SUBTREE PRUNE AND RE-GRAFT: A REVERSIBLE REAL TREE VALUED MARKOV PROCESS

STEVEN N. EVANS AND ANITA WINTER

Abstract. We use Dirichlet form methods to construct and analyze a reversible Markov process, the stationary distribution of which is the Brownian continuum random tree. This process is inspired by the subtree prune and re-graft (SPR) Markov chains that appear in phylogenetic analysis.

A key technical ingredient in this work is the use of a novel Gromov–Hausdorff type distance to metrize the space whose elements are compact real trees equipped with a probability measure. Also, the investigation of the Dirichlet form hinges on a new path decomposition of the Brownian excursion.

Short title: Subtree prune and re-graft

1. Introduction

Markov chains that move through a space of finite trees are an important ingredient for several algorithms in phylogenetic analysis, particularly in Markov chain Monte Carlo algorithms for simulating distributions on spaces of trees in Bayesian tree reconstruction and in simulated annealing algorithms in maximum likelihood and maximum parsimony[1] tree reconstruction (see, for example, [Fel03] for a comprehensive overview of the field). Usually, such chains are based on a set of simple rearrangements that transform a tree into a "neighboring" tree. One widely used set of moves is the *nearest neighbor interchanges* (NNI) (see, for example, [Fel03, BRST02, BHV01, AS01]). Two other standard sets of moves that are implemented in several phylogenetic software packages but seem to have received less theoretical attention are the *subtree prune and re-graft* (SPR) moves and the *tree bisection and reconnection* (TBR) moves that were first described in [SO90] and are further discussed in [Fel03, AS01, SS03]. We note that an NNI move is a particular

Date: September 30, 2005.

2000 *Mathematics Subject Classification.* Primary: 60J25, 60J75; Secondary: 92B10.

Key words and phrases. Dirichlet form, continuum random tree, Brownian excursion, phylogenetic tree, Markov chain Monte Carlo, simulated annealing, path decomposition, excursion theory, Gromov-Hausdorff metric, Prohorov metric.

SNE supported in part by NSF grants DMS-0071468 and DMS-0405778.

[1] *Maximum parsimony* tree reconstruction is based on finding the phylogenetic tree and inferred ancestral states that minimize the total number of obligatory inferred substitution events on the edges of the tree.

type of SPR move and that an SPR move is particular type of TBR move, and, moreover, that every TBR operation is either a single SPR move or the composition of two such moves (see, for example, Section 2.6 of [SS03]). Chains based on other moves are investigated in [DH02, Ald00, Sch02].

In an SPR move, a binary tree T (that is, a tree in which all non-leaf vertices have degree three) is cut "in the middle of an edge" to give two subtrees, say T' and T''. Another edge is chosen in T', a new vertex is created "in the middle" of that edge, and the cut edge in T'' is attached to this new vertex. Lastly, the "pendant" cut edge in T' is removed along with the vertex it was attached to in order to produce a new binary tree that has the same number of vertices as T.

As remarked in [AS01],

> The SPR operation is of particular interest as it can be used to model biological processes such as horizontal gene transfer[2] and recombination.

Section 2.7 of [SS03] provides more background on this point as well as a comment on the role of SPR moves in the two phenomena of lineage sorting and gene duplication and loss.

In this paper we investigate the asymptotics of the simplest possible tree-valued Markov chain based on the SPR moves, namely the chain in which the two edges that are chosen for cutting and for re-attaching are chosen uniformly (without replacement) from the edges in the current tree. Intuitively, the continuous time Markov process we discuss arises as limit when the number of vertices in the tree goes to infinity, the edge lengths are re-scaled by a constant factor so that initial tree converges in a suitable sense to a continuous analogue of a combinatorial tree (more specifically, a compact real tree), and the time scale of the Markov chain is sped up by an appropriate factor.

We do not, in fact, prove such a limit theorem. Rather, we use Dirichlet form techniques to establish the existence of a process that has the dynamics one would expect from such a limit. Unfortunately, although Dirichlet form techniques provide powerful tools for constructing and analyzing symmetric Markov processes, they are notoriously inadequate for proving convergence theorems (as opposed to generator or martingale problem characterizations of Markov processes, for example). We therefore leave the problem of establishing a limit theorem to future research.

The Markov process we construct is a pure jump process that is reversible with respect to the distribution of Aldous's continuum random tree (that is, the random tree which arises as the re-scaling limit of uniform random trees with n vertices when $n \to \infty$ and which is also, up to a constant scaling factor, the random tree associated naturally with the standard Brownian

[2] *Horizontal gene transfer* is the transfer of genetic material from one species to another. It is a particularly common phenomenon among bacteria.

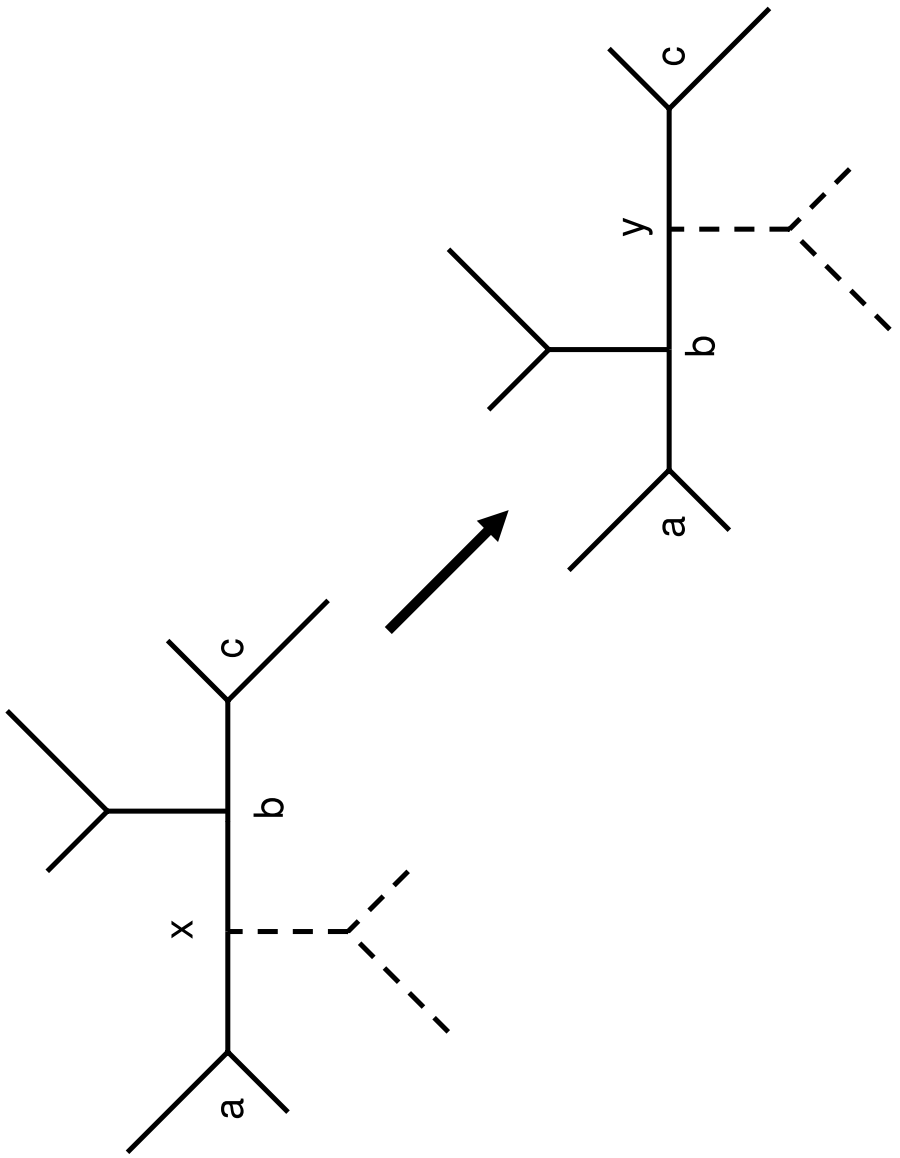

FIGURE 1. An SPR move. The dashed subtree tree attached to vertex x in the top tree is re-attached at a new vertex y that is inserted into the edge (b, c) in the bottom tree to make two edges (b, y) and (y, c). The two edges (a, x) and (b, x) in the top tree are merged into a single edge (a, b) in the bottom tree.

excursion – see Section 4 for more details about the continuum random tree, its connection with Brownian excursion, and references to the literature).

Somewhat more precisely, but still rather informally, the process we construct has the following description.

To begin with, Aldous's continuum random tree has two natural measures on it that can both be thought of as arising from the measure on an approximating finite tree with n vertices that places a unit mass at each vertex. If we re-scale the mass of this measure to get a probability measure, then in the limit we obtain a probability measure on the continuum random tree that happens to assign all of its mass to the leaves with probability one. We call this probability measure the *weight* on the continuum tree. On the other hand, we can also re-scale the measure that places a unit mass at each vertex to obtain in the limit a σ-finite measure on the continuum tree that restricts to one-dimensional Lebesgue measure if we restrict to any path through the continuum tree. We call this σ-finite measure the *length*.

The continuum random tree is a random compact *real tree* of the sort investigated in [EPW04] (we define real trees and discuss some of their properties in Section 2). Any compact real tree has an analogue of the length measure on it, but in general there is no canonical analogue of the weight measure. Consequently, the process we construct has as its state space the set of pairs (T, ν), where T is a compact real tree and ν is a probability measure on T. Let μ be the length measure associated with T. Our process jumps away from T by first choosing a pair of points $(u, v) \in T \times T$ according to the rate measure $\mu \otimes \nu$ and then transforming T into a new tree by cutting off the subtree rooted at u that does not contain v and re-attaching this subtree at v. This jump kernel (which typically has infinite total mass – so that jumps are occurring on a dense countable set) is precisely what one would expect for a limit (as the number of vertices goes to infinity) of the particular SPR Markov chain on finite trees described above in which the edges for cutting and re-attachment are chosen uniformly at each stage.

The framework of Dirichlet forms allows us to translate this description into rigorous mathematics. An important preliminary step that we accomplish in Section 2 is to show that it is possible to equip the space of pairs of compact real trees and their accompanying weights with a nice Gromov–Hausdorff like metric that makes this space complete and separable. We note that a Gromov–Hausdorff like metric on more general metric spaces equipped with measures was introduced in [Stu04]. The latter metric is based on the Wasserstein L^2 distance between measures, whereas ours is based on the Prohorov distance. Moreover, we need to understand in detail the Dirichlet form arising from the combination of the jump kernel with the continuum random tree distribution as a reference measure, and we accomplish this in Sections 5 and 6, where we establish the relevant facts from what appears to be a novel path decomposition of the standard Brownian excursion. We construct the Dirichlet form and the resulting process in Section 7. We use potential theory for Dirichlet forms to show in Section 8 that

from almost all starting points (with respect to the continuum random tree reference measure) our process does not hit the trivial tree consisting of a single point.

We remark that excursion path valued Markov processes that are reversible with respect to the distribution of standard Brownian excursion and have *continuous* sample paths have been investigated in [Zam03, Zam02, Zam01], and that these processes can also be thought of as real tree valued diffusion processes that are reversible with respect to the distribution of the continuum random tree. However, we are unaware of a description in which these latter processes arise as limits of natural processes on spaces of finite trees.

2. Weighted $\mathbb{R}$-trees

A metric space (X, d) is a *real tree* ($\mathbb{R}$-tree) if it satisfies the following axioms.

Axiom 0 (Completeness) The space (X, d) is complete.

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 $\psi([0, 1]) = \phi_{\psi(0),\psi(1)}([0, d(\psi(0), \psi(1))])$.

Axiom 1 says simply that there is a unique "unit speed" path between any two points, whereas Axiom 2 implies that the image of any injective path connecting two points coincides with the image of the unique unit speed path, so that it can be re-parameterized to become the unit speed path. Thus, Axiom 1 is satisfied by many other spaces such as $\mathbb{R}^d$ with the usual metric, whereas Axiom 2 expresses the property of "treeness" and is only satisfied by $\mathbb{R}^d$ when $d = 1$. We refer the reader to ([Dre84, DT96, DMT96, Ter97, Chi01]) for background on $\mathbb{R}$-trees. In particular, [Chi01] shows that a number of other definitions are equivalent to the one above. 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,

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

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

Let $\mathbf{T}$ denote the set of isometry classes of compact $\mathbb{R}$-*trees*. In order to equip $\mathbf{T}$ with a metric, recall that the *Hausdorff distance* between two closed subsets A, B of a metric space (X, d) is defined as

$$d_{\mathrm{H}}(A, B) := \inf\{\varepsilon > 0 : A \subseteq U_\varepsilon(B) \text{ and } B \subseteq U_\varepsilon(A)\}, \tag{2.2}$$

where $U_\epsilon(C) := \{x \in X : d(x, C) \leqslant \varepsilon\}$. Based on this notion of distance between closed sets, we define the *Gromov-Hausdorff distance*, $d_{\mathrm{GH}}(X, Y)$, between two metric spaces (X, d_X) and (Y, d_Y) as the infimum of the Hausdorff distance $d_{\mathrm{H}}(X', Y')$ over all metric spaces X' and Y' that are isomorphic to X and Y, respectively, and that are subspaces of some common metric space Z (compare [Gro99, BH99, BBI01]).

A direct application of the previous definition requires an optimal embedding into a space Z which it is not possible to obtain explicitly in most examples. We therefore give an equivalent reformulation which allows us to get estimates on the distance by looking for "matchings" between the two spaces that preserve the two metrics up to an additive error. In order to be more explicit, we require some more notation. A subset $\Re \subseteq X \times Y$ is said to be a *correspondence* between sets X and Y if for each $x \in X$ there exists at least one $y \in Y$ such that $(x, y) \in \Re$, and for each $y \in Y$ there exists at least one $x \in X$ such that $(x, y) \in \Re$. The *distortion* of $\Re$ is defined by

$$\mathrm{dis}(\Re) := \sup\{|d_X(x_1, x_2) - d_Y(y_1, y_2)| : (x_1, y_1), (x_2, y_2) \in \Re\}. \tag{2.3}$$

Then

$$d_{\mathrm{GH}}((X, d_X), (Y, d_Y)) = \frac{1}{2} \inf_{\Re} \mathrm{dis}(\Re), \tag{2.4}$$

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

It is shown in Theorem 1 in [EPW04] that the metric space $(\mathbf{T}, d_{\mathrm{GH}})$ is complete and separable.

In the following we will be interested in compact $\mathbb{R}$-trees $(T, d) \in \mathbf{T}$ equipped with a probability measure ν on the Borel σ-field $\mathcal{B}(T)$. We call such objects weighted compact $\mathbb{R}$-trees and write $\mathbf{T}^{\mathrm{wt}}$ for the space of weight-preserving isometry classes of weighted compact $\mathbb{R}$-trees, where we say that two weighted, compact $\mathbb{R}$-trees (X, d, ν) and (X', d', ν') are weight-preserving isometric if there exists an isometry ϕ between X and X' such that the *push-forward* of ν by ϕ is ν':

$$\nu' = \phi_* \nu := \nu \circ \phi^{-1}. \tag{2.5}$$

It is clear that the property of being weight-preserving isometric is an equivalence relation.

We want to equip $\mathbf{T}^{\mathrm{wt}}$ with a Gromov-Hausdorff type of distance which incorporates the weights on the trees, but first we need to introduce some notions that will be used in the definition.

An *ε-(distorted) isometry* between two metric spaces (X, d_X) and (Y, d_Y) is a (possibly non-measurable) map $f : X \to Y$ such that

$$\mathrm{dis}(f) := \sup\{|d_X(x_1, x_2) - d_Y(f(x_1), f(x_2))| : x_1, x_2 \in X\} \leqslant \varepsilon \tag{2.6}$$

and $f(X)$ is an ε-net in Y.

It is easy to see that if for two metric spaces (X, d_X) and (Y, d_Y) and $\varepsilon > 0$ we have $d_{\mathrm{GH}}\big((X, d_X), (Y, d_Y)\big) < \varepsilon$, then there exists an 2ε-isometry from X to Y (compare Lemma 7.3.28 in [BBI01]). The following Lemma states that we may choose the distorted isometry between X and Y to be measurable if we allow a slightly bigger distortion.

Lemma 2.1. *Let (X, d_X) and (Y, d_Y) be two compact real trees such that $d_{\mathrm{GH}}\big((X, d_X), (Y, d_Y)\big) < \varepsilon$ for some $\varepsilon > 0$. Then there exists a measurable 3ε-isometry from X to Y.*

Proof. If $d_{\mathrm{GH}}\big((X, d_X), (Y, d_Y)\big) < \varepsilon$, then by (2.4) there exists a correspondence $\Re$ between X and Y such that $\mathrm{dis}(\Re) < 2\varepsilon$. Since (X, d_X) is compact there exists a finite ε-net in X. We claim that for each such finite ε-net $S^{X,\varepsilon} = \{x_1, ..., x_{N^\varepsilon}\} \subseteq X$, any set $S^{Y,\varepsilon} = \{y_1, ..., y_{N^\varepsilon}\} \subseteq Y$ such that $(x_i, y_i) \in \Re$ for all $i \in \{1, 2, ..., N^\varepsilon\}$ is an 3ε-net in Y. To see this, fix $y \in Y$. We have to show the existence of $i \in \{1, 2, ..., N^\varepsilon\}$ with $d_Y(y_i, y) < 3\varepsilon$. For that choose $x \in X$ such that $(x, y) \in \Re$. Since $S^{X,\varepsilon}$ is an ε-net in X there exists an $i \in \{1, 2, ..., N^\varepsilon\}$ such that $d_X(x_i, x) < \varepsilon$. $(x_i, y_i) \in \Re$ implies therefore that $|d_X(x_i, x) - d_Y(y_i, y)| \leqslant \mathrm{dis}(\Re) < 2\varepsilon$, and hence $d_Y(y_i, y) < 3\varepsilon$.

Furthermore we may decompose X into N^ε possibly empty measurable disjoint subsets of X by letting $X^{1,\varepsilon} := \mathcal{B}(x_1, \varepsilon)$, $X^{2,\varepsilon} := \mathcal{B}(x_2, \varepsilon) \backslash X^{1,\varepsilon}$, and so on, where $\mathcal{B}(x, r)$ is the open ball $\{x' \in X : d_X(x, x') < r\}$. Then f defined by $f(x) = y_i$ for $x \in X^{i,\varepsilon}$ is obviously a measurable 3ε-isometry from X to Y. □

We also need to recall the definition of the Prohorov distance between two probability measures (see, for example, [EK86]). Given two probability measures μ and ν on a metric space (X, d) with the corresponding collection of closed sets denoted by $\mathcal{C}$, the Prohorov distance between them is

$$d_{\mathrm{P}}(\mu, \nu) := \inf\{\varepsilon > 0 : \mu(C) \leqslant \nu(C^\varepsilon) + \varepsilon \text{ for all } C \in \mathcal{C}\},$$

where $C^\varepsilon := \{x \in X : \inf_{y \in C} d(x, y) < \varepsilon\}$. The Prohorov distance is a metric on the collection of probability measures on X. The following result shows that if we push measures forward with a map having a small distortion, then Prohorov distances can't increase too much.

Lemma 2.2. *Suppose that (X, d_X) and (Y, d_Y) are two metric spaces, $f : X \to Y$ is a measurable map with $\mathrm{dis}(f) \leqslant \varepsilon$, and μ and ν are two probability measures on X. Then*

$$d_{\mathrm{P}}(f_*\mu, f_*\nu) \leqslant d_{\mathrm{P}}(\mu, \nu) + \varepsilon.$$

Proof. Suppose that $d_{\mathrm{P}}(\mu,\nu) < \delta$. By definition, $\mu(C) \leqslant \nu(C^\delta) + \delta$ for all closed sets $C \in \mathcal{C}$. If D is a closed subset of Y, then

$$\begin{aligned} f_*\mu(D) &= \mu(f^{-1}(D)) \\ &\leqslant \mu(\overline{f^{-1}(D)}) \\ &\leqslant \nu(\overline{f^{-1}(D)}^\delta) + \delta \\ &= \nu(f^{-1}(D)^\delta) + \delta. \end{aligned} \tag{2.7}$$

Now $x' \in f^{-1}(D)^\delta$ means there is $x'' \in X$ such that $d_X(x',x'') < \delta$ and $f(x'') \in D$. By the assumption that $\mathrm{dis}(f) \leqslant \varepsilon$, we have $d_Y(f(x'),f(x'')) < \delta + \varepsilon$, and hence $f(x') \in D^{\delta+\varepsilon}$. Thus

$$f^{-1}(D)^\delta \subseteq f^{-1}(D^{\delta+\varepsilon}) \tag{2.8}$$

and we have

$$f_*\mu(D) \leqslant \nu(f^{-1}(D^{\delta+\varepsilon})) + \delta = f_*\nu(D^{\delta+\varepsilon}) + \delta, \tag{2.9}$$

so that $d_{\mathrm{P}}(f_*\mu, f_*\nu) \leqslant \delta + \varepsilon$, as required. □

We are now in a position to define the weighted Gromov-Hausdorff distance between the two compact, weighted $\mathbb{R}$-trees (X, d_X, ν_X) and (Y, d_Y, ν_Y). For $\varepsilon > 0$, set

$$F^\varepsilon_{X,Y} := \big\{ \text{ measurable } \varepsilon\text{-isometries from } X \text{ to } Y \big\}. \tag{2.10}$$

Put

$$\begin{aligned} &\Delta_{\mathrm{GH}^{\mathrm{wt}}}(X,Y) \\ &:= \inf \left\{ \varepsilon > 0 : \begin{array}{c} \text{exist } f \in F^\varepsilon_{X,Y}, g \in F^\varepsilon_{Y,X} \text{ such that} \\ d_{\mathrm{P}}(f_*\nu_X, \nu_Y) \leqslant \varepsilon,\ d_{\mathrm{P}}(\nu_X, g_*\nu_Y) \leqslant \varepsilon \end{array} \right\}. \end{aligned} \tag{2.11}$$

Note that the set on the right hand side is non-empty because X and Y are compact, and hence bounded. It will turn out that $\Delta_{\mathrm{GH}^{\mathrm{wt}}}$ satisfies all the properties of a metric except the triangle inequality. To rectify this, let

$$d_{\mathrm{GH}^{\mathrm{wt}}}(X,Y) := \inf \left\{ \sum_{i=1}^{n-1} \Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_i, Z_{i+1})^{\frac{1}{4}} \right\}, \tag{2.12}$$

where the infimum is taken over all finite sequences of compact, weighted $\mathbb{R}$-trees $Z_1, \dots Z_n$ with $Z_1 = X$ and $Z_n = Y$.

Lemma 2.3. *The map $d_{\mathrm{GH}^{\mathrm{wt}}} : \mathbf{T}^{\mathrm{wt}} \times \mathbf{T}^{\mathrm{wt}} \to \mathbb{R}_+$ is a metric on $\mathbf{T}^{\mathrm{wt}}$. Moreover,*

$$\frac{1}{2}\Delta_{\mathrm{GH}^{\mathrm{wt}}}(X,Y)^{\frac{1}{4}} \leqslant d_{\mathrm{GH}^{\mathrm{wt}}}(X,Y) \leqslant \Delta_{\mathrm{GH}^{\mathrm{wt}}}(X,Y)^{\frac{1}{4}}$$

for all $X, Y \in \mathbf{T}^{\mathrm{wt}}$.

Proof. It is immediate from (2.11) that the map $\Delta_{\mathrm{GH}^{\mathrm{wt}}}$ is symmetric.

We next claim that

$$\Delta_{\mathrm{GH}^{\mathrm{wt}}}\big((X,d_X,\nu_X),(Y,d_Y,\nu_Y)\big)=0, \tag{2.13}$$

if and only if (X,d_X,ν_X) and (Y,d_Y,ν_Y) are weight-preserving isometric. The "if" direction is immediate. Note first for the converse that (2.13) implies that for all $\varepsilon>0$ there exists an ε-isometry from X to Y, and therefore, by Lemma 7.3.28 in [BBI01], $d_{\mathrm{GH}}\big((X,d_X),(Y,d_Y)\big)<2\varepsilon$. Thus $d_{\mathrm{GH}}\big((X,d_X),(Y,d_Y)\big)=0$, and it follows from Theorem 7.3.30 of [BBI01] that (X,d_X) and (Y,d_Y) are isometric. Checking the proof of that result, we see that we can construct an isometry $f:X\to Y$ by taking any dense countable set $S\subset X$, any sequence of functions (f_n) such that f_n is an ε_n-isometry with $\varepsilon_n\to 0$ as $n\to\infty$, and letting f be $\lim_k f_{n_k}$ along any subsequence such that the limit exists for all $x\in S$ (such a subsequence exists by the compactness of Y). Therefore, fix some dense subset $S\subset X$ and suppose without loss of generality that we have an isometry $f:X\to Y$ given by $f(x)=\lim_{n\to\infty}f_n(x)$, $x\in S$, where $f_n\in F^{\varepsilon_n}_{X,Y}$, $d_{\mathrm{P}}(f_{n*}\nu_X,\nu_Y)\leqslant\varepsilon_n$, and $\lim_{n\to\infty}\varepsilon_n=0$. We will be done if we can show that $f_*\nu_X=\nu_Y$. If μ_X is a discrete measure with atoms belonging to S, then

$$\begin{aligned} d_{\mathrm{P}}(f_*\nu_X,\nu_Y) &\leqslant \limsup_n \Big[d_{\mathrm{P}}(f_{n*}\nu_X,\nu_Y)+d_{\mathrm{P}}(f_{n*}\mu_X,f_{n*}\nu_X)\\ &\qquad + d_{\mathrm{P}}(f_*\mu_X,f_{n*}\mu_X)+d_{\mathrm{P}}(f_*\nu_X,f_*\mu_X)\Big]\\ &\leqslant 2d_{\mathrm{P}}(\mu_X,\nu_X), \end{aligned} \tag{2.14}$$

where we have used Lemma 2.2 and the fact that $\lim_{n\to\infty}d_{\mathrm{P}}(f_*\mu_X,f_{n*}\mu_X)=0$ because of the pointwise convergence of f_n to f on S. Because we can choose μ_X so that $d_{\mathrm{P}}(\mu_X,\nu_X)$ is arbitrarily small, we see that $f_*\nu_X=\nu_Y$, as required.

Now consider three spaces (X,d_X,ν_X), (Y,d_Y,ν_Y), and (Z,d_Z,ν_Z) in $\mathbf{T}^{\mathrm{wt}}$, and constants $\varepsilon,\delta>0$, such that $\Delta_{\mathrm{GH}^{\mathrm{wt}}}\big((X,d_X,\nu_X),(Y,d_Y,\nu_Y)\big)<\varepsilon$ and $\Delta_{\mathrm{GH}^{\mathrm{wt}}}\big((Y,d_Y,\nu_Y),(Z,d_Z,\nu_Z)\big)<\delta$. Then there exist $f\in F^{\varepsilon}_{X,Y}$ and $g\in F^{\delta}_{Y,Z}$ such that $d_{\mathrm{P}}(f_*\nu_X,\nu_Y)<\varepsilon$ and $d_{\mathrm{P}}(g_*\nu_Y,\nu_Z)<\delta$. Note that $g\circ f\in F^{\varepsilon+\delta}_{X,Z}$. Moreover, by Lemma 2.2

$$d_{\mathrm{P}}((g\circ f)_*\nu_X,\nu_Z)\leqslant d_{\mathrm{P}}(g_*\nu_Y,\nu_Z)+d_{\mathrm{P}}(g_*f_*\nu_X,g_*\nu_Y)<\delta+\varepsilon+\delta. \tag{2.15}$$

This, and a similar argument with the roles of X and Z interchanged, shows that

$$\Delta_{\mathrm{GH}^{\mathrm{wt}}}(X,Z)\leqslant 2\left[\Delta_{\mathrm{GH}^{\mathrm{wt}}}(X,Y)+\Delta_{\mathrm{GH}^{\mathrm{wt}}}(Y,Z)\right]. \tag{2.16}$$

The second inequality in the statement of the lemma is clear. In order to see the first inequality, it suffices to show that for any $Z_1,\dots Z_n$ we have

$$\Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_1,Z_n)^{\frac14}\leqslant 2\sum_{i=1}^{n-1}\Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_i,Z_{i+1})^{\frac14}. \tag{2.17}$$

We will establish (2.17) by induction. The inequality certainly holds when $n = 2$. Suppose it holds for $2, \ldots, n-1$. Write S for the value of the sum on the right hand side of (2.17). Put

$$k := \max\left\{1 \leqslant m \leqslant n-1 : \sum_{i=1}^{m-1} \Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_i, Z_{i+1})^{\frac{1}{4}} \leqslant S/2\right\}. \tag{2.18}$$

By the inductive hypothesis and the definition of k,

$$\Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_1, Z_k)^{\frac{1}{4}} \leqslant 2 \sum_{i=1}^{k-1} \Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_i, Z_{i+1})^{\frac{1}{4}} \leqslant 2(S/2) = S. \tag{2.19}$$

Of course,

$$\Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_k, Z_{k+1})^{\frac{1}{4}} \leqslant S \tag{2.20}$$

By definition of k,

$$\sum_{i=1}^{k} \Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_i, Z_{i+1})^{\frac{1}{4}} > S/2, \tag{2.21}$$

so that once more by the inductive hypothesis,

$$\begin{aligned} \Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_{k+1}, Z_n)^{\frac{1}{4}} &\leqslant 2 \sum_{i=k+1}^{n-1} \Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_i, Z_{i+1})^{\frac{1}{4}} \\ &= 2S - 2\sum_{i=1}^{k} \Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_i, Z_{i+1})^{\frac{1}{4}} \\ &\leqslant S. \end{aligned} \tag{2.22}$$

From (2.19), (2.20), (2.22) and two applications of (2.16) we have

$$\begin{aligned} \Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_1, Z_n)^{\frac{1}{4}} &\leqslant \{4[\Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_1, Z_k) + \Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_k, Z_{k+1}) \\ &\quad + \Delta_{\mathrm{GH}^{\mathrm{wt}}}(Z_{k+1}, Z_n)]\}^{\frac{1}{4}} \\ &\leqslant (4 \times 3 \times S^4)^{\frac{1}{4}} \\ &\leqslant 2S, \end{aligned} \tag{2.23}$$

as required.

It is obvious by construction that $d_{\mathrm{GH}^{\mathrm{wt}}}$ satisfies the triangle inequality. The other properties of a metric follow from the corresponding properties we have already established for $\Delta_{\mathrm{GH}^{\mathrm{wt}}}$ and the bounds in the statement of the lemma which we have already established. □

The procedure we used to construct the weighted Gromov-Hausdorff metric $d_{\mathrm{GH}^{\mathrm{wt}}}$ from the semi-metric $\Delta_{\mathrm{GH}^{\mathrm{wt}}}$ was adapted from a proof in [Kel75] of the celebrated result of Alexandroff and Urysohn on the metrizability of uniform spaces. That proof was, in turn, adapted from earlier work of Frink and Bourbaki. The choice of the power $\frac{1}{4}$ is not particularly special, any sufficiently small power would have worked.

Theorem 2.5 below says that the metric space $(\mathbf{T}^{\mathrm{wt}}, d_{\mathrm{GH}^{\mathrm{wt}}})$ is complete and separable and hence is a reasonable space on which to do probability theory. In order to prove this result, we need a compactness criterion that will be useful in its own right.

Proposition 2.4. *A subset* $\mathbf{D}$ *of* $(\mathbf{T}^{\mathrm{wt}}, d_{\mathrm{GH}^{\mathrm{wt}}})$ *is relatively compact if and only if the subset* $\mathbf{E} := \{(T,d) : (T,d,\nu) \in \mathbf{D}\}$ *in* $(\mathbf{T}, d_{\mathrm{GH}})$ *is relatively compact.*

Proof. The "only if" direction is clear. Assume for the converse that $\mathbf{E}$ is relatively compact. Suppose that $((T_n, d_{T_n}, \nu_{T_n}))_{n\in\mathbb{N}}$ is a sequence in $\mathbf{D}$. By assumption, $((T_n, d_{T_n}))_{n\in\mathbb{N}}$ has a subsequence converging to some point (T, d_T) of $(\mathbf{T}, d_{\mathrm{GH}})$. For ease of notation, we will renumber and also denote this subsequence by $((T_n, d_{T_n}))_{n\in\mathbb{N}}$. For brevity, we will also omit specific mention of the metric on a real tree when it is clear from the context.

By Proposition 7.4.12 in [BBI01], for each $\varepsilon > 0$ there is a finite ε-net T^ε in T and for each $n \in \mathbb{N}$ a finite ε-net $T_n^\varepsilon := \{x_n^{\varepsilon,1}, ..., x_n^{\varepsilon,\#T_n^\varepsilon}\}$ in T_n such that $d_{\mathrm{GH}}(T_n^\varepsilon, T^\varepsilon) \to 0$ as $n \to \infty$. Moreover, we take $\#T_n^\varepsilon = \#T^\varepsilon = N^\varepsilon$, say, for n sufficiently large, and so, by passing to a further subsequence if necessary, we may assume that $\#T_n^\varepsilon = \#T^\varepsilon = N^\varepsilon$ for all $n \in \mathbb{N}$. We may then assume that T_n^ε and T^ε have been indexed so that that $\lim_{n\to\infty} d_{T_n}(x_n^{\varepsilon,i}, x_n^{\varepsilon,j}) = d_T(x^{\varepsilon,i}, x^{\varepsilon,j})$ for $1 \leqslant i, j \leqslant N^\varepsilon$.

We may begin with the balls of radius ε around each point of T_n^ε and decompose T_n into N^ε possibly empty, disjoint, measurable sets $\{T_n^{\varepsilon,1}, ..., T_n^{\varepsilon,N^\varepsilon}\}$ of radius no greater than ε. Define a measurable map $f_n^\varepsilon : T_n \to T_n^\varepsilon$ by $f_n^\varepsilon(x) = x_n^{\varepsilon,i}$ if $x \in T_n^{\varepsilon,i}$ and let g_n^ε be the inclusion map from T_n^ε to T_n. By construction, f_n^ε and g_n^ε are measurable ε-isometries. Moreover, $d_{\mathrm{P}}\left((g_n^\varepsilon)_*(f_n^\varepsilon)_*\nu_n, \nu_n\right) < \varepsilon$ and, of course, $d_{\mathrm{P}}\left((f_n^\varepsilon)_*\nu_n, (f_n^\varepsilon)_*\nu_n\right) = 0$. Thus,

$$\Delta_{\mathrm{GH}^{\mathrm{wt}}}\left((T_n^\varepsilon, (f_n^\varepsilon)_*\nu_n), (T_n, \nu_n)\right) \leqslant \varepsilon.$$

By similar reasoning, if we define $h_n^\varepsilon : T_n^\varepsilon \to T^\varepsilon$ by $x_n^{\varepsilon,i} \mapsto x^{\varepsilon,i}$, then

$$\lim_{n\to\infty} \Delta_{\mathrm{GH}^{\mathrm{wt}}}\left((T_n^\varepsilon, (f_n^\varepsilon)_*\nu_n), (T^\varepsilon, (h_n^\varepsilon)_*\nu_n)\right) = 0.$$

Since T^ε is finite, by passing to a subsequence (and relabeling as before) we have

$$\lim_{n\to\infty} d_{\mathrm{P}}\left((h_n^\varepsilon)_*\nu_n, \nu^\varepsilon\right) = 0$$

for some probability measure ν^ε on T^ε, and hence

$$\lim_{n\to\infty} \Delta_{\mathrm{GH}^{\mathrm{wt}}}\left((T^\varepsilon, (h_n^\varepsilon)_*\nu_n), (T^\varepsilon, \nu^\varepsilon)\right) = 0.$$

Therefore, by Lemma 2.3,

$$\limsup_{n\to\infty} d_{\mathrm{GH}^{\mathrm{wt}}}\left((T_n, \nu_n), (T^\varepsilon, (h_n^\varepsilon)_*\nu_n)\right) \leqslant \varepsilon^{\frac{1}{4}}.$$

Now, since (T, d_T) is compact, the family of measures $\{\nu^\varepsilon : \varepsilon > 0\}$ is relatively compact, and so there is a probability measure ν on T such

that ν^ε converges to ν in the Prohorov distance along a subsequence $\varepsilon \downarrow 0$ and hence, by arguments similar to the above, along the same subsequence $\Delta_{\mathrm{GH}^{\mathrm{wt}}}((T^\varepsilon, \nu^\varepsilon), (T, \nu))$ converges to 0. Again applying Lemma 2.3, we have that $d_{\mathrm{GH}^{\mathrm{wt}}}((T^\varepsilon, \nu^\varepsilon), (T, \nu))$ converges to 0 along this subsequence.

Combining the foregoing, we see that by passing to a suitable subsequence and relabeling, $d_{\mathrm{GH}^{\mathrm{wt}}}((T_n, \nu_n), (T, \nu))$ converges to 0, as required. □

Theorem 2.5. *The metric space* $(\mathbf{T}^{\mathrm{wt}}, d_{\mathrm{GH}^{\mathrm{wt}}})$ *is complete and separable.*

Proof. Separability follows readily from separability of $(\mathbf{T}, d_{\mathrm{GH}})$ (see Theorem 1 in [EPW04]), and the separability with respect to the Prohorov distance of the probability measures on a fixed complete, separable metric space (see, for example, [EK86]), and Lemma 2.3.

It remains to establish completeness. By a standard argument, it suffices to show that any Cauchy sequence in $\mathbf{T}^{\mathrm{wt}}$ has a convergent subsequence. Let $(T_n, d_{T_n}, \nu_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathbf{T}^{\mathrm{wt}}$. Then $(T_n, d_{T_n})_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbf{T}$ by Lemma 2.3. By Theorem 1 in [EPW04] there is a $T \in \mathbf{T}$ such that $d_{\mathrm{GH}}(T_n, T) \to 0$, as $n \to \infty$. In particular, the sequence $(T_n, d_{T_n})_{n\in\mathbb{N}}$ is relatively compact in $\mathbf{T}$, and therefore, by Proposition 2.4, $(T_n, d_{T_n}, \nu_n)_{n\in\mathbb{N}}$ is relatively compact in $\mathbf{T}^{\mathrm{wt}}$. Thus $(T_n, d_{T_n}, \nu_n)_{n\in\mathbb{N}}$ has a convergent subsequence, as required. □

We conclude this section by giving a necessary and sufficient condition for a subset of $(\mathbf{T}, d_{\mathrm{GH}})$ to be relatively compact, and hence, by Proposition 2.4, a necessary and sufficient condition for a subset of $(\mathbf{T}^{\mathrm{wt}}, d_{\mathrm{GH}^{\mathrm{wt}}})$ to be relatively compact.

Fix $(T, d) \in \mathbf{T}$, and, as usual, denote the Borel-σ-algebra on T by $\mathcal{B}(T)$. Let

$$T^o = \bigcup\nolimits_{a,b\in T}]a, b[\tag{2.24}$$

the *skeleton* of T. Observe that if $T' \subset T$ is a dense countable set, then (2.24) 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 exists a unique σ-finite measure μ^T on T, called *length measure*, such that $\mu^T(T \backslash T^o) = 0$ and

$$\mu^T(]a, b[) = d(a, b), \quad \forall\, a, b \in T. \tag{2.25}$$

Such a measure may be constructed as the trace onto T^o of one-dimensional Hausdorff measure on T, and a standard monotone class argument shows that this is the unique measure with property (2.25).

For $\varepsilon > 0$, $T \in \mathbf{T}$, and $\rho \in T$ write

$$R_\varepsilon(T, \rho) := \{x \in T :\, \exists\, y \in T, [\rho, y] \ni x,\, d_T(x, y) \geqslant \varepsilon\} \cup \{\rho\}. \tag{2.26}$$

for the *ε-trimming relative to the root ρ* of the compact $\mathbb{R}$-tree T. Then set

$$R_\varepsilon(T) := \begin{cases} \bigcap_{\rho \in T} R_\varepsilon(T,\rho), & \operatorname{diam}(T) > \varepsilon, \\ \text{singleton}, & \operatorname{diam}(T) \leqslant \varepsilon, \end{cases} \tag{2.27}$$

where by *singleton* we mean the trivial $\mathbb{R}$-tree consisting of one point. The tree $R_\varepsilon(T)$ is called the *ε-trimming* of the compact $\mathbb{R}$-tree T.

Lemma 2.6. *A subset $\mathbf{E}$ of $(\mathbf{T}, d_{\mathrm{GH}})$ is relatively compact if and only if for all $\varepsilon > 0$,*

$$\sup\{\mu^T(R_\varepsilon(T)) : T \in \mathbf{E}\} < \infty. \tag{2.28}$$

Proof. The "only if" direction follows from the fact that $T \mapsto \mu^T(R_\varepsilon(T))$ is continuous, which is essentially Lemma 7.3 of [EPW04].

Conversely, suppose that (2.28) holds. Given $T \in \mathbf{E}$, an ε-net for $R_\varepsilon(T)$ is a 2ε-net for T. By Lemma 2.7 below, $R_\varepsilon(T)$ has an ε-net of cardinality at most $\left[(\frac{\varepsilon}{2})^{-1}\mu^T(R_\varepsilon(T))\right]\left[(\frac{\varepsilon}{2})^{-1}\mu^T(R_\varepsilon(T)) + 1\right]$. By assumption, the last quantity is uniformly bounded in $T \in \mathbf{E}$. Thus $\mathbf{E}$ is uniformly totally bounded and hence is relatively compact by Theorem 7.4.15 of [BBI01]. □

Lemma 2.7. *Let $T \in \mathbf{T}$ be such that $\mu^T(T) < \infty$. For each $\varepsilon > 0$ there is an ε-net for T of cardinality at most $\left[(\frac{\varepsilon}{2})^{-1}\mu^T(T)\right]\left[(\frac{\varepsilon}{2})^{-1}\mu^T(T) + 1\right]$*

Proof. Note that an $\frac{\varepsilon}{2}$-net for $R_{\frac{\varepsilon}{2}}(T)$ will be an ε-net for T. The set $T \backslash R_{\frac{\varepsilon}{2}}(T)$ is a collection of disjoint subtrees, one for each leaf of $R_{\frac{\varepsilon}{2}}(T)$, and each such subtree is of diameter at least $\frac{\varepsilon}{2}$. Thus the number of leaves of $R_{\frac{\varepsilon}{2}}(T)$ is at most $(\frac{\varepsilon}{2})^{-1}\mu^T(T)$. Enumerate the leaves of $R_{\frac{\varepsilon}{2}}(T)$ as $x_0, x_1, \ldots, x_n$. Each arc $[x_0, x_i]$, $1 \leqslant i \leqslant n$, of $R_{\frac{\varepsilon}{2}}(T)$ has an $\frac{\varepsilon}{2}$-net of cardinality at most $(\frac{\varepsilon}{2})^{-1}d_T(x_0, x_i) + 1 \leqslant (\frac{\varepsilon}{2})^{-1}\mu^T(T) + 1$. Therefore, by taking the union of these nets, $R_{\frac{\varepsilon}{2}}(T)$ has an $\frac{\varepsilon}{2}$-net of cardinality at most $\left[(\frac{\varepsilon}{2})^{-1}\mu^T(T)\right]\left[(\frac{\varepsilon}{2})^{-1}\mu^T(T) + 1\right]$. □

Remark 2.8. The bound in Lemma 2.7 is far from optimal. It can be shown that T has an ε-net with a cardinality that is of order $\mu^T(T)/\varepsilon$. This is clear for finite trees (that is, trees with a finite number of branch points), where we can traverse the tree with a unit speed path and hence think of the tree as an image of the interval $[0, 2\mu^T(T)]$ by a Lipschitz map with Lipschitz constant 1, so that a covering of the interval $[0, 2\mu^T(T)]$ by ε-balls gives a covering of T by ε-balls. This argument can be extended to arbitrary finite length $\mathbb{R}$-trees, but the details are tedious and so we have contented ourselves with the above simpler bound.

3. Trees and continuous paths

For the sake of completeness and to establish some notation we recall some facts about the connection between continuous excursion paths and trees (see [Ald93, LG99, DLG02] for more on this connection).

Write $C(\mathbb{R}_+)$ for the space of continuous functions from $\mathbb{R}_+$ into $\mathbb{R}$. For $e \in C(\mathbb{R}_+)$, put $\zeta(e) := \inf\{t > 0 : e(t) = 0\}$ and write

$$U := \left\{ e \in C(\mathbb{R}_+) : \begin{array}{c} e(0) = 0,\, \zeta(e) < \infty, \\ e(t) > 0 \text{ for } 0 < t < \zeta(e), \\ \text{and } e(t) = 0 \text{ for } t \geqslant \zeta(e) \end{array} \right\} \tag{3.1}$$

for the space of positive excursion paths. Set $U^\ell := \{e \in U : \zeta(e) = \ell\}$.

We associate each $e \in U^1$ with a compact $\mathbb{R}$-tree as follows. Define an equivalence relation $\sim_e$ on $[0,1]$ by letting

$$u_1 \sim_e u_2, \quad \text{iff} \quad e(u_1) = \inf_{u \in [u_1 \wedge u_2, u_1 \vee u_2]} e(u) = e(u_2). \tag{3.2}$$

Consider the following pseudo-metric on $[0,1]$

$$d_{T_e}(u_1, u_2) := e(u_1) - 2 \inf_{u \in [u_1 \wedge u_2, u_1 \vee u_2]} e(u) + e(u_2), \tag{3.3}$$

which becomes a true metric on the quotient space $T_e := \mathbb{R}_+\big|_{\sim_e} = [0,1]\big|_{\sim_e}$.

Lemma 3.1. *For each $e \in U^1$ the metric space (T_e, d_{T_e}) is a compact $\mathbb{R}$-tree.*

Proof. It is straightforward to check that the quotient map from $[0,1]$ onto T_e is continuous with respect to d_{T_e}. Thus (T_e, d_{T_e}) is path-connected and compact as the continuous image of a metric space with these properties. In particular, (T_e, d_{T_e}) is complete.

To complete the proof, it therefore suffices to verify the *four point condition* (2.1). However, for $u_1, u_2, u_3, u_4 \in T_e$ we have

$$\begin{aligned} \max\{d_{T_e}(u_1, u_3) + d_{T_e}(u_2, u_4), d_{T_e}(u_1, u_4) + d_{T_e}(u_2, u_3)\} \\ \geqslant d_{T_e}(u_1, u_2) + d_{T_e}(u_3, u_4), \end{aligned} \tag{3.4}$$

where strict inequality holds if and only if

$$\begin{aligned} &\min_{i \neq j} \inf_{u \in [u_i \wedge u_j, u_i \vee u_j]} e(u) \\ &\quad \notin \left\{ \inf_{u \in [u_1 \wedge u_2, u_1 \vee u_2]} e(u), \inf_{u \in [u_3 \wedge u_4, u_3 \vee u_4]} e(u) \right\}. \end{aligned} \tag{3.5}$$

□

Remark 3.2. Any compact $\mathbb{R}$-tree T is isometric to T_e for some $e \in U^1$. To see this, fix a root $\rho \in T$. Recall $R_\varepsilon(T, \rho)$, the ε-trimming of T with respect to ρ defined in (2.26). Let $\bar{\mu}$ be a probability measure on T that is equivalent to the length measure μ^T. Because μ^T is σ-finite, such a probability

measure always exists, but one can construct $\bar{\mu}$ explicitly as follows: set $H := \max_{u \in T} d(\rho, u)$, and put

$$\bar{\mu} := 2^{-1} \frac{\mu^T(R(\rho, 2^{-1}H) \cap \cdot)}{\mu^T(R(\rho, 2^{-1}H))} + \sum_{i \geqslant 2} 2^{-i} \frac{\mu^T(R(\rho, 2^{-i}H) \backslash R(\rho, 2^{-i+1}H) \cap \cdot)}{\mu^T(R(\rho, 2^{-i}H) \backslash R(\rho, 2^{-i+1}H))}.$$

For all $0 < \varepsilon < H$ there is a continuous path

$$f_\varepsilon : [0, 2\mu^T(R_\varepsilon(T, \rho))] \to R_\varepsilon(T, \rho)$$

such that h_ε defined by $h_\varepsilon(t) := d(\rho, f_\varepsilon(t))$ belongs to $U^{2\mu^T(R_\varepsilon(T,\rho))}$ (in particular, $f_\varepsilon(0) = f_\varepsilon(2\mu^T(R_\varepsilon(T, \rho))) = \rho$), h_ε is piecewise linear with slopes ± 1, and T_{h_ε} is isometric to $R_\varepsilon(T, \rho)$. Moreover, these paths may be chosen consistently so that if $\varepsilon' \leqslant \varepsilon''$, then

$$f_{\varepsilon''}(t) = f_{\varepsilon'}\left(\inf\{s > 0 : |\{0 \leqslant r \leqslant s \,:\, f_{\varepsilon'}(r) \in R_{\varepsilon''}(T, \rho)\}| > t\}\right),$$

where $|\cdot|$ denotes Lebesgue measure. Now define $e_\varepsilon \in U^{\bar{\mu}(R_\varepsilon(T,\rho))}$ to be the absolutely continuous path satisfying

$$\frac{\mathrm{d}e_\varepsilon(t)}{\mathrm{d}t} = 2 \frac{\mathrm{d}\mu^T}{\mathrm{d}\bar{\mu}}(f_\varepsilon(t)) \frac{\mathrm{d}h_\varepsilon(t)}{\mathrm{d}t}.$$

It can be shown that e_ε converges uniformly to some $e \in U^1$ as $\varepsilon \downarrow 0$ and that T_e is isometric to T.

From the connection we have recalled between excursion paths and real trees, it should be clear that the analogue of an SPR move for a real tree arising from an excursion path is the excision and re-insertion of a sub-excursion. Figure 2 illustrates such an operation.

Each tree coming from a path in U^1 has a natural weight on it: for $e \in U^1$, we equip (T_e, d_{T_e}) with the weight ν_{T_e} given by the push-forward of Lebesgue measure on $[0, 1]$ by the quotient map.

We finish this section with a remark about the natural length measure on a tree coming from a path. Given $e \in U^1$ and $a \geqslant 0$, let

$$\mathcal{G}_a := \left\{ t \in [0,1] : \begin{array}{c} e(t) = a \text{ and, for some } \varepsilon > 0, \\ e(u) > a \text{ for all } u \in]t, t+\varepsilon[, \\ e(t+\varepsilon) = a. \end{array} \right\} \tag{3.6}$$

denote the countable set of starting points of excursions of the function e above the level a. Then μ^{T_e}, the length measure on T_e, is just the push-forward of the measure $\int_0^\infty \mathrm{d}a \sum_{t \in \mathcal{G}_a} \delta_t$ by the quotient map. Alternatively, write

$$\Gamma_e := \{(s, a) : \, s \in]0, 1[, \, a \in [0, e(s)[\} \tag{3.7}$$

for the region between the time axis and the graph of e, and for $(s, a) \in \Gamma_e$ denote by $\underline{s}(e, s, a) := \sup\{r < s : e(r) = a\}$ and $\bar{s}(e, s, a) := \inf\{t > s : e(t) = a\}$ the start and finish of the excursion of e above level a that

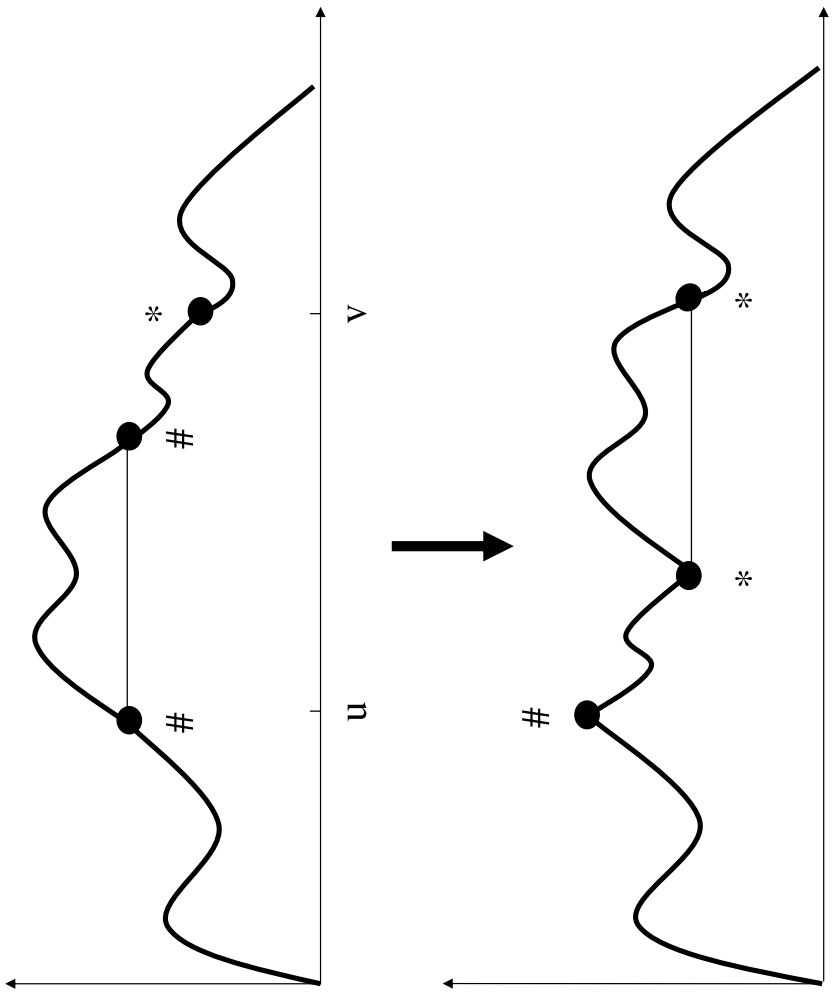

FIGURE 2. A subtree prune and re-graft operation on an excursion path: the excursion starting at time u in the top picture is excised and inserted at time v, and the resulting gap between the two points marked # is closed up. The two points marked # (resp. $*$) in the top (resp. bottom) picture correspond to a single point in the associated $\mathbb{R}$-tree.

straddles time s. Then μ^{T_e} is the push-forward of the measure $\int_{\Gamma_e} \mathrm{d}s \otimes \mathrm{d}a \, \frac{1}{\bar{s}(e,s,a)-\underline{s}(e,s,a)} \delta_{\underline{s}(e,s,a)}$ by the quotient map. We note that the measure μ^{T_e} appears in [AS02].

4. Uniform random weighted compact $\mathbb{R}$-trees: the continuum random tree

In this section we will recall the definition of Aldous's continuum random tree, which can be thought of as a uniformly chosen random weighted compact $\mathbb{R}$-tree.

Consider the Itô excursion measure for excursions of standard Brownian motion away from 0. This σ-finite measure is defined subject to a normalization of Brownian local time at 0, and we take the usual normalization of local times at each level which makes the local time process an occupation density in the spatial variable for each fixed value of the time variable. The excursion measure is the sum of two measures, one which is concentrated on non-negative excursions and one which is concentrated on non-positive excursions. Let $\mathbb{N}$ be the part which is concentrated on non-negative excursions. Thus, in the notation of Section 3, $\mathbb{N}$ is a σ-finite measure on U, where we equip U with the σ-field $\mathcal{U}$ generated by the coordinate maps.

Define a map $v : U \to U^1$ by $e \mapsto \frac{e(\zeta(e)\cdot)}{\sqrt{\zeta(e)}}$. Then

$$\mathbb{P}(\Gamma) := \frac{\mathbb{N}\{v^{-1}(\Gamma) \cap \{e \in U : \zeta(e) \geqslant c\}\}}{\mathbb{N}\{e \in U : \zeta(e) \geqslant c\}}, \quad \Gamma \in \mathcal{U}, \tag{4.1}$$

does not depend on $c > 0$ (see, for example, Exercise 12.2.13.2 in [RY99]). The probability measure $\mathbb{P}$ is called the law of normalized non-negative Brownian excursion. We have

$$\mathbb{N}\{e \in U : \zeta(e) \in \mathrm{d}c\} = \frac{\mathrm{d}c}{2\sqrt{2\pi c^3}} \tag{4.2}$$

and, defining $\mathcal{S}_c : U^1 \to U^c$ by

$$\mathcal{S}_c e := \sqrt{c} e(\cdot/c) \tag{4.3}$$

we have

$$\int \mathbb{N}(\mathrm{d}e)\, G(e) = \int_0^\infty \frac{\mathrm{d}c}{2\sqrt{2\pi c^3}} \int_{U^1} \mathbb{P}(\mathrm{d}e)\, G\left(\mathcal{S}_c e\right) \tag{4.4}$$

for a non-negative measurable function $G : U \to \mathbb{R}$.

Recall from Section 3 how each $e \in U^1$ is associated with a weighted compact $\mathbb{R}$-tree $(T_e, d_{T_e}, \nu_{T_e})$. Let $\mathbf{P}$ be the probability measure on $(\mathbf{T}^{\mathrm{wt}}, d_{\mathrm{GH}^{\mathrm{wt}}})$ that is the push-forward of the normalized excursion measure by the map $e \mapsto (T_{2e}, d_{T_{2e}}, \nu_{T_{2e}})$, where $2e \in U^1$ is just the excursion path $t \mapsto 2e(t)$.

The probability measure $\mathbf{P}$ is the distribution of an object consisting of Aldous's *continuum random tree* along with a natural measure on this tree (see, for example, [Ald91a, Ald93]). The continuum random tree arises as the limit of a uniform random tree on n vertices when $n \to \infty$ and edge lengths

are rescaled by a factor of $1/\sqrt{n}$. The appearance of $2e$ rather than e in the definition of $\mathbf{P}$ is a consequence of this choice of scaling. The associated probability measure on each realization of the continuum random tree is the measure that arises in this limiting construction by taking the uniform probability measure on realizations of the approximating finite trees. The probability measure $\mathbf{P}$ can therefore be viewed informally as the "uniform distribution" on $(\mathbf{T}^{\mathrm{wt}}, d_{\mathrm{GH}^{\mathrm{wt}}})$.

5. Campbell measure facts

For the purposes of constructing the Markov process that is of interest to us, we need to understand picking a random weighted tree (T, d_T, ν_T) according to the continuum random tree distribution $\mathbf{P}$, picking a point u according to the length measure μ^T and another point v according to the weight ν_T, and then decomposing T into two subtrees rooted at u – one that contains v and one that does not (we are being a little imprecise here, because μ^T will be an infinite measure, $\mathbf{P}$ almost surely).

In order to understand this decomposition, we must understand the corresponding decomposition of excursion paths under normalized excursion measure. Because subtrees correspond to sub-excursions and because of our observation in Section 3 that for an excursion e the length measure μ^{T_e} on the corresponding tree is the push-forward of the measure $\int_{\Gamma_e} \mathrm{d}s \otimes \mathrm{d}a \, \frac{1}{\overline{s}(e,s,a)-\underline{s}(e,s,a)} \delta_{\underline{s}(e,s,a)}$ by the quotient map, we need to understand the decomposition of the excursion e into the excursion above a that straddles s and the "remaining" excursion when when e is chosen according to the standard Brownian excursion distribution $\mathbb{P}$ and (s,a) is chosen according to the σ-finite measure $\mathrm{d}s \otimes \mathrm{d}a \, \frac{1}{\overline{s}(e,s,a)-\underline{s}(e,s,a)}$ on Γ_e – see Figure 3.

Given an excursion $e \in U$ and a level $a \geqslant 0$ write:

- $\zeta(e) := \inf\{t > 0 : e(t) = 0\}$ for the "length"of e,
- $\ell_t^a(e)$ for the local time of e at level a up to time t,
- $e^{\downarrow a}$ for e time-changed by the inverse of $t \mapsto \int_0^t \mathrm{d}s \, 1\{e(s) \leqslant a\}$ (that is, $e^{\downarrow a}$ is e with the sub-excursions above level a excised and the gaps closed up),
- $\ell_t^a(e^{\downarrow a})$ for the local time of $e^{\downarrow a}$ at the level a up to time t,
- $U^{\uparrow a}(e)$ for the set of sub-excursion intervals of e above a (that is, an element of $U^{\uparrow a}(e)$ is an interval $I = [g_I, d_I]$ such that $e(g_I) = e(d_I) = a$ and $e(t) > a$ for $g_I < t < d_I$),
- $\mathcal{N}^{\uparrow a}(e)$ for the counting measure that puts a unit mass at each point (s', e'), where, for some $I \in U^{\uparrow a}(e)$, $s' := \ell_{g_I}^a(e)$ is the amount of local time of e at level a accumulated up to the beginning of the sub-excursion I and $e' \in U$ is given by

$$e'(t) = \begin{cases} e(g_I + t) - a, & 0 \leqslant t \leqslant d_I - g_I, \\ 0, & t > d_I - g_I, \end{cases} \tag{5.1}$$

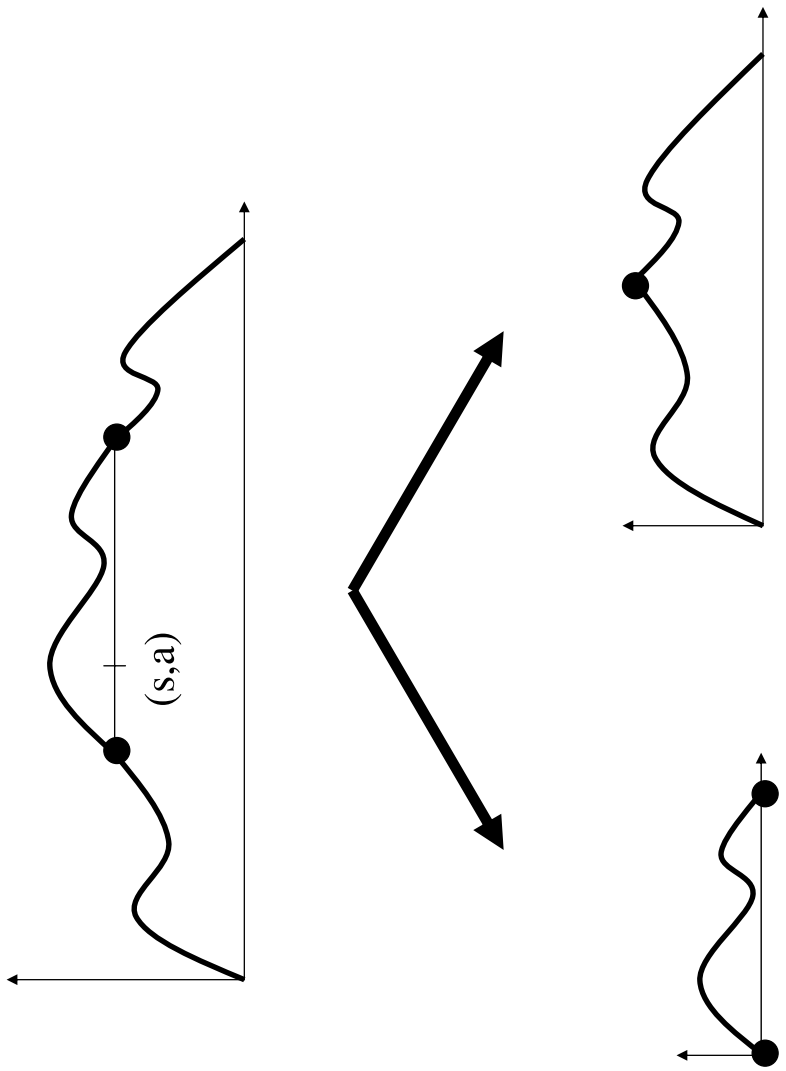

FIGURE 3. The decomposition of the excursion e (top picture) into the excursion $\hat{e}^{s,a}$ above level a that straddles time s (bottom left picture) and the "remaining" excursion $\check{e}^{s,a}$ (bottom right picture).

is the corresponding piece of the path e shifted to become an excursion above the level 0 starting at time 0,

- $\hat{e}^{s,a} \in U$ and $\check{e}^{s,a} \in U$, for the subexcursion "above" $(s,a) \in \Gamma_e$, that is,

$$\hat{e}^{s,a}(t) := \begin{cases} e(\underline{s}(e,s,a)+t) - a, & 0 \leqslant t \leqslant \bar{s}(e,s,a) - \underline{s}(e,s,a), \\ 0, & t > \bar{s}(e,s,a) - \underline{s}(e,s,a), \end{cases} \tag{5.2}$$

respectively "below" $(s,a) \in \Gamma_e$, that is,

$$\check{e}^{s,a}(t) := \begin{cases} e(t), & 0 \leqslant t \leqslant \underline{s}(e,s,a), \\ e(t + \bar{s}(e,s,a) - \underline{s}(e,s,a)), & t > \underline{s}(e,s,a). \end{cases} \tag{5.3}$$

- $\sigma^a_s(e) := \inf\{t \geqslant 0 : \ell^a_t(e) \geqslant s\}$ and $\tau^a_s(e) := \inf\{t \geqslant 0 : \ell^a_t(e) > s\}$,
- $\tilde{e}^{s,a} \in U$ for e with the interval $]\sigma^a_s(e), \tau^a_s(e)[$ containing an excursion above level a excised, that is,

$$\tilde{e}^{s,a}(t) := \begin{cases} e(t), & 0 \leqslant t \leqslant \sigma^a_s(e), \\ e(t + \tau^a_s(e) - \sigma^a_s(e)), & t > \sigma^a_s(e). \end{cases} \tag{5.4}$$

The following path decomposition result under the σ-finite measure $\mathbb{N}$ is preparatory to a decomposition under the probability measure $\mathbb{P}$, Corollary 5.2, that has a simpler intuitive interpretation.

Proposition 5.1. *For non-negative measurable functions F on $\mathbb{R}_+$ and G, H on U,*

$$\begin{aligned} &\int \mathbb{N}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}s \otimes \mathrm{d}a}{\bar{s}(e,s,a) - \underline{s}(e,s,a)} F(\underline{s}(e,s,a)) G(\hat{e}^{s,a}) H(\check{e}^{s,a}) \\ &\quad = \int \mathbb{N}(\mathrm{d}e) \int_0^\infty \mathrm{d}a \int \mathcal{N}^{\uparrow a}(e)(\mathrm{d}(s',\mathrm{e}'))\, F(\sigma^a_{s'}(e)) G(e') H(\tilde{e}^{s',a}) \\ &\quad = \mathbb{N}\big[G\big]\, \mathbb{N}\Big[H \int_0^\zeta \mathrm{d}s\, F(s)\Big]. \end{aligned}$$

Proof. The first equality is just a change in the order of integration and has already been remarked upon in Section 3.

Standard excursion theory (see, for example, [RW00, RY99, Ber96]) says that under $\mathbb{N}$, the random measure $e \mapsto \mathcal{N}^{\uparrow a}(e)$ conditional on $e \mapsto e^{\downarrow a}$ is a Poisson random measure with intensity measure $\lambda^{\downarrow a}(e) \otimes \mathbb{N}$, where $\lambda^{\downarrow a}(e)$ is Lebesgue measure restricted to the interval $[0, \ell^a_\infty(e)] = [0, 2\ell^a_\infty(e^{\downarrow a})]$.

Note that $\tilde{e}^{s',a}$ is constructed from $e^{\downarrow a}$ and $\mathcal{N}^{\uparrow a}(e) - \delta_{(s',e')}$ in the same way that e is constructed from $e^{\downarrow a}$ and $\mathcal{N}^{\uparrow a}(e)$. Also, $\sigma^a_{s'}(\tilde{e}^{s',a}) = \sigma^a_{s'}(e)$. Therefore, by the Campbell-Palm formula for Poisson random measures (see,

for example, Section 12.1 of [DVJ88]),

$$\begin{aligned}
&\int \mathbb{N}(\mathrm{d}e) \int_0^\infty \mathrm{d}a \int \mathcal{N}^{\uparrow a}(e)(\mathrm{d}(s', \mathrm{e}'))\, F(\sigma^a_{s'}(e)) G(e') H(\tilde{e}^{s',a}) \\
&\quad = \int \mathbb{N}(\mathrm{d}e) \int_0^\infty \mathrm{d}a\, \mathbb{N}\Big[\int \mathcal{N}^{\uparrow a}(e)(\mathrm{d}(s', \mathrm{e}'))\, F(\sigma^a_{s'}(e)) G(e') H(\tilde{e}^{s',a}) \,\Big|\, e^{\downarrow a} \Big] \\
&\quad = \int \mathbb{N}(\mathrm{d}e) \int_0^\infty \mathrm{d}a\, \mathbb{N}[G]\, \mathbb{N}\Big[\Big\{ \int_0^{\ell^a_\infty(e)} \mathrm{d}s'\, F(\sigma^a_{s'}(e)) \Big\} H \,\Big|\, e^{\downarrow a} \Big] \\
&\quad = \mathbb{N}[G] \int_0^\infty \mathrm{d}a \int \mathbb{N}(\mathrm{d}e) \left(\Big\{ \int \mathrm{d}\ell^a_s(e)\, F(s) \Big\} H(e) \right) \\
&\quad = \mathbb{N}[G] \int \mathbb{N}(\mathrm{d}e) \left(\Big\{ \int_0^\infty \mathrm{d}a \int \mathrm{d}\ell^a_s(e)\, F(s) \Big\} H(e) \right) \\
&\quad = \mathbb{N}[G]\, \mathbb{N}\Big[H \int_0^\zeta \mathrm{d}s\, F(s) \Big].
\end{aligned}$$

□

The next result says that if we pick an excursion e according to the standard excursion distribution $\mathbb{P}$ and then pick a point $(s, a) \in \Gamma_e$ according to the σ-finite length measure corresponding to the length measure μ^{T_e} on the associated tree T_e (see the end of Section 3), then the following objects are independent:

(a) the length of the excursion above level a that straddles time s,
(b) the excursion obtained by taking the excursion above level a that straddles time s, turning it (by a shift of axes) into an excursion $\hat{e}^{s,a}$ above level zero starting at time zero, and then Brownian re-scaling $\hat{e}^{s,a}$ to produce an excursion of unit length,
(c) the excursion obtained by taking the excursion $\check{e}^{s,a}$ that comes from excising $\hat{e}^{s,a}$ and closing up the gap, and then Brownian re-scaling $\check{e}^{s,a}$ to produce an excursion of unit length,
(d) the starting time $\underline{s}(e, s, a)$ of the excursion above level a that straddles time s rescaled by the length of $\check{e}^{s,a}$ to give a time in the interval $[0, 1]$.

Moreover, the length in (a) is "distributed" according to the σ-finite measure

$$\frac{1}{2\sqrt{2\pi}} \frac{\mathrm{d}\rho}{\sqrt{(1-\rho)\rho^3}}, \quad 0 \leqslant \rho \leqslant 1, \tag{5.5}$$

the unit length excursions in (b) and (c) are both distributed as standard Brownian excursions (that is, according to $\mathbb{P}$), and the time in (d) is uniformly distributed on the interval $[0, 1]$.

Recall from (4.3) that $\mathcal{S}_c : U^1 \to U^c$ is the Brownian re-scaling map defined by

$$\mathcal{S}_c e := \sqrt{c} e(\cdot/c).$$

Corollary 5.2. *For non-negative measurable functions F on $\mathbb{R}_+$ and K on $U \times U$,*

$$\begin{aligned}
&\int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}s \otimes \mathrm{d}a}{\bar{s}(e,s,a) - \underline{s}(e,s,a)} F\Big(\frac{\underline{s}(e,s,a)}{\zeta(\check{e}^{s,a})}\Big) K(\hat{e}^{s,a}, \check{e}^{s,a}) \\
&= \Big\{ \int_0^1 \mathrm{d}u\, F(u) \Big\} \int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}s \otimes \mathrm{d}a}{\bar{s}(e,s,a) - \underline{s}(e,s,a)} K(\hat{e}^{s,a}, \check{e}^{s,a}) \\
&= \Big\{ \int_0^1 \mathrm{d}u\, F(u) \Big\} \frac{1}{2\sqrt{2\pi}} \int_0^1 \frac{\mathrm{d}\rho}{\sqrt{(1-\rho)\rho^3}} \int \mathbb{P}(\mathrm{d}e') \otimes \mathbb{P}(\mathrm{d}e'')\, K(\mathcal{S}_\rho e', \mathcal{S}_{1-\rho} e'').
\end{aligned}$$

Proof. For a non-negative measurable function L on $U \times U$, it follows straightforwardly from Proposition 5.1 that

$$\begin{aligned}
&\int \mathbb{N}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}s \otimes \mathrm{d}a}{\bar{s}(e,s,a) - \underline{s}(e,s,a)} F\Big(\frac{\underline{s}(e,s,a)}{\zeta(\check{e}^{s,a})}\Big) L(\hat{e}^{s,a}, \check{e}^{s,a}) \\
&\qquad = \Big\{ \int_0^1 \mathrm{d}u\, F(u) \Big\} \int \mathbb{N}(\mathrm{d}e') \otimes \mathbb{N}(\mathrm{d}e'')\, L(e', e'') \zeta(e'').
\end{aligned} \tag{5.6}$$

The left-hand side of equation (5.6) is, by (4.4),

$$\int_0^\infty \frac{\mathrm{d}c}{2\sqrt{2\pi c^3}} \int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_{\mathcal{S}_c e}} \mathrm{d}s \otimes \mathrm{d}a\, \frac{F\left(\frac{\underline{s}(\mathcal{S}_c e, s, a)}{\zeta(\widetilde{\mathcal{S}_c e}^{s,a})}\right) L(\widehat{\mathcal{S}_c e}^{s,a}, \widetilde{\mathcal{S}_c e}^{s,a})}{\bar{s}(\mathcal{S}_c e, s, a) - \underline{s}(\mathcal{S}_c e, s, a)}. \tag{5.7}$$

If we change variables to $t = s/c$ and $b = a/\sqrt{c}$, then the integral for (s,a) over $\Gamma_{\mathcal{S}_c e}$ becomes an integral for (t,b) over Γ_e. Also,

$$\begin{aligned}
\underline{s}(\mathcal{S}_c e, ct, \sqrt{c}b) &= \sup\Big\{ r < ct : \sqrt{c} e\left(\frac{r}{c}\right) < \sqrt{c} b \Big\} \\
&= c \sup\{ r < t : e(r) < b \} \\
&= c \underline{s}(e, t, b),
\end{aligned} \tag{5.8}$$

and, by similar reasoning,

$$\bar{s}(\mathcal{S}_c e, ct, \sqrt{c}b) = c \bar{s}(e, t, b) \tag{5.9}$$

and

$$\zeta(\widetilde{\mathcal{S}_c e}^{ct, \sqrt{c}b}) = c \zeta(\check{e}^{t,b}). \tag{5.10}$$

Thus (5.7) is

$$\int_0^\infty \frac{\mathrm{d}c}{2\sqrt{2\pi c^3}} \int \mathbb{P}(\mathrm{d}e)\, \sqrt{c} \int_{\Gamma_e} \mathrm{d}t \otimes \mathrm{d}b\, \frac{F\big(\frac{\underline{s}(e,t,b)}{\zeta(\check{e}^{t,b})}\big) L(\widehat{\mathcal{S}_c e}^{ct, \sqrt{c}b}, \widetilde{\mathcal{S}_c e}^{ct, \sqrt{c}b})}{\bar{s}(e,t,b) - \underline{s}(e,t,b)}. \tag{5.11}$$

Now suppose that L is of the form

$$L(e', e'') = K(\mathcal{R}_{\zeta(e') + \zeta(e'')} e', \mathcal{R}_{\zeta(e') + \zeta(e'')} e'')\, \frac{M(\zeta(e') + \zeta(e''))}{\sqrt{\zeta(e') + \zeta(e'')}}, \tag{5.12}$$

where, for ease of notation, we put for $e \in U$, and $c > 0$,

$$\mathcal{R}_c e := \mathcal{S}_{c^{-1}} e = \frac{1}{\sqrt{c}} e(c\,\cdot). \tag{5.13}$$

Then (5.11) becomes

$$\int_0^\infty \frac{\mathrm{d}c}{2\sqrt{2\pi c^3}} \int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_e} \mathrm{d}t \otimes \mathrm{d}b \, \frac{F\left(\frac{\underline{s}(e,t,b)}{\zeta(\check{e}^{t,b})}\right) K(\hat{e}^{t,b}, \check{e}^{t,b})\, M(c)}{\bar{s}(e,t,b) - \underline{s}(e,t,b)}. \tag{5.14}$$

Since (5.14) was shown to be equivalent to the left hand side of (5.6), it follows from (4.4) that

$$\begin{aligned} &\int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}t \otimes \mathrm{d}b}{\bar{s}(e,t,b) - \underline{s}(e,t,b)} \, F\big(\tfrac{\underline{s}(e,t,b)}{\zeta(\check{e}^{t,b})}\big)\, K(\hat{e}^{t,b}, \check{e}^{t,b}) \\ &\qquad = \frac{\int_0^1 \mathrm{d}u\, F(u)}{\mathbb{N}[M]} \int \mathbb{N}(\mathrm{d}e') \otimes \mathbb{N}(\mathrm{d}e'')\, L(e', e'')\, \zeta(e''), \end{aligned} \tag{5.15}$$

and the first equality of the statement follows.

We have from the identity (5.15) that, for any $C > 0$,

$$\begin{aligned} &\mathbb{N}\{\zeta(e) > C\} \int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}s \otimes \mathrm{d}a}{\bar{s}(e,s,a) - \underline{s}(e,s,a)} \, K(\hat{e}^{s,a}, \check{e}^{s,a}) \\ &= \int \mathbb{N}(\mathrm{d}e') \otimes \mathbb{N}(\mathrm{d}e'')\, K(\mathcal{R}_{\zeta(e')+\zeta(e'')} e', \mathcal{R}_{\zeta(e')+\zeta(e'')} e'') \frac{1\{\zeta(e') + \zeta(e'') > C\}}{\sqrt{\zeta(e') + \zeta(e'')}} \zeta(e'') \\ &= \int_0^\infty \frac{\mathrm{d}c'}{2\sqrt{2\pi c'^3}} \int_0^\infty \frac{\mathrm{d}c''}{2\sqrt{2\pi c''}} \\ &\qquad \int \mathbb{P}(\mathrm{d}e') \otimes \mathbb{P}(\mathrm{d}e'')\, K(\mathcal{R}_{c'+c''} \mathcal{S}_{c'} e', \mathcal{R}_{c'+c''} \mathcal{S}_{c''} e'') \frac{1\{c' + c'' > C\}}{\sqrt{c' + c''}}. \end{aligned}$$

Make the change of variables $\rho = \frac{c'}{c'+c''}$ and $\xi = c' + c''$ (with corresponding Jacobian factor ξ) to get

$$\begin{aligned} &\int_0^\infty \frac{\mathrm{d}c'}{2\sqrt{2\pi c'^3}} \int_0^\infty \frac{\mathrm{d}c''}{2\sqrt{2\pi c''}} \\ &\qquad \int \mathbb{P}(\mathrm{d}e') \otimes \mathbb{P}(\mathrm{d}e'')\, K(\mathcal{R}_{c'+c''} \mathcal{S}_{c'} e', \mathcal{R}_{c'+c''} \mathcal{S}_{c''} e'') \frac{1\{c' + c'' > C\}}{\sqrt{c' + c''}} \\ &= \left(\frac{1}{2\sqrt{2\pi}}\right)^2 \int_0^\infty d\xi \int_0^1 \frac{\mathrm{d}\rho\, \xi}{\sqrt{\rho^3(1-\rho)\xi^4}} \frac{1\{\xi > C\}}{\sqrt{\xi}} \\ &\qquad \int \mathbb{P}(\mathrm{d}e') \otimes \mathbb{P}(\mathrm{d}e'')\, K(\mathcal{S}_\rho e', \mathcal{S}_{1-\rho} e'') \\ &= \left(\frac{1}{2\sqrt{2\pi}}\right)^2 \left\{\int_C^\infty \frac{d\xi}{\sqrt{\xi^3}}\right\} \int_0^1 \frac{\mathrm{d}\rho}{\sqrt{\rho^3(1-\rho)}} \\ &\qquad \int \mathbb{P}(\mathrm{d}e') \otimes \mathbb{P}(\mathrm{d}e'')\, K(\mathcal{S}_\rho e', \mathcal{S}_{1-\rho} e''), \end{aligned}$$

and the corollary follows upon recalling (4.2). □

Corollary 5.3. (i) *For $x > 0$,*

$$\int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}s \otimes \mathrm{d}a}{\bar{s}(e,s,a) - \underline{s}(e,s,a)} 1\{\max_{0 \leqslant t \leqslant \zeta(\hat{e}^{s,a})} \hat{e}^{s,a} > x\}$$
$$= 2 \sum_{n=1}^{\infty} nx \, \exp(-2n^2x^2)$$

(ii) *For $0 < p \leqslant 1$,*

$$\int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}s \otimes \mathrm{d}a}{\bar{s}(e,s,a) - \underline{s}(e,s,a)} 1\{\zeta(\hat{e}^{s,a}) > p\} = \sqrt{\frac{1-p}{2\pi p}}.$$

Proof. (i) Recall first of all from Theorem 5.2.10 in [Kni81] that

$$\mathbb{P}\left\{e \in U^1 : \max_{0 \leqslant t \leqslant 1} e(t) > x\right\} = 2 \sum_{n=1}^{\infty} (4n^2x^2 - 1) \exp(-2n^2x^2). \tag{5.16}$$

By Corollary 5.2 applied to $K(e', e'') := 1\{\max_{t \in [0, \zeta(e')]} e'(t) \geqslant x\}$ and $F \equiv 1$,

$$\int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}s \otimes \mathrm{d}a}{\bar{s}(e,s,a) - \underline{s}(e,s,a)} 1\{\max_{0 \leqslant t \leqslant \zeta(\hat{e}^{s,a})} \hat{e}^{s,a} > x\}$$
$$= \frac{1}{2\sqrt{2\pi}} \int_0^1 \frac{\mathrm{d}\rho}{\sqrt{\rho^3(1-\rho)}} \, \mathbb{P}\left\{\max_{t \in [0,\rho]} \sqrt{\rho} e(t/\rho) > x\right\}$$
$$= \frac{1}{2\sqrt{2\pi}} \int_0^1 \frac{\mathrm{d}\rho}{\sqrt{\rho^3(1-\rho)}} \, \mathbb{P}\left\{\max_{t \in [0,1]} e(t) > \frac{x}{\sqrt{\rho}}\right\}$$
$$= \frac{1}{2\sqrt{2\pi}} \int_0^1 \frac{\mathrm{d}\rho}{\sqrt{\rho^3(1-\rho)}} \, 2 \sum_{n=1}^{\infty} \left(4n^2 \frac{x^2}{\rho} - 1\right) \exp\left(-2n^2 \frac{x^2}{\rho}\right)$$
$$= 2 \sum_{n=1}^{\infty} nx \, \exp(-2n^2x^2),$$

as claimed.
(ii) Corollary 5.2 applied to $K(e', e'') := 1\{\zeta(e') \geqslant p\}$ and $F \equiv 1$ immediately yields

$$\int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}s \otimes \mathrm{d}a}{\bar{s}(e,s,a) - \underline{s}(e,s,a)} 1\{\zeta(\hat{e}^{s,a}) > p\}$$
$$= \frac{1}{2\sqrt{2\pi}} \int_p^1 \frac{\mathrm{d}\rho}{\sqrt{\rho^3(1-\rho)}} = \sqrt{\frac{1-p}{2\pi p}}.$$

□

We conclude this section by calculating the expectations of some functionals with respect to $\mathbf{P}$ (the "uniform distribution" on $(\mathbf{T}^{\mathrm{wt}}, d_{\mathrm{GH}^{\mathrm{wt}}})$ as introduced in the end of Section 4).

For $T \in \mathbf{T}^{\mathrm{wt}}$, and $\rho \in T$, recall $R_c(T,\rho)$ from (2.26), and the length measure μ^T from (2.25). Given $(T,d) \in \mathbf{T}^{\mathrm{wt}}$ and $u, v \in T$, let

$$S^{T,u,v} := \{w \in T : u \in]v,w[\}, \tag{5.17}$$

denote the subtree of T that differs from its closure by the point u, which can be thought of as its root, and consists of points that are on the "other side" of u from v (recall $]v,w[$ is the open arc in T between v and w).

Lemma 5.4. (i) *For $x > 0$,*

$$\begin{aligned}
&\mathbf{P}\left[\mu^T \otimes \nu_T \left\{(u,v) \in T \times T : \mathrm{height}(S^{T,u,v}) > x\right\}\right] \\
&\quad = \mathbf{P}\Big[\int_T \nu_T(\mathrm{d}v)\, \mu^T(R_x(T,v))\Big] \\
&\quad = 2\sum_{n=1}^{\infty} nx\, \exp(-n^2x^2/2).
\end{aligned}$$

(ii) *For $1 < \alpha < \infty$,*

$$\begin{aligned}
&\mathbf{P}\left[\int_T \nu_T(\mathrm{d}v) \int_T \mu^T(\mathrm{d}u)\, \left(\mathrm{height}(S^{T,u,v})\right)^{\alpha}\right] \\
&\quad = 2^{\frac{\alpha+1}{2}}\, \alpha\, \Gamma\big(\frac{\alpha+1}{2}\big)\, \zeta(\alpha),
\end{aligned}$$

where, as usual, $\zeta(\alpha) := \sum_{n \geqslant 1} n^{-\alpha}$.

(iii) *For $0 < p \leqslant 1$,*

$$\mathbf{P}\left[\mu^T \otimes \nu_T\{(u,v) \in T \times T :\, \nu_T(S^{T,u,v}) > p\}\right] = \sqrt{\frac{2(1-p)}{\pi p}}.$$

(iv) *For $\frac{1}{2} < \beta < \infty$,*

$$\mathbf{P}\left[\int_T \nu_T(\mathrm{d}v) \int_T \mu^T(\mathrm{d}u)\, \left(\nu_T\left(S^{T,u,v}\right)\right)^{\beta}\right] = 2^{-\frac{1}{2}}\, \frac{\Gamma\left(\beta - \frac{1}{2}\right)}{\Gamma(\beta)}.$$

Proof. (i) The first equality is clear from the definition of $R_x(T,v)$ and Fubini's theorem.

Turning to the equality of the first and last terms, first recall that $\mathbf{P}$ is the push-forward on $(\mathbf{T}^{\mathrm{wt}}, d_{\mathrm{GH}^{\mathrm{wt}}})$ of the normalized excursion measure $\mathbb{P}$ by the map $e \mapsto (T_{2e}, d_{T_{2e}}, \nu_{T_{2e}})$, where $2e \in U^1$ is just the excursion path $t \mapsto 2e(t)$. In particular, T_{2e} is the quotient of the interval $[0,1]$ by the equivalence relation defined by $2e$. By the invariance of the standard Brownian excursion under random re-rooting (see Section 2.7 of [Ald91b]), the point in T_{2e} that corresponds to the equivalence class of $0 \in [0,1]$ is distributed according to $\nu_{T_{2e}}$ when e is chosen according to $\mathbb{P}$. Moreover, recall from the end of Section 3 that for $e \in U^1$, the length measure μ^{T_e} is the push-forward of the measure $\mathrm{d}s \otimes \mathrm{d}a\, \frac{1}{\overline{s}(e,s,a) - \underline{s}(e,s,a)} \delta_{\underline{s}(e,s,a)}$ on the sub-graph Γ_e by the quotient map defined in (3.2).

It follows that if we pick T according to $\mathbf{P}$ and then pick $(u,v) \in T \times T$ according to $\mu^T \otimes \nu_T$, then the subtree $S^{T,u,v}$ that arises has the same σ-finite law as the tree associated with the excursion $2\hat{e}^{s,a}$ when e is chosen according to $\mathbb{P}$ and (s,a) is chosen according to the measure $\mathrm{d}s \otimes \mathrm{d}a \, \frac{1}{\overline{s}(e,s,a) - \underline{s}(e,s,a)} \delta_{\underline{s}(e,s,a)}$ on the sub-graph Γ_e.

Therefore, by part (i) of Corollary 5.3,

$$\begin{aligned}
&\mathbf{P}\left[\int_T \nu_T(\mathrm{d}v) \int_T \mu^T(\mathrm{d}u) 1\left\{\mathrm{height}(S^{T,u,v}) > x\right\}\right] \\
&\quad = 2\int \mathbb{P}(\mathrm{d}e) \int_{\Gamma_e} \frac{\mathrm{d}s \otimes \mathrm{d}a}{\overline{s}(e,s,a) - \underline{s}(e,s,a)} 1\left\{\max_{0 \leqslant t \leqslant \zeta(\hat{e}^{s,a})} \hat{e}^{s,a} > \frac{x}{2}\right\} \\
&\quad = 2\sum_{n=1}^{\infty} nx \, \exp(-n^2x^2/2).
\end{aligned}$$

Part (ii) is a consequence of part (i) and some straightforward calculus.

Part (iii) follows immediately from part(ii) of Corollary 5.3.

Part (iv) is a consequence of part (iii) and some more straightforward calculus. □

6. A symmetric jump measure on $(\mathbf{T}^{\mathrm{wt}}, d_{\mathrm{GH}^{\mathrm{wt}}})$

In this section we will construct and study a measure on $\mathbf{T}^{\mathrm{wt}} \times \mathbf{T}^{\mathrm{wt}}$ that is related to the decomposition discussed at the beginning of Section 5.

Define a map Θ from $\{((T,d),u,v) : T \in \mathbf{T},\, u \in T,\, v \in T\}$ into $\mathbf{T}$ by setting $\Theta((T,d),u,v) := (T, d^{(u,v)})$ where letting

$$d^{(u,v)}(x,y) := \begin{cases} d(x,y), & \text{if } x,y \in S^{T,u,v}, \\ d(x,y), & \text{if } x,y \in T \backslash S^{T,u,v}, \\ d(x,u) + d(v,y), & \text{if } x \in S^{T,u,v}, y \in T \backslash S^{T,u,v}, \\ d(y,u) + d(v,x), & \text{if } y \in S^{T,u,v}, x \in T \backslash S^{T,u,v}. \end{cases} \tag{6.1}$$

That is, $\Theta((T,d),u,v)$ is just T as a set, but the metric has been changed so that the subtree $S^{T,u,v}$ with root u is now pruned and re-grafted so as to have root v.

If $(T,d,\nu) \in \mathbf{T}^{\mathrm{wt}}$ and $(u,v) \in T \times T$, then we can think of ν as a weight on $(T, d^{(u,v)})$, because the Borel structures induces by d and $d^{(u,v)}$ are the same. With a slight misuse of notation we will therefore write $\Theta((T,d,\nu),u,v)$ for $(T, d^{(u,v)}, \nu) \in \mathbf{T}^{\mathrm{wt}}$. Intuitively, the mass contained in $S^{T,u,v}$ is transported along with the subtree.

Define a kernel κ on $\mathbf{T}^{\mathrm{wt}}$ by

$$\kappa((T,d_T,\nu_T),\mathbf{B}) := \mu^T \otimes \nu_T\big\{(u,v) \in T \times T : \Theta(T,u,v) \in \mathbf{B}\big\} \tag{6.2}$$

for $\mathbf{B} \in \mathcal{B}(\mathbf{T}^{\mathrm{wt}})$. Thus $\kappa((T,d_T,\nu_T),\cdot)$ is the jump kernel described informally in the Introduction.

Remark 6.1. It is clear that $\kappa((T,d_T,\nu_T),\cdot)$ is a Borel measure on $\mathbf{T}^{\mathrm{wt}}$ for each $(T,d_T,\nu_T) \in \mathbf{T}^{\mathrm{wt}}$. In order to show that $\kappa(\cdot,\mathbf{B})$ is a Borel function on

$\mathbf{T}^{\mathrm{wt}}$ for each $\mathbf{B} \in \mathcal{B}(\mathbf{T}^{\mathrm{wt}})$, so that κ is indeed a kernel, it suffices to observe for each bounded continuous function $F : \mathbf{T}^{\mathrm{wt}} \to \mathbb{R}$ that

$$\int F(\Theta(T,u,v))\,\mu^{T}(\mathrm{d}u)\nu_T(\mathrm{d}v) = \lim_{\varepsilon \downarrow 0} \int F(\Theta(T,u,v))\,\mu^{R_\varepsilon(T)}(\mathrm{d}u)\nu_T(\mathrm{d}v)$$

and that

$$(T, d_T, \nu_T) \mapsto \int F(\Theta(T,u,v))\,\mu^{R_\varepsilon(T)}(\mathrm{d}u)\nu_T(\mathrm{d}v)$$

is continuous for all $\varepsilon > 0$ (the latter follows from an argument similar to that in Lemma 7.3 of [EPW04], where it is shown that the $(T, d_T, \nu_T) \mapsto \mu^{R_\varepsilon(T)}(T)$ is continuous). We have only sketched the argument that κ is a kernel, because κ is just a device for defining the measure J on $\mathbf{T}^{\mathrm{wt}} \times \mathbf{T}^{\mathrm{wt}}$ in the next paragraph. It is actually the measure J that we use to define our Dirichlet form, and the measure J can be constructed directly as the push-forward of a measure on $U^1 \times U^1$ – see the proof of Lemma 6.2.

We show in part (i) of Lemma 6.2 below that the kernel κ is reversible with respect to the probability measure $\mathbf{P}$. More precisely, we show that if we define a measure J on $\mathbf{T}^{\mathrm{wt}} \times \mathbf{T}^{\mathrm{wt}}$ by

$$J(\mathbf{A} \times \mathbf{B}) := \int_{\mathbf{A}} \mathbf{P}(\mathrm{d}T)\,\kappa(T, \mathbf{B}) \tag{6.3}$$

for $\mathbf{A}, \mathbf{B} \in \mathcal{B}(\mathbf{T}^{\mathrm{wt}})$, then J is symmetric.

Lemma 6.2. (i) *The measure J is symmetric.*

(ii) *For each compact subset $\mathbf{K} \subset \mathbf{T}^{\mathrm{wt}}$ and open subset $\mathbf{U}$ such that $\mathbf{K} \subset \mathbf{U} \subseteq \mathbf{T}^{\mathrm{wt}}$,*

$$J(\mathbf{K}, \mathbf{T}^{\mathrm{wt}} \backslash \mathbf{U}) < \infty.$$

(iii) *The function $\Delta_{\mathrm{GH}^{\mathrm{wt}}}$ is square-integrable with respect to J, that is,*

$$\int_{\mathbf{T}^{\mathrm{wt}} \times \mathbf{T}^{\mathrm{wt}}} J(\mathrm{d}T, \mathrm{d}S)\,\Delta^2_{\mathrm{GH}^{\mathrm{wt}}}(T,S) < \infty.$$

Proof. (i) Given $e', e'' \in U^1$, $0 \leqslant u \leqslant 1$, and $0 < \rho \leqslant 1$, define $e^\circ(\cdot; e', e'', u, \rho) \in U^1$ by

(6.4)

$$e^\circ(t; e', e'', u, \rho) := \begin{cases} \mathcal{S}_{1-\rho} e''(t), & 0 \leqslant t \leqslant (1-\rho)u, \\ \mathcal{S}_{1-\rho} e''((1-\rho)u) + \mathcal{S}_\rho e'(t - (1-\rho)u), & (1-\rho)u \leqslant t \leqslant (1-\rho)u + \rho, \\ \mathcal{S}_{1-\rho} e''(t-\rho), & (1-\rho)u + \rho \leqslant t \leqslant 1. \end{cases}$$

That is, $e^\circ(\cdot; e', e'', u, \rho)$ is the excursion that arises from Brownian re-scaling e' and e'' to have lengths ρ and $1-\rho$, respectively, and then inserting the re-scaled version of e' into the re-scaled version of e'' at a position that is a fraction u of the total length of the re-scaled version of e''.

Define a measure $\mathbb{J}$ on $U^1 \times U^1$ by

$$\begin{aligned}&\int_{U^1\times U^1} \mathbb{J}(\mathrm{d}e^*, \mathrm{d}e^{**})K(e^*, e^{**})\\ &:= \int_{[0,1]^2} \mathrm{d}u \otimes \mathrm{d}v \frac{1}{2\sqrt{2\pi}} \int_0^1 \frac{\mathrm{d}\rho}{\sqrt{(1-\rho)\rho^3}} \int \mathbb{P}(\mathrm{d}e') \otimes \mathbb{P}(\mathrm{d}e'')\\ &\quad \times K\left(e^\circ(\cdot; e', e'', u, \rho), e^\circ(\cdot; e', e'', v, \rho)\right).\end{aligned} \tag{6.5}$$

Clearly, the measure $\mathbb{J}$ is symmetric. It follows from the discussion at the beginning of the proof of part (i) of Lemma 5.4 and Corollary 5.2 that the measure J is the push-forward of the symmetric measure $2\mathbb{J}$ by the map

$$U^1\times U^1 \ni (e^*, e^{**}) \mapsto ((T_{2e^*}, d_{T_{2e^*}}, \nu_{T_{2e^*}}), (T_{2e^{**}}, d_{T_{2e^{**}}}, \nu_{T_{2e^{**}}})) \in \mathbf{T}^{\mathrm{wt}}\times\mathbf{T}^{\mathrm{wt}},$$

and hence J is also symmetric.
(ii) The result is trivial if $\mathbf{K} = \varnothing$, so we assume that $\mathbf{K} \neq \varnothing$. Since $\mathbf{T}^{\mathrm{wt}}\backslash\mathbf{U}$ and $\mathbf{K}$ are disjoint closed sets and $\mathbf{K}$ is compact, we have that

$$c := \inf_{T\in\mathbf{K}, S\in\mathbf{U}} \Delta_{\mathrm{GH}^{\mathrm{wt}}}(T, S) > 0. \tag{6.6}$$

Fix $T \in \mathbf{K}$. If $(u, v) \in T \times T$ is such that $\Delta_{\mathrm{GH}^{\mathrm{wt}}}(T, \Theta(T, u, v)) > c$, then $\mathrm{diam}(T) > c$ (so that we can think of $R_c(T)$, recall (2.27), as a subset of T). Moreover, we claim that either

- $u \in R_c(T, v)$ (recall (2.26)), or
- $u \notin R_c(T, v)$ and $\nu_T(S^{T,u,v}) > c$ (recall (5.17)).

Suppose, to the contrary, that $u \notin R_c(T, v)$ and that $\nu_T(S^{T,u,\rho}) \leqslant c$. Because $u \notin R_c(T, v)$, the map $f : T \to \Theta(T, u, v)$ given by

$$f(w) := \begin{cases} u, & \text{if } w \in S^{T,u,v},\\ w, & \text{otherwise.}\end{cases}$$

is a measurable c-isometry. There is an analogous measurable c-isometry $g : \Theta(T, u, v) \to T$. Clearly,

$$d_P(f_*\nu^T, \nu^{\Theta(T,u,v)}) \leqslant c$$

and

$$d_P(\nu^T, g^*\nu^{\Theta(T,u,v)}) \leqslant c.$$

Hence, by definition, $\Delta_{\mathrm{GH}^{\mathrm{wt}}}(T, \Theta(T, u, v)) \leqslant c$.

Thus we have

$$
\begin{aligned}
&J(\mathbf{K}, \mathbf{T}^{\mathrm{wt}} \backslash \mathbf{U}) \\
&\quad \leqslant \int_{\mathbf{K}} \mathbf{P}\{\mathrm{d}T\}\, \kappa(T, \{S : \Delta_{\mathrm{GH}^{\mathrm{wt}}}(T,S) > c\}) \\
&\quad \leqslant \int_{\mathbf{K}} \mathbf{P}(\mathrm{d}T) \int_T \nu_T(\mathrm{d}v)\, \mu^T(R_c(T,v)) \\
&\qquad + \int_{\mathbf{K}} \mathbf{P}(\mathrm{d}T) \int_T \nu_T(\mathrm{d}v) \mu^T\{u \in T :\, \nu_T(S^{T,u,v}) > c\} \\
&\quad < \infty,
\end{aligned} \tag{6.7}
$$

where we have used Lemma 5.4.

(iii) Similar reasoning yields that

$$
\begin{aligned}
&\int_{\mathbf{T}^{\mathrm{wt}} \times \mathbf{T}^{\mathrm{wt}}} J(\mathrm{d}T, \mathrm{d}S)\, \Delta^2_{\mathrm{GH}^{\mathrm{wt}}}(T,S) \\
&\quad = \int_{\mathbf{T}^{\mathrm{wt}}} \mathbf{P}\{\mathrm{d}T\} \int_0^\infty \mathrm{d}t\; 2t\, \kappa(T, \{S : \Delta_{\mathrm{GH}^{\mathrm{wt}}}(T,S) > t\}) \\
&\quad \leqslant \int_{\mathbf{T}^{\mathrm{wt}}} \mathbf{P}(\mathrm{d}T) \int_0^\infty \mathrm{d}t\; 2t \int_T \nu_T(\mathrm{d}v)\, \mu^T(R_c(T,v)) \\
&\qquad + \int_{\mathbf{T}^{\mathrm{wt}}} \mathbf{P}(\mathrm{d}T) \int_0^\infty \mathrm{d}t\; 2t \int_T \nu_T(\mathrm{d}v) \mu^T\{u \in T :\, \nu_T\{S^{T,u,v}\} > t\} \\
&\quad \leqslant \int_0^\infty \mathrm{d}t\; 2t \int_{\mathbf{T}^{\mathrm{wt}}} \mathbf{P}(\mathrm{d}T) \int_T \nu_T(\mathrm{d}v)\, \mu^T(R_c(T,v)) \\
&\qquad + \int_{\mathbf{T}^{\mathrm{wt}}} \mathbf{P}(\mathrm{d}T) \int_T \nu_T(\mathrm{d}v) \int_T \mu^T(\mathrm{d}u)\, \nu_T^2(S^{T,u,v}) \\
&\quad < \infty,
\end{aligned} \tag{6.8}
$$

where we have applied Lemma 5.4 once more. □

7. Dirichlet forms

Consider the bilinear form

$$
\begin{aligned}
&\mathcal{E}(f,g) \\
&\quad := \int_{\mathbf{T}^{\mathrm{wt}} \times \mathbf{T}^{\mathrm{wt}}} J(\mathrm{d}T, \mathrm{d}S) \big(f(S) - f(T)\big) \big(g(S) - g(T)\big),
\end{aligned} \tag{7.1}
$$

for f, g in the domain

$$
\mathcal{D}^*(\mathcal{E}) := \{f \in L^2(\mathbf{T}^{\mathrm{wt}}, \mathbf{P}) :\, f \text{ is measurable, and } \mathcal{E}(f,f) < \infty\}, \tag{7.2}
$$

(here as usual, $L^2(\mathbf{T}^{\mathrm{wt}}, \mathbf{P})$ is equipped with the inner product $(f,g)_{\mathbf{P}} := \int \mathbf{P}(\mathrm{d}x)\, f(x) g(x)$). By the argument in Example 1.2.1 in [FOT94] and Lemma 6.2, $(\mathcal{E}, \mathcal{D}^*(\mathcal{E}))$ is well-defined, symmetric and Markovian.

Lemma 7.1. *The form $(\mathcal{E}, \mathcal{D}^*(\mathcal{E}))$ is closed. That is, if $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{D}^*(\mathcal{E})$ such that*

$$\lim_{m,n\to\infty} (\mathcal{E}(f_n - f_m, f_n - f_m) + (f_n - f_m, f_n - f_m)_{\mathbf{P}}) = 0,$$

then there exists $f \in \mathcal{D}^*(\mathcal{E})$ such that

$$\lim_{n\to\infty} (\mathcal{E}(f_n - f, f_n - f) + (f_n - f, f_n - f)_{\mathbf{P}}) = 0.$$

Proof. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence such that $\lim_{m,n\to\infty} \mathcal{E}(f_n - f_m, f_n - f_m) + (f_n - f_m, f_n - f_m)_{\mathbf{P}} = 0$ (that is, $(f_n)_{n\in\mathbb{N}}$ is Cauchy with respect to $\mathcal{E}(\cdot,\cdot) + (\cdot,\cdot)_{\mathbf{P}}$). There exists a subsequence $(n_k)_{k\in\mathbb{N}}$ and $f \in L_2(\mathbf{T}^{\mathrm{wt}}, \mathbf{P})$ such that $\lim_{k\to\infty} f_{n_k} = f$, $\mathbf{P}$-a.s, and $\lim_{k\to\infty}(f_{n_k} - f, f_{n_k} - f)_{\mathbf{P}} = 0$. By Fatou's Lemma,

$$\int J(\mathrm{d}T, \mathrm{d}S)\big((f(S) - f(T)\big)^2 \leqslant \liminf_{k\to\infty} \mathcal{E}(f_{n_k}, f_{n_k}) < \infty, \tag{7.3}$$

and so $f \in \mathcal{D}^*(\mathcal{E})$. Similarly,

$$\begin{aligned} &\mathcal{E}(f_n - f, f_n - f) \\ &\quad = \int J(\mathrm{d}T, \mathrm{d}S) \lim_{k\to\infty} \big((f_n - f_{n_k})(S) - (f_n - f_{n_k})(T)\big)^2 \\ &\quad \leqslant \liminf_{k\to\infty} \mathcal{E}(f_n - f_{n_k}, f_n - f_{n_k}) \to 0 \end{aligned} \tag{7.4}$$

as $n \to \infty$. Thus $(f_n)_{n\in\mathbb{N}}$ has a subsequence that converges to f with respect to $\mathcal{E}(\cdot,\cdot)+(\cdot,\cdot)_{\mathbf{P}}$, but, by the Cauchy property, this implies that $(f_n)_{n\in\mathbb{N}}$ itself converges to f. □

Let $\mathcal{L}$ denote the collection of functions $f : \mathbf{T}^{\mathrm{wt}} \to \mathbb{R}$ such that

$$\sup_{T\in\mathbf{T}^{\mathrm{wt}}} |f(T)| < \infty \tag{7.5}$$

and

$$\sup_{S,T\in\mathbf{T}^{\mathrm{wt}},\, S\neq T} \frac{|f(S) - f(T)|}{\Delta_{\mathrm{GH}^{\mathrm{wt}}}(S,T)} < \infty. \tag{7.6}$$

Note that $\mathcal{L}$ consists of continuous functions and contains the constants. It follows from (2.16) that $\mathcal{L}$ is both a vector lattice and an algebra. By Lemma 7.2 below, $\mathcal{L} \subseteq \mathcal{D}^*(\mathcal{E})$. Therefore, the closure of $(\mathcal{E}, \mathcal{L})$ is a Dirichlet form that we will denote by $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$.

Lemma 7.2. *Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of functions from $\mathbf{T}^{\mathrm{wt}}$ into $\mathbb{R}$ such that*

$$\sup_{n\in\mathbb{N}} \sup_{T\in\mathbf{T}^{\mathrm{wt}}} |f_n(T)| < \infty,$$

$$\sup_{n\in\mathbb{N}} \sup_{S,T\in\mathbf{T}^{\mathrm{wt}},\, S\neq T} \frac{|f_n(S) - f_n(T)|}{\Delta_{\mathrm{GH}^{\mathrm{wt}}}(S,T)} < \infty,$$

and

$$\lim_{n\to\infty} f_n = f, \qquad \mathbf{P}\text{-}a.s.$$

for some $f : \mathbf{T}^{\mathrm{wt}} \to \mathbb{R}$. *Then* $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{D}^*(\mathcal{E})$, $f \in \mathcal{D}^*(\mathcal{E})$, *and*

$$\lim_{n\to\infty} \left(\mathcal{E}(f_n - f, f_n - f) + (f_n - f, f_n - f)_{\mathbf{P}}\right) = 0.$$

Proof. By the definition of the measure J (see (6.3)) and the symmetry of J (Lemma 6.2(i)), we have that $f_n(x) - f_n(y) \to f(x) - f(y)$ for J-almost every pair (x, y). The result then follows from part (iii) of Lemma 6.2 and the dominated convergence theorem. □

Before showing that $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$ is the Dirichlet form of a nice Markov process, we remark that $\mathcal{L}$, and hence $\mathcal{D}(\mathcal{E})$ is quite a rich class of functions: we show in the proof of Theorem 7.3 below that $\mathcal{L}$ separates points of $\mathbf{T}^{\mathrm{wt}}$ and hence if $\mathbf{K}$ is any compact subset of $\mathbf{T}^{\mathrm{wt}}$, then, by the Arzela-Ascoli theorem, the set of restrictions of functions in $\mathcal{L}$ to $\mathbf{K}$ is uniformly dense in the space of real-valued continuous functions on $\mathbf{K}$.

The following theorem states that there is a well-defined Markov process with the dynamics we would expect for a limit of the subtree prune and re-graft chains.

Theorem 7.3. *There exists a recurrent* $\mathbf{P}$*-symmetric Hunt process* $X = (X_t, \mathbb{P}^T)$ *on* $\mathbf{T}^{\mathrm{wt}}$ *whose Dirichlet form is* $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$.

Proof. We will check the conditions of Theorem 7.3.1 in [FOT94] to establish the existence of X.

Because $\mathbf{T}^{\mathrm{wt}}$ is complete and separable (recall Theorem 2.5) there is a sequence $\mathbf{H}_1 \subseteq \mathbf{H}_2 \subseteq \ldots$ of compact subsets of $\mathbf{T}^{\mathrm{wt}}$ such that $\mathbf{P}(\bigcup_{k\in\mathbb{N}} \mathbf{H}_k) = 1$. Given $\alpha, \beta > 0$, write $\mathcal{L}_{\alpha,\beta}$ for the subset of $\mathcal{L}$ consisting of functions f such that

$$\sup_{T\in\mathbf{T}^{\mathrm{wt}}} |f(T)| \leqslant \alpha \tag{7.7}$$

and

$$\sup_{S,T\in\mathbf{T}^{\mathrm{wt}},\, S\neq T} \frac{|f(S) - f(T)|}{\Delta_{\mathrm{GH}^{\mathrm{wt}}}(S, T)} \leqslant \beta. \tag{7.8}$$

By the separability of the continuous real-valued functions on each $\mathbf{H}_k$ with respect to the supremum norm, it follows that for each $k \in \mathbb{N}$ there is a countable set $L_{\alpha,\beta,k} \subseteq \mathcal{L}_{\alpha,\beta}$ such that for every $f \in \mathcal{L}_{\alpha,\beta}$

$$\inf_{g\in L_{\alpha,\beta,k}} \sup_{T\in H_k} |f(T) - g(T)| = 0. \tag{7.9}$$

Set $L_{\alpha,\beta} := \bigcup_{k\in\mathbb{N}} L_{\alpha,\beta,k}$. Then for any $f \in \mathcal{L}_{\alpha,\beta}$ there exists a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $L_{\alpha,\beta}$ such that $\lim_{n\to\infty} f_n = f$ pointwise on $\bigcup_{k\in\mathbb{N}} \mathbf{H}_k$, and hence $\mathbf{P}$-almost surely. By Lemma 7.2, the countable set $\bigcup_{m\in\mathbb{N}} L_{m,m}$ is dense in $\mathcal{L}$, and hence also dense in $\mathcal{D}(\mathcal{E})$, with respect to $\mathcal{E}(\cdot,\cdot) + (\cdot,\cdot)_{\mathbf{P}}$.

Now fix a countable dense subset $\mathbf{S} \subset \mathbf{T}^{\mathrm{wt}}$. Let M denote the countable set of functions of the form

$$T \mapsto p + q(\Delta_{\mathrm{GH}^{\mathrm{wt}}}(S,T) \wedge r) \tag{7.10}$$

for some $S \in \mathbf{S}$ and $p, q, r \in \mathbb{Q}$. Note that $M \subseteq \mathcal{L}$, that M separates the points of $\mathbf{T}^{\mathrm{wt}}$, and, for any $T \in \mathbf{T}^{\mathrm{wt}}$, that there is certainly a function $f \in M$ with $f(T) \neq 0$.

Consequently, if $\mathcal{C}$ is the algebra generated by the countable set $M \cup \bigcup_{m\in\mathbb{N}} L_{m,m}$, then it is certainly the case that $\mathcal{C}$ is dense in $\mathcal{D}(\mathcal{E})$ with respect $\mathcal{E}(\cdot,\cdot) + (\cdot,\cdot)_{\mathbf{P}}$, that $\mathcal{C}$ separates the points of $\mathbf{T}^{\mathrm{wt}}$, and, for any $T \in \mathbf{T}^{\mathrm{wt}}$, that there is a function $f \in \mathcal{C}$ with $f(T) \neq 0$.

All that remains in verifying the conditions of Theorem 7.3.1 in [FOT94] is to check the tightness condition that there exist compact subsets $\mathbf{K}_1 \subseteq \mathbf{K}_2 \subseteq \ldots$ of $\mathbf{T}^{\mathrm{wt}}$ such that $\lim_{n\to\infty} \mathrm{Cap}(\mathbf{T}^{\mathrm{wt}} \backslash \mathbf{K}_n) = 0$ where Cap is the capacity associated with the Dirichlet form – see Remark 7.4 below for a definition. This convergence, however, is the content of Lemma 7.7 below.

Finally, because constants belongs to $\mathcal{D}(\mathcal{E})$, it follows from Theorem 1.6.3 in [FOT94] that X is recurrent. □

Remark 7.4. In the proof of Theorem 7.3 we used the capacity associated with the Dirichlet form $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$. We remind the reader that for an open subset $\mathbf{U} \subseteq \mathbf{T}^{\mathrm{wt}}$,

$$\mathrm{Cap}(\mathbf{U}) := \inf\left\{\mathcal{E}(f,f) + (f,f)_{\mathbf{P}} : f \in \mathcal{D}(\mathcal{E}),\, f(T) \geqslant 1,\, \mathbf{P}\text{−a.e.}T \in \mathbf{U}\right\},$$

and for a general subset $\mathbf{A} \subseteq \mathbf{T}^{\mathrm{wt}}$

$$\mathrm{Cap}(\mathbf{A}) := \inf\left\{\mathrm{Cap}(\mathbf{U}) : \mathbf{A} \subseteq \mathbf{U} \text{ is open}\right\}.$$

We refer the reader to Section 2.1 of [FOT94] for details and a proof that Cap is a Choquet capacity.

The following results were needed in the proof of Theorem 7.3

Lemma 7.5. *For $\varepsilon, a, \delta > 0$, put $\mathbf{V}_{\varepsilon,a} := \{T \in \mathbf{T} : \mu^T(R_\varepsilon(T)) > a\}$ and, as usual, $\mathbf{V}^\delta_{\varepsilon,a} := \{T \in \mathbf{T} : d_{\mathrm{GH}}(T, \mathbf{V}_{\varepsilon,a}) < \delta\}$. Then, for fixed $\varepsilon > 3\delta$,*

$$\bigcap_{a>0} \mathbf{V}^\delta_{\varepsilon,a} = \varnothing.$$

Proof. Fix $S \in \mathbf{T}$. If $S \in \mathbf{V}^\delta_{\varepsilon,a}$, then there exists $T \in \mathbf{V}_{\varepsilon,a}$ such that $d_{\mathrm{GH}}(S,T) < \delta$. Observe that $R_\varepsilon(T)$ is not the trivial tree consisting of a single point because it has total length greater than a. Write $\{y_1, \ldots, y_n\}$ for the leaves of $R_\varepsilon(T)$. For all $i = 1, \ldots, n$, the connected component of $T \backslash R_\varepsilon(T)^o$ that contains y_i contains a point z_i such that $d_T(y_i, z_i) = \varepsilon$.

Let $\Re$ be a correspondence between S and T with $\mathrm{dis}(\Re) < 2\delta$. Pick $x_1, \ldots, x_n \in S$ such that $(x_i, z_i) \in \Re$, and hence $|d_S(x_i, x_j) - d_T(z_i, z_j)| < 2\delta$ for all i, j.

The distance in $R_\varepsilon(T)$ from the point y_k to the arc $[y_i, y_j]$ is

$$\frac{1}{2}\big(d_S(y_k, y_i) + d_S(y_k, y_j) - d_S(y_i, y_j)\big). \tag{7.11}$$

Thus the distance from y_k, $3 \leqslant k \leqslant n$, to the subtree spanned by $y_1, \ldots, y_{k-1}$ is

$$\bigwedge_{1 \leqslant i \leqslant j \leqslant k-1} \frac{1}{2}\big(d_T(y_k, y_i) + d_T(y_k, y_j) - d_T(y_i, y_j)\big), \tag{7.12}$$

and hence

(7.13)

$$\mu^T(R_\varepsilon(T)) = d_T(y_1, y_2) + \sum_{k=3}^{n} \bigwedge_{1 \leqslant i \leqslant j \leqslant k-1} \frac{1}{2}\left(d_T(y_k, y_i) + d_T(y_k, y_j) - d_T(y_i, y_j)\right).$$

Now the distance in S from the point x_k to the arc $[x_i, x_j]$ is

$$\begin{aligned} &\frac{1}{2}\left(d_S(x_k, x_i) + d_S(x_k, x_j) - d_S(x_i, x_j)\right) \\ &\quad \geqslant \frac{1}{2}\left(d_T(z_k, z_i) + d_T(z_k, z_j) - d_T(z_i, z_j) - 3 \times 2\delta\right) \\ &\quad = \frac{1}{2}\left(d_T(y_k, y_i) + 2\varepsilon + d_T(y_k, y_j) + 2\varepsilon - d_T(y_i, y_j) - 2\varepsilon - 6\delta\right) \\ &\quad > 0 \end{aligned} \tag{7.14}$$

by the assumption that $\varepsilon > 3\delta$. In particular, $x_1, \ldots, x_n$ are leaves of the subtree spanned by $\{x_1, \ldots, x_n\}$, and $R_\gamma(S)$ has at least n leaves when $0 < \gamma < 2\varepsilon - 6\delta$. Fix such a γ.

Now

$$\begin{aligned} &\mu^S(R_\gamma(S)) \\ &\quad \geqslant d_S(x_1, x_2) - 2\gamma \\ &\qquad + \sum_{k=3}^{n} \bigwedge_{1 \leqslant i \leqslant j \leqslant k-1} \left[\frac{1}{2}\left(d_S(x_k, x_i) + d_S(x_k, x_j) - d_S(x_i, x_j)\right) - \gamma\right] \\ &\quad \geqslant \mu^T(R_\varepsilon(T)) + (2\varepsilon - 2\delta - 2\gamma) + (n-2)(\varepsilon - 3\delta - \gamma) \\ &\quad \geqslant a + (2\varepsilon - 2\delta - 2\gamma) + (n-2)(\varepsilon - 3\delta - \gamma). \end{aligned} \tag{7.15}$$

Because $\mu^S(R_\gamma(S))$ is finite, it is apparent that S cannot belong to $\mathbf{V}^\delta_{\varepsilon,a}$ when a is sufficiently large. □

Lemma 7.6. *For $\varepsilon, a > 0$, let $\mathbf{V}_{\varepsilon,a}$ be as in Lemma 7.5. Set $\mathbf{U}_{\varepsilon,a} := \{(T, \nu) \in \mathbf{T}^{\mathrm{wt}} : T \in \mathbf{V}_{\varepsilon,a}\}$. Then, for fixed ε,*

$$\lim_{a \to \infty} \mathrm{Cap}(\mathbf{U}_{\varepsilon,a}) = 0. \tag{7.16}$$

Proof. Observe that $(T, d_T, \nu_T) \mapsto \mu^{R_\varepsilon(T)}(T)$ is continuous (this is essentially Lemma 7.3 of [EPW04]), and so $\mathbf{U}_{\varepsilon,a}$ is open.

Choose $\delta > 0$ such that $\varepsilon > 3\delta$. Suppressing the dependence on ε and δ, define $u_a : \mathbf{T}^{\mathrm{wt}} \to [0,1]$ by

$$u_a((T,\nu)) := \delta^{-1}(\delta - d_{\mathrm{GH}}(T, \mathbf{V}_{\varepsilon,a}))_+. \tag{7.17}$$

Note that u_a takes the value 1 on the open set $\mathbf{U}_{\varepsilon,a}$, and so $\mathrm{Cap}(\mathbf{U}_{\varepsilon,a}) \leqslant \mathcal{E}(u_a,u_a) + (u_a,u_a)_{\mathbf{P}}$. Also observe that

$$\begin{aligned} |u_a((T',\nu')) - u_a((T'',\nu''))| &\leqslant \delta^{-1} d_{\mathrm{GH}}(T',T'') \\ &\leqslant \delta^{-1} \Delta_{\mathrm{GH}^{\mathrm{wt}}}((T',\nu'),(T'',\nu'')). \end{aligned} \tag{7.18}$$

It therefore suffices by part (iii) of Lemma 6.2 and the dominated convergence theorem to show for each pair $((T',\nu'),(T'',\nu'')) \in \mathbf{T}^{\mathrm{wt}} \times \mathbf{T}^{\mathrm{wt}}$ that $u_a((T',\nu')) - u_a((T'',\nu''))$ is 0 for a sufficiently large and for each $T \in \mathbf{T}^{\mathrm{wt}}$ that $u_a((T,\nu))$ is 0 for a sufficiently large. However, $u_a((T',\nu')) - u_a((T'',\nu'')) \neq 0$ implies that either T' or T'' belong to $\mathbf{V}^\delta_{\varepsilon,a}$, while $u_a((T,\nu)) \neq 0$ implies that T belongs to $\mathbf{V}^\delta_{\varepsilon,a}$. The result then follows from Lemma 7.5. □

Lemma 7.7. *There is a sequence of compact sets $\mathbf{K}_1 \subseteq \mathbf{K}_2 \subseteq \ldots$ such that $\lim_{n\to\infty} \mathrm{Cap}(\mathbf{T}^{\mathrm{wt}} \backslash \mathbf{K}_n) = 0$.*

Proof. By Lemma 7.6, for $n = 1, 2, \ldots$ we can choose a_n so that $\mathrm{Cap}(\mathbf{U}_{2^{-n},a_n}) \leqslant 2^{-n}$. Set

$$\mathbf{F}_n := \mathbf{T}^{\mathrm{wt}} \backslash \mathbf{U}_{2^{-n},a_n} = \{(T,\nu) \in \mathbf{T}^{\mathrm{wt}} : \mu^T(R_{2^{-n}}(T)) \leqslant a_n\} \tag{7.19}$$

and

$$\mathbf{K}_n := \bigcap_{m \geqslant n} \mathbf{F}_m. \tag{7.20}$$

By Proposition 2.4 and Lemma 2.6, each set $\mathbf{K}_n$ is compact. By construction,

$$\begin{aligned} \mathrm{Cap}(\mathbf{T}^{\mathrm{wt}} \backslash \mathbf{K}_n) &= \mathrm{Cap}\left(\bigcup_{m \geqslant n} \mathbf{U}_{2^{-m},a_m} \right) \\ &\leqslant \sum_{m \geqslant n} \mathrm{Cap}(\mathbf{U}_{2^{-m},a_m}) \leqslant \sum_{m \geqslant n} 2^{-m} = 2^{-(n-1)}. \end{aligned} \tag{7.21}$$

□

8. The trivial tree is essentially polar

From our informal picture of the process X evolving via re-arrangements of the initial tree that preserve the total branch length, one might expect that if X does not start at the trivial tree T_0 consisting of a single point, then X will never hit T_0. However, an SPR move can decrease the diameter of a tree, so it is conceivable that, in passing to the limit, there is some probability that an infinite sequence of SPR moves will conspire to collapse the evolving tree down to a single point. Of course, it is hard to imagine from the approximating dynamics how X could recover from such a catastrophe – which it would have to since it is reversible with respect to the continuum random tree distribution.

In this section we will use potential theory for Dirichlet forms to show that X does not hit T_0 from $\mathbf{P}$-almost all starting points; that is, that the set $\{T_0\}$ is *essentially polar*.

Let $\bar{d}$ be the map which sends a weighted $\mathbb{R}$ tree (T, d, ν) to the ν-averaged distance between pairs of points in T. That is,

$$\bar{d}\big((T,d,\nu)\big) := \int_T \int_T \nu(\mathrm{d}x)\nu(\mathrm{d}y)\, d(x,y), \quad (T,d,\nu) \in \mathbf{T}^{\mathrm{wt}}. \tag{8.1}$$

In order to show that T_0 is essentially polar, it will suffice to show that the set

$$\{(T,d,\nu) \in \mathbf{T}^{\mathrm{wt}} : \bar{d}\big((T,d,\nu)\big) = 0\} \tag{8.2}$$

is essentially polar.

Lemma 8.1. *The function $\bar{d}$ belongs to the domain $\mathcal{D}(\mathcal{E})$.*

Proof. If we let $\bar{d}_n\big((T,d,\nu)\big) := \int_T \int_T \nu(\mathrm{d}x)\nu(\mathrm{d}y)\ [d(x,y) \wedge n]$, for $n \in \mathbb{N}$, then $\bar{d}_n \uparrow \bar{d}$, $\mathbf{P}$-a.s. By the triangle inequality,

$$(\bar{d}, \bar{d})_{\mathbf{P}} \leqslant \int \mathbf{P}(\mathrm{d}T)\, (\mathrm{diam}(T))^2 \leqslant \int \mathbb{P}(\mathrm{de})\, \big(4 \sup_{t \in [0,1]} \mathrm{e}(t)\big)^2 < \infty, \tag{8.3}$$

and hence $\bar{d}_n \to \bar{d}$ as $n \to \infty$ in $L^2(\mathbf{T}^{\mathrm{wt}}, \mathbf{P})$.

Notice, moreover, that for $(T,d,\nu) \in \mathbf{T}^{\mathrm{wt}}$ and $u, v \in T$,

$$\begin{aligned} &\Big(\bar{d}\big((T,d,\nu)\big) - \bar{d}\big(\Theta((T,d,\nu),u,v)\big)\Big)^2 \\ &\qquad = 2\int_{S^{T,u,v}} \int_{T \backslash S^{T,u,v}} \nu(\mathrm{d}x)\nu(\mathrm{d}y)\, \big(d(y,u) - d(y,v)\big)^2 \\ &\qquad = 2\nu_T(S^{T,u,v})\nu(T \backslash S^{T,u,v})\, d^2(u,v). \end{aligned} \tag{8.4}$$

Hence, applying Corollary 5.2 and the invariance of the standard Brownian excursion under random re-rooting (see Section 2.7 of [Ald91b]),

(8.5)
$$\begin{aligned}
&\int_{\mathbf{T}^{\mathrm{wt}}\times\mathbf{T}^{\mathrm{wt}}} J(\mathrm{d}T,\mathrm{d}S)\,\big(\bar{d}(T)-\bar{d}(S)\big)^2\\
&\quad= 2\int_{\mathbf{T}^{\mathrm{wt}}}\mathbf{P}(\mathrm{d}T)\int_{T\times T}\nu_T(\mathrm{d}v)\mu^T(\mathrm{d}u)\nu_T(S^{T,u,v})\nu_T(T\backslash S^{T,u,v})\,d_T^2(u,v)\\
&\quad\leqslant 2\int\mathbb{P}(\mathrm{d}e)\,2\int_{\Gamma_{\mathrm{e}}}\frac{\mathrm{d}s\otimes\mathrm{d}a}{\bar{s}(e,s,a)-\underline{s}(e,s,a)}\zeta(\hat{e}^{s,a})\zeta(\check{e}^{s,a})(2a)^2\\
&\quad= \frac{8}{\sqrt{2\pi}}\int_0^1\frac{\mathrm{d}\rho}{\sqrt{(1-\rho)\rho^3}}\int\mathbb{P}(\mathrm{d}e')\otimes\mathbb{P}(\mathrm{d}e'')\,\rho(1-\rho)\big(\sup\mathcal{S}_{1-\rho}e''\big)^2\\
&\quad= \frac{8}{\sqrt{2\pi}}\int_0^1\frac{\mathrm{d}\rho}{\sqrt{(1-\rho)\rho^3}}\rho(1-\rho)^2\int\mathbb{P}(de)\left(\sup_{t\in[0,1]}e(t)\right)^2<\infty.
\end{aligned}$$

Consequently, by dominated convergence, $\mathcal{E}(\bar{d}-\bar{d}_n,\bar{d}-\bar{d}_n)\to 0$ as $n\to\infty$.

It is therefore enough to verify that $\bar{d}_n\in\mathcal{L}$ for all $n\in\mathbb{N}$. Obviously,

$$\sup_{T\in\mathbf{T}^{\mathrm{wt}}}\bar{d}_n(T)\leqslant n, \tag{8.6}$$

and so the boundedness condition (7.5) holds. To show that the "Lipschitz" property (7.6) holds, fix $\varepsilon>0$, and let $(T,\nu_T),(S,\nu_S)\in\mathbf{T}^{\mathrm{wt}}$ be such that $\Delta_{\mathrm{GH}^{\mathrm{wt}}}\big((T,\nu_T),(S,\nu_S)\big)<\varepsilon$. Then there exist $f\in F^{\varepsilon}_{T,S}$ and $g\in F^{\varepsilon}_{S,T}$ such that $d_{\mathrm{P}}(\nu_T,g_*\nu_S)<\varepsilon$ and $d_{\mathrm{P}}(f_*\nu_T,\nu_S)<\varepsilon$ (recall $F^{\varepsilon}_{T,S}$ from (2.10)). Hence

$$\begin{aligned}
&\Big|\bar{d}_n\big((T,\nu_T)\big)-\bar{d}_n\big((S,\nu_S)\big)\Big|\\
&\quad\leqslant\Big|\int_T\int_T\nu_T(\mathrm{d}x)\nu_T(\mathrm{d}y)\,(d_T(x,y)\wedge n)\\
&\qquad-\int_{g(S)}\int_{g(S)}g_*\nu_S(\mathrm{d}x)g_*\nu_S(\mathrm{d}y)\,(d_T(x,y)\wedge n)\Big|\\
&\qquad+\Big|\int_{g(S)}\int_{g(S)}g_*\nu_S(\mathrm{d}x)g_*\nu_S(\mathrm{d}y)\,(d_T(x,y)\wedge n)\\
&\qquad-\int_S\int_S\nu_S(\mathrm{d}x')\nu_S(\mathrm{d}y')\,(d_S(x',y')\wedge n)\Big|.
\end{aligned}\tag{8.7}$$

For the first term on the right hand side of (8.7) we get

$$
\begin{aligned}
&\Big| \int_T \int_T \nu_T(\mathrm{d}x)\nu_T(\mathrm{d}y)\,(d_T(x,y) \wedge n) \\
&\qquad - \int_{g(S)} \int_{g(S)} g_*\nu_S(\mathrm{d}x) g_*\nu_S(\mathrm{d}y)\,(d_T(x,y) \wedge n) \Big| \\
&\leqslant \Big| \int_T \int_T \nu_T(\mathrm{d}x)\nu_T(\mathrm{d}y)\,(d_T(x,y) \wedge n) \\
&\qquad - \int_T \int_{g(S)} \nu_T(\mathrm{d}x) g_*\nu_S(\mathrm{d}y)\,(d_T(x,y) \wedge n) \Big| \\
&\quad + \Big| \int_{S(g)} \int_T g_*\nu_S(\mathrm{d}x)\nu_T(\mathrm{d}y)\,(d_T(x,y) \wedge n) \\
&\qquad - \int_{g(S)} \int_{g(S)} g_*\nu_S(\mathrm{d}x) g_*\nu_S(\mathrm{d}y)\,(d_T(x,y) \wedge n) \Big|.
\end{aligned}
\tag{8.8}
$$

By assumption and Theorem 3.1.2 in [EK86], we can find a probability measure ν on $T \times T$ with marginals ν_T and $g_*\nu_S$ such that

$$
\nu\big\{(x,y) : d_T(x,y) \geqslant \varepsilon\big\} \leqslant \varepsilon. \tag{8.9}
$$

Hence, for all $x \in T$,

$$
\begin{aligned}
&\Big| \int_T \nu_T(\mathrm{d}y)\,(d_T(x,y) \wedge n) - \int_{g(S)} g_*\nu_S(\mathrm{d}y)\,(d_T(x,y) \wedge n) \Big| \\
&\leqslant \int_{T \times g(S)} \nu\big(\mathrm{d}(y,y')\big) \Big| (d_T(x,y) \wedge n) - (d_T(x,y') \wedge n) \Big| \\
&\leqslant \int_{T \times g(S)} \nu\big(\mathrm{d}(y,y')\big)\,(d_T(y,y') \wedge n) \\
&\leqslant \big(1 + (\mathrm{diam}(T) \wedge n)\big) \cdot \varepsilon.
\end{aligned}
\tag{8.10}
$$

For the second term in (8.7) we use the fact that g is an ε-isometry, that is, $|(d_S(x',y') \wedge n) - (d_T(g(x'),g(y')) \wedge n)| < \varepsilon$ for all $x', x'' \in T$. A change of variables then yields that

$$
\begin{aligned}
&\Big| \int_{g(S)} \int_{g(S)} g_*\nu_S(\mathrm{d}x) g_*\nu_S(\mathrm{d}y)\,(d_T(x,y) \wedge n) \\
&\qquad - \int_S \int_S \nu_S(\mathrm{d}x')\nu_S(\mathrm{d}y')\,(d_S(x',y') \wedge n) \Big| \\
&\leqslant \varepsilon + \Big| \int_{g(S)} \int_{g(S)} g_*\nu_S(\mathrm{d}x) g_*\nu_S(\mathrm{d}y)\,(d_T(x,y) \wedge n) \\
&\qquad - \int_S \int_S \nu_S(\mathrm{d}x')\nu_S(\mathrm{d}y')\,(d_T(g(x'),g(y')) \wedge n) \Big| \\
&= \varepsilon.
\end{aligned}
\tag{8.11}
$$

Combining (8.7) through (8.11) yields finally that

$$\sup_{(T,\nu_T)\neq(S,\nu_S)\in\mathbf{T}^{wt}} \frac{\left|\bar{d}_n\big((T,\nu_T)\big) - \bar{d}_n\big((S,\nu_S)\big)\right|}{\Delta_{\mathrm{GH}^{\mathrm{wt}}}\big((T,\nu_T),(S,\nu_S)\big)} \leqslant 3 + 2n. \tag{8.12}$$

□

Proposition 8.2. *The set* $\{T \in \mathbf{T}^{\mathrm{wt}} : \bar{d}(T) = 0\}$ *is essentially polar. In particular, the set* $\{T_0\}$ *consisting of the trivial tree is essentially polar.*

Proof. We need to show that $\mathrm{Cap}(\{T \in \mathbf{T}^{\mathrm{wt}} : \bar{d}(T) = 0\}) = 0$ (see Theorem 4.2.1 of [FOT94]).

For $\varepsilon > 0$ set

$$\mathbf{W}_\varepsilon := \{T \in \mathbf{T}^{\mathrm{wt}} : \bar{d}(T) < \varepsilon\}. \tag{8.13}$$

By the argument in the proof of Lemma 8.1, the function $\bar{d}$ is continuous, and so $\mathbf{W}_\varepsilon$ is open. It suffices to show that $\mathrm{Cap}(\mathbf{W}_\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$.

Put

$$u_\varepsilon(T) := \left(2 - \frac{\bar{d}(T)}{\varepsilon}\right)_+, \quad T \in \mathbf{T}^{\mathrm{wt}}. \tag{8.14}$$

Then $u \in \mathcal{D}(\mathcal{E})$ by Lemma 8.1 and the fact that the domain of a Dirichlet form is closed under composition with Lipschitz functions. Because $u_\varepsilon(T) \geqslant 1$ for $T \in \mathbf{W}_\varepsilon$, it thus further suffices to show

$$\lim_{\varepsilon\downarrow 0} \left(\mathcal{E}(u_\varepsilon, u_\varepsilon) + (u_\varepsilon, u_\varepsilon)_{\mathbf{P}}\right) = 0. \tag{8.15}$$

By elementary properties of the standard Brownian excursion,

$$(u_\varepsilon, u_\varepsilon)_{\mathbf{P}} \leqslant 4\mathbf{P}\{T : \bar{d}(T) < 2\varepsilon\} \to 0 \tag{8.16}$$

as $\varepsilon \downarrow 0$. Estimating $\mathcal{E}(u_\varepsilon, u_\varepsilon)$ will be somewhat more involved.

Let $\hat{E}$ and $\check{E}$ be two independent standard Brownian excursions, and let U and V be two independent random variables that are independent of $\hat{E}$ and $\check{E}$ and uniformly distributed on $[0,1]$. With a slight abuse of notation, we will write $\mathbb{P}$ for the probability measure on the probability space where $\hat{E}$, $\check{E}$, U and V are defined.

Set

$$\begin{aligned}
\hat{D} &:= 4\int_{0\leqslant s<t\leqslant 1} \mathrm{d}s\otimes\mathrm{d}t\,\left[\hat{E}_s+\hat{E}_t-2\inf_{s\leqslant w\leqslant t}\hat{E}_w\right]\\
\hat{H} &:= 2\int_{[0,1]}\mathrm{d}t\,\hat{E}_t\\
\check{D} &:= 4\int_{0\leqslant s<t\leqslant 1} \mathrm{d}s\otimes\mathrm{d}t\,\left[\check{E}_s+\check{E}_t-2\inf_{s\leqslant w\leqslant t}\check{E}_w\right]\\
\check{H}_U &:= 2\int_{[0,1]}\mathrm{d}t\,\left[\check{E}_t+\check{E}_U-2\inf_{U\wedge t\leqslant w\leqslant U\vee t}\check{E}_w\right]\\
\check{H}_V &:= 2\int_{[0,1]}\mathrm{d}t\,\left[\check{E}_t+\check{E}_V-2\inf_{V\wedge t\leqslant w\leqslant V\vee t}\check{E}_w\right].
\end{aligned}\tag{8.17}$$

For $0\leqslant\rho\leqslant 1$ set

$$\begin{aligned}
D_U(\rho) &:= (1-\rho)^2\sqrt{1-\rho}\check{D}+\rho^2\sqrt{\rho}\hat{D}\\
&\quad+2(1-\rho)\rho\sqrt{\rho}\hat{H}+2(1-\rho)\rho\sqrt{1-\rho}\check{H}_U
\end{aligned}\tag{8.18}$$

and

$$\begin{aligned}
D_V(\rho) &:= (1-\rho)^2\sqrt{1-\rho}\check{D}+\rho^2\sqrt{\rho}\hat{D}\\
&\quad+2(1-\rho)\rho\sqrt{\rho}\hat{H}+2(1-\rho)\rho\sqrt{1-\rho}\check{H}_V.
\end{aligned}\tag{8.19}$$

Then

$$\begin{aligned}
&\mathcal{E}(u_\varepsilon,u_\varepsilon)\\
&\quad=\frac{1}{2\sqrt{2\pi}}\mathbb{P}\left[\int_0^1\frac{\mathrm{d}\rho}{\sqrt{(1-\rho)\rho^3}}\left\{\left(2-\frac{D_U(\rho)}{\varepsilon}\right)_+-\left(2-\frac{D_V(\rho)}{\varepsilon}\right)_+\right\}^2\right].
\end{aligned}\tag{8.20}$$

Fix $0<a<\frac{1}{2}$ and write $\bar{a}=1-a$ for convenience. We can write the right-hand side of (8.20) as the sum of three terms $I(\varepsilon,a)$, $II(\varepsilon,a)$, and $III(\varepsilon,a)$, that arise from integrating ρ over the respective ranges

$$\{\rho : D_U(\rho)\vee D_V(\rho)\leqslant 2\varepsilon,\ 0\leqslant\rho\leqslant a\},\tag{8.21}$$

$$\{\rho : D_U(\rho)\wedge D_V(\rho)\leqslant 2\varepsilon\leqslant D_U(\rho)\vee D_V(\rho),\ 0\leqslant\rho\leqslant a\},\tag{8.22}$$

and

$$\{\rho : a<\rho\leqslant 1\}.\tag{8.23}$$

Consider $I(\varepsilon,a)$ first. Note that if $D_U(\rho)\vee D_V(\rho)\leqslant 2\varepsilon$, then

$$\left\{\left(2-\frac{D_U(\rho)}{\varepsilon}\right)_+-\left(2-\frac{D_V(\rho)}{\varepsilon}\right)_+\right\}^2\leqslant 2^2\frac{\rho^2}{\varepsilon^2}\{\check{H}_U-\check{H}_V\}^2.\tag{8.24}$$

Moreover,

$$\begin{aligned}
&\{0 \leqslant \rho \leqslant a : D_U(\rho) \vee D_V(\rho) \leqslant 2\varepsilon\} \\
&\subseteq \left\{0 \leqslant \rho \leqslant a : (1-\rho)^{\frac{5}{2}}\check{D} + 2(1-\rho)^{\frac{3}{2}}\rho(\check{H}_U \vee \check{H}_V) \leqslant 2\varepsilon\right\} \\
&\subseteq \left\{0 \leqslant \rho \leqslant a : \bar{a}^{\frac{5}{2}}\check{D} + 2\bar{a}^{\frac{3}{2}}\rho(\check{H}_U \vee \check{H}_V) \leqslant 2\varepsilon\right\} \qquad (8.25)\\
&= \left\{\rho : 0 \leqslant \rho \leqslant \frac{(2\varepsilon - \bar{a}^{\frac{5}{2}}\check{D})_+}{2\bar{a}^{\frac{3}{2}}(\check{H}_U \vee \check{H}_V)} \wedge a\right\}.
\end{aligned}$$

Thus $I(\varepsilon, a)$ is bounded above by the expectation of the random variable that arises from integrating $2^2\rho^2\{\check{H}_U - \check{H}_V\}^2/\varepsilon^2$ against the measure $\frac{1}{2\sqrt{2\pi}}\frac{\mathrm{d}\rho}{\sqrt{(1-\rho)\rho^3}}$ over the interval $[0,\ (2\varepsilon - \bar{a}^{\frac{5}{2}}\check{D})_+/(2\bar{a}^{\frac{3}{2}}(\check{H}_U \vee \check{H}_V))]$. Note that

$$\int_0^x \frac{\mathrm{d}\rho}{\sqrt{\rho^3}}\rho^\alpha = \frac{1}{\alpha - \frac{1}{2}}x^{\alpha-\frac{1}{2}}, \quad \alpha > \frac{1}{2}. \qquad (8.26)$$

Hence, letting C denote a generic constant with a value that doesn't depend on ε or a and may change from line to line, and

$$\begin{aligned}
I(\varepsilon, a) &\leqslant C\mathbb{P}\left[\left(\frac{(2\varepsilon - \bar{a}^{\frac{5}{2}}\check{D})_+}{\check{H}_U \vee \check{H}_V}\right)^{\frac{3}{2}}\frac{\{\check{H}_U - \check{H}_V\}^2}{\varepsilon^2}\right] \\
&\leqslant \frac{C}{\varepsilon^2}\mathbb{P}\left[(2\varepsilon - \bar{a}^{\frac{5}{2}}\check{D})_+^{\frac{3}{2}}(\check{H}_U \vee \check{H}_V)^{\frac{1}{2}}\right] \\
&\leqslant \frac{C}{\varepsilon^{\frac{1}{2}}}\mathbb{P}\left[(\check{H}_U + \check{H}_V)^{\frac{1}{2}}1\{\check{D} \leqslant 2\bar{a}^{-\frac{5}{2}}\varepsilon\}\right] \qquad (8.27)\\
&\leqslant \frac{C}{\varepsilon^{\frac{1}{2}}}\mathbb{P}\left[\check{D}^{\frac{1}{2}}1\{\check{D} \leqslant 2\bar{a}^{-\frac{5}{2}}\varepsilon\}\right] \\
&\leqslant C\mathbb{P}\{\check{D} \leqslant 2\bar{a}^{-\frac{5}{2}}\varepsilon\},
\end{aligned}$$

where in the second last line we used the fact that

$$\mathbb{P}[\check{H}_U \,|\, \check{E}] = \mathbb{P}[\check{H}_V \,|\, \check{E}] = \check{D}, \qquad (8.28)$$

and Jensen's inequality for conditional expectations to obtain the inequalities $\mathbb{P}[\check{H}_U^{\frac{1}{2}} \,|\, \check{E}] \leqslant \check{D}^{\frac{1}{2}}$ and $\mathbb{P}[\check{H}_V^{\frac{1}{2}} \,|\, \check{E}] \leqslant \check{D}^{\frac{1}{2}}$. Thus, $\lim_{\varepsilon\downarrow 0} I(\varepsilon, a) = 0$ for any value of a.

Turning to $II(\varepsilon, a)$, first note that $\hat{D} \leqslant 4\hat{H}$ and, by the triangle inequality,

$$\check{D} \leqslant 2(\check{H}_U \wedge \check{H}_V). \qquad (8.29)$$

Hence, for some constant K that does not depend on ε or a,

$$|D_U(\rho) \wedge D_V(\rho) - \check{D}| \leqslant K(\hat{H}\rho^{\frac{3}{2}} + (\check{H}_U \wedge \check{H}_V)\rho) \qquad (8.30)$$

and

$$|D_U(\rho) \vee D_V(\rho) - \check{D}| \leqslant K(\hat{H}\rho^{\frac{3}{2}} + (\check{H}_U \vee \check{H}_V)\rho). \qquad (8.31)$$

Combining (8.31) with an argument similar to that which established (8.25) gives, for a suitable constant K^*,

$$\begin{aligned}
&\{0 \leqslant \rho \leqslant a : D_U(\rho) \wedge D_V(\rho) \leqslant 2\varepsilon \leqslant D_U(\rho) \vee D_V(\rho)\} \\
&\quad = \{0 \leqslant \rho \leqslant a : 2\varepsilon \leqslant D_U(\rho) \vee D_V(\rho), \} \\
&\qquad \cap \{0 \leqslant \rho \leqslant a : D_U(\rho) \wedge D_V(\rho) \leqslant 2\varepsilon\} \\
&\quad \subseteq \left\{\rho : \frac{(2\varepsilon - \check{D})_+}{K^*(\hat{H} + \check{H}_U \vee \check{H}_V)} \leqslant \rho \leqslant a\right\} \\
&\qquad \cap \left\{\rho : 0 \leqslant \rho \leqslant \frac{(2\varepsilon - \bar{a}^{\frac{5}{2}}\check{D})_+}{2\bar{a}^{\frac{3}{2}}(\check{H}_U \wedge \check{H}_V)} \wedge a\right\}.
\end{aligned} \tag{8.32}$$

Moreover, by (8.30) and the observation $|(2\varepsilon - x)_+ - (2\varepsilon - y)_+| \leqslant |x - y|$, we have for $D_U(\rho) \wedge D_V(\rho) \leqslant 2\varepsilon \leqslant D_U(\rho) \vee D_V(\rho)$ that,

(8.33)

$$\begin{aligned}
&\left\{\left(2 - \frac{D_U(\rho)}{\varepsilon}\right)_+ - \left(2 - \frac{D_V(\rho)}{\varepsilon}\right)_+\right\}^2 \\
&\quad = \left\{\left(2 - \frac{D_U(\rho) \wedge D_V(\rho)}{\varepsilon}\right)_+\right\}^2 \\
&\quad \leqslant \frac{2}{\varepsilon^2}\left\{\left(2\varepsilon - \check{D}\right)_+\right\}^2 + \frac{2}{\varepsilon^2}\left\{\left(2\varepsilon - D_U(\rho) \wedge D_V(\rho)\right)_+ - \left(2\varepsilon - \check{D}\right)_+\right\}^2 \\
&\quad \leqslant \frac{2}{\varepsilon^2}\left\{\left(2\varepsilon - \check{D}\right)_+\right\}^2 + \frac{2}{\varepsilon^2}\left\{D_U(\rho) \wedge D_V(\rho) - \check{D}\right\}^2 \\
&\quad \leqslant \frac{C}{\varepsilon^2}\left[(2\varepsilon - \check{D})_+^2 + \hat{H}^2\rho^3 + (\check{H}_U \wedge \check{H}_V)^2\rho^2\right].
\end{aligned}$$

for a suitable constant C that doesn't depend on ε or a. It follows from (8.26) and

$$\int_x^a \frac{\mathrm{d}\rho}{\sqrt{\rho^3}}\rho^\beta = \frac{1}{\frac{1}{2} - \beta}\left[x^{\beta - \frac{1}{2}} - a^{\beta - \frac{1}{2}}\right], \quad \beta < \frac{1}{2}, \tag{8.34}$$

that

$$\begin{aligned}
II(\varepsilon, a) \leqslant\ & \frac{C'}{\varepsilon^2}\mathbb{P}\left[(2\varepsilon - \check{D})_+^2\left\{\frac{(2\varepsilon - \check{D})_+}{\hat{H} + \check{H}_U \vee \check{H}_V}\right\}^{-\frac{1}{2}}\right] \\
&+ \frac{C''}{\varepsilon^2}\mathbb{P}\left[\hat{H}^2\left\{\frac{(2\varepsilon - \bar{a}^{\frac{5}{2}}\check{D})_+}{2\bar{a}^{\frac{3}{2}}(\check{H}_U \wedge \check{H}_V)} \wedge a\right\}^{\frac{5}{2}}\right] \\
&+ \frac{C'''}{\varepsilon^2}\mathbb{P}\left[(\check{H}_U \wedge \check{H}_V)^2\left\{\frac{(2\varepsilon - \bar{a}^{\frac{5}{2}}\check{D})_+}{2\bar{a}^{\frac{3}{2}}(\check{H}_U \wedge \check{H}_V)} \wedge a\right\}^{\frac{3}{2}}\right]
\end{aligned} \tag{8.35}$$

for suitable constants C', C'', and C'''.

Consider the first term in (8.35). Using Jensen's inequality for conditional expectations and (8.28) again, this term is bounded above by

$$\frac{1}{\varepsilon^2}\mathbb{P}\left[(2\varepsilon-\check{D})_+^{\frac{3}{2}}\left\{A\check{D}^{\frac{1}{2}}+B\right\}\right] \leqslant \frac{1}{\varepsilon^2}\mathbb{P}\left[(2\varepsilon-\check{D})_+^{\frac{3}{2}}\left\{2^{\frac{1}{2}}A\varepsilon^{\frac{1}{2}}+B\right\}\right] \tag{8.36}$$

for suitable constants A, B. Now, by Jensen's inequality for conditional expectation yet again, along with the invariance of standard Brownian excursion under random re-rooting (see Section 2.7 of [Ald91b]) and the fact that

$$\mathbb{P}\{\check{E}_U \in \mathrm{d}r\} = re^{-\frac{r^2}{2}}\,\mathrm{d}r \tag{8.37}$$

(see Section 3.3 of [Ald91b]), we have

$$\begin{aligned}
&\mathbb{P}\left[(2\varepsilon-\check{D})_+^{\frac{3}{2}}\right] \\
&\quad= \mathbb{P}\left[\left(\mathbb{P}\left[2\varepsilon-2\left\{\check{E}_U+\check{E}_V-2\inf_{U\wedge V\leqslant t\leqslant U\vee V}\check{E}_t\right\}\,\middle|\,\check{E}\right]\right)_+^{\frac{3}{2}}\right] \\
&\quad\leqslant \mathbb{P}\left[\left(2\varepsilon-2\left\{\check{E}_U+\check{E}_V-2\inf_{U\wedge V\leqslant t\leqslant U\vee V}\check{E}_t\right\}\right)_+^{\frac{3}{2}}\right] \qquad (8.38)\\
&\quad= \mathbb{P}\left[\left(2\varepsilon-2\check{E}_U\right)_+^{\frac{3}{2}}\right] = \int_0^\infty \mathrm{d}r\, re^{-\frac{r^2}{2}}\,(2\varepsilon-2r)_+^{\frac{3}{2}} \\
&\quad\leqslant \int_0^\varepsilon \mathrm{d}r\, r\,(2\varepsilon-2r)^{\frac{3}{2}} = 2^{\frac{3}{2}}\varepsilon^{\frac{7}{2}}\int_0^1 \mathrm{d}s\, s(1-s)^{\frac{3}{2}}.
\end{aligned}$$

Thus the limit as $\varepsilon \downarrow 0$ of the first term in (8.35) is 0 for each a.

For the second term in (8.35), first observe by Jensen's inequality for conditional expectation and (8.37) that

$$\begin{aligned}
\mathbb{P}\{\check{D}\leqslant r\} &\leqslant \mathbb{P}\left[\left(2-\frac{\check{D}}{r}\right)_+\right] \leqslant \mathbb{P}\left[\left(2-\frac{\check{E}_U}{r}\right)_+\right] \qquad (8.39)\\
&\leqslant 2\mathbb{P}\left\{\check{E}_U\leqslant 2r\right\} \leqslant 2\frac{(2r)^2}{2} = 4r^2.
\end{aligned}$$

Combining this observation with (8.29) and integrating by parts gives

$$\begin{aligned}
&\mathbb{P}\left[\left\{\frac{(2\varepsilon-\bar{a}^{\frac{5}{2}}\check{D})_+}{2\bar{a}^{\frac{3}{2}}(\check{H}_U\wedge\check{H}_V)}\wedge a\right\}^{\frac{5}{2}}\right]\\
&\quad\leqslant\mathbb{P}\left[\left\{\frac{(2\varepsilon-\bar{a}^{\frac{5}{2}}\check{D})_+}{\bar{a}^{\frac{3}{2}}\check{D}}\wedge a\right\}^{\frac{5}{2}}\right]\\
&\quad=\int_0^{2\varepsilon/\bar{a}^{\frac{5}{2}}}\mathbb{P}\{\check{D}\in\mathrm{d}r\}\left(\bar{a}^{-\frac{3}{2}}\left(\frac{2\varepsilon}{r}-\bar{a}^{\frac{5}{2}}\right)\wedge a\right)^{\frac{5}{2}} \qquad (8.40)\\
&\quad\leqslant\int_{2\varepsilon/(a\bar{a}^{\frac{1}{2}}+\bar{a}^{\frac{5}{2}})}^{2\varepsilon/\bar{a}^{\frac{5}{2}}}\mathrm{d}r\,4r^2\bar{a}^{-\frac{15}{4}}\frac{5}{2}\left(\frac{2\varepsilon}{r}-\bar{a}^{\frac{5}{2}}\right)^{\frac{3}{2}}\frac{2\varepsilon}{r^2}\\
&\quad=40\varepsilon^2\bar{a}^{-\frac{15}{4}}\int_{1/(a\bar{a}^{\frac{1}{2}}+\bar{a}^{\frac{5}{2}})}^{1/\bar{a}^{\frac{5}{2}}}\mathrm{d}s\left(\frac{1}{s}-\bar{a}^{\frac{5}{2}}\right)^{\frac{3}{2}}.
\end{aligned}$$

If we denote the rightmost term by $L(\varepsilon,a)$, then it is clear that

$$\lim_{a\downarrow 0}\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon^2}L(\varepsilon,a)=0. \qquad (8.41)$$

From (8.28) and Jensen's inequality for conditional expectations, the third term in (8.35) is bounded above by

$$\begin{aligned}
\frac{C}{\varepsilon^2}\mathbb{P}\left[(\check{H}_U\wedge\check{H}_V)^{\frac{1}{2}}\left(2\varepsilon-\bar{a}^{\frac{5}{2}}\check{D}\right)_+^{\frac{3}{2}}\right]&\leqslant\frac{C}{\varepsilon^2}\mathbb{P}\left[\check{D}^{\frac{1}{2}}\left(2\varepsilon-\bar{a}^{\frac{5}{2}}\check{D}\right)_+^{\frac{3}{2}}\right]\\
&\leqslant\frac{C}{\varepsilon^{\frac{3}{2}}}\mathbb{P}\left[\left(2\varepsilon-\bar{a}^{\frac{5}{2}}\check{D}\right)_+^{\frac{3}{2}}\right], \qquad (8.42)
\end{aligned}$$

and the calculation in (8.38) shows that the rightmost term converges to zero as $\varepsilon\downarrow 0$ for each a.

Putting together the observations we have made on the three terms in (8.35), we see that

$$\lim_{a\downarrow 0}\lim_{\varepsilon\downarrow 0}II(\varepsilon,a)=0. \qquad (8.43)$$

It follows from the dominated convergence theorem that

$$\lim_{\varepsilon\downarrow 0}III(\varepsilon,a)=0 \qquad (8.44)$$

for all a, and this completes the proof. □

Acknowledgment The authors thank David Aldous, Peter Pfaffelhuber, Jim Pitman and Tandy Warnow for helpful discussions, and the referee and associate editor for extremely close readings of the paper and several very helpful suggestions. Part of the research was conducted while the authors

were visiting the Pacific Institute for the Mathematical Sciences in Vancouver, Canada.

References

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

[Ald91b] David Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge Univ. Press, Cambridge, 1991.

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

[Ald00] David J. Aldous. Mixing time for a Markov chain on cladograms. *Combin. Probab. Comput.*, 9(3):191–204, 2000.

[AS01] Benjamin L. Allen and Mike Steel. Subtree transfer operations and their induced metrics on evolutionary trees. *Ann. Comb.*, 5(1):1–15, 2001.

[AS02] Romain Abraham and Laurent Serlet. Poisson snake and fragmentation. *Electron. J. Probab.*, 7:no. 17, 15 pp. (electronic), 2002.

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

[Ber96] Jean Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.

[BH99] Martin R. Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[BHV01] Louis J. Billera, Susan P. Holmes, and Karen Vogtmann. Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.*, 27(4):733–767, 2001.

[BRST02] Oliver Bastert, Dan Rockmore, Peter F. Stadler, and Gottfried Tinhofer. Landscapes on spaces of trees. *Appl. Math. Comput.*, 131(2-3):439–459, 2002.

[Chi01] Ian Chiswell. *Introduction to Λ-trees*. World Scientific Publishing Co. Inc., River Edge, NJ, 2001.

[DH02] Persi Diaconis and Susan P. Holmes. Random walks on trees and matchings. *Electron. J. Probab.*, 7:no. 6, 17 pp. (electronic), 2002.

[DLG02] Thomas Duquesne and Jean-François Le Gall. Random trees, Lévy processes and spatial branching processes. *Astérisque*, (281):vi+147, 2002.

[DMT96] Andreas W.M. Dress, V. Moulton, and W.F. Terhalle. T-theorie. *Europ. J. Combinatorics*, 17, 1996.

[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.*, 53:321–402, 1984.

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

[DVJ88] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*. Springer Series in Statistics. Springer-Verlag, New York, 1988.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[EPW04] Steven N. Evans, Jim Pitman, and Anita Winter. Rayleigh processes, real trees, and root growth with re-grafting. Technical Report 654, U.C. Berkeley Department of Statistics, 2004. To appear in *Probability Theory and Related Fields*.

[Fel03] Joseph Felsenstein. *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachusetts, 2003.

[FOT94] Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda. *Dirichlet Forms and Symmetric Markov Processes*. de Gruyter, 1994.

[Gro99] Misha Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. 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 Michael Bates.

[Kel75] John L. Kelley. *General topology*. Springer-Verlag, New York, 1975. Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.], Graduate Texts in Mathematics, No. 27.

[Kni81] Frank B. Knight. *Essentials of Brownian motion and diffusion*, volume 18 of *Mathematical Surveys*. American Mathematical Society, Providence, R.I., 1981.

[LG99] Jean-François Le Gall. *Spatial branching processes, random snakes and partial differential equations*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1999.

[RW00] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[Sch02] Jason Schweinsberg. An $O(n^2)$ bound for the relaxation time of a Markov chain on cladograms. *Random Structures Algorithms*, 20(1):59–70, 2002.

[SO90] D.L. Swofford and G.J. Olsen. Phylogeny reconstruction. In D.M. Hillis and G. Moritz, editors, *Molecular Systematics*, pages 411–501. Sinauer Associates, Sunderland, Massachusetts, 1990.

[SS03] Charles Semple and Mike Steel. *Phylogenetics*, volume 24 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2003.

[Stu04] Karl-Theodor Sturm. On the geometry of metric measure spaces. Technical Report 203, SFB 611, Universität Bonn, 2004.

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

[Zam01] Lorenzo Zambotti. A reflected stochastic heat equation as symmetric dynamics with respect to the 3-d Bessel bridge. *J. Funct. Anal.*, 180(1):195–209, 2001.

[Zam02] Lorenzo Zambotti. Integration by parts on Bessel bridges and related stochastic partial differential equations. *C. R. Math. Acad. Sci. Paris*, 334(3):209–212, 2002.

[Zam03] Lorenzo Zambotti. Integration by parts on δ-Bessel bridges, $\delta > 3$ and related SPDEs. *Ann. Probab.*, 31(1):323–348, 2003.

Steven N. Evans, 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

URL: http://www.stat.berkeley.edu/users/evans/

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

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

URL: http://www.mi.uni-erlangen.de/~winter/