf{E}[f(D_{\mathscr{T}}, \lhd_{\mathscr{T}})]$ can be written as the sum of expressions of the same type, with functions of the form (2.5) with L replaced by $\#\overline{\Lambda} < L$. This finishes the induction step and therefore the proof of the lemma. $\qquad\square$

Proof of Lemma 2.4. — Since $\mathfrak{T}_1^\ell$ is compact by Lemma 2.6 and $\varphi(\mathfrak{T}_1^\ell)$ is Hausdorff, it is enough to show that φ is bijective and continuous [Dug66, Th. XI.2.1]. Continuity follows from Lemma 2.7, and bijectivity follows from Lemma 2.5 and the fact that DM is injective by Fact A.10. $\qquad\square$

Proof of Proposition 2.1. — Follows immediately from Lemmas 2.3 and 2.4. $\qquad\square$

3. The map $\mathscr{D}$: infinite trees (proof of Theorem 3)

In this section, we prove Theorem 3. The proof uses a series of lemmas. The first one concerns the continuity of the map $\mathscr{D}$:

LEMMA 3.1. — *The map* $\mathscr{D} : \mathfrak{T}_e^\ell \to \mathscr{M}_1(\mathbb{T}_e)$ *is continuous.*

In view of Lemma 3.1 and Section B, we can extend $\mathscr{D}$ to a continuous map $\mathscr{D} : \mathscr{M}_1(\mathfrak{T}_e^\ell) \to \mathscr{M}_1(\mathbb{T}_e)$. The next lemma shows that this extension is injective.

LEMMA 3.2. — *The map* $\mathscr{D} : \mathfrak{T}_e^\ell \to \mathscr{M}_1(\mathbb{T}_e)$ *and its extension* $\mathscr{D} : \mathscr{M}_1(\mathfrak{T}_e^\ell) \to \mathscr{M}_1(\mathbb{T}_e)$ *are injective.*

The next lemma will be used to prove that the image $\mathscr{D}$ is closed in $\mathscr{M}_1(\mathbb{T}_e)$:

LEMMA 3.3. — *If* $\mathscr{T}_1, \mathscr{T}_2, \cdots \in \mathfrak{T}_e^\ell$ *(resp.,* $\mathscr{M}_1(\mathfrak{T}_e^\ell)$*), such that* $\mathscr{D}(\mathscr{T}_n)$ *converges in law to a random tree T supported on* $\mathbb{T}_e$*, then there exists* $\mathscr{T} \in \mathfrak{T}_e^\ell$ *(resp.,* $\mathscr{M}_1(\mathfrak{T}_e^\ell)$*), such that* $\mathscr{D}(\mathscr{T}) = T$ *and* $\mathscr{T}_n \to \mathscr{T}$*.*

The proof of Theorem 3 now directly follows from the previous results:

Proof of Theorem 3. — Continuity, injectivity and closedness of the map $\mathscr{D}$ and its extension are exactly Lemmas 3.1, 3.2 and 3.3, respectively. $\qquad\square$

One ingredient for Lemma 3.3 is the following lemma which is of independent interest:

LEMMA 3.4. — $\mathfrak{T}_e^\ell$, $\mathfrak{T}_f^\ell$ *and* $\mathfrak{T}_1^\ell$ *are closed subspaces of* $\mathfrak{T}_e$, $\mathfrak{T}_f$ *and* $\mathfrak{T}_1$*, respectively.*

The proofs of the four lemmas follow.

Proof of Lemma 3.1. — The continuity on the space $\mathfrak{T}_1^\ell$ follows from Proposition 2.1. There is the map:

$$\mathfrak{T}_f^\ell \longrightarrow \mathfrak{T}_1^\ell \times \mathbb{R}^+, \quad \mathscr{T} \longmapsto \left(\mathscr{S}_{\mu(\mathscr{V})^{-1}}(\mathscr{T}), \mu(\mathscr{T})\right)$$

that is continuous, therefore the continuity extends to the space $\mathfrak{T}_f^\ell$. We now turn to $\mathfrak{T}_e^\ell$. We consider $\mathscr{T}_n \in \mathfrak{T}_e^\ell$ with limit $\mathscr{T}$, and we need to prove that

$$(3.1) \qquad\qquad \mathscr{D}(\mathscr{T}_n)^{\leqslant m} \longrightarrow \mathscr{D}(\mathscr{T})^{\leqslant m}, \quad \text{for every } m \in \mathbb{N}.$$

The convergence of $\mathscr{T}_n$ implies by A.13 for each $R \geqslant 0$ the existence of a sequence $R_n \geqslant R$ converging to R and such that $\mathscr{T}_n^{\leqslant R_n} \to \mathscr{T}^{\leqslant R}$. The convergence

$$(3.2) \qquad \mathscr{D}(\mathscr{T}_n^{\leqslant R_n}) \longrightarrow \mathscr{D}(\mathscr{T}^{\leqslant R})$$

follows since the truncated trees $\mathscr{T}_n^{\leqslant R_n}, \mathscr{T}^{\leqslant R}$ belong to $\mathfrak{T}_f^\ell$, on which the map $\mathscr{D}$ is continuous. Fix $m \in \mathbb{N}$ and $\varepsilon > 0$. Throughout the rest of the proof, we write $\mathscr{T}_\infty = \mathscr{T}$ and let n take values in $\mathbb{N} \cup \{\infty\}$. The convergence in (3.1) now follows from (3.2), $R_n \geqslant R$ and the following claim: there exist $R, n_0 < \infty$, such that

$$(3.3) \qquad \mathbf{P}(\mathscr{D}(\mathscr{T}_n)^{\leqslant m} \subset \mathscr{D}(\mathscr{T}_n^{\leqslant R})) > 1 - \varepsilon \quad \text{for } n \geqslant n_0.$$

To establish the claim, consider the paths in $\mathscr{T}_n$ of length R starting from the root. The number of these paths that have a distinct restriction to a distance r from the root, $r \leqslant R$, is $N_{r,R}(\mathscr{T}_n)$ defined before Lemma A.14. By that lemma, for every $r \geqslant 0$, there exist $R, n_0 < \infty$ such that $N_{r,R}(\mathscr{T}_n) \leqslant N$ for $n \geqslant n_0$, where N only depends on the family $(\mathscr{T}_n)_{n \in \mathbb{N} \cup \{\infty\}}$ and not on m, r, R, n_0. Now, the number of points in V_0 in a path of length r (in the discretization of $\mathscr{T}_n$) is a Poi(r)-distributed random variable X, and we can take r large enough, so that $\mathbf{P}(X < m) < \varepsilon/N$. A union bound then shows that, with probability at least $1 - \varepsilon$, the restriction to distance r of any path in $\mathscr{T}_n$ of length R starting from the root has at least m points in V_0, for $n \geqslant n_0$. This proves (3.3), and (3.1) follows. $\qquad\square$

Proof of Lemma 3.2. — Since $\mathscr{D}(\mathscr{T}) = \mathscr{D}(\delta_\mathscr{T})$ for every $\mathscr{T} \in \mathfrak{T}_e^\ell$, it suffices to show the injectivity of the extension only. The map DM is injective on the space $\mathscr{M}_1(\mathfrak{T}_1^\ell)$ by Fact A.10. By Lemma 1.10, the extension of the map φ to $\mathscr{M}_1(\mathfrak{T}_1^\ell)$ then is injective as well. Furthermore, the bijective correspondence between $\varphi(\mathscr{T})$ and $\mathscr{D}(\mathscr{T})$ for $\mathscr{T} \in \mathfrak{T}_1^\ell$ described before Lemma 2.4 readily extends to random trees. This immediately shows injectivity of $\mathscr{D}$ on $\mathscr{M}_1(\mathfrak{T}_1^\ell)$.

The injectivity of $\mathscr{D}$ on $\mathscr{M}_1(\mathfrak{T}_f^\ell)$ follows using the rescaling argument in the proof of Lemma 3.1.

To prove the injectivity of $\mathscr{D}$ on $\mathscr{M}_1(\mathfrak{T}_e^\ell)$, we introduce two pruning operations, on $\mathbb{R}$-trees and on discrete trees, that commute with $\mathscr{D}$. Let $\mathscr{T}$ be a random tree in $\mathfrak{T}_e^\ell$. For $\lambda > 0$, let $\mathscr{T}^\lambda$ be the random tree obtained by cutting the tree $\mathscr{T}$ at the atoms of a Poisson process with intensity $\lambda \ell_\mathscr{T}|_{\mathrm{Spine}(\mathscr{T})}$, and keeping the component containing the root. Also, let T be a random tree in $\mathbb{T}_e$. For $\lambda > 0$, let T^λ be the random tree obtained from T by first marking every vertex on the spine independently with probability $\lambda/(1 + \lambda)$, then removing all the marked vertices (together with the adjacent edges) and keeping the component containing the root. There is the equality in distribution[11]

$$\mathscr{D}(\mathscr{T}^\lambda) = \mathscr{D}(\mathscr{T})^\lambda.$$

---

[11]This relies on the fact that in the discretization operation the spine vertices are necessarily in V_0 and thus have been sampled according to the length measure $\ell_\mathscr{T}$.

We are now ready to complete the proof. Consider $\mathscr{T}_1$ and $\mathscr{T}_2$ two random trees in $\mathfrak{T}_e^\ell$ that satisfy $\mathscr{D}(\mathscr{T}_1) = \mathscr{D}(\mathscr{T}_2)$. We then have:

$$(3.4) \qquad \mathscr{D}(\mathscr{T}_1^\lambda) = \mathscr{D}(\mathscr{T}_1)^\lambda = \mathscr{D}(\mathscr{T}_2)^\lambda = \mathscr{D}(\mathscr{T}_2^\lambda).$$

Also, $\mathscr{T}_1^\lambda$ and $\mathscr{T}_2^\lambda$ are elements of $\mathscr{M}_1(\mathfrak{T}_f^\ell)$. This may be justified as follows (for $\mathscr{T}_1$, say): the tree $\mathscr{D}(\mathscr{T}_1)$ has an almost surely finite number of infinite paths starting from the root and every one of them is truncated in $\mathscr{D}(\mathscr{T}_1)^\lambda$ at a finite distance from the root, therefore the tree $\mathscr{D}(\mathscr{T}_1)^\lambda$ has a finite diameter and, being locally finite, it has finitely many vertices. Since the number of non-root vertices in $\mathscr{D}(\mathscr{T}_1)^\lambda$ is a $\mathrm{Poi}(\mu(\mathscr{T}_1^\lambda))$-distributed random variable, this entails that $\mu(\mathscr{T}_1^\lambda) < \infty$ almost surely, and since $\mu \geqslant \ell_{\mathscr{T}_1}$, that $\mathscr{T}_1^\lambda \in \mathscr{M}_1(\mathfrak{T}_f^\ell)$ almost surely. From (3.4) and the injectivity of $\mathscr{D}$ on $\mathscr{M}_1(\mathfrak{T}_f^\ell)$, we deduce that $\mathscr{T}_1^\lambda = \mathscr{T}_2^\lambda$, and the identity $\mathscr{T}_1 = \mathscr{T}_2$ follows taking the $\lambda \to 0$ limit. $\qquad\square$

Proof of Lemma 3.4. — The space $\mathfrak{T}_1^\ell$ is compact according to Lemma 2.6. In particular, it is closed in $\mathfrak{T}_1$. It remains to prove that $\mathfrak{T}_e^\ell$, $\mathfrak{T}_f^\ell$ are closed subspaces of $\mathfrak{T}_e$, $\mathfrak{T}_f$. For $\mathfrak{T}_f^\ell$, this follows by rescaling as in the proof of Lemma 3.1. For $\mathfrak{T}_e^\ell$, we use truncation: If $\mathscr{T}_n \in \mathfrak{T}_e^\ell$ converge to $\mathscr{T} \in \mathfrak{T}_e$, then, by A.13, for every $R > 0$, there exists a sequence $R_n \geqslant R$ converging to R such that $\mathscr{T}_n^{\leqslant R_n} \to \mathscr{T}^{\leqslant R}$, with $\mathscr{T}_n^{\leqslant R_n} \in \mathfrak{T}_f^\ell$ and $\mathscr{T}^{\leqslant R} \in \mathfrak{T}_f$. Since $\mathfrak{T}_f^\ell$ is closed in $\mathfrak{T}_f$, we have that $\mathscr{T}^{\leqslant R} \in \mathfrak{T}_f^\ell$ for every R, whence μ dominates $\ell_{\mathscr{T}}$ on every ball of radius R around the root in $\mathscr{T}$. Hence, $\mu \geqslant \ell_{\mathscr{T}}$ everywhere and thus $\mathscr{T} \in \mathfrak{T}_e^\ell$. $\qquad\square$

Proof of Lemma 3.3. — We only prove the Lemma for $\mathscr{M}_1(\mathfrak{T}_e^\ell)$, the proof for $\mathfrak{T}_e^\ell$ is similar. Assume that $(\mathscr{T}_n = (\mathscr{V}_n, d_n, \rho_n, \mu_n))_{n \geqslant 0}$ is a sequence of random trees taking values in $\mathscr{M}_1(\mathfrak{T}_e^\ell)$, such that $\mathscr{D}(\mathscr{T}_n)$ converges as $n \to \infty$ to a random tree in $\mathbb{T}_e$. It is enough to show that the sequence $(\mathscr{T}_n)_{n \geqslant 0}$ is precompact in $\mathscr{M}_1(\mathfrak{T}_e)$. Suppose indeed this holds. Let $\mathscr{T}^* \in \mathscr{M}_1(\mathfrak{T}_e)$ be a limit point of the sequence $(\mathscr{T}_n)_{n \geqslant 0}$. By Lemma 3.4, we actually have that $\mathscr{T}^* \in \mathscr{M}_1(\mathfrak{T}_e^\ell)$. By Lemma 3.1, $\mathscr{D}(\mathscr{T}^*) = \lim \mathscr{D}(\mathscr{T}_n)$. Lemma 3.2 now gives that $\mathscr{T}^*$ is unique, whence $\mathscr{T}_n \to \mathscr{T}^*$ in law as $n \to \infty$, which was to be proven.

We now show precompactness of $(\mathscr{T}_n)_{n \geqslant 0}$ in $\mathscr{M}_1(\mathfrak{T}_e)$. From the characterization of precompactness in Lemmata A.8 and A.14, there are two points to show:

(1) For every $r \geqslant 0$, the family of random variables $\mu_n(\mathscr{T}_n^{\leqslant r})$ is tight.

(2) For every $r \geqslant 0$, there exist $R = R(r)$ and $n_0 = n_0(r)$, such that the family of random variables $(N_{r,R(r)}(\mathscr{T}_n))_{r \geqslant 0, n \geqslant n_0(r)}$ is tight.

We prove these two points by contradiction.

First point. — Assume that there exists $t \geqslant 0$ such that the family of random variables $M_n := \mu_n(\mathscr{T}_n^{\leqslant r})$ is not tight. Restricting to a subsequence and conditioning on suitable events, we may assume that $\lim_{n \to \infty} M_n = +\infty$, in law. Under this assumption, we aim to show that there exists an integer t' such that

$$(3.5) \qquad \text{the sequence } (\#\mathscr{D}(\mathscr{T}_n)^{\leqslant t'})_{n \in \mathbb{N}} \text{ is not tight.}$$

By Lemma A.8, this will lead to the required contradiction.

Denote the set of vertices of the tree $\mathscr{D}(\mathscr{T}_n^{\leqslant t})$ by U_n and set $N_n = \#U_n$. We label the vertices in U_n at random by elements of $\{1, \ldots, N_n\}$, and define, for $k \in \mathbb{N}$, the random variable I_k by: $I_k = 1$ if $k \leqslant N_n$ and the vertex k has more than $4t$ non-root ancestors (that is, ancestors distinct from the root) below, and $I_k = 0$ otherwise. The sum $N_n^I = \sum_k I_k$ then gives the number of those vertices that have more than $4t$ non-root ancestors. The number of non-root ancestors of a vertex in U_n is, conditionally given N_n, dominated by a $\mathrm{Bin}(N_n - 1, t/M_n)$-distributed random variable, since the length measure of the path from this vertex to ρ in $\mathscr{T}_n^{\leqslant t}$ cannot exceed t by definition. We first compute the conditional expectation of N_n^I given N_n:

(3.6)
$$\mathbf{E}(N_n^I | N_n) = N_n \, \mathbf{E}(I_1 | N_n) \leqslant N_n \, \mathbf{P}\Big(\mathrm{Bin}\Big(N_n - 1, \frac{t}{M_n}\Big) \geqslant 4\,t \,\Big|\, N_n\Big)$$
$$\leqslant N_n \frac{(N_n - 1)}{4\,M_n},$$

using the conditional Markov inequality for the last estimate. Another application of the same inequality now gives:

$$\mathbf{P}(N_n^I > N_n/2) = \mathbf{E}(\mathbf{P}(N_n^I > N_n/2 | N_n)) \leqslant 2\,\mathbf{E}\Big(\frac{\mathbf{E}(N_n^I | N_n)}{N_n}\Big) \leqslant 1/2,$$

using (3.6) and the fact that N_n is a $\mathrm{Poi}(M_n)$-distributed random variable for the last inequality. Notice this inequality is equivalent to:

$$\mathbf{P}(N_n - N_n^I \geqslant N_n/2) = \mathbf{P}(N_n^I \leqslant N_n/2) \geqslant 1/2.$$

Now, fix an arbitrary $k \in \mathbb{N}$. Our assumption that M_n diverges in law implies that the sequence $\mathbf{P}(N_n/2 \geqslant k)$ has limit 1, and in particular is larger than $3/4$ for n large enough. This gives, for these values of n,

$$\mathbf{P}(N_n - N_n^I \geqslant k) \geqslant \mathbf{P}(N_n - N_n^I \geqslant N_n/2,\, N_n/2 \geqslant k)$$
$$\geqslant \mathbf{P}(N_n - N_n^I \geqslant N_n/2) + \mathbf{P}(N_n/2 \geqslant k) - 1$$
$$\geqslant 1/4,$$

and, together with the inequality $\#\mathscr{D}(\mathscr{T}_n)^{\leqslant t'} \geqslant N_n - N_n^I$ valid for $t' = 4t + 1$, this proves (3.5).

Second point. — Assume that for all choices of $R = R(r)$ and $n_0 = n_0(r)$ the family of random variables $(N_{r,R}(\mathscr{T}_n))_{r \geqslant 0, n \geqslant n_0}$ is not tight. This means that there exists $c > 0$, such that for every $N \in \mathbb{N}$, for some $r = r(N)$,

$$\limsup_{n \to \infty} \mathbf{P}(N_{r,R}(\mathscr{T}_n) \geqslant N) > c \quad \forall R \geqslant r,$$

where we used the fact that $N_{r,R}$ is decreasing in R. Since $N_{r,R}$ is also increasing in r, we may assume without loss of generality that $r(N) \to \infty$ as $N \to \infty$. Under this assumption, we show that for large N,

(3.7)
$$\limsup_{n \to \infty} \mathbf{P}(N_{2r,R/2}(\mathscr{D}(\mathscr{T}_n)) \geqslant N/2) \geqslant c/2, \quad \forall R \geqslant 8r.$$

By Proposition A.4, one readily checks that this implies that $\mathscr{D}(\mathscr{T}_n)$ is not precompact in $\mathscr{M}_1(\mathbb{T}_e)$, which will yield the required contradiction.

In order to show (3.7), it is enough to show that for large r and N, for every tree $\mathscr{T} \in \mathfrak{T}_e^\ell$,

$$(3.8) \quad \forall R \geqslant r : N_{r,R}(\mathscr{T}) \geqslant N \quad \Longrightarrow \quad \forall R \geqslant 8r : \mathbf{P}(N_{2r,R/2}(\mathscr{D}(\mathscr{T})) \geqslant N/2) \geqslant 1/2.$$

For this, fix $R \geqslant 8r$ and consider the vertices of $\mathscr{T}$ at distance r from the root that contribute to the quantity $\mathscr{N} := N_{r,R}(\mathscr{T})$. To each such vertex v, associate a single path (among possibly many) of length R that links the root ρ to a vertex at distance R, and contains v. Let F_v be the event that this path has less than $2r$ vertices in V_0 at distance $\leqslant r$ from the root, but more than $R/2$ vertices in V_0 at distance $\leqslant R$ from the root. For a vertex v, the probability $\mathbf{P}(F_v)$ is independent of v, and arbitrary close to 1 for large r, so in particular larger than $3/4$ for r large enough.

Now let $\mathscr{N}^{\overline{F}}$ be the number of vertices v contributing to $\mathscr{N}$ that satisfy the complementary event $\overline{F}_v$. Then $\mathbf{E}(\mathscr{N}^{\overline{F}}|\mathscr{N}) = \mathscr{N}(1 - \mathbf{P}(F_v))$, and by the Markov inequality, for $\mathscr{N}$ and r large enough,

$$\mathbf{P}(\mathscr{N}^{\overline{F}} \geqslant \mathscr{N}/2) \leqslant 2(1 - \mathbf{P}(F_v)) \leqslant 1/2 \text{ or, equivalently, } \mathbf{P}(\mathscr{N} - \mathscr{N}^{\overline{F}} > \mathscr{N}/2) \geqslant 1/2.$$

Together with the pointwise inequality $N_{2r,R/2}(\mathscr{D}(\mathscr{T}_n)) \geqslant \mathscr{N} - \mathscr{N}^{\overline{F}}$ this proves (3.8) and finishes the proof. $\qquad\square$

4. Tree rescaling: proofs of Theorems 1 and 2

In this section, we will prove a coupling lemma (Lemma 4.1 below), which, together with Theorem 3, will yield Theorems 2 and 1.

The space of discrete trees $\mathbb{T}_e$ is naturally embedded into the space $\mathfrak{T}_e^\ell$ via the following embedding ι: given $T \in \mathbb{T}_e$, we define $\iota(T) = (\mathscr{V}, d, \rho, \mu)$ as follows:

– The set of vertices is given by $\mathscr{V} = \{(i,x), i \in T \smallsetminus \{\rho\}, x \in [0,1)\} \cup \{(\rho, 0)\}$.
– The distance is defined by $d((i,x),(j,y)) = d_T(i,j) - x - y$ if $(i \npreceq_T j$ and $j \npreceq_T i)$ and $d((i,x),(j,y)) = d_T(i,j) - x + y$ if $j \preceq_T i$, and it is symmetric in its arguments.
– μ is the Lebesgue measure $\ell_{\mathscr{T}}$.

Informally, the tree $\iota(T)$ is defined from T by adding segments of length 1 between adjacent vertices of T, and the element (i,x) of $\mathscr{V}$ is at distance x of $(i,0)$ on the path $[\![\rho, i]\!]$ to the root.

LEMMA 4.1. — *Let $(p_n)_{n\geqslant 0}$ and $(q_n)_{n\geqslant 0}$ be sequences of positive numbers such that $p_n \to 0$ and $q_n \to 0$ as $n \to \infty$. Let $T_n \in \mathscr{M}_1(\mathbb{T}_e)$ be a sequence of random rooted trees. Then the following statements are equivalent:*

(1) *The sequence $(\mathscr{C}_{p_n,q_n}(T_n))_{n\geqslant 0}$ is tight in*[12] $\mathbb{T}$.
(2) *The sequence $(\mathscr{D}(\mathscr{S}_{p_n,q_n}(\iota(T_n))))_{n\geqslant 0}$ is tight in $\mathbb{T}$.*

---

[12] Here, and below, we say that a family of random trees on $\mathbb{T}_e$ is *tight in* $\mathbb{T}$ if the family of their laws, seen as laws on $\mathbb{T}$, are tight.

In this case, for every $m \geqslant 1$ and $\varepsilon > 0$, there exists an integer n_0 and for every $n \geqslant n_0$ a coupling between $\mathscr{C}_{p_n,q_n}(T_n)$ and $\mathscr{D}(\mathscr{S}_{p_n,q_n}(\iota(T_n)))$, such that

$$(4.1) \qquad \mathbf{P}\left([\mathscr{C}_{p_n,q_n}(T_n)]^{\leqslant m} = [\mathscr{D}(\mathscr{S}_{p_n,q_n}(\iota(T_n)))]^{\leqslant m}\right) > 1 - \varepsilon.$$

Proof. — Recall that by definition of the topology of local convergence on $\mathbb{T}_e$, a sequence of random discrete trees $\widetilde{T}_n$ in $\mathscr{M}_1(\mathbb{T}_e)$ is tight in $\mathbb{T}$ if and only if for every $m \geqslant 1$, the sequence $(\#V(\widetilde{T}_n^{\leqslant m}))_{n \geqslant 1}$ is tight. In order to apply this, we first need to define a suitable coupling between the operations $\mathscr{C}_{p,q}$ and $\mathscr{S}_{p,q}$ and certain one-dimensional stochastic processes.

Let $T \in \mathbb{T}_e$. Recall that in the construction of $\mathscr{C}_{p,q}(T)$, the vertices are colored with two colors, say black and white; the black vertices (and the root) are retained, whereas the white vertices are discarded. Given a tree $T \in \mathbb{T}_e$ we can couple this operation with two infinite sequences $(B_i)_{i \geqslant 1}$ and $(B'_i)_{i \geqslant 1}$ of iid random variables distributed according to the Bernoulli distribution with parameter p and q, respectively: we perform a breadth-first traversal of the vertices of the tree starting from the root and color the vertex visited at step i according to B_i ($1 = $ black and $0 = $ white) if it is an off-spine vertex, and according to B'_i if it is a spine vertex. Given an integer $m \geqslant 1$, we add an additional rule: When a black vertex is visited which has exactly m black ancestors, then the subtree above it is subsequently ignored by the algorithm. The restriction of the tree T to the black vertices then exactly gives the tree $\mathscr{C}_{p,q}(T)^{\leqslant m}$. Note that this algorithm terminates almost surely since $\mathscr{C}_{p,q}(T)$ is locally finite by definition, hence $\mathscr{C}_{p,q}(T)^{\leqslant m}$ is finite and the breadth-first traversal only has to go to a certain (random, but finite) depth in the tree T.

In order to construct the tree $\mathscr{D}(\mathscr{S}_{p,q}(\iota(T)))^{\leqslant m}$ one can proceed in a similar manner, but using now two Poisson processes $(P_t)_{t \geqslant 0}$ and $(P'_t)_{t \geqslant 0}$ with intensity p and q (respectively) defined on the $\mathbb{R}$-tree $\iota(T)$. We omit the details.

For every $M \geqslant 0$, we now couple the sequences $(B_i)_{i \geqslant 1}$ and $(B'_i)_{i \geqslant 1}$ with the Poisson processes $(P_t)_{t \geqslant 0}$ and $(P'_t)_{t \geqslant 0}$ in such a way that with probability $1 - O(M(p \vee q))$, for every $i \leqslant M/p$, $B_i = 1$ if and only if $P_i - P_{i-1} = 1$ and for every $i \leqslant M/q$, $B'_i = 1$ if and only if $P'_i - P'_{i-1} = 1$. It is now easy to check that for large M and small p and q, on the event that (either) $\#V(\mathscr{C}_{p,q}(T)^{\leqslant m}) \leqslant M/2$ or $\#V(\mathscr{D}(\mathscr{S}_{p,q}(\iota(T)))^{\leqslant m}) \leqslant M/2$, both trees agree with high probability. This implies that if the first or second statement of the lemma holds, then (4.1) is true, which in turn implies equivalence of the two statements. This finishes the proof of the lemma. $\qquad\square$

Proof of Theorem 2. — We start with the easy direction: Let $\mathscr{T} \in \mathscr{M}_1(\mathfrak{T}_1^\ell)$. The family $T_n = \mathscr{D}(\mathscr{T}, n)$ is a compatible family of random rooted trees since, for $n \geqslant m \geqslant 1$:

$$\mathscr{C}(T_n, m) = \mathscr{C}(\mathscr{D}(\mathscr{T}, n), m) = \mathscr{D}(\mathscr{T}, m) = T_m.$$

Now let $(T_n)_{n \in \mathbb{N}}$ be a compatible family of random trees and let $(B_n)_{n \in \mathbb{N}}$ be a sequence of $\mathrm{Bin}(n, 1/n)$-distributed random variables, independent of $(T_n)_{n \in \mathbb{N}}$. We then have the following equality,

$$\mathscr{C}(T_n, B_n) = T_{B_n}.$$

On the RHS, there is convergence towards T_B as $n \to \infty$, where B is a Poi(1)-distributed random variable independent of $(T_n)_{n\in\mathbb{N}}$. On the LHS, we have $\mathscr{C}(T_n, B_n) = \mathscr{C}_{1/n}(T_n)$, and by Lemma 4.1 there exists a coupling such that the equality

$$\mathscr{C}_{1/n}(T_n) = \mathscr{D}(\mathscr{S}_{1/n}(\iota(T_n)))$$

holds with high probability as $n \to \infty$. The space $\mathfrak{T}_1^\ell$ is compact by Lemma 2.6, therefore the sequence $(\mathscr{S}_{1/n}(\iota(T_n)))_{n\geqslant 1} \in \mathscr{M}_1(\mathfrak{T}_1^\ell)$ is tight. Since the map $\mathscr{D}$ is continuous and injective by Theorem 3, every subsequential limit $\mathscr{T}$ has $T_B = \mathscr{D}(\mathscr{T})$ and is thus unique. Furthermore, conditioning on $B = m$ yields $T_m = \mathscr{D}(\mathscr{T}, m)$ for each $m \in \mathbb{N}$. $\qquad\square$

Proof of Theorem 1. — One direction is obvious: Let $\mathscr{T} \in \mathscr{M}_1(\mathfrak{T}_e^\ell)$ be a (p, q)-self-similar random $\mathbb{R}$-tree. Lemma 1.10 then yields

$$\mathscr{C}_{p,q}(\mathscr{D}(\mathscr{T})) = \mathscr{D}(\mathscr{S}_{p,q}(\mathscr{T})) = \mathscr{D}(\mathscr{T}),$$

whence the discrete tree $\mathscr{D}(\mathscr{T})$ is (p, q)-self-similar as well.

For the other direction, let $T \in \mathscr{M}_1(\mathbb{T}_e)$ be a (p, q)-self-similar random rooted tree, i.e., $\mathscr{C}_{p,q}(T) = T$. Iterating this equality yields for each integer $n \geqslant 1$,

$$(4.2) \qquad\qquad \mathscr{C}_{p^n,q^n}(T) = \mathscr{C}_{p,q}^n(T) = T.$$

In particular, the sequence $(\mathscr{C}_{p^n,q^n}(T))_{n\geqslant 1}$ is tight in $\mathbb{T}$. Lemma 4.1 now yields that the sequence $(\mathscr{D}(\mathscr{S}_{p^n,q^n}(\iota(T))))_{n\geqslant 1}$ is tight in $\mathbb{T}$ and that every subsequential limit equals T. By Theorem 3, the sequence $(\mathscr{S}_{p^n,q^n}(\iota(T)))_{n\geqslant 1}$ then converges in law to a random tree $\mathscr{T}$ taking values in $\mathfrak{T}_e^\ell$ and such that $\mathscr{D}(\mathscr{T}) = T$. This uniquely determines the tree $\mathscr{T}$ by injectivity of the map $\mathscr{D}$. Furthermore, by Lemma 1.10 and the self-similarity of T, we have $\mathscr{D}(\mathscr{S}_{p,q}(\mathscr{T})) = \mathscr{C}_{p,q}(T) = T = \mathscr{D}(\mathscr{T})$, whence, again by injectivity of $\mathscr{D}$, $\mathscr{S}_{p,q}(\mathscr{T}) = \mathscr{T}$. This finishes the proof of the theorem. $\quad\square$

5. Examples of self-similar trees

In this section, we construct some examples of (p, q)-self-similar $\mathbb{R}$-trees. We do not believe it is possible to completely characterize this family, similarly to the situation for self-similar real-valued processes. For simplicity, we restrict ourselves to trees whose spine consists of a single infinite ray only.

1. **Subordination of a real-valued self-similar process.** — If $(\mathscr{T}, d, \rho, \mu)$ is a (p, q)-self-similar $\mathbb{R}$-tree, then let $\mathscr{V}_t$ be the subset consisting of the vertices whose most recent ancestor on the spine is at distance at most t from the root. Setting

$$(5.1) \qquad\qquad X(t) = \mu(\mathscr{V}_t) - t$$

defines a (semi)-self-similar real-valued non-decreasing process with Hurst exponent[13] $H = \log p / \log q$, i.e.

$$(5.2) \qquad (q^{-H} X(qt), t \geqslant 0) \stackrel{\text{law}}{=} (X(t), t \geqslant 0).$$

On the other hand, if we are given such a process $X(t)$, we can construct from it a (p, q)-self-similar $\mathbb{R}$-tree $\mathscr{T} \in \mathscr{M}_1(\mathfrak{T}_e^\ell)$. Write $X_c(t)$ for its continuous part and $X_j(t)$ for its jump part. Let $\mathscr{T}' \in \mathscr{M}_1(\mathfrak{T}_1^\ell)$ be an arbitrary random tree. The random tree $\mathscr{T}$ is then constructed as follows:

– Spine($\mathscr{T}$) consists of a single infinite ray, and $\mu(\text{Spine}(\mathscr{T}) \cap [0, t]) = X_c(t) + t$,
– For every jump time t of X_j, we attach an independent copy of $\mathscr{T}'$ to the spine at distance t of the root, rescaled by the size of the jump $X_j(t) - X_j(t-)$.

It is easy to show that the resulting tree is indeed (p, q)-self-similar.

One can easily generalize the above construction. For example, instead of attaching independent rescaled copies of the same tree $\mathscr{T}'$ to the spine, one can take a $|\log q|$-stationary process $(\mathscr{T}'(s), s \in \mathbb{R})$ of rooted, probability-measured $\mathbb{R}$-trees ($|\log q|$-stationary means that $(\mathscr{T}(s), s \in \mathbb{R})$ is equal in law to $(\mathscr{T}(s + |\log q|), s \in \mathbb{R})$), and attach a rescaled copy of $\mathscr{T}'(\log t)$ at the point t on the spine. One can also introduce a stronger dependency between the process $X(t)$ and the trees. For example, let $X(t)$ be as above and suppose for simplicity that it is a pure-jump process. Let $\mathscr{R}$ be the set of its *record jumps*, i.e. $r \in \mathscr{R}$ if and only if $X(r) - X(r-) > X(s) - X(s-)$ for all $s < r$. Write $\mathbb{R} = \{\cdots < r_{-1} < r_0 < r_1 < \ldots\}$, with $r_0 \leqslant 1 < r_1$ and let $\mathscr{T}'_n$ be a sequence of iid copies of a probability-measured rooted $\mathbb{R}$-tree. We then construct a (p, q)-self-similar tree as follows: For every t, let $N(t)$ be such that $r_{N(t)} \leqslant t < r_{N(t)+1}$. Then for each jump time t of $X(t)$, add the tree $\mathscr{T}'_{N(t)}$ to the spine, rescaled by the size $X(t) - X(t-)$ of the jump. One readily checks that the resulting process is (p, q)-self-similar.

2. Rescaling along the spine. — Given a (p, q)-self-similar $\mathbb{R}$-tree $\mathscr{T}$, one can easily construct a whole family of self-similar $\mathbb{R}$-trees: Let $\beta > 0$. First, one can rescale the tree along the spine: define a new $\mathbb{R}$-tree $\mathscr{T}'$ obtained from $\mathscr{T}$ by mapping a point t on the spine to t^β. The mass process of this new tree is $X^\beta(t) = X(t^{1/\beta})$, and therefore $(p^{-1} X^\beta(q^\beta t); t \geqslant 0) \stackrel{\text{law}}{=} (X^\beta(t); t \geqslant 0)$. Since we have not changed the structure of the subtrees, it follows that the resulting tree is (p, q^β)-self-similar.

For $\gamma \in \mathbb{R}$, one can also define a new mass process by setting $X^\gamma(t) = \int_0^t s^\gamma dX(s)$, as long as this quantity is finite for some (hence, any) $t > 0$. The tree defined by this process in the canonical way (i.e., by rescaling the subtrees of the spine and the measure μ on the spine) is then (pq^γ, q)-self-similar.

Finally, one can apply the previous scaling to the continuous part $X_c(t)$ of the mass process only, and scale the jump process $X_j(t)$ instead by setting $X_j^\delta(t) = \sum_{s \leqslant t} (X_j(s) - X_j(s-))^\delta$ for some $\delta > 0$. If δ and γ are such that $pq^\gamma = p^\delta$, then the

---

[13]The exponent H is called the Hurst exponent after the study in hydrology [HBS65], see also the paper [MVN68].

mass process $X^{\gamma,\delta}(t) = X_c^{\gamma}(t) + X_j^{\delta}(t)$, if it exists, defines a (p^{δ}, q)-self-similar tree in the canonical way.

6. Translation invariant self-similar trees

In this section, we study self-similar trees which are invariant under translation along the spine. For simplicity, we restrict ourselves to one-ended self-similar trees. We denote the corresponding subspaces of $\mathbb{T}_e$ and $\mathfrak{T}_e^{\ell}$ by $\overline{\mathbb{T}}$ and $\overline{\mathfrak{T}}^{\ell}$, respectively. Throughout the section, $T \in \mathcal{M}_1(\overline{\mathbb{T}})$ will denote a one-ended (p,q)-self-similar tree and $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathcal{M}_1(\overline{\mathfrak{T}}^{\ell})$ the limiting $\mathbb{R}$-tree obtained from Theorem 1. We define a shift operator $\Theta : \mathcal{M}_1(\overline{\mathbb{T}}) \to \mathcal{M}_1(\overline{\mathbb{T}})$ which maps T to the subtree rooted at the vertex on the spine at distance 1 from the root. We say T is *translation invariant*, if $\Theta T \overset{\text{law}}{=} T$. Similarly, for $t \geqslant 0$, we define a shift operator $\Theta_t : \mathcal{M}_1(\overline{\mathfrak{T}}^{\ell}) \to \mathcal{M}_1(\overline{\mathfrak{T}}^{\ell})$ mapping $\mathscr{T}$ to its subtree rooted at the vertex on the spine at distance t from the root. Note that $(\Theta_t)_{t \geqslant 0}$ is a semigroup. We then say that $\mathscr{T}$ is translation invariant if $\Theta_t \mathscr{T} = \mathscr{T}$ for every $t \geqslant 0$. The following proposition says that the two notions are equivalent:

Proposition 6.1. — *T is translation invariant if and only if $\mathscr{T}$ is translation invariant.*

Lemma 6.2. — *The semigroup $(\Theta_t)_{t \geqslant 0}$ is strongly left-continuous, i.e., for every $\mathscr{T} \in \mathcal{M}_1(\overline{\mathfrak{T}}^{\ell})$, the function $t \mapsto \Theta_t(\mathscr{T})$ is left-continuous.*

Proof. — It is enough to check the lemma for a deterministic tree $\mathscr{T} \in \overline{\mathfrak{T}}^{\ell}$. Consider the subtrees $\mathscr{T}_1 = \Theta_{t-\varepsilon}(\mathscr{T})$ and $\mathscr{T}_2 = \Theta_t(\mathscr{T})$ rooted at the spine vertices at distance $t - \varepsilon$ and t from ρ, respectively denoted by ρ_1 and ρ_2. From the definition (A.6) of the GHP distance and the dominated convergence Theorem, it is enough to prove that the GHP distance between the compact trees $\mathscr{T}_1^{\leqslant r}$ and $\mathscr{T}_2^{\leqslant r}$ has a null $\varepsilon \to 0$ limit for any fixed value of $r > 0$. We let μ_1 and μ_2 be the measures associated with $\mathscr{T}_1^{\leqslant r}$ and $\mathscr{T}_2^{\leqslant r}$ respectively. We have $d(\rho_1, \rho_2) = \varepsilon$, and, with $\mathscr{V}^{<r}$ the restriction of $\mathscr{V}$ to the *open* ball of radius r centered at ρ, it holds

$$d_P^c(\mu_1, \mu_2) \leqslant \mu(\mathscr{V}^{<t} \smallsetminus \mathscr{V}^{<t-\varepsilon}) + \mu(\mathscr{V}^{<t+r} \smallsetminus \mathscr{V}^{<t+r-\varepsilon}),$$

and the last expression has a null $\varepsilon \to 0$ limit since the two sets on the RHS decrease to the null set. These elements combined with Lemma A.9 now allow to conclude since:

$$d_{GHP}^c(\mathscr{T}_1^{\leqslant r}, \mathscr{T}_2^{\leqslant r}) \leqslant 3\, d_{GP}^c(\mathscr{T}_1^{\leqslant r}, \mathscr{T}_2^{\leqslant r}) \leqslant d(\rho_1, \rho_2) + d_P^c(\mu_1, \mu_2). \qquad \square$$

Proof of Proposition 6.1. — By Theorem 1, the tree T is obtained from $\mathscr{T}$ by sampling vertices according to a Poisson process on $\mathscr{T}$ with intensity μ. By definition of the sampling procedure, the vertices on the spine are those which are sampled according to ℓ_{Spine}, where $\ell_{\text{Spine}} = \ell_{\mathscr{T}}|_{\text{Spine}(\mathscr{T})}$ is the restriction of the length measure $\ell_{\mathscr{T}}$ to the spine. It follows that

$$(6.1) \qquad\qquad \Theta^n T = \mathscr{D}(\Theta_{E_1 + \cdots + E_n} \mathscr{T}),$$

where $E_1, \ldots, E_n$ are independent exponential random variables with parameter 1, independent from $\mathscr{T}$. Setting $n = 1$ in (6.1) directly yields the "if" statement of the proposition.

Now suppose that T is translation invariant. Then, for every $n \geqslant 0$ and $k \geqslant 0$, we have

$$
\begin{aligned}
\mathscr{D}(\mathscr{T}) &= T & &\text{by hypothesis on } T \text{ and } \mathscr{T} \\
&= \mathscr{C}_{p,q}^k T & &\text{by the self-similarity of } T \\
&= \mathscr{C}_{p,q}^k \Theta^n T & &\text{by translation invariance of } T \\
&= \mathscr{C}_{p,q}^k \mathscr{D}(\Theta_{E_1 + \cdots + E_n} \mathscr{T}) & &\text{by (6.1)} \\
&= \mathscr{D}(\mathscr{S}_{p,q}^k \Theta_{E_1 + \cdots + E_n} \mathscr{T}) & &\text{by Lemma 1.10} \\
&= \mathscr{D}(\Theta_{q^k E_1 + \cdots + q^k E_n} \mathscr{S}_{p,q}^k \mathscr{T}) & &\text{by definition} \\
&= \mathscr{D}(\Theta_{q^k E_1 + \cdots + q^k E_n} \mathscr{T}) & &\text{by the self-similarity of } \mathscr{T}.
\end{aligned}
$$

Since $\mathscr{D}$ is injective by Lemma 3.2, this implies for every $n \geqslant 0$ and $k \geqslant 0$,

$$
\mathscr{T} = \Theta_{q^k E_1 + \cdots + q^k E_n} \mathscr{T}.
$$

By the law of large numbers, for any $t \geqslant 0$, we can now let $k, n \to \infty$ in such a way that $q^k E_1 + \cdots + q^k E_n$ converges from the left to t almost surely. By Lemma 6.2, this yields the statement. $\qquad\square$

We now show that for translation invariant trees, the range of possible values for p and q is restricted. The mass process $X(t)$ defined in 5.1 is real-valued, increasing, and semi-self-similar process with Hurst exponent $H = \log p / \log q$. Recall $X_c(t)$ and $X_j(t)$ denote respectively the continuous part and the jump part of $X(t)$, and note that $X_c(t) = \mu(\mathrm{Spine}(\mathscr{T}) \cap [0, t]) - t$.

PROPOSITION 6.3. — *Suppose $\mathscr{T} \in \mathscr{M}_1(\overline{\mathfrak{T}}^\ell)$ is (p, q)-self-similar, translation invariant and non-degenerate (i.e., it is not isometric to $\mathbb{R}_+$). Then $q \geqslant p$. Moreover, if $q > p$, then $X_c \equiv 0$ almost surely, and if $q = p$, then $X_j \equiv 0$ almost surely and $X_c(t) = X(1) t$ for every $t \geqslant 0$.*

Proof. — This proposition is essentially a corollary of results of Vervaat [Ver85]. X satisfies the hypotheses 1.4 in [Ver85], except for one: Equation (5.2) does not hold for *every* $p, q > 0$ with $\log p / \log q = H$ in our case. This is the only missing assumption in Theorem 3.3 in [Ver85]. We claim that the conclusion of that theorem still hold under (5.2). To justify this claim, we observe that the proof of Theorem 3.3 relies on Theorem 3.1, items a and b for $H \leqslant 1$, and on Theorem 3.5 for $H > 1$. The proof of the latter is not affected by our assumption. We therefore have to verify that the conclusions of Vervaat's Theorem 3.1, items a and b still hold under (5.2).

Let $\mathscr{I}$ be the invariant σ-field of the stationary sequence $(X(t) - X(t-1), t \in \mathbb{N})$. In the proof of Theorem 3.1, Vervaat establishes the following for a *truly* semi-self-similar process:

$$t^{H-1}X(1) \overset{\text{law}}{=} \frac{X(t)}{t} = \frac{X(t) - X(0)}{t} = \frac{1}{t} \sum_{1 \leqslant s \leqslant t} X(s) - X(s-1)$$
$$\longrightarrow \mathbf{E}(X(1) - X(0) \mid \mathscr{I}) = \mathbf{E}(X(1) \mid \mathscr{I}),$$

using respectively semi-self-similarity, $X(0) = 0$ a.s., and Birkhoff's ergodic theorem in a version that allows for infinite means, like the one presented in Theorem 3.7 of [Ver85]. Because the integer t is not necessarily of the form q^{-m}, the left-most equality has to be replaced in our case by:

$$\frac{X(q^m t)}{p^m t} \overset{\text{law}}{=} \frac{X(t)}{t}.$$

Now, the sequence of integers $t = t(m) = \lceil q^{-m} \rceil$ satisfies $q^m t \geqslant 1$ and $q^m t \to 1$ as $m \to \infty$. Since the process X has càdlàg sample paths (this may be seen using arguments similar as in the proof of Lemma 6.2), $X(q^m t)$ converges to $X(1)$ a.s. as $m \to \infty$. From that point on, the proof of Theorem 3.1, items a and b, follows unchanged.

Therefore the conclusion of Theorem 3.3 holds, and that conclusion implies the statement of our proposition, since the event A (in that theorem) that X has locally bounded variation has probability 1 in our case due to the monotonicity of the sample paths of X. $\square$

We now want to study the (p, q)-self-similar discrete trees T which are translation invariant *and* for which the subtrees along the spine are independent (hence, iid).

PROPOSITION 6.4. — *The subtrees of T along the spine are iid if and only if $q \geqslant p$ and $\mathscr{T}$ is constructed as follows:*

– If $q > p$, then $\mu_{|\,\mathrm{Spine}} = \ell_{\mathbb{R}_+}$, and the point process with atoms $(t, \mathscr{T}_t)$, where $\mathscr{T}_t$ is the subtree rooted at the spine vertex at distance t from the root, is a Poisson point process with intensity $dt \otimes \nu(d\mathscr{T})$, with the measure ν decomposing as follows: There exists a measurable family $(\sigma_x)_{x>0}$ of probability measures on the space $\mathfrak{T}_1^\ell$ such that $\sigma_x = \sigma_{px}$ for every x, and a measure $\Lambda(dx)$ on $(0, \infty)$ satisfying $\Lambda(A) = q\Lambda(pA)$ for every Borel $A \subset (0, \infty)$, such that ν decomposes as the semi-direct product $\nu = \Lambda(dx)\sigma_x^x$, where σ_x^x is the push-forward of the measure σ_x under the map $\mathscr{S}_x$.

– If $q = p$, then $\mathscr{T}$ is the non-negative real line with a (deterministic) multiple of Lebesgue measure: $\mathscr{V} = \mathbb{R}_+$, $\rho = 0$ and there exists $\lambda \geqslant 1$ such that $\mu = \lambda \ell_{\mathbb{R}_+}$.

Proof. — By Proposition 6.3, we either have $q > p$ and the mass process $X(t)$ is a pure-jump process, or $q = p$ and the mass process is continuous with $X(t) = tX(1)$. The statement in the case $q = p$ now follows from the fact that the subtrees of T are independent if and only if $X(1)$ is a deterministic constant.

Now consider the case $q > p$. By Proposition 6.3, the mass process $X(t)$ is then a pure-jump process, i.e., the restriction of μ to the spine equals the length measure. Denote by T^i for every $i \geqslant 0$ the off-spine subtree of the i-th vertex on the spine of T. Furthermore, for $s \leqslant t$, denote by $\mathscr{T}^{s,t}$ the *concatenation* of the off-spine subtrees of the tree $\mathscr{T}$ rooted at vertices v on the spine with $s \leqslant d(0,v) < t$. Here, the *concatenation* of a collection of rooted trees is defined to be the rooted tree obtained from the disjoint union of the trees by identification of the roots. By the definition of the discretization operation $\mathscr{D}$, we then have

$$(6.2) \qquad (T^0, T^1, T^2, \dots) \overset{\text{law}}{=} (\mathscr{D}(\mathscr{T}^{\xi_0, \xi_1}), \mathscr{D}(\mathscr{T}^{\xi_1, \xi_2}), \dots),$$

where $\xi_0 = 0$ and for $n \in \mathbb{N}^*$, $\xi_n = \sum_{k=1}^{n} E_k$, with $E_1, E_2, \dots$ a sequence of independent exponential random variables with parameter 1, independent from $\mathscr{T}$. Equation (6.2) now shows that if the point process with atoms $(t, \mathscr{T}_t)$ is a translation invariant Poisson process, then the trees $(T^i)_{i \geqslant 0}$ are iid.

Now assume that the trees $(T^i)_{i \geqslant 0}$ are iid. Fix $0 = t_0 < t_1 < t_2 < \cdots$. We first notice that:

$$
\begin{aligned}
\mathscr{D}(\mathscr{T}^{t_0, t_1}) &\overset{\text{law}}{=} \mathscr{D}(\mathscr{S}_{p,q}^k(\mathscr{T})^{t_0, t_1}) && \text{by the self-similarity of } \mathscr{T} \\
&= \mathscr{D}(\mathscr{S}_{p,q}^k(\mathscr{T}^{t_0 q^{-k}, t_1 q^{-k}})) && \text{by definition} \\
(6.3) \qquad &\overset{\text{law}}{=} \mathscr{C}_{p,q}^k(\mathscr{D}(\mathscr{T}^{t_0 q^{-k}, t_1 q^{-k}})) && \text{by Lemma 1.10.}
\end{aligned}
$$

Fix an integer i, and real numbers $t'_1, \dots, t'_i$ satisfying $t_{j-1} < t'_j < t_j$ for every $j \in \{1, \dots i\}$. Let j be in this set. By the law of large numbers, we may find integers $n_j(k)$ such that, for k large enough,

$$(6.4) \qquad t'_j q^{-k} < \xi_{n_j} < t_j q^{-k},$$

with the (ξ_n) distributed as above. Set also $n_0 = 0$. Applying (6.3) with t_0 and t_1 replaced by t_{j-1} and t'_j, and recalling (6.4) and (6.2), we deduce that that the discretized tree $\mathscr{D}(\mathscr{T}^{t_{j-1}, t'_j})$ is a function of $T^{n_{j-1}}, \dots, T^{n_j}$, for $j \in \{1, \dots, i\}$. Therefore the collection of trees $(\mathscr{T}^{t_0, t'_1}, \dots, \mathscr{T}^{t_{i-1}, t'_i})$ is independent. The numbers $t'_1, \dots, t'_i$ being arbitrary, this implies the collection $(\mathscr{T}^{t_0, t_1}, \dots, \mathscr{T}^{t_{i-1}, t_i})$ is independent, therefore the point process $(t, \mathscr{T}_t)$ is a Poisson process. By translation invariance, its intensity measure is of the form $dt \otimes \nu$ for some measure ν. By the (p, q)-self-similarity, $\nu = q^{-1} \nu^p$, where ν^p is the push-forward of the measure under the map $\mathscr{T} \mapsto p\mathscr{T}$. Disintegrating the measure ν with respect to the mass of the tree yields the decomposition stated in the theorem. □

In case $q = p$ and the subtrees of T along the spine are iid, their common distribution is that of a rooted tree with a Geo($1/\lambda$)-distributed number[14] of edges adjacent to the root. In case $q > p$, there is the following corollary that follows from Proposition 6.4 and standard properties of Poisson processes. Recall that we define

---

[14] Our geometric distribution starts at 0, $\mathbf{P}(\text{Geo}(\gamma) = k) = (1 - \gamma)^k \gamma$, for each $k \in \mathbb{N}$, for $\gamma \in (0, 1]$.

the *concatenation* of a collection of rooted trees to be the rooted tree obtained from the disjoint union of the trees by identification of the roots.

Corollary 6.5. — *In case $q > p$ and the subtrees of T along the spine are iid, their distribution is characterized as follows: There exists a constant $\gamma \in (0,1]$, as well as a measurable family of probability measures σ_x and a measure $\Lambda(dx)$ as in the statement of Proposition 6.4, such that*

— *T^0 is the concatenation of copies of* Geo(c)-*distributed number of independent copies of a tree T', where $c = 1/(1+d)$ and $d = \int_0^\infty (1 - e^{-x})\Lambda(dx)$, and*

— *$\mathbf{E}[F(T')] = d^{-1} \int_0^\infty \Lambda(dx) \int \sigma_x(d\mathscr{T})\mathbf{E}[F(\mathscr{D}(\mathscr{T}, \mathrm{Poi}(x)))]$ for every bounded measurable function F with $F(\rho) = 0$.*

If a measure $\Lambda(dx)$ satisfies the condition in Proposition 6.4, any measure proportional to $\Lambda(dx)$ again satisfies this condition, so the parameter c of the Geometric random variable in Corollary 6.5 may indeed take arbitrary values in $(0,1]$.

Remark 6.6. — The number of edges in the tree T' that appears in the statement of Corollary 6.5 follows the law

$$\mathbf{P}[N(T') = k] = d^{-1} \int_0^\infty \mathbf{P}(\mathrm{Poi}(x) = k)\Lambda(dx), \ \forall k \geqslant 1.$$

The laws of this form with d and Λ as in the statement of Proposition 6.4 are exactly the quasi-stationary distributions of the Markov chain $(Z_{(\log p^{-1})n}; n = 0, 1, 2, \ldots)$, where $(Z_t; t \geqslant 0)$ is the standard pure death process (i.e., $Z_{t-} \to Z_t - 1$ with rate Z_{t-}) killed at 0 [Maiar]. If P denotes the substochastic transition matrix (on $\mathbb{N}^*$) of this discrete time Markov chain, they are the distributions η on $\mathbb{N}^*$ that satisfy:

$$\eta P = q\eta.$$

7. A different approach to Theorem 2

We now explain a proof idea of Theorem 2 that we initially pursued but then dropped in favour of the current proof. The approach builds on an idea, popularized by David Aldous [Ald93], of using exchangeability to construct concrete representations of (*a priori* abstractly defined) inverse (or projective) limits of compatible sequences.[15]

There is a correspondence between a partial order $\preceq_{T_n}$ on $\{1, \ldots, n\}$ and a (graph-theoretic) rooted tree T_n with n labelled non-root vertices: assuming w.l.o.g. that the non-root vertices of T_n are labelled by $\{1, \ldots, n\}$, this correspondence writes:

$$i \preceq_{T_n} j \text{ iff vertex } i \text{ is an ancestor of vertex } j \text{ in } T_n.$$

Let $(T_n)_{n \in \mathbb{N}}$ be a sequence of random trees that is compatible in the sense of Definition 1.11, and assume that the n non-root vertices of T_n are labelled by an independent random permutation of $\{1, \ldots, n\}$. The induced random partial orders $\preceq_{T_n}$

---

[15]For examples of other settings see e.g. [Jan11] and [FHP11] and the references therein.

on the label set $\{1, \ldots, n\}$ are then exchangeable and compatible (for the restriction of a partial order) as n varies. By Kolmogorov's extension theorem, there exists an exchangeable partial order $\preceq$ on $\mathbb{N}^* := \{1, 2, \ldots\}$ whose restriction to $\{1, \ldots, n\}$ is $\preceq_{T_n}$ for each $n \in \mathbb{N}^*$.

Call $i \wedge j$ the most recent common ancestor of i and j for $\preceq$, for two integers i and j: De Finetti's theorem then implies the existence of the almost sure limit

$$d(i, j) := \lim_{n \to \infty} \frac{1}{n} \sum_{k \leqslant n, k \neq i, j} \mathbb{1}_{k \preceq i, k \preceq j, k \npreceq i \wedge j}.$$

The map $d(., .)$ a.s. is a random pseudo-metric on $\mathbb{N}^*$, and the metric completion of $(\mathbb{N}^*, d)$ then defines a random $\mathbb{R}$-tree $(\mathscr{V}, d)$ that is naturally rooted at its minimal element ρ. This metric space also supports a probability measure μ given by:

$$\mu([\![i, j]\!]) := \lim_{n \to \infty} \frac{1}{n} \sum_{k \leqslant n, k \neq i, j} \mathbb{1}_{d(i,k) + d(k,j) = d(i,j)},$$

that dominates the length measure in the sense that $\mu([\![i, j]\!]) \geqslant d(i, j)$ for any $(i, j) \in (\mathbb{N}^*)^2$. The resulting random $\mathbb{R}$-tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu)$ defines an element of $\mathscr{M}_1(\mathfrak{T}_1^\ell)$, whose discretization $\mathscr{D}(\mathscr{T}, n)$ to a discrete tree with $n + 1$ vertices is distributed as T_n: this is the correspondence exposed in Theorem 2.

The construction outlined has several advantages: it is direct, in the sense that it quickly points to the right objects to be used like $\mathfrak{T}_1^\ell$, whose introduction may otherwise seem *ad-hoc* to the reader. Sticking closer to the objects of interest than our abstract topological considerations, it is also arguably more natural. Last, it may be pushed to build scaling limits of (infinite) random rooted trees of $\mathbb{T}_e$ that are invariant under the contraction operation $\mathscr{C}_{p,q}$.

Implementing this program, we encountered several difficulties, one of which being that the proper definition of the completion of a random metric space requires to identify an appropriate σ-field (and a certain subspace of $\mathbb{R}$-trees) with respect to which the completion is measurable. Once the subspace of $\mathbb{R}$-trees and the σ-field were identified, compactness was in easy reach, and the compactness property provides us with a candidate for the limit of a compatible sequence of random trees without effort—in particular, without using the de Finetti theorem, so crucial to the above construction, in an extensive way. These reflexions convinced us to finally drop the exchangeability arguments above in favour of the topological considerations developed in Section 2, 3 and 4: besides being more precise and rigorous, these considerations are also comparatively more robust, with statements concerning compactness and closure that could be useful to attack further problems in the domain.

Appendix A. Tree spaces

In this appendix, we collect some properties of the tree spaces that we work with in this paper, namely, the space of locally finite (graph-theoretic) trees and the space of locally compact measured $\mathbb{R}$-trees.

A.1. Discrete trees. — In this section, we consider rooted, locally finite trees $T = (V, E, \rho)$ in the graph-theoretic sense. We recall from the introduction that two such trees are called *equivalent* if there exists a root-preserving graph isomorphism between them. For simplicity, we will always identify an equivalence class with its representatives. We then denote by $\mathbb{T}$ the space of (equivalence classes of) trees and by $\mathbb{T}_f \subset \mathbb{T}$ the subspace of finite trees, both endowed with the topology of local convergence (see introduction). Also recall that a tree $T \in \mathbb{T}$ determines and is determined by a partial order $\preceq_T$ on its vertex set $V(T)$ called the *ancestral relation*.

For a tree $T \in \mathbb{T}$, denote by $V_R(T)$, $R \geqslant 0$ the set of vertices at (graph) distance at most R from the origin. The following well-known precompactness criterion is easily proven by a diagonalization argument:

Proposition A.1. — *A family $\mathfrak{S} \subset \mathbb{T}$ is precompact in $\mathbb{T}$ if and only if for every $R \geqslant 0$,*

$$\sup_{T \in \mathfrak{S}} \#V_R(T) < \infty.$$

From Proposition A.1, it is easy to see that the space $\mathbb{T}$ is not locally compact. However, it is topologically complete, as can be seen by defining the following metric:

$$d_{\mathbb{T}}(T_1, T_2) = \sum_{k=0}^{\infty} 2^{-k} \mathbb{1}_{(T_1^{\leqslant k} \neq T_2^{\leqslant k})}.$$

Proposition A.2. — *The metric $d_{\mathbb{T}}$ generates the topology of local convergence in $\mathbb{T}$. Furthermore, the space $\mathbb{T}$ is complete and separable.*

Proof. — Let $(T_n)_{n \geqslant 0}$ be a sequence of trees in $\mathbb{T}$ and $T \in \mathbb{T}$. By definition, $d_{\mathbb{T}}(T_n, T) \to 0$ as $n \to \infty$ if and only if for every $k \geqslant 0$, $T_n^{\leqslant k} \to T^{\leqslant k}$ in $\mathbb{T}_f$. This shows the first statement. For the separability, we note that by definition, $T^{\leqslant k} \to T$ in $\mathbb{T}$ as $k \to \infty$ for every $T \in \mathbb{T}$. Since every $T \in \mathbb{T}$ is locally finite, $T^{\leqslant k} \in \mathbb{T}_f$ for every $k \geqslant 0$. Hence, the space $\mathbb{T}_f$ is dense in $\mathbb{T}$ and obviously countable, which proves separability. As for the completeness, let $(T_n)_{n \geqslant 0}$ be a Cauchy sequence in $\mathbb{T}$. By definition of the metric $d_{\mathbb{T}}$, for each $k \geqslant 0$, there exists $N(k) < \infty$, such that $T_n^{\leqslant k} = T_{n'}^{\leqslant k}$ for every $n, n' \geqslant N(k)$. By diagonalization, we can construct $T \in \mathbb{T}$, such that for every $k \geqslant 0$, $T_n^{\leqslant k} = T^{\leqslant k}$ for all $n \geqslant N(k)$. This implies that $T_n \to T$ as $n \to \infty$ and proves completeness of the space $\mathbb{T}$.   □

For two trees $T, T' \in \mathbb{T}$, write $T \hookrightarrow T'$, if there exists a root-preserving graph homomorphism from T to T'. By restricting to balls around the root, one easily sees that $T = T'$ if and only if $T \hookrightarrow T'$ and $T' \hookrightarrow T$.

An *end* of a tree $T = (V, E, \rho) \in \mathbb{T}$ is by definition an infinite path from the root, i.e., a sequence of pairwise distinct vertices $v_0, v_1, \ldots$, such that $v_0 = \rho$ and $\{v_n, v_{n+1}\} \in E$ for all n. The *spine* of the tree $\mathscr{T}$, denoted by $\mathrm{Spine}(T)$ is then defined to be the set of vertices that lie on an end, with $\mathrm{Spine}(T) = \{\rho\}$ if the tree has no end. We also set

$$\mathbb{T}_e = \{T \in \mathbb{T} : T \text{ has finitely many ends}\}.$$

The following proposition is included for completeness:

Proposition A.3. — *The space $\mathbb{T}_e$ is not topologically complete.*

Proof. — By [Dug66, Th. XIV.8.3], a subspace of a complete metric space is topologically complete if and only if it is a countable intersection of open sets. Assume that this is the case and let $G_1, G_2, \ldots$ be open subsets of $\mathbb{T}$, such that $\mathbb{T}_e = \bigcap_n G_n$. Assume w.l.o.g. that the sequence $(G_n)_n$ is decreasing, otherwise set $G'_n = \bigcap_{k=1}^n G_k$. In order to get a contradiction, we will construct a tree T which is an element of every G_n but with an infinite number of ends.

For a tree $T \in \mathbb{T}$ and $r > 0$, denote by $B_r(T)$ the (open) ball of radius r around T, i.e.,

$$B_r(T) = \{T' \in \mathbb{T} : d_{\mathbb{T}}(T', T) < r\}.$$

Note by the definition of $d_{\mathbb{T}}$,

(A.1) $\qquad \forall r > 0 \ \exists R \in \mathbb{N} : B_r(T) = \{T' \in \mathbb{T} : (T')^{\leqslant R} = T^{\leqslant R}\} =: A_R(T).$

We now construct the tree T mentioned above by diagonalization: start with the tree T_1 which consists of a single infinite ray. Since $T_1 \in \mathbb{T}_e \subset G_1$, and since G_1 is open, there exists by (A.1) $R_1 \in \mathbb{N}$, such that $A_{R_1}(T_1) \subset G_1$. Construct the tree T_2 from T_1 by gluing an infinite ray to the vertex at distance R_1 from the root. Then $T_2 \in A_{R_1}(T_1)$ and obviously $T_2 \in \mathbb{T}_e$. This construction can be repeated *ad infinitum*: given the tree $T_n \in \mathbb{T}_e$, consisting of n infinite rays glued together, let $R_n \in \mathbb{N}$ such that $A_{R_n}(T_n) \subset G_n$. Then construct a tree T_{n+1} by gluing an infinite ray to a vertex at distance R_n from the root, such that $T_{n+1} \in A_{R_n}(T_n)$. Note that we can and will assume that the sequence $(R_n)_n$ is increasing. This gives a sequence $(T_n)_n$ of trees in $\mathbb{T}_e$, such that

$$\forall n \in \mathbb{N} \ \forall k, l \geqslant n : T_k^{\leqslant R_n} = T_l^{\leqslant R_n}.$$

By diagonalization, this sequence now defines a tree T with $T^{\leqslant R_n} = T_n^{\leqslant R_n}$ for all $n \in \mathbb{N}$. Hence, $T \in A_{R_n}(T_n) \subset G_n$ for all $n \in \mathbb{N}$, so that $T \in \bigcap_n G_n = \mathbb{T}_e$. However, by construction, the number of ends in the tree T is infinite, such that $T \notin \mathbb{T}_e$. This is the contradiction mentioned above and therefore finishes the proof. □

Since the space $\mathbb{T}_e$ is not topologically complete, we cannot make use of Prokhorov's theorem for measures on $\mathbb{T}_e$. For this reason, we formulate in the following proposition a precompactness criterion for a family of such measures. For a tree $T \in \mathbb{T}$ and $0 \leqslant r \leqslant R$, denote by $N_{r,R}(T)$ the number of vertices at distance r of the root in T that have a descendant at distance R from the root. Note that $N_{r,R}(T)$ is increasing in r and decreasing in R, with $N_{r,R}(T) \to N_{r,\infty}(T)$ as $R \to \infty$, where $N_{r,\infty}(T)$ denotes the number of spine vertices at distance r of the root in T.

Proposition A.4. — *A sequence of random trees $T_1, T_2, \cdots \in \mathscr{M}_1(\mathbb{T}_e)$ is precompact in $\mathscr{M}_1(\mathbb{T}_e)$ if and only if it is precompact in $\mathscr{M}_1(\mathbb{T})$ and for every $r \in \mathbb{N}$ there exist $R = R(r)$ and $n_0 = n_0(r)$, such that*

the family of random variables $(N_{r,R(r)}(T_n))_{r \in \mathbb{N}, n \geqslant n_0(r)}$ is tight.

Proof. — Let $T_1, T_2, \ldots$ be a sequence of random trees converging to a limit $T \in \mathscr{M}_1(\mathbb{T})$. In order to show the proposition it suffices to show that T is supported on $\mathbb{T}_e$ if and only if the second condition of the statement holds. Note that by Skorokhod's representation theorem we can and will assume that the convergence holds almost surely on a suitable probability space. In particular, this implies that

$$(\text{A.2}) \qquad \forall \varepsilon > 0 \; \forall R > 0 \; \exists n_0 \in \mathbb{N} \; \forall n \geqslant n_0 : \mathbf{P}(T_n^{\leqslant R} = T^{\leqslant R}) > 1 - \varepsilon.$$

We first show the "only if" statement, i.e. we assume that T is supported on $\mathbb{T}_e$. Fix $\varepsilon > 0$ and $r \in \mathbb{N}$. Since $N_{r,R}(T) \to N_{r,\infty}(T)$ as $R \to \infty$, there exists $R = R(r)$, such that

$$(\text{A.3}) \qquad \mathbf{P}(N_{r,R}(T) = N_{r,\infty}(T)) > 1 - \varepsilon.$$

Together with (A.2), this yields the existence of $n_0 = n_0(r)$, such that

$$(\text{A.4}) \qquad \forall n \geqslant n_0 : \mathbf{P}(N_{r,R}(T_n) = N_{r,\infty}(T)) > 1 - 2\varepsilon.$$

Since $N_{r,\infty}(T)$ is bounded by the number of ends of T for every r, the family of random variables $(N_{r,\infty}(T))_{r \in \mathbb{N}}$ is tight. Together with (A.4) this shows that the family of random variables $(N_{r,R}(T_n))_{r \in \mathbb{N}, n \geqslant n_0}$ is tight, which proves the "only if" statement.

In order to show the "if" statement, assume that for every $r \in \mathbb{N}$ there exist $R = R(r)$ and $n_0 = n_0(r)$ such that the family of random variables $(N_{r,R}(T_n))_{r \in \mathbb{N}, n \geqslant n_0}$ is tight. This entails that the family $(N_{r,R(r)}(T))_{r \in \mathbb{N}}$ is tight by (A.2), whence the family $(N_{r,\infty}(T))_{r \in \mathbb{N}}$ is tight as well, since $N_{r,R}(T) \geqslant N_{r,\infty}(T)$ for every r. But since $N_{r,\infty}(T)$ converges to the number of ends in T as $r \to \infty$, this number must be almost surely finite, whence T is supported on $\mathbb{T}_e$. This finishes the proof of the "if" statement and of the proposition. $\qquad \square$

A.2. Measured $\mathbb{R}$-trees. — There are several equivalent definitions of an $\mathbb{R}$-tree, see [Dre84, Eva08]. We follow here the treatment in [ADH13],

Definition A.5. — An $\mathbb{R}$-tree is a metric space $(\mathscr{V}, d)$ with the following properties:

(1) It is *geodesically linear*, i.e., for every $x, y \in \mathscr{V}$, there is a unique isometry $f_{x,y} : [0, d(x,y)] \to \mathscr{V}$ such that $f_{x,y}(0) = x$ and $f_{x,y}(d(x,y)) = y$.

(2) It is "without loops", i.e., for every $x, y \in \mathscr{V}$, if r and q are continuous injective maps from $[0,1]$ to $\mathscr{V}$ such that $q(0) = x$ and $q(1) = y$, and $r(0) = x$ and $r(1) = y$, then $q([0,1]) = r([0,1])$.

Elements of $\mathscr{V}$ are called the *vertices* of the $\mathbb{R}$-tree. If $x, y \in \mathscr{V}$, we use the notation $[\![x, y]\!]$, respectively $[\![x, y[\![$, to denote the image of $[0, d(x,y)]$, resp. $[0, d(x,y))$, under the map $f_{x,y}$.

A rooted $\mathbb{R}$-tree $\mathscr{T} = (\mathscr{V}, d, \rho)$ is an $\mathbb{R}$-tree $(\mathscr{V}, d)$ together with a distinguished vertex $\rho \in \mathscr{V}$ called the *root*. In this context, a partial order $\preceq_{\mathscr{T}}$ on $\mathscr{V}$ is defined by: $x \preceq_{\mathscr{T}} y$ if and only if $x \in [\![\rho, y]\!]$, in which case x is an *ancestor* of y, and y is a *descendant* of x. Also we write $x \prec_{\mathscr{T}} y$ when $x \preceq_{\mathscr{T}} y$ and $x \neq y$.

A rooted *measured* $\mathbb{R}$-tree is a quadruple $\mathscr{T} = (\mathscr{V}(\mathscr{T}), d_{\mathscr{T}}, \rho_{\mathscr{T}}, \mu_{\mathscr{T}})$ where $(\mathscr{V}, d, \rho)$ is a rooted $\mathbb{R}$-tree and μ is a Borel measure on $\mathscr{V}$. Two measured rooted $\mathbb{R}$-trees are said to be *equivalent* if there exists a root- and measure-preserving isometry between them. As usual, we identify a tree with its equivalence class.

DEFINITION A.6. — We define $\mathfrak{T}$ the space of (equivalence classes of) rooted measured $\mathbb{R}$-trees $\mathscr{T} = (\mathscr{V}, d, \rho, \mu)$, where

    (1) the metric space $(\mathscr{V}, d)$ is complete and locally compact,
    (2) the measure μ is boundedly finite, i.e. $\mu(A) < \infty$ for every bounded Borel set A.

We further denote by $\mathfrak{T}_f$ the subspace of *compact* trees (in particular, the measure μ is then finite), and by $\mathfrak{T}_1$ the subspace of $\mathfrak{T}_f$ where μ is a probability measure.

Note that by the Hopf–Rinow–Cohn–Vossen theorem, every $\mathscr{T} \in \mathfrak{T}$ is a *proper metric space*, i.e., every bounded closed set is compact. In particular, every $\mathscr{T} \in \mathfrak{T}$ is separable (it is well-known and easy to show that this is true for every proper metric space). Also, a measure on $\mathscr{T}$ is boundedly finite if and only if it is *locally finite* (i.e., every point has a neighbourhood of finite measure), but we won't need this fact here.

We now define a metric on the space $\mathfrak{T}$, which will be called the Gromov–Hausdorff–Prokhorov (GHP) metric d_{GHP}. We first recall the definition of the Hausdorff (pseudo-) metric between two subsets of a metric space (Z, d^Z):

$$d_H^Z(A, B) = \inf\{\varepsilon > 0 : A \subset B^\varepsilon \text{ and } B \subset A^\varepsilon\}, \quad A, B \subset Z,$$

where for $A \subset Z$ we define

$$A^\varepsilon = \{x \in Z : d^Z(x, A) < \varepsilon\}, \quad \text{where } d^Z(x, A) = \inf_{y \in A} d^Z(x, y).$$

Furthermore, we recall the definition of the Prokhorov metric on the space $\mathscr{M}_f(Z)$ of finite Borel measures on Z:

$$d_P^Z(\mu, \nu) = \inf\{\varepsilon > 0 : \mu(F) \leqslant \nu(F^\varepsilon) + \varepsilon \text{ and } \nu(F) \leqslant \mu(F^\varepsilon) + \varepsilon \text{ for all closed } F \subset Z\}.$$

We can now define the GHP metric on the space $\mathfrak{T}_f$ of *compact* trees. For $\mathscr{T} = (\mathscr{V}, d, \rho, \mu), \mathscr{T}' = (\mathscr{V}', d', \rho', \mu') \in \mathfrak{T}_f$, set

$$(A.5) \quad d_{GHP}^c(\mathscr{T}, \mathscr{T}')$$
$$= \inf_{\varphi, \varphi', Z} \left[d^Z(\varphi(\rho), \varphi'(\rho')) + d_H^Z(\varphi(\mathscr{V}), \varphi'(\mathscr{V}')) + d_P^Z(\varphi_*\mu, \varphi'_*\mu') \right].$$

Here, the infimum is taken over all isometric embeddings $\varphi : \mathscr{V} \hookrightarrow Z, \varphi' : \mathscr{V}' \hookrightarrow Z$ into some common complete separable metric space (Z, d^Z) and $\varphi_*\mu$ is the push-forward of the measure μ by the map φ.

We now extend the GHP metric to the space $\mathfrak{T}$. For a tree $\mathscr{T} \in \mathfrak{T}$ and $r \geqslant 0$, denote by $\mathscr{T}^{\leqslant r}$ its restriction to the closed ball of radius r around the root (i.e., restriction of the underlying metric space as well as of the measure). Then $\mathscr{T}^{\leqslant r}$ is compact as explained above. We define the GHP metric d_{GHP} on $\mathfrak{T}$ by

$$(A.6) \qquad d_{GHP}(\mathscr{T}_1, \mathscr{T}_2) = \int_0^\infty e^{-r} \left(1 \wedge d_{GHP}^c\left(\mathscr{T}_1^{\leqslant r}, \mathscr{T}_2^{\leqslant r} \right) \right) dr.$$

The following facts are, respectively, Corollary 3.2 and Proposition 2.10 in [ADH13].

Fact A.7

(1) *The space $(\mathfrak{T}, d_{GHP})$ is a complete separable metric space.*

(2) *The metrics d_{GHP} and d_{GHP}^c induce the same topology on $\mathfrak{T}_f$ and $\mathfrak{T}_1$.*

In what follows, we will always endow the space $\mathfrak{T}$ and its subspaces with the metric d_{GHP} and its induced topology and Borel σ-algebra, unless mentioned otherwise.

The *length measure* of a rooted $\mathbb{R}$-tree $\mathscr{T} = (\mathscr{V}, d, \rho)$ is by definition the unique σ-finite measure $\ell_{\mathscr{T}}$ on $\mathscr{V}$, such that [Eva08, §4.3.5]

(A.7) $\forall a, b \in \mathscr{V} : \ell_{\mathscr{T}}(]\!]a, b[\![) = d(a, b)$ and $\ell_{\mathscr{T}}(\mathscr{V} \smallsetminus \mathscr{V}^\circ) = 0$, where $\mathscr{V}^\circ = \bigcup\limits_{x \in \mathscr{V}} [\![\rho, x[\![$.

We then define the spaces $\mathfrak{T}^\ell$, $\mathfrak{T}_f^\ell$ and $\mathfrak{T}_1^\ell$ as follows:

$$\mathfrak{T}^\ell = \{\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T} : \mu \geqslant \ell_{\mathscr{T}}\}, \quad \mathfrak{T}_f^\ell = \mathfrak{T}_f \cap \mathfrak{T}^\ell, \quad \mathfrak{T}_1^\ell = \mathfrak{T}_1 \cap \mathfrak{T}^\ell.$$

The following lemma collects some properties of the space $\mathfrak{T}^\ell$. Note that a refined version of the second part is proved in the main text (Lemma 3.4). In the lemma, $\mathscr{V}^{\leqslant r}$ denotes the restriction of $\mathscr{V}$ to the closed ball of radius r around the root.

Lemma A.8

(1) *A family $\mathfrak{S} \subset \mathfrak{T}^\ell$ is precompact in $\mathfrak{T}$ if and only if for every $r \geqslant 0$,*

$$\sup_{\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{S}} \mu(\mathscr{V}^{\leqslant r}) < \infty.$$

(2) *Let $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}$ be a limit point of a sequence in $\mathfrak{T}_1^\ell$. Then μ is a probability measure (hence $\mathscr{T} \in \mathfrak{T}_1$) and it has full support (i.e., $\operatorname{supp} \mu = \mathscr{V}$), furthermore, the convergence holds with respect to the d_{GHP}^c metric as well.*

Proof. — The first part follows from combining Theorems 2.11 or 2.6 in [ADH13] and Lemma 4.37 in [Eva08].

For the second part, let $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}$ be a limit point of a sequence $\mathscr{T}_n = (\mathscr{V}_n, d_n, \rho_n, \mu_n) \in \mathfrak{T}_1^\ell$. Observe that every tree $\mathscr{T}_n$, $n \in \mathbb{N}$, has diameter bounded by one, so that the convergence holds with respect to the d_{GHP}^c metric. This readily implies that the masses of the measures μ_n converge to the mass of μ, which thus has mass one.

We now show that the measure μ has full support. For this, it is enough to prove that, for $0 < \varepsilon \leqslant 2$ and $x \in \mathscr{V}$, the closed ball $D_\varepsilon(x)$ of radius ε in $\mathscr{V}$ centered at x verifies

(A.8) $$\mu(D_\varepsilon(x)) \geqslant \varepsilon/4.$$

To prove this, we fix $\varepsilon \in (0, 2]$ and n large enough so that $d_{GHP}^c(\mathscr{T}_n, \mathscr{T}) < \varepsilon/4$. There is an embedding of $\mathscr{T}_n$ and $\mathscr{T}$ into a common metric space (Z, d) such that the following two properties hold (for simplicity, we do not distinguish the trees from their embedding): (1) the Hausdorff distance satisfies $d_H(\mathscr{V}_n, \mathscr{V}) < \varepsilon/4$ and (2) the

Prokhorov distance satisfies $d_P(\mu, \mu_n) < \varepsilon/4$. The following chain of inequalities then holds:

$$\varepsilon/2 \leqslant \mu_n(D_{\varepsilon/2}(x_n)) \leqslant \mu_n(D_{3\varepsilon/4}(x)) \leqslant \mu((D_{3\varepsilon/4})^{\varepsilon/4}(x)) + \varepsilon/4 \leqslant \mu(D_\varepsilon(x)) + \varepsilon/4.$$

For the first inequality, we distinguish according to whether the set $\mathscr{V} \cap (Z \smallsetminus D_{\varepsilon/2}(x_n))$ is empty or not. In the first case, $\mathscr{T} \subseteq D_{\varepsilon/2}(x_n)$, therefore $\mu_n(D_{\varepsilon/2}(x_n)) = 1 \geqslant \varepsilon/2$ given our choice of ε. In the second case, there is a path of length at least $\varepsilon/2$ from x_n to the boundary of $D_{\varepsilon/2}(x_n)$, and the μ-measure of this path is larger than or equal to its length, therefore $\mu_n(D_{\varepsilon/2}(x_n)) \geqslant \varepsilon/2$ is again valid. The second inequality follows since $d(x, x_n) < \varepsilon/4$ according to (1). The third inequality is a consequence of (2), and the fourth inequality is plain. Subtracting $\varepsilon/4$ then proves (A.8). $\qquad\square$

On the previously defined spaces, the Hausdorff distance appearing in the GHP metric is actually unnecessary. This is shown by the following lemma, that can also be seen as a consequence of the general results exposed in [ALW16] (this article appeared after a first version of our work paper was published on the arXiv).

Define the Gromov–Prokhorov (GP) metric d_{GP}^c on $\mathfrak{T}_f^\ell$ by

$$d_{GP}^c(\mathscr{T}, \mathscr{T}') = \inf_{\varphi, \varphi', Z} \left[d^Z(\varphi(\rho), \varphi'(\rho')) + d_P^Z(\varphi_*\mu, \varphi'_*\mu') \right],$$

where the infimum is over φ, φ' and Z as specified in (A.5).

Lemma A.9. — *We have $d_{GP}^c \leqslant d_{GHP}^c \leqslant 3d_{GP}^c$ on $\mathfrak{T}_f^\ell$.*

Proof. — The first inequality is immediate. For the second one, we fix $\varepsilon > 0$ and consider $\mathscr{T} = (\mathscr{V}, d, \rho, \mu)$ and $\mathscr{T}' = (\mathscr{V}', d', \rho', \mu')$ two element of $\mathfrak{T}_f^\ell$ such that $d_{GP}^c(\mathscr{T}, \mathscr{T}') < \varepsilon$. Without loss of generality, we may assume that the trees $\mathscr{T}$ and $\mathscr{T}'$ are subsets of a complete separable metric space (Z, d), on which: $d(\rho, \rho') + d_P(\mu, \mu') < \varepsilon$. We call η a number such that $d_P(\mu, \mu') < \eta < \varepsilon$. It is enough, from the definition of d_{GHP}^c, to show that

$$\text{(A.9)} \qquad\qquad\qquad d_H(\mathscr{V}, \mathscr{V}') \leqslant 2\varepsilon$$

to prove the second inequality. We set $\mathscr{W} = \mathscr{V} \cap (Z \smallsetminus \mathscr{V}'^{2\varepsilon})$, and prove by contradiction $\mathscr{W} = \varnothing$. By symmetry, the same statement will then hold changing the role of $\mathscr{V}$ and $\mathscr{V}'$, and (A.9) will be proved. If $\mathscr{W} \neq \varnothing$, there exists $x \in \mathscr{W}$, that is $x \in \mathscr{V}$ such that $D_{\varepsilon+\eta}(x) \cap \mathscr{V}' = \varnothing$. Using the bound on the Prokhorov distance and the fact that μ' has support $\mathscr{V}'$, we deduce $\mu(D_\varepsilon(x)) \leqslant \mu'(D_{\varepsilon+\eta}(x)) + \eta \leqslant \eta < \varepsilon$. But we also have $\varepsilon \leqslant \mu(D_\varepsilon(x))$, reasoning as in the proof of Lemma A.8. This is a contradiction. $\qquad\square$

The metric d_{GP}^c gives rise to a topology on $\mathfrak{T}_f$ called the Gromov–Prokhorov topology. We now give an equivalent definition of this topology on the space $\mathfrak{T}_1$. We follow [GPW09] (which is influenced by Chapter $3\frac{1}{2}$ in [Gro07]). They consider the case of unrooted trees (or metric spaces), but the results can be easily generalized to the rooted case, for example by identifying the rooted, probability measured tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_1$ with the unrooted, probability measured tree $(\mathscr{V}, d, \frac{1}{3}\mu + \frac{2}{3}\delta_\rho)$.

Given a tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_1$, we define a *distance matrix distribution* $\mathrm{DM}(\mathscr{T})$, i.e., a probability measure on $[0,1]^{\mathbb{N} \times \mathbb{N}}$, as the push-forward of the probability measure $\delta_\rho \otimes \mu^{\otimes \mathbb{N}^*}$ by the map

$$(x_i)_{0 \leqslant i \leqslant n} \longmapsto d(x_i, x_j)_{0 \leqslant i,j \leqslant n}.$$

Proposition 2.6 and Corollary 3.1 in [GPW09] now imply the following:

Fact A.10. — *A tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}_1$ (resp., a random tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathscr{M}_1(\mathfrak{T}_1)$) whose measure μ has full support (resp., has full support almost surely) is uniquely determined by its distance-matrix distribution $\mathrm{DM}(\mathscr{T})$.*

Furthermore, by Theorem 5 in [GPW09], we have:

Fact A.11. — *Let $\mathscr{T}, \mathscr{T}_1, \mathscr{T}_2, \cdots \in \mathfrak{T}_1$. Then $d^c_{GP}(\mathscr{T}_n, \mathscr{T}) \to 0$ if and only if $\mathrm{DM}(\mathscr{T}_n) \to \mathrm{DM}(\mathscr{T})$ as $n \to \infty$ (the latter convergence is weak convergence on $\mathscr{M}_1(\mathbb{R}^{\mathbb{N} \times \mathbb{N}})$ and $\mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ is equipped with the product topology).*

Fact A.11 and Lemma A.9 yield the following corollary:

Corollary A.12. — *Let $\mathscr{T}, \mathscr{T}_1, \mathscr{T}_2, \cdots \in \mathfrak{T}_1^\ell$. Then $d^c_{GHP}(\mathscr{T}_n, \mathscr{T}) \to 0$ if and only if $\mathrm{DM}(\mathscr{T}_n) \to \mathrm{DM}(\mathscr{T})$ as $n \to \infty$.*

For $r \geqslant 0$, the restriction map $\mathscr{T} \to \mathscr{T}^{\leqslant r}$ is not necessarily continuous in the GHP topology. However, there is the following

Lemma A.13. — *Let $r \geqslant 0$ and let $\mathscr{T}, \mathscr{T}_1, \mathscr{T}_2, \cdots \in \mathfrak{T}$ satisfy $d_{GHP}(\mathscr{T}_n, \mathscr{T}) \to 0$. There exists a sequence $(r_n)_{n \geqslant 1}$ that satisfies:*

$$r_n \geqslant r, \ r_n \longrightarrow r \ \text{and} \ d^c_{GHP}(\mathscr{T}_n^{\leqslant r_n}, \mathscr{T}^{\leqslant r}) \longrightarrow 0 \ \text{as} \ n \longrightarrow \infty.$$

Proof. — Define a sequence of functions on $(0, \infty)$ by $f_n(s) = d^c_{GHP}(\mathscr{T}_n^{\leqslant s}, \mathscr{T}^{\leqslant s}) \wedge 1$. Fix $r \geqslant 0$ and set, for $n \geqslant 1$, $\varepsilon(n) := (e^{r+1} \int_{(0,\infty)} f_n(s) e^{-s} \, ds)^{1/2}$. First, the definition (A.6) of d_{GHP} gives $\lim_{n \to \infty} \varepsilon(n) = 0$. Second, for $n \geqslant 1$, we have $\int_{(r,r+1)} f_n(s) \, ds \leqslant \varepsilon(n)^2$ and $f_n \geqslant 0$, thus there exists r_n satisfying

$$r \leqslant r_n \leqslant (r + \varepsilon(n)) \wedge (r+1) \ \text{and} \ 0 \leqslant f_n(r_n) \leqslant \varepsilon(n) \vee \varepsilon(n)^2.$$

The sequence $(r_n)_{n \geqslant 1}$ then satisfies the first two requirements of the lemma; also the sequence $d^c_{GHP}(\mathscr{T}_n^{\leqslant r_n}, \mathscr{T}^{\leqslant r_n})$ has a null limit. To control the remaining $d^c_{GHP}(\mathscr{T}^{\leqslant r_n}, \mathscr{T}^{\leqslant r})$ term, we observe that the function $r \to \mathscr{T}^{\leqslant r}$ is right-continuous: this follows from Lemma 5.2 in [ADH13] and the right-continuity of $r \to \mu(\mathscr{T}^{\leqslant r})$. □

Analogously to discrete trees, we call an *end* of a tree $\mathscr{T} = (\mathscr{V}, d, \rho, \mu) \in \mathfrak{T}$ an infinite ray starting from the root, i.e., the union $\bigcup_n [\![\rho, x_n]\!]$, where $x_1, x_2, \cdots \in \mathscr{V}$ are such that $x_n \preceq_T x_{n+1}$ for all n and $d(\rho, x_n) \to \infty$ as $n \to \infty$. For a tree $\mathscr{T} \in \mathfrak{T}$, we denote by $\mathrm{Spine}(\mathscr{T})$ the union of its ends, called the *spine*. We further define the subspaces $\mathfrak{T}_e \subset \mathfrak{T}$ and $\mathfrak{T}_e^\ell \subset \mathfrak{T}^\ell$ of trees having only a finite number of ends.

If $0 \leqslant r \leqslant R$, we denote by $N_{r,R}(\mathscr{T})$ the number of vertices at distance r of the root in $\mathscr{T}$ that have a descendant at distance R from the root. The analogue to Proposition A.4 for $\mathbb{R}$-trees is the following, whose proof we omit:

Lemma A.14. — *A sequence of random trees $\mathscr{T}_1, \mathscr{T}_2, \dots \in \mathscr{M}_1(\mathfrak{T}_e)$ is precompact in $\mathscr{M}_1(\mathfrak{T}_e)$ if and only if it is precompact in $\mathscr{M}_1(\mathfrak{T})$ and for every $r \geqslant 0$ there exist $R = R(r)$ and $n_0 = n_0(r)$, such that*

the family of random variables $(N_{r,R(r)}(\mathscr{T}_n))_{r \in \mathbb{N}, n \geqslant n_0(r)}$ is tight.

Appendix B. An extension of continuous maps

Let X and Y be separable metric spaces endowed with the Borel σ-field and denote by $\mathscr{M}_1(X)$ and $\mathscr{M}_1(Y)$ the spaces of probability measures on X and Y, respectively, endowed with the topology of weak convergence, i.e., $\mu_n \Rightarrow \mu$ in $\mathscr{M}_1(X)$ if and only if

$$\int f(x)\, \mu_n(dx) \longrightarrow \int f(x)\, \mu(dx), \quad \forall f : X \longrightarrow \mathbb{R} \text{ bounded, continuous.}$$

The following basic fact is used several times in the article and mentioned for completeness: every continuous function $g : X \to \mathscr{M}_1(Y)$, $x \mapsto g_x$ can be naturally extended to a continuous function $\widehat{g} : \mathscr{M}_1(X) \to \mathscr{M}_1(Y)$, $\mu \mapsto \widehat{g}_\mu$, where

$$\int f(y)\, \widehat{g}_\mu(dy) = \int \mu(dx) \left(\int f(y)\, g_x(dy) \right), \quad \forall f : Y \longrightarrow \mathbb{R} \text{ bounded, measurable.}$$

The fact that this map is well defined and continuous directly follows from the above definition of weak convergence.

References

[ADH13] R. Abraham, J.-F. Delmas & P. Hoscheit – "A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces", *Electron. Comm. Probab.* **18** (2013), no. 14, p. 1–21.

[Ald93] D. Aldous – "The continuum random tree III", *Ann. Probability* **21** (1993), no. 1, p. 248–289.

[AP98] D. Aldous & J. Pitman – "Tree-valued Markov chains derived from Galton-Watson processes", *Ann. Inst. H. Poincaré Probab. Statist.* **34** (1998), no. 5, p. 637–686.

[ALW16] S. Athreya, W. Löhr & A. Winter – "The gap between Gromov-vague and Gromov–Hausdorff-vague topology", *Stochastic Processes Appl.* **126** (2016), no. 9, p. 2527–2553.

[Dre84] A. 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 Math.* **53** (1984), no. 3, p. 321–402.

[Drm09] M. Drmota – *Random trees*, SpringerWienNewYork, Vienna, 2009.

[Dug66] J. Dugundji – *Topology*, Allyn and Bacon, Inc., Boston, 1966.

[Eva08] S. N. Evans – *Probability and real trees*, Lect. Notes in Math., vol. 1920, Springer, Berlin, 2008.

[EPW06] S. N. Evans, J. Pitman & A. Winter – "Rayleigh processes, real trees, and root growth with re-grafting", *Probab. Theory Related Fields* **134** (2006), no. 1, p. 81–126.

[FS09] Ph. Flajolet & R. Sedgewick – *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.

[FHP11] N. Forman, C. Haulk & J. Pitman – "A representation of exchangeable hierarchies by sampling from real trees", arXiv:1101.5619, 2011.

[GPW09] A. Greven, P. Pfaffelhuber & A. Winter – "Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)", *Probab. Theory Related Fields* **145** (2009), no. 1-2, p. 285–322.

[Gro07] M. Gromov – *Metric structures for Riemannian and non-Riemannian spaces*, Modern Birkhäuser Classics, Birkhäuser Boston Inc., Boston, MA, 2007.

[HBS65] H. E. Hurst, R. P. Black & Y. M. Simaika – *Long-term storage: an experimental study*, Constable, London, 1965.

[Jan11] S. Janson – "Poset limits and exchangeable random posets", *Combinatorica* **31** (2011), no. 5, p. 529–563.

[LS06] L. Lovász & B. Szegedy – "Limits of dense graph sequences", *J. Combinatorial Theory Ser. B* **96** (2006), no. 6, p. 933–957.

[Maiar] P. Maillard – "The λ-invariant measures of subcritical Bienaymé–Galton–Watson processes", *Bernoulli* (to appear), arXiv:1508.00845.

[MVN68] B. B. Mandelbrot & J. W. Van Ness – "Fractional Brownian motions, fractional noises and applications", *SIAM Rev.* **10** (1968), no. 4, p. 422–437.

[OV85] G. L. O'Brien & W. Vervaat – "Self-similar processes with stationary increments generated by point processes", *Ann. Probability* **13** (1985), no. 1, p. 28–52.

[Rém85] J.-L. Rémy – "Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire", *RAIRO Inform. Théor.* **19** (1985), no. 2, p. 179–195.

[Sta97] R. P. Stanley – *Enumerative combinatorics. Vol. 1*, Cambridge Studies in Advanced Mathematics, vol. 49, Cambridge University Press, Cambridge, 1997.

[Ver85] W. Vervaat – "Sample path properties of self-similar processes with stationary increments", *Ann. Probability* **13** (1985), no. 1, p. 1–27.

Olivier Hénard, Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, CNRS, Université Paris-Saclay
91405 Orsay, France
E-mail : olivier.henard@gmail.com
Url : http://www.math.u-psud.fr/~henard/

Pascal Maillard, Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, CNRS, Université Paris-Saclay
91405 Orsay, France
E-mail : pascal.maillard@u-psud.fr
Url : http://www.proba.jussieu.fr/dw/doku.php?id=users:maillard:index