RAYLEIGH PROCESSES, REAL TREES, AND ROOT GROWTH WITH RE-GRAFTING

STEVEN N. EVANS, JIM PITMAN, AND ANITA WINTER

Abstract. The real trees form a class of metric spaces that extends the class of trees with edge lengths by allowing behavior such as infinite total edge length and vertices with infinite branching degree. Aldous's Brownian continuum random tree, the random tree-like object naturally associated with a standard Brownian excursion, may be thought of as a random compact real tree. The continuum random tree is a scaling limit as $N \to \infty$ of both a critical Galton-Watson tree conditioned to have total population size N as well as a uniform random rooted combinatorial tree with N vertices. The Aldous–Broder algorithm is a Markov chain on the space of rooted combinatorial trees with N vertices that has the uniform tree as its stationary distribution. We construct and study a Markov process on the space of all rooted compact real trees that has the continuum random tree as its stationary distribution and arises as the scaling limit as $N \to \infty$ of the Aldous–Broder chain. A key technical ingredient in this work is the use of a pointed Gromov–Hausdorff distance to metrize the space of rooted compact real trees.

Berkeley Statistics Technical Report No. 654 (February 2004), revised October 2004. To appear in *Probability Theory and Related Fields*.

Date: November 17, 2004.

2000 *Mathematics Subject Classification.* Primary: 60B05, 60J27; Secondary: 60J80, 60B99.

Key words and phrases. continuum random tree, Brownian excursion, real tree, Gromov-Hausdorff metric, Hausdorff metric, Aldous-Broder algorithm, piecewise-deterministic Markov process.

SNE supported in part by NSF grants DMS-0071468 and DMS-0405778, and a Miller Institute for Basic Research in Science research professorship.

JP supported in part by NSF grants DMS-0071448 and DMS-0405779.

AW supported by a DFG Forschungsstipendium.

Contents

1. Motivation and background

It is shown in [Ald91a, Ald91b, Ald93] (see also [LG99, Pit02a]) that a suitably rescaled family of Galton-Watson trees, conditioned to have total population size n, converges as $n \to \infty$ to the *Brownian continuum random tree (CRT)*, which can be thought of as the tree inside a *standard Brownian excursion* (more precisely, twice a standard Brownian excursion if one follows Aldous's choice of re-scaling). Aldous describes a procedure for representing trees as closed subsets of ℓ^1, and convergence is here in the sense of weak convergence of probability measures on the this space of closed subsets equipped with the Hausdorff distance.

The Brownian CRT can be obtained as an almost sure limit by growing finite trees (that is, trees with finitely many leaves and finite total branch length) in continuous time as follows (at all times $t \geq 0$ the procedure will produce a rooted tree $\mathcal{R}_t$ with total edge length t):

- Write $\tau_1, \tau_2, \ldots$ for the successive arrival times of an inhomogeneous Poisson process with arrival rate t at time $t \geq 0$. Call τ_n the n^{th} *cut time.*
- Start at time 0 with the 1-tree (that is a line segment with two ends), $\mathcal{R}_0$, of length zero ($\mathcal{R}_0$ is "really" the *trivial tree* that consists of one point only, but thinking this way helps visualize the dynamics more clearly for this semi-formal description). Identify one end of $\mathcal{R}_0$ as the root.
- Let this line segment grow at unit speed until the first cut time τ_1.
- At time τ_1 pick a point uniformly on the segment that has been grown so far. Call this point the *first cut point.*
- Between time τ_1 and time τ_2, evolve a tree with 3 ends by letting a new branch growing away from the first cut point at unit speed.
- Proceed inductively: Given the n-tree (that is, a tree with $n+1$ ends), $\mathcal{R}_{\tau_n-}$, pick the *n-th cut point* uniformly on $\mathcal{R}_{\tau_n-}$ to give an $n+1$-tree, $\mathcal{R}_{\tau_n}$, with one edge of length zero, and for $t \in [\tau_n, \tau_{n+1}[$, let $\mathcal{R}_t$ be the tree obtained from $\mathcal{R}_{\tau_n}$ by letting a branch grow away from the n^{th} cut point with unit speed.

The tree $\mathcal{R}_{\tau_n-}$ has the same distribution as the subtree of the CRT that arises from sampling twice the standard Brownian excursion at n i.i.d. uniform points on the unit interval, and the Brownian CRT is the limit (with respect to a suitable notion of convergence) of the increasing family of rooted finite trees $(\mathcal{R}_t)_{t\geq 0}$.

Again using the cut times $\{\tau_1, \tau_2, \ldots\}$, a closely related way of growing rooted trees $(\mathcal{T}_t)_{t\geq 0}$ can be described as follows (here again the tree at time t will have total edge length t):

- Start with the 1-tree (with one end identified as the *root* and the other as a *leaf*), $\mathcal{T}_0$, of length zero.
- Let this segment grow at unit speed on the time interval $[0, \tau_1[$, and for $t \in [0, \tau_1[$ let $\mathcal{T}_t$ be the rooted 1-tree that has its points labeled

by the interval $[0, t]$ in such a way that the root is t and the leaf is 0.

- At time τ_1 sample the *first cut point* uniformly along the tree $\mathcal{T}_{\tau_1-}$, prune off the piece of $\mathcal{T}_{\tau_1-}$ that is above the cut point (that is, prune off the interval of points that are further away from the root t than the first cut point).
- Re-graft the pruned segment such that its cut end and the root are glued together. Just as we thought of $\mathcal{T}_0$ as a tree with two points, (a leaf and a root) connected by an edge of length zero, we take $\mathcal{T}_{\tau_1}$ to be the the rooted 2-tree obtained by "ramifying" the root $\mathcal{T}_{\tau_1-}$ into two points (one of which we keep as the root) that are joined by an edge of length zero.
- Proceed inductively: Given the labeled and rooted n-tree, $\mathcal{T}_{\tau_{n-1}}$, for $t \in [\tau_{n-1}, \tau_n[$, let $\mathcal{T}_t$ be obtained by letting the edge containing the root grow at unit speed so that the points in $\mathcal{T}_t$ correspond to the points in the interval $[0, t]$ with t as the root. At time τ_n, the n^{th} cut point is sampled randomly along the edges of the n-tree, $\mathcal{T}_{\tau_n-}$, and the subtree above the cut point (that is the subtree of points further away from the root than the cut point) is pruned off and re-grafted so that its cut end and the root are glued together. The root is then "ramified" as above to give an edge of length zero leading from the root to the rest of the tree.

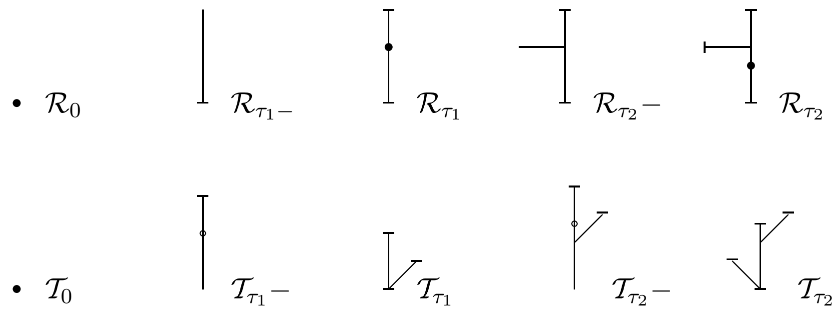

FIGURE 1 *illustrates how the tree-valued processes* $(\mathcal{R}_t;\, t \geq 0)$ *and* $(\mathcal{T}_t;\, t \geq 0)$ *evolve. (The bold dots re-present an edge of length zero, while the small dots indicate the position of the cut point that is going to show up at the next moment.)*

The link between these two dynamics for growing trees is provided by Proposition 4.1, where we show that $\mathcal{T}_{\tau_n-}$ has the same law as $\mathcal{R}_{\tau_n-}$ for each n.

The process $(\mathcal{T}_t)_{t\geq 0}$ is clearly a time-homogeneous Markov process and we can run its dynamics (which we will refer to as *root growth with re-grafting*) starting with any finite tree. The resulting process evolves via alternating deterministic root growth and random jumps due to re-grafting and is an example of a *piecewise-deterministic Markov process*. A general framework for such processes was introduced in [Dav84] as an abstraction of numerous examples in queueing and control theory, and this line of research was extensively developed in the subsequent monograph [Dav93]. A more general formulation in terms of martingales and additive functionals can be found in [JS96]. Some other appearances of such processes are [EP98, CDP01, DC99, Cai93, Cos90]. We note also that pruning and re-grafting operations such as the one we consider here play an important role in algorithms that attempt to reconstruct optimal phylogenetic trees from data by moving through tree space as part of a hill-climbing or simulated annealing procedure (see, for example, [Fel03]).

The crucial feature of the root growth with re-grafting dynamics is that they have a simple projective structure: If one follows the evolution of the points in a rooted subtree of the initial tree along with that of the points added at later times due to root growth, then these points together form a rooted subtree at each period in time and this subtree evolves autonomously according to the root growth with re-grafting dynamics.

The presence of this projective structure suggests that one can make sense of the notion of running the root growth with re-grafting dynamics starting from an initial "tree" that has exotic behavior such as infinitely many leaves, points with infinite branching, and infinite total edge length – provided that this "tree" can be written as the increasing limit of a sequence of finite trees in some appropriate sense. Moreover, by the remarks above about the relationship between the processes $\mathcal{R}$ and $\mathcal{T}$, this extended process should have a stationary distribution that is related to the Brownian CRT, and the stationary distribution should be the limiting distribution for the extended process starting from any initial state.

One of our main objectives is to give rigorous statements and proofs of these and related facts.

Once the extended process has been constructed, we gain a new perspective on objects such as standard Brownian excursion and the associated random triangulation of the circle (see [Ald94a, Ald94b, Ald00]). For example, suppose we follow the height (that is, distance from the root) of some point in the initial tree. It is clear that this height evolves autonomously as a one-dimensional piecewise-deterministic Markov process that:

- increases linearly at unit speed (due to growth at the root),
- makes jumps at rate x when it is in state x (due to cut points falling on the path that connects the root to the point we are following),
- jumps from state x to a point that is uniformly distributed on $[0, x]$ (due to re-grafting at the root).

We call such a process a *Rayleigh process* because, as we will show in Section 8, this process converges to the *standard Rayleigh* stationary distribution $\mathbf{R}$ on $\mathbb{R}_+$ given by

$$\mathbf{R}(]x, \infty[) = e^{-x^2/2}, \quad x \geq 0,$$

(thus $\mathbf{R}$ is also the distribution of the Euclidean length of a two-dimensional standard Gaussian random vector or, up to a scaling constant, the distribution of the distance to the closest point to the origin in a standard planar Poisson process). Now, if $B^{\mathrm{ex}} := \{B_u^{\mathrm{ex}}; u \in [0,1]\}$ is standard Brownian excursion and U is an independent uniform random variable on $[0,1]$, then there is a valid sense in which $2B_U^{ex}$ has the law of the height of a randomly sampled leaf of the Brownian CRT, and this accords with the well-known result

$$\mathbf{P}\{2B_U^{\mathrm{ex}} \in dx\} = \mathbf{R}(dx). \tag{1.1}$$

In order to extend the root growth with re-grafting dynamics to infinite trees we will need to fix on a suitable class of infinite trees and a means of measuring distances between them. Our path to extending the definition of a tree to accommodate the "exotic" behaviors mentioned above will be the one followed in the so-called *T-theory* (see [Dre84, DMT96, Ter97]). T-theory takes finite trees to be just metric spaces with certain characteristic properties and then defines a more general class of tree-like metric spaces called *real-trees* or $\mathbb{R}$*-trees*. We note that one of the primary impetuses for the development of T-theory was to provide mathematical tools for concrete problems in the reconstruction of phylogenies. We also note that $\mathbb{R}$-trees have been objects of intensive study in geometric group theory (see, for example, the surveys [Sha87, Mor92, Sha91, Bes02] and the recent book [Chi01]). Diffusions on an $\mathbb{R}$-tree were investigated in [Eva00]. Some of the results on the space of $\mathbb{R}$-trees obtained in this paper have already been found useful in the study scaling limits of Galton–Watson branching processes in [DLG04].

Once we have an extended notion of trees as just particular abstract metric spaces (or, more correctly, isometry classes of metric spaces), we need a means of assigning a distance between two metric spaces, and this is provided by the *Gromov-Hausdorff distance.* This distance originated in geometry as a means of making sense of intuitive notions such as the convergence to Euclidean space of a re-scaled integer lattice as the grid size approaches zero or the convergence to Euclidean space of a sphere when viewed from a fixed point (for example, the North Pole) as the radius approaches infinity. Our approach is thus rather different to Aldous's in which trees are viewed as closed subsets of ℓ^1 via a particular choice of embedding and distances are measured using the familiar Hausdorff metric. Although our use of the Gromov-Hausdorff distance to metrize the space of $\mathbb{R}$-trees turns out to quite elegant and easy to work with, it also appears to be rather novel, and thus much of our work in this paper is directed towards establishing facts about the structure of this space. However, we think that the resulting mathematics is interesting in its own right and potentially useful in other investigations where trees have hitherto been coded as other objects such as paths (for example, [LG99, DLG02]). We remark in passing that the papers [Pau89, Pau88] are an application of the Gromov-Hausdorff distance to the study of $\mathbb{R}$-trees that is quite different to ours.

We note that there is quite a large literature on other approaches to "geometrizing" and "coordinatizing" spaces of trees. The first construction of codes for labeled trees without edge-length goes back to 1918: Prüfer [Prü18] sets up a bijection between labeled trees of size n and the points of $\{1, 2, \ldots, n\}^{n-2}$. Phylogenetic trees are identified with points in matching polytopes in [DH98], and [BHV01] equips the space of finite phylogenetic trees with a fixed number of leaves with a metric that makes it a cell-complex with non-positive curvature.

The plan of the rest of the paper is as follows. We collect some results on the set of isometry classes of rooted compact $\mathbb{R}$-trees and the properties of the Gromov-Hausdorff distance in Section 2. We construct the extended root growth with re-grafting process in Section 3 via a procedure that is roughly analogous to building a discontinuous Markov process in Euclidean space as the solution of a stochastic differential equation with respect to a sufficiently rich Poisson noise. This approach is particularly well-suited to establishing the strong Markov property. In Section 4 we establish the fact claimed above that $\mathcal{T}_{\tau_n -}$ has the same law as $\mathcal{R}_{\tau_n -}$ for each n. We prove in Section 5 that the extended root growth with re-grafting process is recurrent and convergent to the continuum random tree stationary distribution. We verify that the extended process has a Feller semigroup in Section 6, and show in Section 7 that it is a re-scaling limit of the Markov chain appearing in the Aldous–Broder algorithm for simulating a uniform rooted tree on some finite number of vertices. We devote Section 8 to a discussion of the Rayleigh process described above.

2. $\mathbb{R}$-trees

2.1. **Unrooted trees.** A complete metric space (X, d) is said to be an $\mathbb{R}$-*tree* if it satisfies the following axioms:

Axiom 1 (Unique geodesics) For all $x, y \in X$ there exists a unique isometric embedding $\phi_{x,y} : [0, d(x, y)] \to X$ such that $\phi_{x,y}(0) = x$ and $\phi_{x,y}(d(x, y)) = y$.

Axiom 2 (Loop-free) For every injective continuous map $\psi : [0, 1] \to X$ one has $\psi([0, 1]) = \phi_{\psi(0),\psi(1)}([0, d(\psi(0), \psi(1))])$.

We refer the reader to ([Dre84, DT96, DMT96, Ter97]) for background on $\mathbb{R}$-trees. A particularly useful fact is that a metric space (X, d) is an $\mathbb{R}$-tree if and only if it is complete, path-connected, and satisfies the so-called *four point condition*, that is,

$$\begin{aligned}(2.1)\qquad & d(x_1,x_2)+d(x_3,x_4)\\ &\quad\le \max\{d(x_1,x_3)+d(x_2,x_4),\ d(x_1,x_4)+d(x_2,x_3)\}\end{aligned}$$

for all $x_1,\dots,x_4 \in X$.

Recall that the *Hausdorff distance* between two subsets A_1, A_2 of a metric space (X,d) is defined as

$$(2.2)\qquad d_{\mathrm{H}}(A_1,A_2) := \inf\{\varepsilon>0;\ A_1 \subseteq U_\varepsilon(A_2) \text{ and } A_2 \subseteq U_\varepsilon(A_1)\},$$

where

$$(2.3)\qquad U_\epsilon(A) := \{x \in X;\ d(x,A) \le \varepsilon\}.$$

Based on this notion of distance between closed sets, we define the *Gromov-Hausdorff distance*, $d_{\mathrm{GH}}(X_1,X_2)$, between two metric spaces (X_1,d_{X_1}) and (X_2,d_{X_2}) as the infimum of $d_{\mathrm{H}}(X_1',X_2')$ over all metric spaces X_1' and X_2' that are isomorphic to X_1 and X_2, respectively, and that are subspaces of some common metric space Z (compare [Gro99, BH99, BBI01]). The Gromov-Hausdorff distance defines a finite metric on the space of all isometry classes of compact metric spaces (see, for example, Theorem 7.3.30 in [BBI01]).

Let $(\mathbf{T},d_{\mathrm{GH}})$ be the metric space of isometry classes of compact $\mathbb{R}$-trees equipped with d_{GH}. We will elaborate $\mathbf{T}$ slightly to incorporate the notion of rooted trees, and this latter space of rooted trees will be the state space of the Markov process having root growth with re-grafting dynamics that we are going to construct. We will be a little loose and sometimes refer to an $\mathbb{R}$-tree as an element of $\mathbf{T}$ rather than as a class representative of an element.

Remark As we remarked in the Introduction, Aldous's approach to formalizing the intuitive notion of an infinite tree and putting a metric structure on the resulting class of objects is to work with particular closed subsets of ℓ^1 and to measure distances using the Hausdorff metric on closed sets. Seen in the light of our approach, Aldous's approach uses the distance between two particular representative elements for the isometry classes of a pair of trees rather than the two that minimize the Hausdorff distance. In general this leads to greater distances and hence a topology that is stronger than ours.

The following results says that, at the very least, $\mathbf{T}$ equipped with the Gromov-Hausdorff distance is a "reasonable" space on which to do probability theory.

Theorem 1. *The metric space* $(\mathbf{T},d_{\mathrm{GH}})$ *is complete and separable.*

Before we prove Theorem 1, we point out that a direct application of the stated definition of the Gromov-Hausdorff distance requires an optimal embedding into a new metric space Z. While this definition is conceptually appealing and builds on the more familiar Hausdorff distance between sets, it turns out to often not be so useful for explicit computations in concrete examples. A re-formulation of the Gromov-Hausdorff distance is suggested by the following observation. Suppose that two spaces (X_1, d_{X_1}) and (X_2, d_{X_2}) are close in the Gromov-Hausdorff distance as witnessed by isometric embeddings f_1 and f_2 into some common space Z. The map that associates each point in $x_1 \in X_1$ to a point in $x_2 \in X_2$ such that $d_Z(f_1(x_1), f_2(x_2))$ is minimal should then be close to an isometry onto its image, and a similar remark holds with the roles of X_1 and X_2 reversed.

In order to quantify the observation of the previous paragraph, we require some more notation. A subset $\Re \subseteq X_1 \times X_2$ is said to be a *correspondence* between sets X_1 and X_2 if for each $x_1 \in X_1$ there exists at least one $x_2 \in X_2$ such that $(x_1, x_2) \in \Re$, and for each $y_2 \in X_2$ there exists at least one $y_1 \in X_1$ such that $(y_1, y_2) \in \Re$. Given metrics d_{X_1} and d_{X_2} on X_1 and X_2, respectively, the *distortion* of $\Re$ is defined by

$$\mathrm{dis}(\Re) := \sup\{|d_{X_1}(x_1, y_1) - d_{X_2}(x_2, y_2)|;\ (x_1, x_2), (y_1, y_2) \in \Re\}. \tag{2.4}$$

Then

$$d_{\mathrm{GH}}((X_1, d_{X_1}), (X_2, d_{X_2})) = \frac{1}{2} \inf_{\Re} \mathrm{dis}(\Re), \tag{2.5}$$

where the infimum is taken over all correspondences $\Re$ between X_1 and X_2 (see, for example, Theorem 7.3.25 in [BBI01]).

The following result is also useful in the proof of Theorem 1.

Lemma 2.1. *The set* $\mathbf{T}$ *of compact* $\mathbb{R}$*-trees is a closed subset of the space of compact metric spaces equipped with the Gromov-Hausdorff distance.*

Proof. It suffices to note that the limit of a sequence in $\mathbf{T}$ is path-connected (see, for example, Theorem 7.5.1 in [BBI01]) and satisfies the four point condition (2.1), (indeed, as remarked after Proposition 7.4.12 in [BBI01], there is a "meta–theorem" that if a feature of a compact metric space can be formulated as a continuous property of distances among finitely many points, then this feature is preserved under Gromov-Hausdorff limits). □

Proof of Theorem 1 We start by showing *separability*. Given a compact $\mathbb{R}$-tree, T, and $\varepsilon > 0$, let S_ε be a finite ε-net in T. For $a, b \in T$, let

$$[a, b[:= \phi_{a,b}(\,[0, d(a, b)[\,) \quad \text{and} \quad]a, b[:= \phi_{a,b}(\,]0, d(a, b)[\,) \tag{2.6}$$

be the unique half open and open, respectively, *arc* between them, and write T_ε for the *subtree of* T *spanned by* S_ε, that is,

$$T_\varepsilon := \bigcup\nolimits_{x,y \in S_\varepsilon} [x, y] \quad \text{and} \quad d_{T_\varepsilon} := d\big|_{T_\varepsilon}. \tag{2.7}$$

Obviously, T_ε is still an ε-net for T, and hence $d_{\mathrm{GH}}(T_\varepsilon, T) \le d_H(T_\varepsilon, T) \le \varepsilon$.

Now each T_ε is just a "finite tree with edge-lengths" and can clearly be approximated arbitrarily closely in the d_{GH}-metric by trees with the same tree topology (that is, "shape"), and rational edge-lengths. The set of isometry types of finite trees with rational edge-lengths is countable, and so $(\mathbf{T}, d_{\mathrm{GH}})$ is separable.

It remains to establish *completeness*. It suffices by Lemma 2.1 to show that any Cauchy sequence in $\mathbf{T}$ converges to some compact metric space, or, equivalently, any Cauchy sequence in $\mathbf{T}$ has a subsequence that converges to some metric space.

Let $(T_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathbf{T}$. By Exercise 7.4.14 and Theorem 7.4.15 in [BBI01], a sufficient condition for this sequence to have a subsequential limit is that for every $\varepsilon > 0$ there exists a positive number $N = N(\varepsilon)$ such that every T_n contains an ε-net of cardinality N.

Fix $\varepsilon > 0$ and $n_0 = n_0(\varepsilon)$ such that $d_{\mathrm{GH}}(T_m, T_n) < \varepsilon/2$ for $m, n \geq n_0$. Let S_{n_0} be a finite $(\varepsilon/2)$-net for T_{n_0} of cardinality N. Then by (2.5) for each $n \geq n_0$ there exists a correspondence $\Re_n$ between T_{n_0} and T_n such that $\mathrm{dis}(\Re_n) < \varepsilon$. For each $x \in T_{n_0}$, choose $f_n(x) \in T_n$ such that $(x, f_n(x)) \in \Re_n$. Since for any $y \in T_n$ with $(x, y) \in \Re_n$, $d_{T_n}(y, f_n(x)) \leq \mathrm{dis}(\Re_n)$, for all $n \geq n_0$, the set $f_n(S_{n_0})$ is an ε-net of cardinality N for T_n, $n \geq n_0$. □

2.2. **Unrooted trees with 4 leaves.** For the sake of reference and establishing some notation, we record here some well-known facts about reconstructing trees from a knowledge of the distances between the leaves. We remark that the fact that trees can be reconstructed from their collection of leaf-to-leaf distances (plus also the leaf-to-root distances for rooted trees) is of huge practical importance in so-called *distance methods* for inferring phylogenetic trees from DNA sequence data, and the added fact that one can build such trees by building subtrees for each collection of four leaves is the starting point for the sub-class of distance methods called *quartet methods.* We refer the reader to [Fel03, SS03] for an extensive description of these techniques and their underlying theory.

Lemma 2.2. *The isometry class of an unrooted tree (T, d) with four leaves is uniquely determined by the distances between the leaves of T.*

Proof. Let $\{x_1, x_2, x_3, x_4\}$ be the set of leaves of T. The tree T has one of four possible shapes:

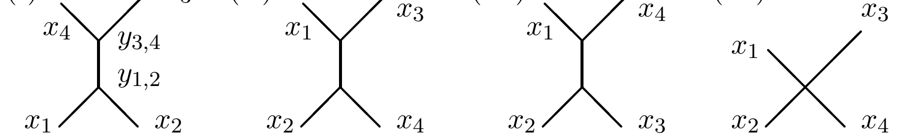

FIGURE 2 *shows the* 4 *different shapes of a labeled tree with* 4 *leaves.*

Consider case (I), and let $y_{1,2}$ be the uniquely determined branch point on the tree that lies on the arcs $[x_1, x_2]$ and $[x_1, x_3]$, and $y_{3,4}$ be the uniquely determined branch point on the tree that lies on the arcs $[x_3, x_4]$ and $[x_1, x_3]$. Observe that

$$
\begin{aligned}
d(x_1, y_{1,2}) &= \frac{1}{2}(d(x_1, x_2) + d(x_1, x_3) - d(x_2, x_3)) \\
d(x_2, y_{1,2}) &= \frac{1}{2}(d(x_1, x_2) + d(x_2, x_3) - d(x_1, x_3)) \\
d(x_3, y_{3,4}) &= \frac{1}{2}(d(x_3, x_4) + d(x_1, x_3) - d(x_1, x_4)) \\
d(x_4, y_{3,4}) &= \frac{1}{2}(d(x_3, x_4) + d(x_1, x_4) - d(x_1, x_3)) \\
d(y_{1,2}, y_{3,4}) &= \frac{1}{2}(d(x_1, x_4) + d(x_2, x_3) - d(x_1, x_2) - d(x_3, x_4)),
\end{aligned}
\tag{2.8}
$$

Similar observations for the other cases show that if we know the shape of the tree, then we can determine its edge-lengths from leaf-to-leaf distances. Note also that

$$
\begin{aligned}
\chi_{(I)}(T) &:= \frac{1}{2}\left(d(x_1, x_3) + d(x_2, x_4) - d(x_1, x_2) - d(x_3, x_4)\right) \\
&\begin{cases} > 0 & \text{for shape (I)}, \\ < 0 & \text{for shape (II)}, \\ = 0 & \text{for shapes (III) and (IV)} \end{cases} .
\end{aligned}
\tag{2.9}
$$

This and analogous inequalities for the quantities that reconstruct the length of the "internal" edge in shapes (II) and (III), respectively, show that the shape of the tree can also be reconstructed from leaf-to-leaf distances. □

2.3. **Rooted $\mathbb{R}$-trees.** Since we are mainly interested in rooted trees, we extend our definition as follows: A *rooted* $\mathbb{R}$*-tree*, (X, d, ρ), is an $\mathbb{R}$-tree (X, d) with a distinguished point $\rho \in X$ that we call the *root*. It is helpful to use genealogical terminology and think of ρ as a common ancestor and $h(x) := d(\rho, x)$ as the real-valued generation to which $x \in X$ belongs ($h(x)$ is also called the *height* of x). We define a partial order $\leq$ on X by declaring (using the notation introduced in (2.6)) that $x \leq y$ if $x \in [\rho, y]$, so that x is an ancestor of y. Each pair $x, y \in X$ has a well-defined *greatest common lower bound*, $x \wedge y$, in this partial order that we think of as the most recent common ancestor of x and y.

Let $\mathbf{T}^{\text{root}}$ denote the collection of all root-invariant isometry classes of rooted compact $\mathbb{R}$-trees, where we define a root-invariant isometry to be an isometry $\xi : (X_1, d_{X_1}, \rho_1) \to (X_2, d_{X_2}, \rho_2)$ with $\xi(\rho_1) = \rho_2$.

We want to equip $\mathbf{T}^{\text{root}}$ with a Gromov-Hausdorff type distance that incorporates the special status of the root. We define the *rooted Gromov-Hausdorff distance*, $d_{\text{GH}^{\text{root}}}((X_1, \rho_1), (X_2, \rho_2))$, between two rooted $\mathbb{R}$-trees

(X_1, ρ_1) and (X_2, ρ_2) as the infimum of $d_{\mathrm{H}}(X'_1, X'_2) \vee d_Z(\rho'_1, \rho'_2)$ over all rooted $\mathbb{R}$-trees (X'_1, ρ'_1) and (X'_2, ρ'_2) that are root-invariant isomorphic to (X_1, ρ_1) and (X_2, ρ_2), respectively, and that are (as unrooted trees) subspaces of a common metric space (Z, d_Z).

As in (2.5), we can compute $d_{\mathrm{GH}^{\mathrm{root}}}((X_1, d_{X_1}, \rho_1), (X_2, d_{X_2}, \rho_2))$ by comparing distances within X_1 to distances within X_2, provided that the distinguished status of the root is respected.

Lemma 2.3. *For two rooted trees (X_1, d_{X_1}, ρ_1), and (X_2, d_{X_2}, ρ_2),*

$$d_{\mathrm{GH}^{\mathrm{root}}}((X_1, d_{X_1}, \rho_1), (X_2, d_{X_2}, \rho_2)) = \frac{1}{2} \inf_{\Re^{\mathrm{root}}} \mathrm{dis}(\Re^{\mathrm{root}}), \tag{2.10}$$

where now the infimum is taken over all correspondences $\Re^{\mathrm{root}}$ between X_1 and X_2 with $(\rho_1, \rho_2) \in \Re^{\mathrm{root}}$.

Proof. Indeed, for any root-invariant isometric copies (X'_1, ρ'_1) and (X'_2, ρ'_2) embedded in Z, and $r > d_{\mathrm{GH}^{\mathrm{root}}}((X_1, \rho_1), (X_2, \rho_2))$,

$$\Re^{\mathrm{root}} := \{(x_1, x_2);\, x_1 \in X'_1, x_2 \in X'_2,\, d_Z(x_1, x_2) < r\} \tag{2.11}$$

gives a correspondence between X_1 and X_2 containing (ρ_1, ρ_2) such that $\mathrm{dis}(\Re^{\mathrm{root}}) < 2r$.

On the other hand, given a correspondence $\Re^{\mathrm{root}}$ between X_1 and X_2 containing (ρ_1, ρ_2), define a metric $d_{X_1 \amalg X_2}$ on the disjoint union $X_1 \amalg X_2$ such that the restriction of $d_{X_1 \amalg X_2}$ to X_i is d_{X_i}, for $i = 1, 2$, and for $x_1 \in X_1$, $x_2 \in X_2$, by

$$\begin{aligned} &d_{X_1 \amalg X_2}(x_1, x_2) \\ &\quad := \inf\{d_{X_1}(x_1, y_1) + d_{X_2}(x_2, y_2) + \frac{1}{2}\mathrm{dis}(\Re^{\mathrm{root}}) \,:\, (y_1, y_2) \in \Re^{\mathrm{root}}\} \end{aligned} \tag{2.12}$$

– in particular, if the pair (x_1, x_2) actually belongs to the correspondence $\Re^{\mathrm{root}}$, then $d_{X_1 \amalg X_2}(x_1, x_2) = \frac{1}{2} dis(\Re^{\mathrm{root}})$. We leave it to the reader to check that $d_{X_1 \amalg X_2}$ is, indeed, a metric. Then (computing Hausdorff distance within $X_1 \amalg X_2$ using $d_{X_1 \amalg X_2}$) we have

$$d_{\mathrm{H}}(X_1, X_2) \vee d_{X_1 \amalg X_2}(\rho_1, \rho_2) \le \frac{1}{2}\mathrm{dis}(\Re^{\mathrm{root}}). \tag{2.13}$$

□

We state an analogue of Theorem 1 for rooted compact $\mathbb{R}$-trees.

Theorem 2. *The metric space $(\mathbf{T}^{\mathrm{root}}, d_{GH^{\mathrm{root}}})$ is complete and separable.*

Before we can prove Theorem 2 we need two preparatory results. The first is a the counterpart of Corollary 7.3.28 in [BBI01] and presents convenient upper and lower estimates for $d_{\mathrm{GH}^{\mathrm{root}}}$ that differ by a multiplicative constant.

Let (X_1, ρ_1) and (X_2, ρ_2) be two rooted compact $\mathbb{R}$-trees, and take $\varepsilon > 0$. A map f is called a *root-invariant ε-isometry* from (X_1, ρ_1) to (X_2, ρ_2) if

$f(\rho_1) = \rho_2$, $\operatorname{dis}(f) := \sup\{|d_{X_1}(x,y) - d_{X_2}(f(x), f(y))| : x, y \in X_1\} < \varepsilon$ and $f(X_1)$ is an ε-net for X_2.

Lemma 2.4. *Let (X_1, ρ_1) and (X_2, ρ_2) be two rooted compact $\mathbb{R}$-trees, and take $\varepsilon > 0$. Then the following hold.*

(i) *If $d_{\mathrm{GH}^{\mathrm{root}}}((X_1, \rho_1), (X_2, \rho_2)) < \varepsilon$, then there exists a root-invariant 2ε-isometry from (X_1, ρ_1) to (X_2, ρ_2).*

(ii) *If there exists a root-invariant ε-isometry from (X_1, ρ_1) to (X_2, ρ_2), then*
$$d_{\mathrm{GH}^{\mathrm{root}}}((X_1, \rho_1), (X_2, \rho_2)) \le \frac{3}{2}\varepsilon.$$

Proof. (i) Let $d_{\mathrm{GH}^{\mathrm{root}}}((X_1, \rho_1), (X_2, \rho_2)) < \varepsilon$. By Lemma 2.3 there exists a correspondence $\Re^{\mathrm{root}}$ between X_1 and X_2 such that $(\rho_1, \rho_2) \in \Re^{\mathrm{root}}$ and $\operatorname{dis}(\Re^{\mathrm{root}}) < 2\varepsilon$. Define $f : X_1 \to X_2$ by setting $f(\rho_1) = \rho_2$, and choosing $f(x)$ such that $(x, f(x)) \in \Re^{\mathrm{root}}$ for all $x \in X_1 \setminus \{\rho_1\}$. Clearly, $\operatorname{dis}(f) \le \operatorname{dis}(\Re^{\mathrm{root}}) < 2\varepsilon$. To see that $f(X_1)$ is an 2ε-net for X_2, let $x_2 \in X_2$, and choose $x_1 \in X_1$ such that $(x_1, x_2) \in \Re^{\mathrm{root}}$. Then $d_{X_2}(f(x_1), x_2) \le d_{X_1}(x_1, x_1) + \operatorname{dis}(\Re^{\mathrm{root}}) < 2\varepsilon$.

(ii) Let f be a root-invariant ε-isometry from (X_1, ρ_1) to (X_2, ρ_2). Define a correspondence $\Re_f^{\mathrm{root}} \subseteq X_1 \times X_2$ by

$$\Re_f^{\mathrm{root}} := \{(x_1, x_2) : d_{X_2}(x_2, f(x_1)) \le \varepsilon\}. \tag{2.14}$$

Then $(\rho_1, \rho_2) \in \Re_f^{\mathrm{root}}$ and $\Re_f^{\mathrm{root}}$ is indeed a correspondence since $f(X_1)$ is a ε-net for X_2. If $(x_1, x_2), (y_1, y_2) \in \Re_f^{\mathrm{root}}$, then

$$\begin{aligned} |d_{X_1}(x_1, y_1) - d_{X_2}(x_2, y_2)| &\le |d_{X_2}(f(x_1), f(y_1)) - d_{X_1}(x_1, y_1)| \\ &\quad + d_{X_2}(x_2, f(x_1)) + d_{X_2}(f(x_1), y_2) \\ &< 3\varepsilon. \end{aligned} \tag{2.15}$$

Hence $\operatorname{dis}(\Re_f^{\mathrm{root}}) < 3\varepsilon$ and, by (2.10), $d_{\mathrm{GH}^{\mathrm{root}}}((X_1, \rho_1), (X_2, \rho_2)) \le \frac{3}{2}\varepsilon$. $\square$

The second preparatory result we need is the following compactness criterion, which is the analogue of Theorem 7.4.15 in [BBI01] (note also Exercise 7.4.14 in [BBI01]) and can be proved the same way, using Lemma 2.4 in place of Corollary 7.3.28 in [BBI01] and noting that the analogue of Lemma 2.1 holds for $\mathbf{T}^{\mathrm{root}}$.

Lemma 2.5. *A subset $\mathcal{T} \subset \mathbf{T}^{\mathrm{root}}$ is pre-compact if for every $\varepsilon > 0$ there exists a positive integer $N(\varepsilon)$ such that each $T \in \mathcal{T}$ has an ε-net with at most $N(\varepsilon)$ points.*

Proof of Theorem 2 The proof follows very much the same lines as that of Theorem 1. The proof of *separability* is almost identical. The key step in establishing *completeness* is again to show that a Cauchy sequence in $\mathbf{T}^{\mathrm{root}}$ has a subsequential limit. This can be shown in the same manner as in the proof of Theorem 1, with an appeal to Lemma 2.5 replacing one to Theorem 7.4.15 and Exercise 7.4.14 in [BBI01]. $\square$

2.4. **Length measure.** Recall that the root growth with re-grafting dynamics involve points being chosen uniformly at random on a finite tree. In order to extend the dynamics to general rooted compact $\mathbb{R}$-trees, we will require the fact that rooted compact $\mathbb{R}$-trees are associated with a natural length measure as follows. Fix $(T, d, \rho) \in \mathbf{T}^{\mathrm{root}}$, and denote the Borel-$\sigma$-field on T by $\mathcal{B}(T)$. For $a, b \in T$, recall the half open arc $[a, b[$ from (2.6), and let

$$T^o := \bigcup_{b \in T} [\rho, b[\tag{2.16}$$

the *skeleton* of T. Observe that if $T' \subset T$ is a dense countable set, then (2.16) holds with T replaced by T'. In particular, $T^o \in \mathcal{B}(T)$ and $\mathcal{B}(T)\big|_{T^o} = \sigma(\{]a, b[;\ a, b \in T'\})$, where $\mathcal{B}(T)\big|_{T^o} := \{A \cap T^o;\ A \in \mathcal{B}(T)\}$. Hence there exist a unique σ-finite measure μ on T, called *length measure*, such that $\mu(T \setminus T^o) = 0$ and

$$\mu(]a, b[) = d(a, b), \quad \forall\, a, b \in T. \tag{2.17}$$

In particular, μ is the trace onto T^o of one-dimensional Hausdorff measure on T.

Remark The terminology *skeleton* might seem somewhat derisory, since for finite trees the difference between the skeleton and the whole tree is just a finite number of points. However, it is not difficult to produce $\mathbb{R}$-trees for which the difference between the skeleton and the whole tree is a set with Hausdorff dimension greater than one (the Brownian CRT will almost surely be such a tree). This explains our requirement that μ is carried by the skeleton.

Remark Elements of $\mathbf{T}^{\mathrm{root}}$ are really equivalence classes of trees rather than trees themselves, so what we are describing here is a way of associating a measure to each element of the equivalence class. However, this procedure respects the equivalence relation in that if T' and T'' are two representatives of the same equivalence class and are related by a root-invariant isometry $f : T' \to T''$, then the associated length measures μ' and μ'' are such that μ'' is the push-forward of μ' by f and μ' is the push-forward of μ'' by the inverse of f (that is, $\mu''(A'') = \mu'(f^{-1}(A''))$ and $\mu'(A') = \mu''(f(A'))$ for Borel sets A' and A'' of T' and T'', respectively).

2.5. **Rooted subtrees and trimming.** Recall from the Introduction that our strategy for extending the root growth with re-grafting dynamics from finite trees will involve a limiting procedure in which a general rooted compact $\mathbb{R}$-tree is approximated "from the inside" by an increasing sequence of finite subtrees. We therefore need to establish some facts about such approximations.

To begin with, we require a notation for one tree being a subtree of another, with both trees sharing the same root. We need to incorporate the fact that we are dealing with equivalence classes of trees rather than trees themselves. A *rooted subtree* of $(T, d, \rho) \in \mathbf{T}^{\mathrm{root}}$ is an element $(T^*, d^*, \rho^*), \in \mathbf{T}^{\mathrm{root}}$ that has a class representative that is a subspace of a class representative of (T, d, ρ), with the two roots coincident. Equivalently, any class representative of (T^*, d^*, ρ^*) can be isometrically embedded into any class representative of (T, d, ρ) via an isometry that maps roots to roots. We write $T^* \preceq^{\mathrm{root}} T$ and note that $\preceq^{\mathrm{root}}$ is an partial order on $\mathbf{T}^{\mathrm{root}}$.

All of the "wildness" in a compact $\mathbb{R}$-tree happens "at the leaves". For example, if $T \in \mathbf{T}^{\mathrm{root}}$ has a point x at which infinite branching occurs (so that the removal of x would disconnect T into infinitely many components), then any open neighborhood of x must contain infinitely many leaves, while for each $\eta > 0$ there are only finitely many leaves y such that $x \in [\rho, y]$ with $d(x, y) > \eta$. A natural way in which to produce a finite subtree that approximates a given tree is thus to fix $\eta > 0$ and trim off the fringe of the tree by removing those points that are not at least distance η from at least one leaf. Formally, for $\eta > 0$ define $R_\eta : \mathbf{T}^{\mathrm{root}} \to \mathbf{T}^{\mathrm{root}}$ to be the map that assigns to $(T, \rho) \in \mathbf{T}^{\mathrm{root}}$ the rooted subtree $(R_\eta(T), \rho)$ that consists of ρ and points $a \in T$ for which the subtree

$$S^{T,a} := \{x \in T : a \in [\rho, x[\} \tag{2.18}$$

(that is, the *subtree above* a) has height greater than or equal to η. Equivalently,

$$R_\eta(T) := \{x \in T : \exists\, y \in T\ x \in [\rho, y],\ d_T(x, y) \geq \eta\} \cup \{\rho\}. \tag{2.19}$$

In particular, if T has height at most η, then $R_\eta(T)$ is just the trivial tree consisting of the root ρ.

Remark Notice that the map described in (2.19) maps a metric space into a sub-space. However, since isometric spaces are mapped into isometric sub-spaces, we may think of R_η as a map from $\mathbf{T}^{\mathrm{root}}$ into $\mathbf{T}^{\mathrm{root}}$.

Lemma 2.6. (i) *The range of R_η consists of finite rooted trees.*
(ii) *The map R_η is continuous.*
(iii) *The family of maps $(R_\eta)_{\eta>0}$ is a semigroup; that is, $R_{\eta'} \circ R_{\eta''} = R_{\eta'+\eta''}$ for $\eta', \eta'' > 0$. In particular, $R_{\eta'}(T) \preceq^{\mathrm{root}} R_{\eta''}(T)$ for $\eta' \geq \eta'' > 0$.*
(iv) *For any $(T, \rho) \in \mathbf{T}^{\mathrm{root}}$, $d_{\mathrm{GH}^{\mathrm{root}}}((T, \rho), (R_\eta(T), \rho)) \leq d_{\mathrm{H}}(T, R_\eta(T)) \leq \eta$.*

Proof. (i) Fix $(T, d, \rho) \in \mathbf{T}^{\mathrm{root}}$. Let $E \subset R_\eta(T)$ be the leaves of R_η, that is, the points that have no subtree above them. We have to show that E is finite. However, if $a_1, a_2, \ldots$ are infinitely many points in $E \setminus \{\rho\}$, then we can find points $b_1, b_2, \ldots$ in T such that b_i is in the subtree above a_i and $d(a_i, b_i) \geq \eta$. It follows that $\inf_{i \neq j} d(b_i, b_j) \geq 2\eta$, which contradicts the compactness of T.

(ii) Suppose that (T', d', ρ') and (T'', d'', ρ'') are two compact trees with

$$d_{\mathrm{GH}^{\mathrm{root}}}((T', \rho'), (T'', \rho'')) < \varepsilon.$$

By part (i) of Lemma 2.4 there exists a root-invariant 2ϵ-isometry $f : T' \to T''$. Recall that this means, $f(\rho') = \rho''$, $\mathrm{dis}(f) < 2\varepsilon$, and $f(T')$ is an 2ε-net for T''.

For $a \in R_\eta(T')$, let $\bar{f}(a)$ be the unique point in $R_\eta(T'')$ that is closest to $f(a)$. We will show that $\bar{f} : R_\eta(T') \to R_\eta(T'')$ is a root-invariant 25ε-isometry and hence, by part (ii) of Lemma 2.4, $d_{\mathrm{GH}^{\mathrm{root}}}(R_\eta(T'), R_\eta(T'')) \leq \frac{3}{2} 25\varepsilon$.

We first show that

$$\sup\{d''(f(a), \bar{f}(a)) : a \in R_\eta(T')\} \leq 8\varepsilon. \tag{2.20}$$

Fix $a \in R_\eta(T')$ and let $b \in T'$ be a point in the subtree above a such that $d'(a, b) \geq \eta$. Denote the most recent common ancestor of $f(a)$ and $f(b)$ on T'' by $f(a) \wedge'' f(b)$.

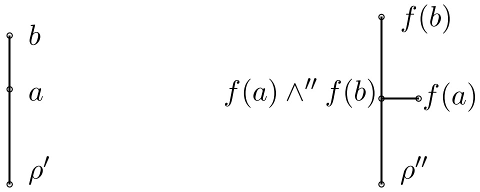

FIGURE 3 *illustrates the shapes of the trees spanned by* $\{\rho', a, b\}$ *and by* $\{\rho'', f(a), f(b)\}$. *The point* $\bar{f}(a)$ *lies somewhere on the arc* $[\rho'', f(a)]$.

Then

$$\begin{aligned}
&d''\big(f(a) \wedge'' f(b), f(a)\big) \\
&\quad = \frac{1}{2}\big(d''(f(a), f(b)) + d''(\rho'', f(a)) - d''(\rho'', f(b))\big) \\
&\quad \leq \frac{1}{2}\big(\big|d''(f(a), f(b)) - d'(a, b)\big| \\
&\qquad\qquad + \big|d''(\rho'', f(a)) - d'(\rho', a)\big| + \big|d''(\rho'', f(b)) - d'(\rho', b)\big|\big) \\
&\quad \leq 3\varepsilon.
\end{aligned} \tag{2.21}$$

If $\bar{f}(a) \in [f(a) \wedge'' f(b), f(a)]$ then we are immediately done. Otherwise, $\bar{f}(a) \in [\rho'', f(a)]$ and $\bar{f}(a)$ is a leaf in $R_\eta(T'')$. Hence $f(b) \notin R_\eta(T'')$, and therefore

$$d''(\bar{f}(a), f(b)) \leq \eta. \tag{2.22}$$

Furthermore,

$$
\begin{aligned}
d''\big(f(a)\wedge'' f(b), f(b)\big) &= d''(f(a), f(b)) - d''\big(f(a)\wedge'' f(b), f(a)\big) \\
&\geq \big(d'(a,b) - 2\varepsilon\big) - 3\varepsilon \\
&\geq \eta - 5\varepsilon.
\end{aligned}
\tag{2.23}
$$

Combining (2.21), (2.22) and (2.23) finally yields that $d''(\bar{f}(a), f(a)) \leq 8\varepsilon$ and completes the proof of (2.20).

It follows from (2.20) that

$$
\begin{aligned}
\operatorname{dis}(\bar{f}) &= \sup\{|d'(a,b) - d''(\bar{f}(a), \bar{f}(b))| : a, b \in R_\eta(T')\} \\
&\leq \sup\{|d'(a,b) - d''(f(a), f(b))| : a, b \in R_\eta(T')\} \\
&\quad + 2\sup\{d''(f(a), \bar{f}(a)) : a \in R_\eta(T')\} \\
&< 2\varepsilon + 2 \times 8\varepsilon \\
&= 18\varepsilon.
\end{aligned}
$$

The proof of (ii) will thus be completed if we can show that $\bar{f}(R_\eta(T'))$ is a 25ε-net in $R_\eta(T'')$. Consider a point $c \in R_\eta(T'')$. We need to show that there is a point $b \in R_\eta(T')$ such that

$$
d''(\bar{f}(b), c) < 25\varepsilon. \tag{2.24}
$$

If $d''(\rho'', c) < 7\varepsilon$, then we are done, because we can take $b = \rho'$ (recall that $\bar{f}(\rho') = \rho''$). Assume, therefore, that $d''(\rho'', c) \geq 7\varepsilon$. We can then find points $c_-, c_+ \in T''$ such that $\rho'' \leq c_- \leq c \leq c+$ with $d''(c_-, c) = 7\varepsilon$ and $d''(c, c_+) \geq \eta$. There are corresponding points $a_-, a, a_+ \in T'$ such that $d''(f(a_-), c_-) < 2\varepsilon$, $d''(f(a), c) < 2\varepsilon$, and $d''(f(a_+), c_+) < 2\varepsilon$. We claim that $b := a_- \wedge' a_+$ (the most recent common ancestor of a_- and a_+ in the tree T') belongs to $R_\eta(T')$ and satisfies (2.24).

Note first of all that

$$
\begin{aligned}
d'(b, a_+) &= d'(a_- \wedge' a_+, a_+) \\
&= \frac{1}{2}\big(d'(a_+, a_-) + d'(\rho', a_+) - d'(\rho', a_-)\big) \\
&\geq \frac{1}{2}\big(d''(f(a_+), f(a_-)) - 2\varepsilon + d''(f(\rho'), f(a_+)) - 2\varepsilon \\
&\quad - d''(f(\rho'), f(a_-)) - 2\varepsilon\big) \\
&\geq \frac{1}{2}\big(d''(c_+, c_-) - 4\varepsilon + d''(\rho'', c_+) - 2\varepsilon - d''(\rho'', c_-) - 2\varepsilon\big) - 3\varepsilon \\
&= d''(c_+, c_-) - 7\varepsilon \\
&= \eta + 7\varepsilon - 7\varepsilon \\
&\eta,
\end{aligned}
$$

and so $b \in R_\eta(T')$.

Furthermore,

$$\begin{aligned}
d''(c, f(b)) &\le d''(c, c_-) + d''(c_-, f(a_-)) + d''(f(a_-), f(b)) \\
&\le 7\varepsilon + 2\varepsilon + d'(a_-, b) + 2\varepsilon \\
&= 11\varepsilon + \frac{1}{2}\left(d'(a_+, a_-) + d'(\rho', a_-) - d'(\rho', a_+)\right) \\
&\le 11\varepsilon + \frac{1}{2}\big(d''(f(a_+), f(a_-)) + 2\varepsilon + d''(f(\rho'), f(a_-)) + 2\varepsilon \\
&\quad - d'(f(\rho'), f(a_+)) + 2\varepsilon\big) \\
&\le 14\varepsilon + \frac{1}{2}\left(d''(c_+, c_-) + 2\varepsilon + d''(\rho'', c_-) + 2\varepsilon - d''(\rho'', c_+) + 2\varepsilon\right) \\
&= 17\varepsilon.
\end{aligned}$$

Therefore, by (2.20),

$$d(c, \bar{f}(b)) \le 17\varepsilon + 8\varepsilon = 25\varepsilon.$$

This completes the proof of (2.24), and thus the proof of part (ii).
Claims (iii) and (iv) are clear. □

Finally, we require the following result, which will be the key to showing that the "projective limit" of a consistent family of tree-valued processes can actually be thought of as a tree-valued process in its own right.

Lemma 2.7. *Consider a sequence $(T_n)_{n\in\mathbb{N}}$ of representatives of isometry classes of rooted compact trees in $(\mathbf{T}, d_{\mathrm{GH}^{\mathrm{root}}})$ with the following properties.*

- *Each set T_n is a subset of some common set U.*
- *Each tree T_n has the same root $\rho \in U$.*
- *The sequence $(T_n)_{n\in\mathbb{N}}$ is nondecreasing, that is, $T_1 \subseteq T_2 \subseteq \cdots \subseteq U$.*
- *Writing d_n for the metric on T_n, for $m < n$ the restriction of d_n to T_m coincides with d_m, so that there is a well-defined metric on $T := \bigcup_{n\in\mathbb{N}} T_n$ given by*

$$d(a,b) = d_n(a,b), \quad a, b \in T_n. \tag{2.25}$$

- *The sequence of subsets $(T_n)_{n\in\mathbb{N}}$ is Cauchy in the Hausdorff distance with respect to d.*

Then the following hold.

(i) *The metric completion $\bar{T}$ of T is a compact $\mathbb{R}$-tree, and $d_{\mathrm{H}}(T_n, \bar{T}) \to 0$ as $n \to \infty$, where the Hausdorff distance is computed with respect to the extension of d to $\bar{T}$. In particular,*

$$\lim_{n\to\infty} d_{\mathrm{GH}^{\mathrm{root}}}((T_n, \rho), (\bar{T}, \rho)) = 0. \tag{2.26}$$

(ii) *The tree $\bar{T}$ has skeleton $\bar{T}^o = \bigcup_{n\in\mathbb{N}} T_n^o$.*

(iii) *The length measure on $\bar{T}$ is the unique measure concentrated on $\bigcup_{n\in\mathbb{N}} T_n^o$ that restricts to the length measure on T_n for each $n \in \mathbb{N}$.*

Proof. (i) Because $\bar{T}$ is a complete metric space, the collection of closed subsets of $\bar{T}$ equipped with the Hausdorff distance is also complete (see, for example, Proposition 7.3.7 of [BBI01]). Therefore the Cauchy sequence $(T_n)_{n\in\mathbb{N}}$ has a limit that is (see Exercise 7.3.4 of [BBI01]) the closure of $\bigcup_{k\in\mathbb{N}} T_k$, i.e, $\bar{T}$ itself. It is clear that the complete space $\bar{T}$ is totally bounded, path-connected, and satisfies the four point condition, and so $\bar{T}$ is a compact $\mathbb{R}$-tree. Finally,

$$d_{\mathrm{GH}^{\mathrm{root}}}((T_n,\rho),(\bar{T},\rho)) \le d_{\mathrm{H}}(T_n,\bar{T}) \vee d(\rho,\rho) = d_{\mathrm{H}}(T_n,\bar{T}) \to 0, \tag{2.27}$$

as $n \to \infty$.

Claims (ii) and (iii) are obvious. □

3. Root growth with re-grafting

3.1. **Beginning the construction.** We are now ready to begin in earnest the construction of the $\mathbf{T}^{\mathrm{root}}$-valued Markov process, X, having the root growth with re-grafting dynamics.

Fix a tree $(T,d,\rho) \in \mathbf{T}^{\mathrm{root}}$. This tree will be the initial state of X. We first recapitulate the strategy outlined in the Introduction. In line with that semi-formal description, the "stochastic inputs" to the construction of X will be a collection of cut times and a corresponding collection of cut points.

- Construct simultaneously for each finite rooted subtree $T^* \preceq^{\mathrm{root}} T$ a process X^{T^*} with $X_0^{T^*} = T^*$ that evolves according the root growth with re-grafting dynamics.
- Carry out this construction in such a way that if T^* and T^{**} are two finite subtrees with $T^* \preceq^{\mathrm{root}} T^{**}$, then $X_t^{T^*} \preceq^{\mathrm{root}} X_t^{T^{**}}$ and the cut points for X^{T^*} are those for $X^{T^{**}}$ that happen to fall on $X_{\tau-}^{T^*}$ for a corresponding cut time τ of $X^{T^{**}}$. Cut times τ for $X^{T^{**}}$ for which the corresponding cut point does not fall on $X_{\tau-}^{T^*}$ are not cut times for X^{T^*}.
- The tree (T,ρ) is a rooted Gromov-Hausdorff limit of finite $\mathbb{R}$-trees with root ρ (indeed, any subtree spanned by a finite ε-net and ρ is finite and has rooted Gromov-Hausdorff distance less than ε from (T,ρ)). In particular, (T,ρ) is the "smallest" rooted compact $\mathbb{R}$-tree that contains all of the finite rooted subtrees of (T,ρ).
- Because of the consistent projective nature of the construction, we can define $X_t := X_t^T$ for $t \ge 0$ as the "smallest" element of $\mathbf{T}^{\mathrm{root}}$ that contains $X_t^{T^*}$, for all finite trees $T^* \preceq^{\mathrm{root}} T$.

It will be convenient for establishing features of the process X such as the strong Markov property to introduce randomness later and work initially in a setting where the cut times and cut points are fixed. There are two types of cut points: those that occur at points which were present in the initial tree T and those that occur at points which were added due to subsequent root growth. Accordingly, we consider two countable subsets $\pi_0 \subset \mathbb{R}^{++} \times T^o$ and $\pi \subset \{(t,x) \in \mathbb{R}^{++} \times \mathbb{R}^{++} : x \le t\}$. (Once again we note that we are moving backwards and forwards between thinking of T as a metric space or as an equivalence class of metric spaces. As we have written things here, we are thinking of π_0 being associated with a particular class representative, but of course π_0 corresponds to a similar set for any representative of the same equivalence class by mapping across using the appropriate root invariant isometry.)

Assumption 3.1. *Suppose that the sets π_0 and π have the following properties.*

(a) *For all $t_0 > 0$, each of the sets $\pi_0 \cap (\{t_0\} \times T^o)$ and $\pi \cap (\{t_0\} \times]0, t_0])$ has at most one point and at least one of these sets is empty.*
(b) *For all $t_0 > 0$ and all finite subtrees $T' \subseteq T$, the set $\pi_0 \cap (]0, t_0] \times T')$ is finite.*
(c) *For all $t_0 > 0$, the set $\pi \cap \{(t, x) \in \mathbb{R}^{++} \times \mathbb{R}^{++} : x \le t \le t_0\}$ is finite.*

Remark Conditions (a)–(c) of Assumption 3.1 will hold almost surely if π_0 and π are realizations of Poisson point processes with respective intensities $\lambda \otimes \mu$ and $\lambda \otimes \lambda$ (where λ is Lebesgue measure), and it is this random mechanism that we will introduce later to produce a stochastic process having the root growth with re-grafting dynamics.

It will be convenient to use the notations π_0 and π to also refer to the integer-valued measures that are obtained by placing a unit point mass at each point of the corresponding set.

Consider a finite rooted subtree $T^* \preceq^{\mathrm{root}} T$. It will avoid annoying circumlocutions about equivalence via root invariant isometries if we work with particular class representatives for T^* and T, and, moreover, suppose that T^* is embedded in T.

Put $\tau_0^* := 0$, and let $0 < \tau_1^* < \tau_2^* < \ldots$ (the cut times for X^{T^*}) be the points of $\{t > 0 : \pi_0(\{t\} \times T^*) > 0\} \cup \{t > 0 : \pi(\{t\} \times \mathbb{R}^{++}) > 0\}$.

An explicit construction of $X_t^{T^*}$ is then given in two steps:

Step 1 (Root growth) At any time $t \ge 0$, $X_t^{T^*}$ as a set is given by the disjoint union $T^* \amalg]0, t]$. The root of $X_t^{T^*}$ is the point $\rho_t := t \in]0, t]$. The metric $d_t^{T^*}$ on $X_t^{T^*}$ is defined inductively as follows. Set $d_0^{T^*}$ to be the metric on $X_0^{T^*} = T^*$; that is, $d_0^{T^*}$ is the restriction of d to T^*. Suppose that $d_t^{T^*}$ has been defined for $0 \le t \le \tau_n^*$. Define $d_t^{T^*}$ for $\tau_n^* < t < \tau_{n+1}^*$ by

$$d_t^{T^*}(a, b) := \begin{cases} d_{\tau_n^*}(a, b), & \text{if } a, b \in X_{\tau_n^*}^{T^*}, \\ |b - a|, & \text{if } a, b \in]\tau_n^*, t], \\ |a - \tau_n^*| + d_{\tau_n^*}(\rho_{\tau_n^*}, b), & \text{if } a \in]\tau_n^*, t],\ b \in X_{\tau_n^*}^{T^*}. \end{cases} \tag{3.1}$$

Step 2 (Re-Grafting) Note that the left-limit $X_{\tau_{n+1}^*-}^{T^*}$ exists in the rooted Gromov-Hausdorff metric. As a set this left-limit is the disjoint union

$$X_{\tau_n^*}^{T^*} \amalg]\tau_n^*, \tau_{n+1}^*] = T^* \amalg]0, \tau_{n+1}^*], \tag{3.2}$$

and the corresponding metric $d_{\tau^*_{n+1}-}$ is given by a prescription similar to (3.1).
Define the $(n+1)^{\rm st}$ cut point for X^{T^*} by

$$p^*_{n+1} := \begin{cases} a \in T^*, & \text{if } \pi_0(\{(\tau^*_{n+1}, a)\}) > 0, \\ x \in]0, \tau^*_{n+1}], & \text{if } \pi(\{(\tau^*_{n+1}, x)\}) > 0. \end{cases} \tag{3.3}$$

Let S^*_{n+1} be the *subtree above* p^*_{n+1} in $X^{T^*}_{\tau^*_{n+1}-}$, that is,

$$S^*_{n+1} := \{b \in X^{T^*}_{\tau^*_{n+1}-} : p^*_{n+1} \in [\rho_{\tau^*_{n+1}-}, b[\,\}. \tag{3.4}$$

Define the metric $d_{\tau^*_{n+1}}$ by

(3.5)

$$d_{\tau^*_{n+1}}(a,b) := \begin{cases} d_{\tau^*_{n+1}-}(a,b), & \text{if } a, b \in S^*_{n+1}, \\ d_{\tau^*_{n+1}-}(a,b), & \text{if } a, b \in X^{T^*}_{\tau^*_{n+1}} \setminus S^*_{n+1}, \\ d_{\tau^*_{n+1}-}(a, \rho_{\tau^*_{n+1}}) + d_{\tau^*_{n+1}-}(p^*_{n+1}, b), & \text{if } a \in X^{T^*}_{\tau^*_{n+1}} \setminus S^*_{n+1},\ b \in S^*_{n+1}. \end{cases}$$

In other words $X^{T^*}_{\tau^*_{n+1}}$ is obtained from $X^{T^*}_{\tau^*_{n+1}-}$ by pruning off the subtree S^*_{n+1} and re-attaching it to the root.

Now consider two other finite, rooted subtrees (T^{**}, ρ) and (T^{***}, ρ) of T such that $T^* \cup T^{**} \subseteq T^{***}$ (with induced metrics). Build $X^{T^{**}}$ and $X^{T^{***}}$ from π_0 and π in the same manner as X^{T^*} (but starting at T^{**} and T^{***}). It is clear from the construction that:

- $X^{T^*}_t$ and $X^{T^{**}}_t$ are rooted subtrees of $X^{T^{***}}_t$ for all $t \geq 0$,
- the Hausdorff distance between $X^{T^*}_t$ and $X^{T^{**}}_t$ as subsets of $X^{T^{***}}_t$ does not depend on T^{***},
- the Hausdorff distance is constant between jumps of X^{T^*} and $X^{T^{**}}$ (when only root growth is occurring in both processes).

The following lemma shows that the Hausdorff distance between $X^{T^*}_t$ and $X^{T^{**}}_t$ as subsets of $X^{T^{***}}_t$ does not increase at jump times.

Lemma 3.2. *Let T be a finite rooted tree with root ρ and metric d, and let T' and T'' be two rooted subtrees of T (both with the induced metrics and root ρ). Fix $p \in T$, and let S be the subtree in T above p (recall (3.4)). Define a new metric $\hat{d}$ on T by putting*

$$\hat{d}(a,b) := \begin{cases} d(a,b), & \text{if } a, b \in S, \\ d(a,b), & \text{if } a, b \in T \setminus S, \\ d(a,p) + d(\rho, b), & \text{if } a \in S,\ b \in T \setminus S. \end{cases}$$

Then the sets T' and T'' are also subtrees of T equipped with the induced metric $\hat{d}$, and the Hausdorff distance between T' and T'' with respect to $\hat{d}$ is not greater than that with respect to d.

Proof. Suppose that the Hausdorff distance between T' and T'' under d is less than some given $\varepsilon > 0$. Given $a \in T'$, there then exists $b \in T''$ such that $d(a,b) < \varepsilon$. Because $d(a, a \wedge b) \leq d(a,b)$ and $a \wedge b \in T''$, we may suppose (by replacing b by $a \wedge b$ if necessary) that $b \leq a$. We claim that $\hat{d}(a,c) < \varepsilon$ for some $c \in T''$. This and the analogous result with the roles of T' and T'' interchanged will establish the result.

If $a, b \in S$ or $a, b \in T \setminus S$, then $\hat{d}(a,b) = d(a,b) < \varepsilon$. The only other possibility is that $a \in S$ and $b \in T \setminus S$, in which case $p \in [b,a]$ (for T equipped with d). Then $\hat{d}(a,\rho) = d(a,p) \leq d(a,b) < \varepsilon$, as required (because $\rho \in T''$). □

Now let $T_1 \subseteq T_2 \subseteq \cdots$ be an increasing sequence of finite subtrees of T such that $\bigcup_{n\in\mathbb{N}} T_n$ is dense in T. Thus $\lim_{n\to\infty} d_{\mathrm{H}}(T_n, T) = 0$. Let $X^1, X^2, \ldots$ be constructed from π_0 and π starting with $T_1, T_2, \ldots$. Applying Lemma 3.2 yields

$$\lim_{m,n\to\infty} \sup_{t\geq 0} d_{\mathrm{GH}^{\mathrm{root}}}(X_t^m, X_t^n) = 0. \tag{3.6}$$

Hence by completeness of $\mathbf{T}^{\mathrm{root}}$, there exists a càdlàg $\mathbf{T}^{\mathrm{root}}$-valued process X such that $X_0 = T$ and

$$\lim_{m\to\infty} \sup_{t\geq 0} d_{\mathrm{GH}^{\mathrm{root}}}(X_t^m, X_t) = 0. \tag{3.7}$$

A priori, the process X could depend on the choice of the approximating sequence of trees $(T_n)_{n\in\mathbb{N}}$. To see that this is not so, consider two approximating sequences $T_1^1 \subseteq T_2^1 \subseteq \cdots$ and $T_1^2 \subseteq T_2^2 \subseteq \cdots$. For $k \in \mathbb{N}$, write T_n^3 for the smallest rooted subtree of T that contains both T_n^1 and T_n^2. As a set, $T_n^3 = T_n^1 \cup T_n^2$. Now let $((X_t^{n,i})_{t\geq 0})_{n\in\mathbb{N}}$ for $i = 1,2,3$ be the corresponding sequences of finite tree-value processes and let $(X_t^{\infty,i})_{t\geq 0}$ for $i = 1,2,3$ be the corresponding limit processes. By Lemma 3.2,

$$\begin{aligned}
d_{\mathrm{GH}^{\mathrm{root}}}(X_t^{n,1}, X_t^{n,2}) &\leq d_{\mathrm{GH}^{\mathrm{root}}}(X_t^{n,1}, X_t^{n,3}) + d_{\mathrm{GH}^{\mathrm{root}}}(X_t^{n,2}, X_t^{n,3}) \\
&\leq d_{\mathrm{H}}(X_t^{n,1}, X_t^{n,3}) + d_{\mathrm{H}}(X_t^{n,2}, X_t^{n,3}) \\
&\leq d_{\mathrm{H}}(T_n^1, T_n^3) + d_{\mathrm{H}}(T_n^2, T_n^3) \\
&\leq d_{\mathrm{H}}(T_n^1, T) + d_{\mathrm{H}}(T_n^2, T) \to 0
\end{aligned} \tag{3.8}$$

as $n \to \infty$. Thus, for each $t \geq 0$ the sequences $(X_t^{n,1})_{n\in\mathbb{N}}$ and $(X_t^{n,2})_{n\in\mathbb{N}}$ do indeed have the same rooted Gromov-Hausdorff limit and the process X does not depend on the choice of approximating sequence for the initial tree T.

3.2. **Finishing the construction.** In 3.1 we constructed a $\mathbf{T}^{\mathrm{root}}$-valued function $t \mapsto X_t$ starting with a fixed triple (T, π_0, π), where $T \in \mathbf{T}^{\mathrm{root}}$ and π_0, π satisfy the conditions of Assumption 3.1. We now want to think of X as a function of time and such triples.

Let Ω^* be the set of triples (T, π_0, π), where T is a rooted compact $\mathbb{R}$-tree (that is, a class representative of an element of $\mathbf{T}^{\mathrm{root}}$) and π_0, π satisfy Assumption 3.1.

The root invariant isometry equivalence relation on rooted compact $\mathbb{R}$-trees extends naturally to an equivalence relation on Ω^* by declaring that two triples (T', π_0', π') and (T'', π_0'', π''), where $\pi_0' = \{(\sigma_i', x_i') : i \in \mathbb{N}\}$ and $\pi_0'' = \{(\sigma_i'', x_i'') : i \in \mathbb{N}\}$, are equivalent if there is a root invariant isometry f mapping T' to T'' and a permutation γ of $\mathbb{N}$ such that $\sigma_i'' = \sigma_{\gamma(i)}'$ and $x_i'' = f(x_{\gamma(i)}')$ for all $i \in \mathbb{N}$. We write Ω for the resulting quotient space of equivalence classes.

In order to do probability, we require that Ω has a suitable measurable structure. We could do this by specifying a metric on Ω, but the following approach is a little less cumbersome and suffices for our needs.

Let Ω^{fin} denote the subset of Ω consisting of triples (T, π_0, π) such that T, π_0 and π are finite. We are going to define a metric on Ω^{fin}. Let (T', π_0', π') and (T'', π_0'', π'') be two points in Ω^{fin}, where $\pi_0' = \{(\sigma_1', x_1'), \ldots, (\sigma_p', x_p')\}$, $\pi' = \{\tau_1', \ldots, \tau_r'\}$, $\pi_0'' = \{(\sigma_1'', x_1''), \ldots, (\sigma_q'', x_q'')\}$, and $\pi'' = \{\tau_1'', \ldots, \tau_s''\}$. Assume that $0 < \sigma_1' < \cdots < \sigma_p'$, $0 < \tau_1' < \cdots < \tau_r'$, $0 < \sigma_1'' < \cdots < \sigma_q''$, and $0 < \tau_1'' < \cdots < \tau_s''$. The distance between (T', π_0', π') and (T'', π_0'', π'') will be 1 if either $p \ne q$ or $r \ne s$. Otherwise, the distance is

$$1 \wedge \left(\frac{1}{2} \inf_{\mathfrak{R}^{\mathrm{root,cuts}}} \mathrm{dis}(\mathfrak{R}^{\mathrm{root,cuts}}) + \max_i |\sigma_i' - \sigma_i''| + \max_j |\tau_j' - \tau_j''| \right), \tag{3.9}$$

where the infimum is over all correspondences between T' and T'' that contain the pairs $(\rho_{T'}, \rho_{T''})$ and (x_i', x_i'') for $1 \le i \le p$.

Equip Ω^{fin} with the Borel σ-field corresponding to this metric. For $t \ge 0$, let $\mathcal{F}_t^o$ be the σ-field on Ω generated by the family of maps from Ω into Ω^{fin} given by $(T, \pi_0, \pi) \mapsto (R_\eta(T), \pi_0 \cap (]0,t] \times (R_\eta(T))^o), \pi \cap \{(s,x) : x \le s \le t\})$ for $\eta > 0$. As usual, set $\mathcal{F}_t^+ := \bigcap_{u>t} \mathcal{F}_u^o$ for $t \ge 0$. Put $\mathcal{F}^o := \bigvee_{t \ge 0} \mathcal{F}_t^o$.

It is straightforward to establish the following result from Lemma 2.6 and the construction of X in Subsection 3.1, and we omit the proof.

Lemma 3.3. *The map $(t, (T, \pi_0, \pi)) \mapsto X_t(T, \pi_0, \pi)$ from $\mathbb{R}^+ \times \Omega$ into $\mathbf{T}^{\mathrm{root}}$ is progressively measurable with respect to the filtration $(\mathcal{F}_t^o)_{t \ge 0}$. (Here, of course, we are equipping $\mathbf{T}^{\mathrm{root}}$ with the Borel σ-field associated with the metric $d_{\mathrm{GH^{root}}}$.)*

Given $T \in \mathbf{T}^{\mathrm{root}}$, let $\mathbf{P}^T$ be the probability measure on Ω defined by the following requirements.

- The measure $\mathbf{P}^T$ assigns all of its mass to the set $\{(T', \pi_0', \pi') \in \Omega : T' = T\}$.
- Under $\mathbf{P}^T$, the random variable $(T', \pi_0', \pi') \mapsto \pi_0'$ is a Poisson point process on the set $\mathbb{R}^{++} \times T^o$ with intensity $\lambda \otimes \mu$, where μ is the length measure on T.

- Under $\mathbf{P}^T$, the random variable $(T', \pi'_0, \pi') \mapsto \pi'$ is a Poisson point process on the set $\{(t, x) \in \mathbb{R}^{++} \times \mathbb{R}^{++} : x \leq t\}$ with intensity $\lambda \otimes \lambda$ restricted to this set.
- The random variables $(T', \pi'_0, \pi') \mapsto \pi'_0$ and $(T', \pi'_0, \pi') \mapsto \pi'$ are independent under $\mathbf{P}^T$.

Of course, the random variable $(T', \pi'_0, \pi') \mapsto \pi'_0$ takes values in a space of equivalence classes of countable sets rather than a space of sets *per se*, so, more formally, this random variable has the law of the image of a Poisson process on an arbitrary class representative under the appropriate quotient map.

For $t \geq 0$, g a bounded Borel function on $\mathbf{T}^{\mathrm{root}}$, and $T \in \mathbf{T}^{\mathrm{root}}$, set

$$P_t g(T) := \mathbf{P}^T[g(X_t)]. \tag{3.10}$$

With a slight abuse of notation, let $\tilde{R}_\eta$ for $\eta > 0$ also denote the map from Ω into Ω that sends (T, π_0, π) to $(R_\eta(T), \pi_0 \cap (\mathbb{R}^{++} \times (R_\eta(T))^o), \pi)$.

Our main construction result is the following.

Theorem 3. (i) *If $T \in \mathbf{T}^{\mathrm{root}}$ is finite, then $(X_t)_{t \geq 0}$ under $\mathbf{P}^T$ is a Markov process that evolves via the root growth with re-grafting dynamics on finite trees.*

(ii) *For all $\eta > 0$ and $T \in \mathbf{T}^{\mathrm{root}}$, the law of $(X_t \circ \tilde{R}_\eta)_{t \geq 0}$ under $\mathbf{P}^T$ coincides with the law of $(X_t)_{t \geq 0}$ under $\mathbf{P}^{R_\eta(T)}$.*

(iii) *For all $T \in \mathbf{T}^{\mathrm{root}}$, the law of $(X_t)_{t \geq 0}$ under $\mathbf{P}^{R_\eta(T)}$ converges as $\eta \downarrow 0$ to that of $(X_t)_{t \geq 0}$ under $\mathbf{P}^T$ (in the sense of convergence of laws on the space of càdlàg $\mathbf{T}^{\mathrm{root}}$-valued paths equipped with the Skorohod topology).*

(iv) *For $g \in \mathrm{b}\mathcal{B}(\mathbf{T}^{\mathrm{root}})$, the map $(t, T) \mapsto P_t g(T)$ is $\mathcal{B}(\mathbb{R}^+) \times \mathcal{B}(\mathbf{T}^{\mathrm{root}})$-measurable.*

(v) *The process $(X_t, \mathbf{P}^T)$ is strong Markov with respect to the filtration $(\mathcal{F}_t^+)_{t \geq 0}$ and has transition semigroup $(P_t)_{t \geq 0}$.*

Proof. (i) This is clear from the definition of the root growth and re-grafting dynamics.

(ii) It is enough to check that the push-forward of the probability measure $\mathbf{P}^T$ under the map $R_\eta : \Omega \to \Omega$ is the measure $\mathbf{P}^{R_\eta(T)}$. This, however, follows from the observation that the restriction of length measure on a tree to a subtree is just length measure on the subtree.

(iii) This is immediate from part (ii), the limiting construction in Subsection 3.1, and part (iv) of Lemma 2.6. Indeed, we have that

$$\sup_{t \geq 0} d_{\mathrm{GH}^{\mathrm{root}}}(X_t, X_t \circ \tilde{R}_\eta) \leq d_{\mathrm{H}}(T, R_\eta(T)) \leq \eta. \tag{3.11}$$

(iv) By a monotone class argument, it is enough to consider the case where the test function g is continuous. It follows from part (iii) that $P_t g(R_\eta(T))$

converges pointwise to $P_t g(T)$ as $\eta \downarrow 0$, and it is not difficult to show using Lemma 2.6 and part (i) that $(t,T) \mapsto P_t g(R_\eta(T))$ is $\mathcal{B}(\mathbb{R}^+) \times \mathcal{B}(\mathbf{T}^{\mathrm{root}})$-measurable. We omit the details, because we will establish an even stronger result in Proposition 6.1.

(v) By construction and part (ii) of Lemma 2.7, we have for $t \geq 0$ and $(T, \pi_0, \pi) \in \Omega$ that, as a set, $X_t^o(T, \pi_0, \pi)$ is the disjoint union $T^o \amalg]0,t]$.

Put

$$\begin{aligned}
&\theta_t(T,\pi_0,\pi) \\
&\quad := \Big(X_t(T,\pi_0,\pi), \{(s,x) \in \mathbb{R}^{++} \times T^o : (t+s,x) \in \pi_0\}, \\
&\qquad \{(s,x) \in \mathbb{R}^{++} \times \mathbb{R}^{++} : (t+s, t+x) \in \pi\}\Big) \\
&\quad = \Big(X_t(T,\pi_0,\pi), \{(s,x) \in \mathbb{R}^{++} \times X_t^o(T,\pi_0,\pi) : (t+s,x) \in \pi_0\}, \\
&\qquad \{(s,x) \in \mathbb{R}^{++} \times \mathbb{R}^{++} : (t+s, t+x) \in \pi\}\Big).
\end{aligned} \tag{3.12}$$

Thus θ_t maps Ω into Ω. Note that $X_s \circ \theta_t = X_{s+t}$ and that $\theta_s \circ \theta_t = \theta_{s+t}$, that is, the family $(\theta_t)_{t\geq 0}$ is a semigroup. It is not hard to show that $(t, (T,\pi_0,\pi)) \mapsto \theta_t(T,\pi_0,\pi)$ is jointly measurable, and we leave this to the reader.

Fix $t \geq 0$ and $(T,\pi_0,\pi) \in \Omega$. Write μ' for the measure on $T^o \amalg]0,t]$ that restricts to length measure on T^o and to Lebesgue measure on $]0,t]$. Write μ'' for the length measure on $X_t^o(T,\pi_0,\pi)$. The strong Markov property will follow from a standard strong Markov property for Poisson processes if we can show that $\mu' = \mu''$. This equality is clear from the construction if T is finite: the tree $X_t(T,\pi_0,\pi)$ is produced from the tree T and the set $]0,t]$ by a finite number of dissections and rearrangements. The equality for general T follows from the construction and part (iii) of Lemma 2.7. □

4. Connection with Aldous's construction of the CRT

Let $(\mathcal{R}_t)_{t\geq 0}$, $(\mathcal{T}_t)_{t\geq 0}$, and $(\tau_n)_{n\in\mathbb{N}}$ be as in the Introduction. Thus $(\mathcal{T}_t)_{t\geq 0}$ has the same law as $(X_t)_{t\geq 0}$ under $\mathbf{P}^{T_0}$, where T_0 is the trivial tree.

Proposition 4.1. *The two random finite rooted trees $\mathcal{R}_{\tau_n-}$ and $\mathcal{T}_{\tau_n-}$ have the same distribution for all $n \in \mathbb{N}$.*

Proof. Let R_n denote the object obtained by taking the rooted finite tree with edge-lengths $\mathcal{R}_{\tau_n-}$ and labeling the leaves with $1,\ldots,n$, in the order that they are added in Aldous's construction. Let T_n be derived similarly from the rooted finite tree with edge-lengths $\mathcal{T}_{\tau_n-}$, by labeling the leaves with $1,\ldots,n$ in the order that they appear in the root growth with re-grafting construction. It will suffice to show that R_n and T_n have the same distribution. Note that both R_n and T_n are rooted, bifurcating trees with n labeled leaves and edge-lengths. Such a tree S_n is uniquely specified by its *shape*, denoted shape(S_n), which is a rooted, bifurcating, leaf-labeled combinatorial tree, and by the list of its $(2n-1)$ edge-lengths in a canonical order determined by its shape, say

$$\mathrm{lengths}(S_n) := (\mathrm{length}(S_n, 1), \dots, \mathrm{length}(S_n, 2n-1)),$$

where the edge-lengths are listed in order of traversal of edges by first working along the path from the root to leaf 1, then along the path joining that path to leaf 2, and so on.

Recall that τ_n is the nth point of a Poisson process on $\mathbb{R}^{++}$ with rate $t\,dt$. We construct R_n and T_n on the same probability space using cuts at points $U_i\tau_i$, $1 \le i \le n-1$, where $U_1, U_2, \dots$ is a sequence of independent random variables uniformly distributed on the interval $]0,1]$ and independent of the sequence $(\tau_n)_{n\in\mathbb{N}}$. Then, by construction, the common collection of edge-lengths of R_n and of T_n is the collection of lengths of the $2n-1$ subintervals of $]0,\tau_n]$ obtained by cutting this interval at the $2n-2$ points

$$\{X_i^{(n)}, 1 \le i \le 2n-2\} := \bigcup_{i=1}^{n-1} \{U_i\tau_i, \tau_i\}$$

where the $X_i^{(n)}$ are indexed to increase in i for each fixed n. Let $X_0^{(n)} := 0$ and $X_{2n-1}^{(n)} := \tau_n$. Then

$$\mathrm{length}(R_n, i) = X_i^{(n)} - X_{i-1}^{(n)}, \quad 1 \le i \le 2n-1, \tag{4.1}$$

$$\mathrm{length}(T_n, i) = \mathrm{length}(R_n, \sigma_{n,i}), \quad 1 \le i \le 2n-1, \tag{4.2}$$

for some almost surely unique random indices $\sigma_{n,i} \in \{1, \dots 2n-1\}$ such that $i \mapsto \sigma_{n,i}$ is almost surely a permutation of $\{1, \dots 2n-1\}$. According to [Ald93, Lemma 21], the distribution of R_n may be characterized as follows:

(i) the sequence $\mathrm{lengths}(R_n)$ is exchangeable, with the same distribution as the sequence of lengths of subintervals obtained by cutting $]0,\tau_n]$ at $2n-2$ uniformly chosen points $\{U_i\tau_n : 1 \le i \le 2n-2\}$;

(ii) $\mathrm{shape}(R_n)$ is uniformly distributed on the set of all $1 \times 3 \times 5 \times \cdots \times (2n-3)$ possible shapes;

(iii) $\mathrm{lengths}(R_n)$ and $\mathrm{shape}(R_n)$ are independent.

In view of this characterization and (4.2), to show that T_n has the same distribution as R_n it is enough to show that

(a) the random permutation $\{i \mapsto \sigma_{n,i} : 1 \le i \le 2n-1\}$ is a function of $\mathrm{shape}(T_n)$;

(b) $\mathrm{shape}(T_n) = \Psi_n(\mathrm{shape}(R_n))$ for some bijective map Ψ_n from the set of all possible shapes to itself.

This is trivial for $n = 1$, so we assume below that $n \ge 2$. Before proving (a) and (b), we recall that (ii) above involves a natural bijection

$$(I_1, \dots, I_{n-1}) \leftrightarrow \mathrm{shape}(R_n) \tag{4.3}$$

where $I_{n-1} \in \{1, \ldots, 2n-3\}$ is the unique i such that $U_{n-1}\tau_{n-1} \in (X_{i-1}^{(n-1)}, X_i^{(n-1)})$. Hence I_{n-1} is the index in the canonical ordering of edges of R_{n-1} of the edge that is cut in the transformation from R_{n-1} to R_n by attachment of an additional edge, of length $\tau_n - \tau_{n-1}$, connecting the cut-point to leaf n. Thus (ii) and (iii) above correspond via (4.3) to the facts that $I_1, \ldots, I_{n-1}$ are independent and uniformly distributed over their ranges, and independent of lengths(R_n). These facts can be checked directly from the construction of $(R_n)_{n\in\mathbb{N}}$ from $(\tau_n)_{n\in\mathbb{N}}$ and $(U_n)_{n\in\mathbb{N}}$ using standard facts about uniform order statistics.

Now (a) and (b) follow from (4.3) and another bijection

$$(I_1, \ldots, I_{n-1}) \leftrightarrow \text{shape}(T_n) \tag{4.4}$$

where each possible value i of I_m is identified with edge $\sigma_{m,i}$ in the canonical ordering of edges of T_m. This is the edge of T_m whose length equals length(R_m, i). The bijection (4.4), and the fact that $\sigma_{n,i}$ depends only on shape(T_n), will now be established by induction on $n \geq 2$. For $n = 2$ the claim is obvious. Suppose for some $n \geq 3$ that the correspondence between $(I_1, \ldots, I_{n-2})$ and shape(T_{n-1}) has been established, and that the length of edge $\sigma_{n-1,i}$ in the canonical ordering of edges of T_{n-1} is equals the length of the ith edge in the canonical ordering of edges of R_{n-1}, for some $\sigma_{n-1,i}$ which is a function of i and shape(T_{n-1}). According to the construction of T_n, if $I_{n-1} = i$ then T_n is derived from T_{n-1} by splitting T_{n-1} into two branches at some point along edge $\sigma_{n-1,i}$ in the canonical ordering of the edges of T_{n-1}, and forming a new tree from the two branches and an extra segment of length $\tau_n - \tau_{n-1}$. Clearly, shape(T_n) is determined by shape(T_{n-1}) and I_{n-1}, and in the canonical ordering of the edge-lengths of T_n the length of the ith edge equals the length of the edge $\sigma_{n,i}$ of R_n, for some $\sigma_{n,i}$ which is a function of shape(T_{n-1}) and I_{n-1}, and hence a function of shape(T_n). To complete the proof, it is enough by the inductive hypothesis to show that the map

$$(\text{shape}(T_{n-1}), I_{n-1}) \to \text{shape}(T_n)$$

just described is invertible. But shape(T_{n-1}) and I_{n-1} can be recovered from shape(T_n) by the following sequence of moves:

- delete the edge attached to the root of shape(T_n)
- split the remaining tree into its two branches leading away from the internal node to which the deleted edge was attached;
- re-attach the bottom end of the branch not containing leaf n to leaf n on the other branch, joining the two incident edges to form a single edge;
- the resulting shape is shape(T_{n-1}), and I_{n-1} is the index such that the joined edge in shape(T_{n-1}) is the edge $\sigma_{n-1,I_{n-1}}$ in the canonical ordering of edges on shape(T_{n-1}).

$\square$

5. Recurrence and convergence to stationarity

Lemma 5.1. *For any $(T, d, \rho) \in \mathbf{T}^{\mathrm{root}}$ we can build on the same probability space two $\mathbf{T}^{\mathrm{root}}$-valued processes X' and X'' such that:*

- *X' has the law of X under $\mathbf{P}^{T_0}$, where T_0 is the trivial tree consisting of just the root,*
- *X'' has the law of X under $\mathbf{P}^{T}$,*
- *for all $t \geq 0$,*

$$d_{\mathrm{GH}^{\mathrm{root}}}(X'_t, X''_t) \leq d_{\mathrm{GH}^{\mathrm{root}}}(T_0, T) = \sup\{d(\rho, x) : x \in T\} \tag{5.1}$$

-

$$\lim_{t \to \infty} d_{\mathrm{GH}^{\mathrm{root}}}(X'_t, X''_t) = 0, \quad \textit{almost surely.} \tag{5.2}$$

Proof. The proof follows almost immediately from construction of X in Section 3 and Lemma 3.2. The only point requiring some comment is (5.2). For that it will be enough to show for any $\varepsilon > 0$ that for $\mathbf{P}^T$-a.e. $(T, \pi_0, \pi) \in \Omega$ there exists $t > 0$ such that the projection of $\pi_0 \cap (]0, t] \times T^o)$ onto T is an ε-net for T.

Note that the projection of $\pi_0 \cap (]0, t] \times T^o)$ onto T is a Poisson process under $\mathbf{P}^T$ with intensity $t\mu$, where μ is the length measure on T. Moreover, T can be covered by a finite collection of ε-balls, each with positive μ-measure. Therefore, the $\mathbf{P}^T$-probability of the set of $(T, \pi_0, \pi) \in \Omega$ such that the projection of $\pi_0 \cap (]0, t] \times T^o)$ onto T is an ε-net for T increases as $t \to \infty$ to 1. □

Proposition 5.2. *For any $T \in \mathbf{T}^{\mathrm{root}}$, the law of X_t under $\mathbf{P}^T$ converges weakly to that of the Brownian CRT as $t \to \infty$.*

Proof. It suffices by Lemma 5.1 to consider the case where T is the trivial tree.

We saw in the Proposition 4.1 that, in the notation of the Introduction, $\mathcal{T}_{\tau_n -}$ has the same distribution as $\mathcal{R}_{\tau_n -}$. Moreover, we recalled in the Introduction that $\mathcal{R}_t$ converges in distribution to the continuum random tree as $t \to \infty$ if we use Aldous's metric on trees that comes from thinking of them as closed subsets of ℓ^1 with the root at the origin and equipped with the Hausdorff distance. By construction, $(\mathcal{T}_t)_{t \geq 0}$ has the root growth with re-grafting dynamics started at the trivial tree. Clearly, the rooted Gromov–Hausdorff distance between $\mathcal{T}_t$ and $\mathcal{T}_{\tau_{n+1} -}$ is at most $\tau_{n+1} - \tau_n$ for $\tau_n \leq t < \tau_{n+1}$. It remains to observe that $\tau_{n+1} - \tau_n \to 0$ in probability as $n \to \infty$. □

Proposition 5.3. *Consider a non-empty open set $U \subseteq \mathbf{T}^{\mathrm{root}}$. For each $T \in \mathbf{T}^{\mathrm{root}}$,*

$$\mathbf{P}^T\{\textit{for all } s \geq 0, \textit{ there exists } t > s \textit{ such that } X_t \in U\} = 1. \tag{5.3}$$

Proof. It is straightforward, but notationally rather tedious, to show that if $B' \subseteq \mathbf{T}^{\mathrm{root}}$ is any ball and T_0 is the trivial tree, then

$$\mathbf{P}^{T_0}\{X_t \in B'\} > 0 \tag{5.4}$$

for all t sufficiently large. Thus, for any ball $B' \subseteq \mathbf{T}^{\mathrm{root}}$ there is, by Lemma 5.1, a ball $B'' \subseteq \mathbf{T}^{\mathrm{root}}$ containing the trivial tree such that

$$\inf_{T \in B''} \mathbf{P}^{T}\{X_t \in B'\} > 0 \tag{5.5}$$

for each t sufficiently large.

By a standard application of the Markov property, it therefore suffices to show for each $T \in \mathbf{T}^{\mathrm{root}}$ and each ball B'' around the trivial tree that

$$\mathbf{P}^{T}\{\text{there exists } t > 0 \text{ such that } X_t \in B''\} = 1. \tag{5.6}$$

By another standard application of the Markov property, equation (5.6) will follow if we can show that there is a constant $p > 0$ depending on B'' such that for any $T \in \mathbf{T}^{\mathrm{root}}$

$$\liminf_{t \to \infty} \mathbf{P}^{T}\{X_t \in B''\} > p.$$

This, however, follows from Proposition 5.2 and the observation that for any $\varepsilon > 0$ the law of the Brownian CRT assigns positive mass to the set of trees with height less than ε (which is just the observation that the law of the Brownian excursion assigns positive mass to the set of excursion paths with maximum less that $\varepsilon/2$). □

Proposition 5.4. *The law of the Brownian CRT is the unique stationary distribution for X. That is, if ξ is the law of the CRT, then $\int \xi(dT)P_t f(T) = \int \xi(dT) f(T)$ for all $t \geq 0$ and $f \in \mathrm{b}\mathcal{B}(\mathbf{T}^{\mathrm{root}})$, and ξ is the unique probability measure on $\mathbf{T}^{\mathrm{root}}$ with this property.*

Proof. This is a standard argument given Proposition 5.2 and the Feller property for the semigroup $(P_t)_{t \geq 0}$ established in Proposition 6.1, but we include the details for completeness.

Consider a test function $f : \mathbf{T}^{\mathrm{root}} \to \mathbb{R}$ that is continuous and bounded. By Proposition 6.1 below, the function $P_t f$ is also continuous and bounded for each $t \geq 0$. Therefore, by Proposition 5.2,

$$\begin{aligned}\int \xi(dT) f(T) &= \lim_{s \to \infty} \int \xi(dT) P_s f(T) = \lim_{s \to \infty} \int \xi(dT) P_{s+t} f(T) \\ &= \lim_{s \to \infty} \int \xi(dT) P_s (P_t f)(T) = \int \xi(dT) P_t f(T)\end{aligned} \tag{5.7}$$

for each $t \geq 0$, and hence ξ is stationary. Moreover, if ζ is a stationary measure, then

$$\begin{aligned}\int \zeta(dT) f(T) &= \int \zeta(dT) P_t f(T) \\ &\to \int \zeta(dT) \left(\int \xi(dT) f(T) \right) = \int \xi(dT) f(T),\end{aligned} \tag{5.8}$$

and $\zeta = \xi$, as claimed. □

6. Feller property

The following result says that the law of X_t under $\mathbf{P}^T$ is weakly continuous in the initial value T for each $t \geq 0$. This property is sometimes referred to as the Feller property of the semigroup $(P_t)_{t\geq 0}$, although this terminology is often restricted to the case of a locally compact state space and transition operators that map the space of continuous functions that vanish at infinity into itself. A standard consequence of this result is that the law of the process $(X_t)_{t\geq 0}$ is weakly continuous in the initial value (when the space of càdlàg $\mathbf{T}^{\mathrm{root}}$-valued paths is equipped with the Skorohod topology).

Proposition 6.1. *If the function $f : \mathbf{T}^{\mathrm{root}} \to \mathbb{R}$ is continuous and bounded, then the function $P_t f$ is also continuous and bounded for each $t \geq 0$.*

We will prove the proposition by a coupling argument that, *inter alia*, builds processes with the law of X under $\mathbf{P}^T$ for two different finite values of T on the same probability space. The key to constructing such a coupling is the following pair of lemmas.

We require the following notion. A *rooted combinatorial tree* is just a connected, acyclic graph with one vertex designated as the root. Equivalently, we can think of a rooted combinatorial tree as a finite rooted tree in which all edges have length one. Thus any finite rooted tree is associated with a unique rooted combinatorial tree by changing all the edge lengths to one, and any two finite rooted trees with the same topology are associated with the same rooted combinatorial tree. If U and V are two rooted combinatorial trees with leaves labeled by $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$, then we say that U and V are isomorphic if there exists a graph isomorphism between U and V that maps the root of U to the root of V and x_i to y_i for $1 \leq i \leq n$.

Lemma 6.2. *Let (T, ρ) be a finite rooted trees with leaves $\{x_1, \ldots, x_n\} = T \setminus T^o$. (recall the definition of the skeleton T^o from (2.16)). Write η for the minimum of the (strictly positive) edge lengths in T. Suppose that (T', ρ') is another finite rooted tree with $d_{\mathrm{GH}^{\mathrm{root}}}((T', \rho'), (T, \rho)) < \delta < \frac{\eta}{16}$. Then there exists a subtree $(T'', \rho') \preceq^{\mathrm{root}} (T', \rho')$ and a map $\bar{f} : T \to T''$ such that:*

(i) *$\bar{f}(\rho) = \rho'$,*
(ii) *T'' is spanned by $\{\bar{f}(x_1), \ldots, \bar{f}(x_n), \rho'\}$,*
(iii) *$d_{\mathrm{H}}(T', T'') < 3\delta$,*
(iv) *$\mathrm{dis}(\bar{f}) < 8\delta$,*
(v) *T'' has leaves $\{\bar{f}(x_1), \ldots, \bar{f}(x_n)\}$,*
(vi) *by possibly deleting some internal edges from the rooted combinatorial tree associated to T'' with leaves labeled by $(\bar{f}(x_1), \ldots, \bar{f}(x_n))$, one can obtain a leaf-labeled rooted combinatorial tree that is isomorphic to the rooted combinatorial rooted tree associated to T with leaves labeled by $(x_1, \ldots, x_n)$.*

Proof. We have from (2.10) that there is a correspondence $\mathfrak{R}^{\mathrm{root}}$ containing (ρ, ρ') between T and T' such that $\mathrm{dis}(\mathfrak{R}^{\mathrm{root}}) < 2\delta$. For $x \in T \setminus \{\rho\}$, choose $f(x) \in T'$ such that $(x, f(x)) \in \mathfrak{R}^{\mathrm{root}}$, and put $f(\rho) := \rho'$. Set T'' to be the subtree of T' spanned by $\{f(x_1), \ldots, f(x_n), \rho'\}$. For $x \in T$ define $\bar{f}(x) \in T''$ to be the point in T'' that has minimum distance to $f(x)$. In particular, $\bar{f}(\rho) = f(\rho) = \rho'$ and $\bar{f}(x_i) = f(x_i)$ for all i, so that (i) and (ii) hold.

For $x' \in T' \setminus \{\rho'\}$ choose $g(x')$ such that $(g(x'), x') \in \mathfrak{R}^{\mathrm{root}}$ and put $g(\rho') := \rho$. Then $f(T)$ and $g(T')$ are 2δ-nets for T' and T, respectively, and $\mathrm{dis}(f) \vee \mathrm{dis}(g) < 2\delta$. For each $i \in \{1, \ldots, n\}$ and $y' \in T'$ we have $(\rho, \rho'), (x_i, f(x_i)), (g(y'), y') \in \mathfrak{R}^{\mathrm{root}}$. Hence $d_{T'}(\rho', y') < d_T(\rho, g(y')) + 2\delta$, and $d_{T'}(y', f(x_i)) < d_T(g(y'), x_i) + 2\delta$. Now fix $y' \in T'$, and choose $i \in \{1, \ldots, n\}$ such that $g(y') \in [\rho, x_i]$. Then

$$
\begin{aligned}
d_{T'}(\rho', f(x_i)) + 2d_{\mathrm{H}}\left(\{y'\}, [\rho', f(x_i)]\right) &= d_{T'}(\rho', y') + d_{T'}(y', f(x_i)) \\
&< d_T(\rho, x_i) + 4\delta \\
&< d_{T'}(\rho', f(x_i)) + 2\delta + 4\delta,
\end{aligned}
\tag{6.1}
$$

and hence $d_{\mathrm{H}}(\{y'\}, T'') < 3\delta$. Thus (iii) holds.

For $x, y \in T$,

$$
\begin{aligned}
&|d_T(x, y) - d_{T''}(\bar{f}(x), \bar{f}(y))| \\
&\quad \le |d_T(x, y) - d_{T'}(f(x), f(y))| + d_{T'}(\bar{f}(x), f(x)) + d_{T'}(\bar{f}(y), f(y)) \\
&\quad \le \mathrm{dis}(f) + 2d_H(T', T'') < 8\delta,
\end{aligned}
\tag{6.2}
$$

and (iv) holds.

In order to establish (v), it suffices to observe for $1 \le i \ne j \le n$ that, by part (iv),

$$
\begin{aligned}
&d_{T''}(\bar{f}(x_i), \bar{f}(x_j)) + d_{T''}(\bar{f}(x_j), \rho') - d_{T''}(\bar{f}(x_i), \rho') \\
&\quad \ge d_T(x_i, x_j) + d_T(x_j, \rho) - d_T(x_i, \rho) - 3\mathrm{dis}(\bar{f}) \\
&\quad > 2\eta - 24\delta \\
&\quad > 0.
\end{aligned}
\tag{6.3}
$$

Similarly, part (vi) follows from part (iv) and the observations in Subsection 2.2 about re-constructing tree shapes from distances between the points in subsets of size four drawn from the leaves and the root of T'' once we observe the inequality $\frac{1}{2}4\mathrm{dis}(\bar{f}) < 16\delta < \eta$. □

Lemma 6.3. *Let (T, ρ) be a finite rooted tree and $\varepsilon > 0$. There exists $\delta > 0$ depending on T and ε such that if (T', ρ') is a finite rooted tree with $d_{\mathrm{GH}^{\mathrm{root}}}((T', \rho'), (T, \rho)) < \delta$, then there exist subtrees $(S, \rho) \preceq^{\mathrm{root}} T$ and $(S', \rho') \preceq^{\mathrm{root}} T'$ for which:*

(i) *$d_{\mathrm{H}}(S, T) < \varepsilon$ and $d_{\mathrm{H}}(S', T') < \varepsilon$,*
(ii) *S and S' have the same total length,*
(iii) *there is a bijective measurable map $\psi : S \to S'$ that preserves length measure and has distortion at most ε,*

(iv) *the length measure of the set of points $a \in S$ such that $\{b' \in S' : \psi(a) \le b'\} \neq \psi(\{b \in S : a \le b\})$ (that is, the set of points a such that the subtree above $\psi(a)$ is not the image under ψ of the subtree above a) is less than ε.*

Proof. As in Lemma 6.2, denote by η the minimum of the (strictly positive) edge lengths of T. Let (T', ρ') be a finite rooted tree with

$$d_{\mathrm{GH}^{\mathrm{root}}}((T', \rho'), (T, \rho)) < \delta < \frac{\eta}{16}, \tag{6.4}$$

where δ depending on T and ε will be chosen later. Set (T'', ρ') and $\bar{f}$ to be a subtree of T' and a function from T to T'' whose existence is guaranteed by Lemma 6.2 for this choice of δ. Let $\{x_1, \dots x_n\}$ denote the leaves of T and write $x'_i := f(x_i) = \bar{f}(x_i)$ for $i = 1, \dots, n$.

Define inductively subtrees $S_1, \dots, S_n$ of T (all with root ρ) and $S'_1, \dots, S'_n$ of $T'' \subseteq T'$ (all with root ρ') as follows. Set $S_1 := [\rho, y_1]$ and $S'_1 := [\rho, y'_1]$, where y_1 and y'_1 are the unique points on the arcs $[\rho, x_1]$ and $[\rho', x'_1]$, respectively, such that

$$d_T(\rho, y_1) = d_{T'}(\rho', y'_1) = d_T(\rho, x_1) \wedge d_{T'}(\rho', x'_1). \tag{6.5}$$

Suppose that $S_1, \dots, S_m$ and $S'_1, \dots, S'_m$ have been defined. Let z_{m+1} and z'_{m+1} be the points on S_m and S'_m closest to x_{m+1} and x'_{m+1}. Put $S_{m+1} := S_m \cup]z_{m+1}, y_{m+1}]$ and $S'_{m+1} := S'_m \cup]z'_{m+1}, y'_{m+1}]$, where y_{m+1} and y'_{m+1} are the unique points on the arcs $]z_{m+1}, x_{m+1}]$ and $]z'_{m+1}, x'_{m+1}]$, respectively, such that

$$\begin{aligned} d_{T'}(z_{m+1}, y_{m+1}) &= d_{T'}(z'_{m+1}, y'_{m+1}) \\ &= d_T(z_{m+1}, x_{m+1}) \wedge d_{T'}(z'_{m+1}, x'_{m+1}). \end{aligned} \tag{6.6}$$

Set $S := S_n$ and $S' := S'_n$.

Put $z_1 := \rho$, and $z'_1 := \rho'$. By construction, the arcs $]z_k, y_k]$, $1 \le k \le n$, are disjoint and their union is $S \setminus \{\rho\}$. Similarly, the arcs $]z'_k, y'_k]$ are disjoint and their union is $S' \setminus \{\rho'\}$. Moreover, the arcs $]z_k, y_k]$ and $]z'_k, y'_k]$ have the same length (in particular, S and S' have the same length and part (ii) holds). We may therefore define a measure-preserving bijection ψ between S and S' by setting $\psi(\rho) = \rho'$ and letting the restriction of ψ to each arc $]z_k, y_k]$ be the obvious length preserving bijection onto $]z'_k, y'_k]$. More precisely, if $a \in]z_k, y_k]$, then $\psi(a)$ is the uniquely determined point on $]z'_k, y'_k]$ such that $d_{S'}(z'_k, \psi(a)) = d_S(z_k, a)$.

We next estimate the distortion of ψ to establish part (iii). We first claim that for $a, b \in S$,

$$|d_S(a, b) - d_{S'}(\psi(a), \psi(b))| \le 5\gamma, \tag{6.7}$$

where

$$\gamma := \max_{1 \le k, m \le n} |d_S(y_k, y_m) - d_{S'}(y'_k, y'_m)| \vee \max_{1 \le k \le n} |d_S(y_k, \rho) - d_{S'}(y'_k, \rho')|. \tag{6.8}$$

To see (6.7), consider $a, b \in S \setminus \{\rho\}$ with $a \in]z_k, y_k]$ and $b \in]z_m, y_m]$ where $k \neq m$. (The case where $a = \rho$ or $b = \rho$ holds "by continuity" and is left to the reader.) Without loss of generality, assume that $k < m$, so that $y_k \wedge y_m \le z_m < b \le y_m$ in the partial order on S and $y'_k \wedge y'_m \le z'_m < \psi(b) \le y'_m$ in the partial order on S'. Note that $y_k \wedge y_m$ and z_k are comparable in the partial order, as are $y'_k \wedge y'_m$ and z'_k. Moreover, by part (vi) of Lemma 6.2, $y_k \wedge y_m \le z_k$ if and only if $y'_k \wedge y'_m \le z'_k$. We then have to consider four cases depending on the relative positions of $y_k \wedge y_m, a$ and $y'_k \wedge y'_m, \psi(a)$.

Case I: $y_k \wedge y_m < a \le y_k$ and $y'_k \wedge y'_m < \psi(a) \le y'_k$.
We have

$$d_S(y_k, y_m) = d_S(y_k, a) + d_S(a, b) + d_S(b, y_m) \tag{6.9}$$

and

$$d_{S'}(y'_k, y'_m) = d_{S'}(y'_k, \psi(a)) + d_{S'}(\psi(a), \psi(b)) + d_{S'}(\psi(b), y'_m). \tag{6.10}$$

By construction,

$$d_S(y_k, a) = d_{S'}(y'_k, \psi(a)) \tag{6.11}$$

and

$$d_S(b, y_m) = d_{S'}(\psi(b), y'_m). \tag{6.12}$$

Hence

$$|d_S(a, b) - d_{S'}(\psi(a), \psi(b))| = |d_S(y_k, y_m) - d_{S'}(y'_k, y'_m)| \le \gamma. \tag{6.13}$$

Case II: $y_k \wedge y_m < a \le y_k$ and $\psi(a) \le y'_k \wedge y'_m < y'_k$.
Note that in this case $z_k \le y_k \wedge y_m$. We again have

$$d_S(y_k, y_m) = d_S(y_k, a) + d_S(a, b) + d_S(b, y_m), \tag{6.14}$$

but now

$$\begin{aligned} d_{S'}(y'_k, y'_m) = {} & d_{S'}(y'_k, \psi(a)) + d_{S'}(\psi(a), \psi(b)) + d_{S'}(\psi(b), y'_m) \\ & - 2 d_{S'}(\psi(a), y'_k \wedge y'_m). \end{aligned} \tag{6.15}$$

Let y_ℓ be such that $z_k = y_\ell \wedge y_k = y_\ell \wedge y_m$ and hence $z'_k = y'_\ell \wedge y'_k = y'_\ell \wedge y'_m$. Observe from Subsection 2.2 that

$$\begin{aligned} & d_{S'}(\psi(a), y'_k \wedge y'_m) \\ & \quad = \frac{1}{2}(d_{S'}(y'_\ell, y'_m) + d_{S'}(y'_k, \rho') - d_{S'}(y'_k, y'_m) - d_{S'}(y'_\ell, \rho')) - d_{S'}(z'_k, \psi(a)) \\ & \quad \le \frac{1}{2}(d_S(y_\ell, y_m) + d_S(y_k, \rho) - d_S(y_k, y_m) - d_S(y_\ell, \rho)) + \frac{1}{2}4\gamma - d_S(z_k, a) \\ & \quad = d_S(z_k, y_k \wedge y_m) - d_S(z_k, a) + 2\gamma \\ & \quad \le 2\gamma, \end{aligned} \tag{6.16}$$

and hence

$$|d_S(a, b) - d_{S'}(\psi(a), \psi(b))| \le 5\gamma. \tag{6.17}$$

Case III: $a \le y_k \wedge y_m < y_k$ and $y'_k \wedge y'_m \le \psi(a) < y'_k$.
Note that in this case, $z'_k \le y'_k \wedge y'_m$. This case is similar to Case II, but we record some of the details for use later in the proof of part (iv). Letting the index ℓ be as in Case II, we have

$$\begin{aligned} d_S(y_k, y_m) = d_S(y_k, a) + d_S(a, b) + d_S(b, y_m) \\ - 2d_S(a, y_k \wedge y_m) \end{aligned} \tag{6.18}$$

and

$$d_{S'}(y'_k, y'_m) = d_{S'}(y'_k, \psi(a)) + d_{S'}(\psi(a), \psi(b)) + d_{S'}(\psi(b), y'_m). \tag{6.19}$$

We have

$$\begin{aligned} &d_S(a, y_k \wedge y_m) \\ &= \frac{1}{2}(d_S(y_\ell, y_m) + d_S(y_k, \rho) - d_S(y_k, y_m) - d_S(y_\ell, \rho)) - d_S(z_k, a) \\ &\le \frac{1}{2}(d_{S'}(y'_\ell, y'_m) + d_{S'}(y'_k, \rho') - d_{S'}(y'_k, y'_m) - d_{S'}(y'_\ell, \rho')) + \frac{1}{2}4\gamma \\ &\quad - d_{S'}(z'_k, \psi(a)) \\ &= d_{S'}(z'_k, y'_k \wedge y'_m) - d_{S'}(z'_k, \psi(a)) + 2\gamma \\ &\le 2\gamma, \end{aligned} \tag{6.20}$$

and hence

$$|d_S(a, b) - d_{S'}(\psi(a), \psi(b))| \le 5\gamma. \tag{6.21}$$

Case IV: $a \le y_k \wedge y_m < y_k$ and $\psi(a) \le y'_k \wedge y'_m < y'_k$.
Letting the index ℓ be as in Case II, we have

$$d_S(z_k, y_m) = d_S(z_k, a) + d_S(a, b) + d_S(b, y_m) \tag{6.22}$$

and

$$d_{S'}(z'_k, y'_m) = d_{S'}(z'_k, \psi(a)) + d_{S'}(\psi(a), \psi(b)) + d_{S'}(\psi(b), y'_m). \tag{6.23}$$

Hence, from Subsection 2.2,

$$\begin{aligned} &d_S(a, b) - d_{S'}(\psi(a), \psi(b)) \\ &\quad = d_S(z_k, y_m) - d_{S'}(z'_k, y'_m) \\ &\quad = d_S(y_\ell \wedge y_m, y_m) - d_{S'}(y'_\ell \wedge y'_m, y'_m) \\ &\quad = \frac{1}{2}(d_S(y_m, \rho) + d_S(y_\ell, y_m) - d_S(y_\ell, \rho)) \\ &\qquad - \frac{1}{2}(d_{S'}(y'_m, \rho') + d_{S'}(y'_\ell, y'_m) - d_{S'}(y'_\ell, \rho')). \end{aligned} \tag{6.24}$$

Thus

$$|d_S(a, b) - d_{S'}(\psi(a), \psi(b))| \le \frac{3}{2}\gamma. \tag{6.25}$$

Combining Cases I–IV, we see that (6.7) holds. We thus require an estimate of γ to complete the estimation of the distortion of ψ . Clearly,

$$\begin{aligned}|d_S(y_k, y_m) - d_{S'}(y'_k, y'_m)| \le |d_T(x_k, x_m) - d_{T''}(x'_k, x'_m)| \\ + d_T(y_k, x_k) + d_T(y_m, x_m) \\ + d_{T''}(y'_k, x'_k) + d_{T''}(y'_m, x'_m).\end{aligned} \tag{6.26}$$

By (6.5),

$$\begin{aligned}d_T(y_1, x_1) \vee d_{T''}(y'_1, x'_1) &= |d_T(\rho, x_1) - d_{T''}(\rho', x'_1)| \\ &\le \mathrm{dis}(\bar{f}) \\ &< 8\delta.\end{aligned} \tag{6.27}$$

For $2 \le k \le n$ there exists by construction an index $i \in \{1, 2, \dots, k-1\}$ such that $z_k \in [z_i, y_i]$ and $z'_k \in [z'_i, y'_i]$. Applying the observations of Subsection 2.2,

$$\begin{aligned}&d_T(y_k, x_k) \vee d_{T''}(y'_k, x'_k) \\ &\quad= |d_T(y_k, x_k) - d_{T''}(y'_k, x'_k)| \\ &\quad= |d_T(z_k, x_k) - d_{T''}(z'_k, x'_k)| \\ &\quad\le \frac{1}{2}\{|d_T(x_i, x_k) - d_{T''}(x'_i, x'_k)| \\ &\qquad+ |d_T(\rho, x_i) - d_{T''}(\rho', x'_i)| + |d_T(\rho, x_k) - d_{T''}(\rho', x'_k)|\} \\ &\quad\le \frac{3}{2}\mathrm{dis}(\bar{f}) \\ &\quad\le 12\delta.\end{aligned} \tag{6.28}$$

Thus, from (6.26),

$$|d_S(y_k, y_m) - d_{S'}(y'_k, y'_m)| < (8 + 4 \times 12)\delta = 56\delta. \tag{6.29}$$

A similar argument shows that $|d_S(y_k, \rho) - d_{S'}(y'_k, \rho')| < (8 + 2 \times 12)\delta = 32\delta$, and hence $\gamma < 56\delta$. Substituting into (6.7) gives

$$\mathrm{dis}(\psi) \le 5\gamma < (5 \times 56)\delta = 280\delta. \tag{6.30}$$

Moving to part (i), apply (6.28) to obtain

$$d_{\mathrm{H}}(S, T) \le \max_{1 \le i \le n} d(y_k, x_k) \le \gamma < 56\delta \tag{6.31}$$

and, by similar arguments,

$$d_{\mathrm{H}}(S', T') \le d_{\mathrm{H}}(S', T'') + d_{\mathrm{H}}(T'', T') < 59\delta. \tag{6.32}$$

Finally, we consider part (iv). Suppose that $a \in S$ is such that the subtree of S' above $\psi(a)$ is not the image under ψ of the subtree of S above a. Let k be the unique index such that $a \in]z_k, y_k]$ (and hence $\psi(a) \in]z'_k, y'_k]$). It follows from the construction of ψ that there must exist an index ℓ such that either $z_k < a \le z_\ell$ and $z'_k < z'_\ell \le \psi(a)$ or $z_k < z_\ell \le a$ and $z'_k < \psi(a) \le z'_\ell$. These two situations have already been considered in Case III and Case II above (in that order): there we represented z_ℓ as $y_k \wedge y_m$ and z'_ℓ as $y'_k \wedge y'_m$. It follows from the inequality (6.20) that the mass of the set of points a that satisfy the first alternative is at most $2\gamma n < 112\delta n$. Similarly, from the inequality (6.16) and the fact that ψ is measure-preserving, the mass of the set of points a that satisfy the second alternative is also at most $2\gamma n < 112\delta n$. Thus the total mass of the set of points of interest is at most $224\delta n$. □

Before completing the proof of Proposition 6.1, we recall the definition of the *Wasserstein metric*. Suppose that (E, d) is a complete, separable metric space. Write B for the set of continuous functions functions $f : E \to \mathbb{R}$ such that $|f(x)| \le 1$ and $|f(x) - f(y)| \le d(x, y)$ for $x, y \in E$. The Wasserstein (sometimes transliterated as Vasershtein) distance between two Borel probability measures α and β on E is given by

$$d_{\mathrm{W}}(\alpha, \beta) := \sup_{f \in B} \left| \int f \, d\alpha - \int f \, d\beta \right|. \tag{6.33}$$

The Wasserstein distance is a genuine metric on the space of Borel probability measures and convergence with respect to this distance implies weak convergence (see, for example, Theorem 3.3.1 and Problem 3.11.2 of [EK86]). If V and W are two E-valued random variables on the same probability space $(\Sigma, \mathcal{A}, \mathbb{P})$ with distributions α and β, respectively, then

$$\begin{aligned} d_{\mathrm{W}}(\alpha, \beta) &\le \sup_{f \in B} |\mathbb{P}[f(V)] - \mathbb{P}[f(W)]| \\ &\le \sup_{f \in B} \mathbb{P}[|f(V) - f(W)|] \le \mathbb{P}[d(V, W)]. \end{aligned} \tag{6.34}$$

Proof of Proposition 6.1 For $(T, \rho) \in \mathbf{T}^{\mathrm{root}}$ and $t \ge 0$, let

$$\mathbf{P}_t((T, \rho), \cdot) := \mathbf{P}^{(T,\rho)}\{X_t \in \cdot\}. \tag{6.35}$$

We need to show that $(T, \rho) \mapsto \mathbf{P}_t((T, \rho), \cdot)$ is weakly continuous for each $t \ge 0$. This is equivalent to showing for each $(T, \rho) \in \mathbf{T}^{\mathrm{root}}$ and $t \ge 0$ that

$$\lim_{(T',\rho') \to (T,\rho)} d_{\mathrm{W}}\left(\mathbf{P}_t((T, \rho), \cdot), \mathbf{P}_t((T', \rho'), \cdot)\right) = 0. \tag{6.36}$$

¿From the coupling argument in the proof of part (iii) of Theorem 3 (in particular, the inequality (3.11)), we have that

$$\begin{aligned} &d_{\mathrm{W}}(\mathbf{P}_t((T, \rho), \cdot), \mathbf{P}_t((T', \rho'), \cdot)) \\ &\quad \le d_{\mathrm{W}}(\mathbf{P}_t((T, \rho), \cdot), \mathbf{P}_t((R_\eta(T), \rho), \cdot)) \\ &\qquad + d_{\mathrm{W}}(\mathbf{P}_t((R_\eta(T), \rho), \cdot), \mathbf{P}_t((R_\eta(T'), \rho'), \cdot)) \\ &\qquad + d_{\mathrm{W}}(\mathbf{P}_t((R_\eta(T'), \rho), \cdot), \mathbf{P}_t((T', \rho'), \cdot)) \\ &\quad \le d_{\mathrm{W}}(\mathbf{P}_t((R_\eta(T), \rho), \cdot), \mathbf{P}_t((R_\eta(T'), \rho'), \cdot)) + 2\eta. \end{aligned} \tag{6.37}$$

By part (ii) of Lemma 2.6, $R_\eta(T')$ converges to $R_\eta(T)$ as (T', ρ') converges to (T, ρ), and so it suffices to establish (6.36) when (T, ρ) and (T', ρ') are finite trees, and so we will suppose this for the rest of the proof.

Fix (T,ρ) and $\varepsilon > 0$. Suppose that $\delta > 0$ depending on (T,ρ) and ε is sufficiently small that the conclusions of Lemma 6.3 hold for any (T',ρ') within distance δ of (T,ρ). Let (S,ρ) and (S',ρ') be the subtrees guaranteed by Lemma 6.3. From the coupling argument in proof of part (iii) of Theorem 3 we have

$$d_{\mathrm{W}}\left(\mathbf{P}_t((T,\rho),\cdot),\mathbf{P}_t((S,\rho),\cdot)\right) < \varepsilon \tag{6.38}$$

and

$$d_{\mathrm{W}}\left(\mathbf{P}_t((T',\rho'),\cdot),\mathbf{P}_t((S',\rho'),\cdot)\right) < \varepsilon. \tag{6.39}$$

It therefore suffices to give a bound on $d_{\mathrm{W}}(\mathbf{P}_t((S,\rho),\cdot),\mathbf{P}_t((S',\rho),\cdot))$ that only depends on ε and converges to zero as ε converges to 0.

Construct on some probability space $(\Sigma,\mathcal{A},\mathbb{P})$ a Poisson point process Π_0 on the set $\mathbb{R}^{++}\times S^o$ with intensity $\lambda\otimes\mu$, where μ is the length measure on S. Construct on the same space another independent Poisson point process on the set $\{(t,x)\in\mathbb{R}^{++}\times\mathbb{R}^{++} : x\le t\}$ with intensity $\lambda\otimes\lambda$ restricted to this set. If we set $\Pi_0' := \{(t,\psi(x)) : (t,x)\in\Pi_0\}\subset\mathbb{R}^{++}\times(S')^o$, then Π_0' is a Poisson process on the set $\mathbb{R}^{++}\times(S')^o$ with intensity $\lambda\otimes\mu'$, where μ' is the length measure on S' (because ψ preserves length measure). Now apply the construction of Subsection 3.1 to realizations of Π_0 and Π (respectively, Π_0' and Π) to get two $\mathbf{T}^{\mathrm{root}}$-valued processes that we will denote by $(Y_t)_{t\ge0}$ and $(Y_t')_{t\ge0}$. We see from the proof of Theorem 3 that Y (respectively, Y') has the same law as X under $\mathbf{P}^{(S,\rho)}$ (respectively, $\mathbf{P}^{(S',\rho')}$).

Define a map ψ_t from $Y_t = S\amalg]0,t]$ to $Y_t' = S'\amalg]0,t]$ by setting the restriction of ψ_t to S be ψ and the restriction of ψ_t to $]0,t]$ be the identity map. Let d_t and d_t' be the metrics on Y_t and Y_t', respectively. We will bound the rooted Gromov-Hausdorff distance between Y_t and Y_t' by bounding the distortion of ψ_t.

The cut-times for Y and Y' coincide. If ξ is a cut-point of Y at some cut-time τ, then the corresponding cut-point for Y' will be $\psi(\xi)$.

It is clear that the distortion of ψ_t is constant between cut-times. Write B_t for the set of points $b\in Y_t$ such that the subtree of Y_t' above $\psi_t(b)$ is not the image under ψ_t of the subtree of Y_t above b. The set B_t is unchanged between cut-times.

Consider a cut-time τ such that the corresponding cut-point ξ is in $Y_{\tau-}\setminus B_{\tau-}$. If x and y are in the subtree above ξ in $Y_{\tau-}$, then they are moved together by the re-grafting operation and their distance apart is unchanged in Y_τ. Also, $\psi_{\tau-}(x)$ and $\psi_{\tau-}(y)$ are in subtree above $\psi_{\tau-}(\xi)$ in $Y_{\tau-}'$ and these two points are also moved together. More precisely,

$$d_\tau(x,y) = d_{\tau-}(x,y) \tag{6.40}$$

and

$$d_\tau'(\psi_\tau(x),\psi_\tau(y)) = d_{\tau-}'(\psi_{\tau-}(x),\psi_{\tau-}(y)). \tag{6.41}$$

The same conclusion holds if neither x or y are in the subtree above ξ in $Y_{\tau-}$. If x is in the subtree above ξ in $Y_{\tau-}$ and y is not, then

$$d_\tau(x,y) = d_{\tau-}(x,\xi) + d_{\tau-}(\tau,y) \tag{6.42}$$

and

$$d'_\tau(\psi_\tau(x),\psi_\tau(y)) = d'_{\tau-}(\psi_{\tau-}(x),\psi_{\tau-}(\xi)) + d'_{\tau-}(\tau,\psi_{\tau-}(y)) \tag{6.43}$$

(where we recall that τ is the root in each of the trees $Y_{\tau-}, Y_\tau, Y'_{\tau-}, Y'_\tau$). Combining these cases, we see that

$$\mathrm{dis}(\psi_\tau) \le 2\mathrm{dis}(\psi_{\tau-}). \tag{6.44}$$

Moreover, if $\xi \in Y_{\tau-} \setminus B_{\tau-}$, then $B_\tau = B_{\tau-}$.

Also, for any $t \ge 0$ we always have the upper bound

$$\begin{aligned}\mathrm{dis}(\psi_t) &\le \mathrm{diam}(Y_t) + \mathrm{diam}(Y'_t)\\ &\le \mathrm{diam}(S) + \mathrm{diam}(S') + 2t\\ &\le \mathrm{diam}(T) + \mathrm{diam}(T') + 2t\\ &\le 2\mathrm{diam}(T) + d_{\mathrm{GH^{root}}}((T,\rho),(T',\rho')) + 2t\\ &\le 2\mathrm{diam}(T) + \delta + 2t\\ &=: D_t\end{aligned} \tag{6.45}$$

Set $N_t := |\Pi_0 \cap (]0,t] \times S^o)| + |\Pi \cap \{(s,x) : 0 < x \le s \le t\}|$ and write I_t for the indicator of the event $\{\Pi_0 \cap (]0,t] \times B_0) \neq \emptyset\}$, which, by the above argument, is the event that $\xi \in B_{\tau-}$ for some (cut-time, cut-point) pairs (τ,ξ) with $0 < \tau \le t$. We have

$$\begin{aligned}&d_{\mathrm{W}}(\mathbf{P}_t((S,\rho),\cdot),\mathbf{P}_t((S',\rho),\cdot)\\ &\quad\le \mathbb{P}[d_{\mathrm{GH^{root}}}(Y_t,Y'_t)]\\ &\quad\le \frac{1}{2}\mathbb{P}[\mathrm{dis}(\psi_t)]\\ &\quad\le \frac{1}{2}\mathbb{P}\left[\varepsilon 2^{N_t} + I_t D_t\right]\\ &\quad= \frac{1}{2}\left\{\varepsilon \exp\left(\mu(T)t + \frac{t^2}{2}\right) + \left[1-\exp\left(-\varepsilon t\right)\right] D_t\right\},\end{aligned} \tag{6.46}$$

and this suffices to complete the proof. □

7. Asymptotics of the Aldous-Broder algorithm

Given an irreducible Markov matrix $\mathbb{P}$ with state space V, there is a natural probability measure on the collection of combinatorial trees with vertices labeled by V that assigns mass

$$C^{-1}\prod \mathbb{P}(x,y) \tag{7.1}$$

to the tree T, where C is a normalization constant and the product is over pairs of adjacent vertices (x,y) in T ordered so that y is on the path from the root to x. For example, if $\mathbb{P}(x,y) \equiv 1/|V|$ for all $x,y \in V$, (so that the associated Markov chain consists of successive uniform random picks from V), then the distribution (7.1) is uniform on the set of $|V|^{|V|-1}$ rooted combinatorial trees labeled by V.

The *Aldous-Broder algorithm* [AT89, Bro89, Ald90] is a tree-valued Markov chain that has the distribution in (7.1) as its stationary distribution. The discrete time version of the algorithm has the following transition dynamics.

- Pick a vertex v at random according to $\mathbb{P}(\rho, \cdot)$, where ρ is the current root.
- If $v = \rho$, do nothing.
- If $v \neq \rho$:
  - Erase the edge connecting v to the unique vertex adjacent to v and on the path from ρ to v.
  - Insert a new edge between v and ρ.
  - Designate v as the new root.

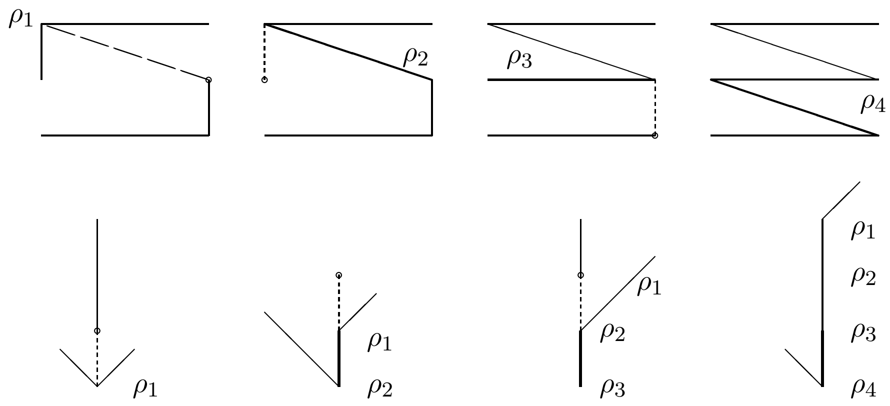

FIGURE 4 *illustrates Aldous-Broder algorithm in discrete time. Here the dots mark the vertices that become the root in the next step. Once the new root is chosen we erase the edge adjacent to the new root on the path from the old to the new root and insert an edge connecting the old to the new root. The bold edges in the picture present the edges introduced recently.*

It will be more convenient for us to work with the continuous time version of this algorithm in which the above transitions are made at the arrival times of an independent Poisson process with rate $|V|/(|V|-1)$ (so that the continuous time chain makes actual jumps at rate 1).

We can associate a rooted compact real tree with a rooted labeled combinatorial tree in the obvious way by thinking of the edges as line segments with length 1. Because we don't record the labeling, the process that arises from mapping the continuous-time Aldous-Broder algorithm in this way won't be Markovian in general. However, this process will be Markovian in the case where $\mathbb{P}$ is the transition matrix for i.i.d. uniform sampling (that is, when $\mathbb{P}(x,y) = 1/|V|$ for all $x, y \in V$) and we assume this from now on. The following result says that if we rescale "space" and time appropriately, then this process converges to the root growth with re-grafting process. If $T = (T, d, \rho)$ is a rooted compact real tree and $c > 0$, we write cT for the tree $(T, c\,d, \rho)$ (that is, $cT = T$ as sets and the roots are the same, but the metric is re-scaled by c).

Proposition 7.1. *Let $Y^n = (Y^n_t)_{t\geq 0}$ be a sequence of Markov processes that take values in the space of rooted compact real trees with integer edge lengths and evolve according to the dynamics associated with the continuous-time Aldous-Broder chain for i.i.d. uniform sampling. Suppose that each tree Y^n_0 is non-random with total branch length N_n, that N_n converges to infinity as $n \to \infty$, and that $N_n^{-1/2} Y^n_0$ converges in the rooted Gromov-Hausdorff metric to some rooted compact real tree T as $n \to \infty$. Then, in the sense of weak convergence of processes on the space of càdlàg paths equipped with the Skorohod topology, $(N_n^{-1/2} Y^n(N_n^{1/2} t))_{t\geq 0}$ converges as $n \to \infty$ to the root growth with re-grafting process X under $\mathbf{P}^T$.*

Proof. Define $Z^n = (Z^n_t)_{t\geq 0}$ by

$$Z^n_t := N_n^{-1/2} Y^n(N_n^{1/2} t). \tag{7.2}$$

For $\eta > 0$, let $Z^{\eta,n}$ be the $\mathbf{T}^{\mathrm{root}}$-valued process constructed as follows.

- Set $Z^{\eta,n}_0 = R_{\eta_n}(Z^n_0)$, where $\eta_n := N_n^{-1/2} \lfloor N_n^{1/2} \eta \rfloor$.
- The value of $Z^{\eta,n}$ is unchanged between jump times of $(Z^n_t)_{t\geq 0}$.
- At a jump time τ for $(Z^n_t)_{t\geq 0}$, the tree $Z^{\eta,n}_\tau$ is the subtree of Z^n_τ spanned by $Z^{\eta,n}_{\tau-}$ and the root of Z^n_τ.

An argument similar to that in the proof of Lemma 3.2 shows that

$$\sup_{t\geq 0} d_{\mathrm{H}}(Z^n_t, Z^{\eta,n}_t) \leq \eta_n, \tag{7.3}$$

and so it suffices to show that $Z^{\eta,n}$ converges weakly as $n \to \infty$ to X under $\mathbf{P}^{R_\eta(T)}$.

Note that $Z^{\eta,n}_0$ converges to $R_\eta(T)$ as $n \to \infty$. Moreover, if Λ is the map that sends a tree to its total length (that is, the total mass of its length measure), then $\lim_{n\to\infty} \Lambda(Z^{\eta,n}_0) = \Lambda \circ R_\eta(T) < \infty$ by Lemma 7.3 below.

The pure jump process $Z^{\eta,n}$ is clearly Markovian. If it is in a state (T', ρ'), then it jumps with the following rates.

- With rate $N_n^{1/2}(N_n^{1/2}\Lambda(T'))/N_n = \Lambda(T')$, one of the $N_n^{1/2}\Lambda(T')$ points in T' that are at distance a positive integer multiple of $N_n^{-1/2}$ from the root ρ' is chosen uniformly at random and the subtree above this point is joined to ρ' by an edge of length $N_n^{-1/2}$. The chosen point becomes the new root and an arc of length $N_n^{-1/2}$ that previously led from the new root toward ρ' is erased. Such a transition results in a tree with the same total length as T'.

- With rate $N_n^{1/2} - \Lambda(T')$, a new root not present in T' is attached to ρ' by an edge of length $N_n^{-1/2}$. This results in a tree with total length $\Lambda(T') + N_n^{-1/2}$.

It is clear that these dynamics converge to those of the root growth with re-grafting process, with the first class of transitions leading to re-graftings in the limit and the second class leading to root growth. □

Lemma 7.2. *Let $(T, d, \rho) \in \mathbf{T}^{\mathrm{root}}$ and suppose that $\{x_0, \dots, x_n\} \subset T$ spans T, so that the root ρ and the leaves of T form a subset of $\{x_0, \dots, x_n\}$. Then the total length of T (that is, the total mass of its length measure) is given by*

$$d(x_0, x_1) + \sum_{k=2}^{n} \bigwedge_{0 \le i < j \le k-1} \frac{1}{2} \left(d(x_k, x_i) + d(x_k, x_j) - d(x_i, x_j)\right).$$

Proof. This follows from the observation that the distance from the point x_k to the arc $[x_i, x_j]$ is

$$\frac{1}{2} \left(d(x_k, x_i) + d(x_k, x_j) - d(x_i, x_j)\right),$$

and so length of the arc connecting x_k, $2 \le k \le n$, to the subtree spanned by $x_0, \dots, x_{k-1}$ is

$$\bigwedge_{0 \le i < j \le k-1} \frac{1}{2} \left(d(x_k, x_i) + d(x_k, x_j) - d(x_i, x_j)\right).$$

□

Lemma 7.3. *Let $\Lambda : \mathbf{T}^{\mathrm{root}} \to \mathbb{R} \cup \{\infty\}$ be the map that sends a tree to its total length. For $\eta > 0$, the map $\Lambda \circ R_\eta$ is continuous.*

Proof. For all $\eta > 0$ we have by Lemma 2.6 that: $R_\eta = R_{\eta/2} \circ R_{\eta/2}$, the map R_η is continuous, and the range of R_η consists of finite trees. It therefore suffices to show for all $\eta > 0$ that if (T, d, ρ) is a fixed finite tree and (T', d', ρ') is any another finite tree sufficiently close to T, then $\Lambda \circ R_\eta(T')$ is close to $\Lambda \circ R_\eta(T)$.

Suppose, therefore, that (T, d, ρ) is a fixed finite tree with leaves $\{x_1, \dots, x_n\}$ and that (T', d', ρ') is another finite tree with

$$d_{\mathrm{GH}^{\mathrm{root}}}((T, d, \rho), (T', d', \rho')) < \delta,$$

where δ is small enough that the conclusions of Lemma 6.2 hold. Consider a rooted subtree (T'', d', ρ') of (T', d', ρ') and a map $\bar{f} : T \to T''$ with the properties guaranteed by Lemma 6.2. Set $x'_k = \bar{f}(x_k)$ for $1 \le k \le n$.

Fix $\kappa > 0$. For $1 \le k \le n$, write $\hat{x}_k \in T$ for the point on the arc $[\rho, x_k]$ that is at distance $\kappa \wedge d(\rho, x_k)$ from x_k. Set $\hat{x}_0 := \rho$. Define $\hat{x}'_0, \dots, \hat{x}'_n \in T''$ similarly. Note that $R_\kappa(T)$ is spanned by $\{\hat{x}_0, \dots, \hat{x}_n\}$ and $R_\kappa(T'')$ is spanned by $\{\hat{x}'_0, \ldots, \hat{x}'_n\}$. By Lemma 7.2,

$$\Lambda \circ R_\kappa(T) = d(\hat{x}_0, \hat{x}_1) + \sum_{k=2}^{n} \bigwedge_{0 \le i < j \le k-1} \frac{1}{2} \left(d(\hat{x}_k, \hat{x}_i) + d(\hat{x}_k, \hat{x}_j) - d(\hat{x}_i, \hat{x}_j) \right).$$

and

$$\Lambda \circ R_\kappa(T'') = d'(\hat{x}'_0, \hat{x}'_1) + \sum_{k=2}^{n} \bigwedge_{0 \le i < j \le k-1} \frac{1}{2} \left(d'(\hat{x}'_k, \hat{x}'_i) + d'(\hat{x}'_k, \hat{x}'_j) - d'(\hat{x}'_i, \hat{x}'_j) \right).$$

Also observe that

$$\begin{aligned} d(\hat{x}_i, \hat{x}_j) = {} & (d(x_0, x_i) - \kappa)_+ + (d(x_0, x_j) - \kappa)_+ \\ & - 2\Big[(d(x_0, x_i) - \kappa)_+ \wedge (d(x_0, x_j) - \kappa)_+ \\ & \wedge \Big\{ \frac{1}{2} \big(d(x_0, x_i) + d(x_0, x_j) - d(x_i, x_j) \big) \Big\} \Big] \end{aligned}$$

and

$$\begin{aligned} d'(\hat{x}'_i, \hat{x}'_j) = {} & (d'(x'_0, x'_i) - \kappa)_+ + (d'(x'_0, x'_j) - \kappa)_+ \\ & - 2\Big[(d'(x'_0, x'_i) - \kappa)_+ \wedge (d'(x'_0, x'_j) - \kappa)_+ \\ & \wedge \Big\{ \frac{1}{2} \big(d'(x'_0, x'_i) + d'(x'_0, x'_j) - d'(x'_i, x'_j) \big) \Big\} \Big]. \end{aligned}$$

Now the function $t \mapsto (t - \kappa)_+$ is Lipschitz with Lipschitz constant 1 for all $\kappa > 0$, and it follows that there is a family of Lipschitz functions F_κ, $\kappa > 0$, with Lipschitz constants uniformly bounded by some constant C such that

$$\Lambda \circ R_\kappa(T) = F_\kappa \left((d(x_i, x_j))_{0 \le i,j \le n} \right)$$

and

$$\Lambda \circ R_\kappa(T'') = F_\kappa \left((d'(x'_i, x'_j))_{0 \le i,j \le n} \right).$$

By construction $|d(x_i, x_j) - d'(x'_i, x'_j)| < 8\delta$, and so

$$|\Lambda \circ R_\kappa(T) - \Lambda \circ R_\kappa(T'')| \le 8\delta C$$

for all $\kappa > 0$.

Because $d_{\mathrm{H}}(T', T'') < 3\delta$, we have

$$\Lambda \circ R_\eta(T'') \le \Lambda \circ R_\eta(T') \le \Lambda \circ R_{\eta - 3\delta}(T'').$$

Thus

$$\Lambda \circ R_\eta(T) - 8\delta C \le \Lambda \circ R_\eta(T') \le \Lambda \circ R_{\eta - 3\delta}(T) + 8\delta C.$$

Since $\lim_{\delta \downarrow 0} \Lambda \circ R_{\eta - 3\delta}(T) = \Lambda \circ R_\eta(T)$, this suffices to establish the result. $\square$

An alternative algorithm for simulating from the distribution in (7.1) in the case of i.i.d. uniform sampling is the complete graph special case of Wilson's loop-erased walk algorithm for generating a uniform spanning tree of a graph [PW98, Wil96, WP96]. Asymptotics of the latter algorithm have been investigated in [Pit02b]. Wilson's algorithm was also used in [PR04] to show that the finite-dimensional distributions of the re-scaled uniform random spanning tree for the d-dimensional discrete torus converges to the Brownian CRT as the number of vertices goes to infinity when $d \geq 5$.

8. Rayleigh process

Suppose that we take the root growth with re-grafting process $(X_t)_{t\geq 0}$ under $\mathbf{P}^T$ for some $T \in \mathbf{T}^{\text{root}}$, we fix a point $x \in T$, and we denote by R_t the distance between x and the root t of X_t (that is, R_t is the height of x in X_t). According to the root growth with re-grafting dynamics, R_t grows deterministically with unit speed between cut-time τ for which the corresponding cut-point falls on the arc $[\tau, x]$. Such cut-times τ come along at intensity $R_{t-}\,dt$ in time, and at $\tau-$ the position of the corresponding cut-point is uniformly distributed on the arc $[\tau, x]$ conditional on the past up to $\tau-$, so that R_τ is uniformly distributed on $[0, R_{\tau-}]$ conditional on the past up to $\tau-$. Consequently, the $\mathbb{R}^+$-valued process $(R_t)_{t\geq 0}$ is autonomously Markovian. In particular, $(R_t)_{t\geq 0}$ is an example of the class of *piecewise deterministic* Markov processes discussed in the Introduction.

In order to describe the properties of $(R_t)_{t\geq 0}$, we need the following definitions. A non-negative random variable R is said to have *standard Rayleigh distribution* if it is distributed as the length of a standard normal vector in $\mathbb{R}^2$, that is,

$$\mathbb{P}\{R > r\} = \exp\left(-\frac{r^2}{2}\right), \quad r \geq 0. \tag{8.1}$$

If R^* is distributed according to the size-biased standard Rayleigh distribution, that is,

$$\mathbb{P}\{R^* \in \mathrm{d}r\} = \frac{r\mathbb{P}\{R \in \mathrm{d}r\}}{\mathbb{P}[R]} = \sqrt{\frac{2}{\pi}}\, r^2 e^{-\frac{1}{2}r^2} \mathrm{d}r, \quad r \geq 0, \tag{8.2}$$

and if U is a uniform random variable that is independent of R^*, then UR^* has the inverse size-biased standard Rayleigh distribution:

$$\mathbb{P}\{UR^* \in \mathrm{d}r\} = \frac{r^{-1}\mathbb{P}\{R \in \mathrm{d}r\}}{\mathbb{P}[R^{-1}]} = \sqrt{\frac{2}{\pi}}\, e^{-\frac{1}{2}r^2} \mathrm{d}r, \quad r \geq 0. \tag{8.3}$$

Thus R^* and UR^* are distributed as the length of a standard normal vector in $\mathbb{R}^3$ and $\mathbb{R}$, respectively.

For reasons that are apparent from Proposition 8.1 below, we call the process $(R_t)_{t\geq 0}$ the *Rayleigh process*. We note that there is a body of literature on stationary processes with Rayleigh one-dimensional marginal distributions that arise as the length process of a vector-valued process in $\mathbb{R}^2$ with coordinate processes that are independent copies of some stationary centered Gaussian process (see, for example, [Has70, MBB58, BS02]).

Proposition 8.1. *Consider the Rayleigh process $(R_t)_{t\geq 0}$. Write $\mathbf{P}^r$ for the law of $(R_t)_{t\geq 0}$ started at $r \geq 0$.*

(i) *The unique stationary distribution of the Rayleigh process is the standard Rayleigh distribution and the total variation distance between $\mathbf{P}^r\{R_t \in \cdot\}$ and the standard Rayleigh distribution converges to 0 as $t \to \infty$.*
(ii) *Under $\mathbf{P}^0$, for each fixed $t > 0$, R_t has the same law as $R \wedge t$, where R has the standard Rayleigh distribution.*
(iii) *For $x > 0$, the mean return time to x is $x^{-1}e^{\frac{1}{2}x^2}$.*
(iv) *If τ_n denotes the nth jump time of $(R_t)_{t\geq 0}$, then as $n \to \infty$ the triple $(R_{\tau_n}, R_{\tau_{n+1}-}, R_{\tau_{n+1}})$ converges in law to the triple $(U'R^*, R^*, U''R^*)$, where U' and U'' are independent uniform random variables on $]0,1[$ independent of R^*, and R^* has the size-biased Rayleigh distribution.*
(v) *The jump counting process $N(t) := |\{n \in \mathbb{N} : \tau_n \leq t\}|$ has asymptotically stationary increments under $\mathbf{P}^r$ for any $r \geq 0$, and*

$$\frac{1}{t}N(t) \to \sqrt{\frac{\pi}{2}}, \quad \mathbf{P}^r - \text{a.s.} \tag{8.4}$$

as $t \to \infty$.

Proof. (i) Let $\bar{\Pi}$ be a Poisson point process in $\mathbb{R}\times\mathbb{R}_+$ with Lebesgue intensity. For $-\infty < t < \infty$ let

$$\bar{R}_t := \inf\{x + (t-s) : (s,x) \in \bar{\Pi},\ s \leq t\}. \tag{8.5}$$

It is clear that $(\bar{R}_t)_{t\in\mathbb{R}}$ is a stationary Markov process with the transition dynamics of the Rayleigh process. Similarly, for $r \in \mathbb{R}_+$ and $t \geq 0$, set

$$R_t^r = (r+t) \wedge \inf\{x + (t-s) : (s,x) \in \bar{\Pi},\ 0 \leq s \leq t\}. \tag{8.6}$$

Then $(R_t^r)_{t\geq 0}$ has the same law as the Rayleigh process under $\mathbf{P}^r$.

Note that the event $\{\bar{R}_t > r\}$ is the event that $\bar{\Pi}$ has no points in the triangle with vertices $(t-r,0),(t,0),(t,r)$ and area $r^2/2$. Thus $\mathbf{P}\{\bar{R}_t > r\} = \exp(-r^2/2)$ and the standard Rayleigh distribution is a stationary distribution for the Rayleigh process.

Let $T^r := \inf\{t \geq 0 : R_t^r = \bar{R}_t\}$. Note that $R_t^r = \bar{R}_t$ for all $t \geq T^r$. Note that $T^r > t$ if and only if either $\bar{R}_0 > r$ and $\bar{\Pi}$ puts no points into the quadrilateral with vertices $(0,0)$, $(t,0)$, $(t, r+t)$, $(0,r)$, or $\bar{R}_0 \leq r$ and $\bar{\Pi}$ puts no points into the quadrilateral with vertices $(0,0)$, $(t,0)$, $(t,\bar{R}_0+t)$, $(0,\bar{R}_0)$. Hence

$$\begin{aligned}\mathbb{P}\{T^r > t\} &= \exp\left(-\frac{r^2}{2}\right)\exp\left(-\frac{1}{2}(r+(r+t))t\right) \\ &\quad + \int_0^r \exp\left(-\frac{1}{2}(x+(x+t))t\right) x\exp\left(-\frac{x^2}{2}\right) dx.\end{aligned} \tag{8.7}$$

By the standard coupling inequality, the total variation between $\mathbf{P}^r\{R_t \in \cdot\}$ and $\mathbb{P}\{\bar{R}_t \in \cdot\}$ is at most $2\mathbb{P}\{T^r > t\}$, which converges to 0 as $t \to \infty$.

This certainly shows that the standard Rayleigh distribution is the unique stationary distribution.

(ii) Note that $R_t^r > x$ if and only if $r + t \geq t > x$ and there are no points of $\bar{\Pi}$ in the triangle with vertices $(t-x, 0), (t, 0), (t, x)$ of area $x^2/2$ or $r + t > x \geq t$ and there are not points of $\bar{\Pi}$ in the quadrilateral with vertices $(0,0), (t,0), (t,x), (0, x-t)$ of area $((x-t)+x)t/2 = x^2/2 - (x-t)^2/2$. In either case,

$$\mathbf{P}^r\{R_t > x\} = 1\{r + t > x\} \exp\left(-\frac{1}{2}x^2 + \frac{1}{2}((x-t)_+)^2\right). \tag{8.8}$$

Taking $r = 0$ gives the result.

(iii) Let $T_y := \inf\{t > 0 : R_t = y\}$. It is obvious from the Poisson construction that, for all $x \geq 0$ and $y > 0$, $\mathbf{P}^x\{0 < T_y < \infty\} = 1$ and $\mathbf{P}^x[\exp(uT_y)] < \infty$ for all u in some neighbourhood of 0. The Laplace transforms $\mathbf{P}^x[\exp -\lambda T_y]$ are determined by standard methods of renewal theory:

$$\mathbf{P}^x[\exp -\lambda T_x] = \frac{U_x(\lambda)}{1 + U_x(\lambda)}, \quad \lambda > 0, \tag{8.9}$$

where by (8.8),

$$\begin{aligned} U_x(\lambda) &:= \int_0^\infty \mathrm{d}t \, e^{-\lambda t} \frac{\mathbf{P}^x\{R_t \in dx\}}{dx} \\ &= \int_0^x \mathrm{d}t \, e^{-\lambda t} \, t \, e^{-xt + t^2/2} + \mathbb{P}\{R \in \mathrm{d}x\} \frac{e^{-\lambda x}}{\lambda}. \end{aligned} \tag{8.10}$$

In particular, it follows easily that the mean return time of state x

$$\mathbf{P}^x[T_x] = -\lim_{\lambda \downarrow 0} \frac{1}{\lambda} \mathbf{P}^x[\exp -\lambda T_x] \tag{8.11}$$

is the inverse of the density of R at x, that is $x^{-1} e^{\frac{1}{2}x^2}$, as claimed.

(iv) Let $\bar{\tau} := \inf\{t > 0 : \bar{R}_t \neq \bar{R}_{t-}\}$. By part (i), the joint distribution of $(R_{\tau_n}, R_{\tau_{n+1}-}, R_{\tau_{n+1}})$ converges to the joint distribution of $(\bar{R}_0, \bar{R}_{\bar{\tau}-}, \bar{R}_{\bar{\tau}})$ conditional on $\bar{R}_0 \neq \bar{R}_{0-}$. Let C denote the intensity of the stationary point process $\{t \in \mathbb{R} : \bar{R}_t \neq \bar{R}_{t-}\}$. Then

$$\begin{aligned} &\mathbb{P}\{\bar{R}_0 \in dx, \bar{R}_{\bar{\tau}-} \in dy, \bar{R}_{\bar{\tau}} \in dz \,|\, \bar{R}_0 \neq \bar{R}_{0-}\} \\ &= C^{-1} \exp\left(-\frac{1}{2}x^2\right) dx \, \exp\left(-\frac{1}{2}(x+y)(y-x)\right) dy \, dz \\ &= \left[\frac{1}{y}\, dx\right] \times \left[C^{-1} y^2 \exp\left(-\frac{1}{2}y^2\right) dy\right] \times \left[\frac{1}{y}\, dz\right] \end{aligned} \tag{8.12}$$

for $x < y$ and $z < y$. The result now follows from (8.2), which also identifies $C = \sqrt{\pi/2}$.

(v) The stationary point process $\{t \in \mathbb{R} : \bar{R}_t \neq \bar{R}_{t-}\}$ is clearly ergodic by construction, and it has intensity $\sqrt{\pi/2}$ from the argument in part (iv). For any $r > 0$, it follows from the argument in part (i) that $R_t^r = \bar{R}_t$ for all t

sufficiently large, and so the result follows from the ergodic theorem applied to $\{t \in \mathbb{R} : \bar{R}_t \neq \bar{R}_{t-}\}$. □

Then the following corollary is a consequence of Proposition 7.1. See [DGR02], where similar scaling limits are derived.

Corollary 8.2. *For each $N \in \mathbb{N}$, let $(\tilde{R}^N_t)_{t\geq 0}$ denote a continuous time Markov chain with state space $\{1, \ldots, N\}$ and infinitesimal generator matrix*

$$\tilde{Q}^N(i,j) := \begin{cases} 1/N, & 1 \leq j \leq i-1, \\ -(N-1)/N, & j = i, \\ (N-i)/N, & j = i+1, \\ 0, & \textit{otherwise}. \end{cases} \tag{8.13}$$

Write $\tilde{\mathbf{P}}^{N,r}$, $r \in \{1, \ldots, N\}$, for the corresponding family of laws. If a sequence $(r_N)_{N\in\mathbb{N}}$, $r_N \in \{1, \ldots, N\}$, is such that $\lim_{N\to\infty} N^{-1/2} r_N = r_\infty$ exists, then the law of $N^{-1/2}\left(\tilde{R}^N_{t\sqrt{N}}\right)_{t\geq 0}$ under $\tilde{\mathbf{P}}^{N,r_N}$ converges to that of the Rayleigh process $(R_t)_{t\geq 0}$ under $\mathbf{P}^{r_\infty}$ in the usual sense of convergence of càdlàg processes with the Skorohod topology.

Acknowledgment: We thank Jean-François Le Gall and an anonymous referee for several helpful comments that improved the presentation and correctness of the paper.

References

[Ald90] David J. Aldous, *The random walk construction of uniform spanning trees and uniform labelled trees*, SIAM J. Discrete Math. **3** (1990), no. 4, 450–465. MR 1069105

[Ald91a] David Aldous, *The continuum random tree I*, Ann. Probab. **19** (1991), 1–28. MR 1085326

[Ald91b] David Aldous, *The continuum random tree. II. An overview*, Stochastic analysis (Durham, 1990), London Math. Soc. Lecture Note Ser., vol. 167, MR 1166406

[Ald93] David Aldous, *The continuum random tree III*, Ann. Probab. **21** (1993), 248–289. MR 1207226

[Ald94a] David Aldous, *Recursive self-similarity for random trees, random triangulations and Brownian excursion*, Ann. Probab. **22** (1994), no. 2, 527–545. MR 1288122

[Ald94b] ———, *Triangulating the circle, at random*, Amer. Math. Monthly **101** (1994), no. 3, 223–233. MR 1264002

[Ald00] David Aldous, *Mixing time for a Markov chain on cladograms*, Combinatorics, Probability, and Computing **9** (2000), 191–204. MR 1774749

[AT89] D.J. Aldous and P. Tsoucas, *A proof of the Markov chain theorem*, Stat. Probab. Letters **8** (1989), 189–192. MR 1017890

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov, *A course in metric geometry*, Graduate studies in mathematics, vol. 33, AMS, Boston, MA, 2001. MR 1835418

[Bes02] Mladen Bestvina, $\mathbb{R}$-*trees in topology, geometry, and group theory*, Handbook of geometric topology, North-Holland, Amsterdam, 2002, MR 1886668

[BH99] Martin R. Bridson and André Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319, Springer-Verlag, Berlin, MR 1744486

[BHV01] Louis J. Billera, Susan P. Holmes, and Karen Vogtman, *Geometry of the space of phylogenetic trees*, Adv. Appl. Math. **27** (2001), 733–767. MR 1867931

[Bro89] A. Broder, *Generating random spanning trees*, 30th IEEE Symp. Found. Comp. Sci., 1989, pp. 442–447.

[BS02] Andrei N. Borodin and Paavo Salminen, *Handbook of Brownian motion—facts and formulae*, second ed., Probability and its Applications, MR 1912205

[Cai93] Haiyan Cai, *Piecewise deterministic Markov processes*, Stochastic Anal. MR 1220886

[CDP01] G. Colombo and P. Dai Pra, *A class of piecewise deterministic Markov processes*, Markov Process. Related Fields **7** (2001), no. 2, MR 1856497

[Chi01] Ian Chiswell, *Introduction to* Λ*-trees*, World Scientific MR 1851337

[Cos90] O. L. V. Costa, *Stationary distributions for piecewise-deterministic Markov processes*, J. Appl. Probab. **27** (1990), no. 1, 60–73. MR 1039184

[Dav84] M. H. A. Davis, *Piecewise-deterministic Markov processes: a general class of nondiffusion stochastic models*, J. Roy. Statist. Soc. Ser. B MR 790622

[Dav93] ———, *Markov models and optimization*, Monographs on Statistics and MR 1283589

[DC99] François Dufour and Oswaldo L. V. Costa, *Stability of piecewise-deterministic Markov processes*, SIAM J. Control Optim. MR 1710229

[DGR02] Vincent Dumas, Fabrice Guillemin, and Philippe Robert, *A Markovian analysis of additive-increase multiplicative-decrease algorithms*, Adv. in MR 1895332

[DH98] Persi Diaconis and Susan Holmes, *Matching and phylogenetic trees*, Proc. Nat. Acad. Sci. U.S.A. **53** (1998), 321–402. MR 1665632

[DLG02] Thomas Duquesne and Jean-François Le Gall, *Random trees, Lévy processes and spatial branching processes*, Astérisque (2002), no. 281, MR 1954248

[DLG04] ———, *Probabilistic and fractal aspects of Lévy trees*, Preprint, 2004.

[DMT96] Andreas Dress, Vincent Moulton, and Werner Terhalle, *T-theory: an overview*, European J. Combin. **17** (1996), no. 2-3, 161–175, MR 1379369

[Dre84] Andreas W.M. Dress, *Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: A note on combinatorical properties of metric spaces*, Adv. Math. **53** (1984), 321–402. MR 753872

[DT96] Andreas W.M. Dress and W.F. Terhalle, *The real tree*, Adv. Math. **120** (1996), 283–301. MR 1397084

[EK86] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes*, John Wiley & Sons Inc., New York, 1986, Characterization and convergence. MR 838085

[EP98] Steven N. Evans and Jim Pitman, *Stationary Markov processes related to stable Ornstein-Uhlenbeck processes and the additive coalescent*, MR 1649003

[Eva00] Steven N. Evans, *Snakes and spiders: Brownian motion on* $\mathbb{R}$*-trees*, Probab. Theory Related Fields **117** (2000), no. 3, MR 1774068

[Fel03] Joseph Felsenstein, *Inferring phylogenies*, Sinauer Associates, 2003.

[Gro99] Misha Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999, Based on the 1981 French original [MR 85e:53051], With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean MR 1699320

[Has70] A. M. Hasofer, *On the derivative and the upcrossings of the Rayleigh process*, Austral. J. Statist. **12** (1970), 150–151. MR 298755

[JS96] Jean Jacod and Anatolii V. Skorokhod, *Jumping Markov processes*, Ann. Inst. Henri Poincaré **32** (1996), 11–67. MR 1373726

[LG99] Jean-François Le Gall, *Spatial branching processes, random snakes and partial differential equations*, Lectures in Mathematics ETH Zürich, MR 1714707

[MBB58] K. S. Miller, R. I. Bernstein, and L. E. Blumenson, *Rayleigh processes*, MR 94862

[Mor92] John W. Morgan, Λ*-trees and their applications*, Bull. Amer. MR 1100579

[Pau88] Frédéric Paulin, *Topologie de Gromov équivariante, structures hyperboliques et arbres réels*, Invent. Math. **94** (1988), no. 1, MR 958589

[Pau89] ______, *The Gromov topology on* $\mathbf{R}$*-trees*, Topology Appl. MR 1007101

[Pit02a] J. Pitman, *Combinatorial stochastic processes*, Tech. Report 621, Dept. Statistics, U.C. Berkeley, 2002, Lecture notes for St. Flour course, July 2002. Available via `http://www.stat.berkeley.edu/tech-reports/`.

[Pit02b] Boris Pittel, *Note on exact and asymptotic distributions of the parameters of the loop-erased random walk on the complete graph*, Mathematics and computer science, II (Versailles, 2002), Trends Math., Birkhäuser, MR 1940151

[PR04] Yuval Peres and David Revelle, *Scaling limits of the uniform spanning tree and loop-erased random walk on finite graphs*, Preprint. arXiv:math.PR/0410430, 2004.

[Prü18] H. Prüfer, *Neuer Beweis eines Satzes über Permutationen*, Arch. Math. Phys. **27** (1918), 742–744.

[PW98] James Gary Propp and David Bruce Wilson, *How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph*, J. Algorithms **27** (1998), no. 2, 170–217, 7th Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996). MR 1622393

[Sha87] Peter B. Shalen, *Dendrology of groups: an introduction*, Essays in group theory, Math. Sci. Res. Inst. Publ., vol. 8, Springer, New York, 1987, MR 919830

[Sha91] ______, *Dendrology and its applications*, Group theory from a geometrical viewpoint (Trieste, 1990), World Sci. Publishing, River Edge, NJ, MR 1170376

[SS03] Charles Semple and Mike Steel, *Phylogenetics*, Oxford Lecture Series in Mathematics and its Applications, vol. 24, Oxford Univ. Press, Oxford, 2003. MR 2060009

[Ter97] W.F. Terhalle, *R-trees and symmetric differences of sets*, Europ. J. Combinatorics **18** (1997), 825–833. MR 1478827

[Wil96] David Bruce Wilson, *Generating random spanning trees more quickly than the cover time*, Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996) (New York), ACM, 1996, MR 1427525

[WP96] David Bruce Wilson and James Gary Propp, *How to get an exact sample from a generic Markov chain and sample a random spanning tree from a directed graph, both within the cover time*, Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996) (New York), MR 1381954

Department of Statistics #3860, University of California at Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860, U.S.A

E-mail address: `evans@stat.Berkeley.EDU`

Department of Statistics #3860, University of California at Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860, U.S.A

E-mail address: `pitman@stat.Berkeley.EDU`

Mathematisches Institut, Universität Erlangen–Nürnberg, Bismarckstrasse $1\frac{1}{2}$, 91054 Erlangen, GERMANY

E-mail address: `winter@mi.uni-erlangen.de`