The Annals of Probability
2017, Vol. 45, No. 2, 625–667
DOI: 10.1214/15-AOP1071

INVARIANCE PRINCIPLE FOR VARIABLE SPEED RANDOM WALKS ON TREES

BY SIVA ATHREYA[*,1], WOLFGANG LÖHR[†] AND ANITA WINTER[†]

Indian Statistical Institute[*] and University of Duisburg-Essen[†]

We consider stochastic processes on complete, locally compact tree-like metric spaces (T, r) on their "natural scale" with boundedly finite speed measure ν. Given a triple (T, r, ν) such a speed-ν motion on (T, r) can be characterized as the unique strong Markov process which if restricted to compact subtrees satisfies for all $x, y \in T$ and all positive, bounded measurable f,

$$\mathbb{E}^x\left[\int_0^{\tau_y} \mathrm{d}s f(X_s)\right] = 2\int_T \nu(\mathrm{d}z) r\big(y, c(x, y, z)\big) f(z) < \infty, \tag{0.1}$$

where $c(x, y, z)$ denotes the branch point generated by x, y, z. If (T, r) is a discrete tree, X is a continuous time nearest neighbor random walk which jumps from v to $v' \sim v$ at rate $\frac{1}{2} \cdot (\nu(\{v\}) \cdot r(v, v'))^{-1}$. If (T, r) is path-connected, X has continuous paths and equals the ν-Brownian motion which was recently constructed in [*Trans. Amer. Math. Soc.* **365** (2013) 3115–3150]. In this paper, we show that speed-ν_n motions on (T_n, r_n) converge weakly in path space to the speed-ν motion on (T, r) provided that the underlying triples of metric measure spaces converge in the Gromov–Hausdorff-vague topology introduced in [*Stochastic Process. Appl.* **126** (2016) 2527–2553].

1. Introduction and main result (Theorem 1). Fifty years ago in [31], Markov processes were considered which have in common that their state spaces are closed subsets of the real line and that their random trajectories "*do not jump over points*." When put in their "*natural scale*" these processes are determined by their "*speed measure*." Stone argues that in some sense the processes depend continuously on the speed measures. The most classical example is the symmetric simple random walk on $\mathbb{Z}$ which, after a suitable rescaling, converges to standard Brownian motion. If you rescale edge lengths by a factor $\frac{1}{\sqrt{n}}$ and speed up time by a factor n, then you might think of the rescaled random walk as such a process with speed measure $\frac{1}{\sqrt{n}} q(\sqrt{n}\cdot)$, where q denotes the counting measure on $\mathbb{Z}$, and of the standard Brownian motion as such a process whose speed measure equals the Lebesgue measure on $\mathbb{R}$.

Received April 2014; revised October 2015.

[1]Supported in part by CPDA grant from the Indian Statistical Institute.

MSC2010 subject classifications. Primary 60J65, 60B05; secondary 60J25, 60J27, 60J80, 60B99, 58J65.

Key words and phrases. Brownian motion, $\mathbb{R}$-tree, Gromov–Hausdorff-vague topology, convergence of Markov chains, diffusions on metric measure trees, speed measure, Dirichlet form.

In the present paper, we want to extend this result from $\mathbb{R}$-valued Markov processes to Markov processes which take values in tree-like metric spaces. Before we state our main result precisely, we do the preliminary work and define the space of rooted metric boundedly finite measure trees equipped with pointed Gromov-vague topology and give our notion of convergence in path space.

DEFINITION 1.1 (Rooted metric boundedly finite measure trees). (i) A *pointed Heine–Borel space* (X, r, ρ) consists of a Heine–Borel space[2] (X, r) and a distinguished point $\rho \in X$.

(ii) A *rooted metric tree* is a pointed Heine–Borel space (T, r, ρ), which is both 0-*hyperbolic*, or equivalently, satisfies the *four point condition*, that is,

$$\begin{aligned} &r(x_1, x_2) + r(x_3, x_4) \\ &\quad \le \max\{r(x_1, x_3) + r(x_2, x_4), r(x_1, x_4) + r(x_2, x_3)\}, \end{aligned} \tag{1.1}$$

holds for all $x_1, x_2, x_3, x_4 \in T$, and *fine*, that is, for all $x_1, x_2, x_3 \in T$ there is a (necessarily unique) point $c(x_1, x_2, x_3) \in T$, such that for $i, j \in \{1, 2, 3\}$, $i \neq j$,

$$r(x_i, c(x_1, x_2, x_3)) + r(x_j, c(x_1, x_2, x_3)) = r(x_i, x_j). \tag{1.2}$$

The point $c(x_1, x_2, x_3)$ is referred to as *branch point*, and the distinguished point $\rho \in T$ as the *root*.

(iii) In a rooted metric tree (T, r, ρ), we define for $a, b \in T$ the intervals

$$[a, b] := \{x \in T : r(a, x) + r(x, b) = r(a, b)\}, \tag{1.3}$$

$(a, b) := [a, b] \setminus \{a, b\}$, $[a, b) := [a, b] \setminus \{b\}$ and $(a, b] := [a, b] \setminus \{a\}$. We say that $x, y \in T$ are connected by an *edge*, in symbols $x \sim_T y$ or simply $x \sim y$, iff

$$x \neq y \quad \text{and} \quad [x, y] = \{x, y\}. \tag{1.4}$$

If $x \sim y$ and $x \in [\rho, y]$, we call the pair (x, y) an *oriented edge* of length $r(x, y)$.

(iv) A *rooted metric boundedly finite measure tree* (T, r, ρ, ν) consists of a rooted metric tree (T, r, ρ) and a measure ν on $(T, \mathcal{B}(T))$ which is finite on bounded sets and has full support, $\mathrm{supp}(\nu) = T$.

REMARK 1.2 ($\mathbb{R}$-trees versus trees with edges). A metric tree is connected (i.e., is an $\mathbb{R}$-tree) if and only if it has no edges. Due to separability, there can be only countably many edges.

We will establish a one-to-one correspondence between rooted metric boundedly finite measure trees (T, r, ρ, ν) and strong Markov processes $X = (X_t)_{t \ge 0}$ with values in (T, r) starting at ρ. When (T, r) is compact such a process can

[2]Recall that a *Heine–Borel space* is a metric space in which every bounded closed subset is compact. Note that every Heine–Borel space is complete, separable and locally compact.

be characterized by the occupation time formula given in (0.1) (see Proposition 5.1). For general rooted metric boundedly finite measure trees, the corresponding Markov process is associated with a regular Dirichlet form (see Definition 2.7). We will refer to this Markov process as *speed-ν motion on* (T, r) or *variable speed motion associated to* ν *on* (T, r). If (T, r) is path-connected, then X has continuous paths and equals the so-called ν-Brownian motion on (T, r), which was recently constructed in [5]. On the other hand, if (T, r) is discrete, X is a continuous time nearest neighbor Markov chain which jumps from v to $v' \sim v$ at rate

$$\gamma_{vv'} := \tfrac{1}{2} \cdot \big(\nu(\{v\}) \cdot r(v, v')\big)^{-1} \tag{1.5}$$

(see Lemma 2.11).

The invariance principle which we are going to state says that a sequence of variable speed motions converges in path space to a limiting variable speed motion whenever the underlying metric measure trees converge in the pointed Gromov–Hausdorff-vague topology which was recently introduced in [6]. In particular, it was shown that convergence in pointed Gromov–Hausdorff-vague topology is equivalent to convergence in pointed Gromov-vague topology together with the *uniform local lower mass-bound property*, that is, for each $\delta, R > 0$,

$$\liminf_{n\to\infty} \inf_{x \in B_n(\rho_n, R)} \nu_n\big(B_n(x, \delta)\big) > 0 \tag{1.6}$$

(see Proposition 3.8). Here, $B_n(x, R) = \{y \in T_n : r_n(x, y) < R\}$ is the ball around x with radius R in the metric space (T_n, r_n). In the Introduction, we recall only the definition of Gromov-vague topology. For a more elaborate discussion of the topology, we refer the reader to Section 3.

We call two rooted metric measure trees (T, r, ρ, ν) and (T', r', ρ', ν') equivalent iff there is an isometry φ between (T, r) and (T', r') such that $\varphi(\rho) = \rho'$ and $\nu \circ \varphi^{-1} = \nu'$. Denote

$$\mathbb{T} := \{\text{equivalence classes of rooted metric boundedly finite measure trees}\}. \tag{1.7}$$

Let $\mathcal{X} := (T, r, \rho, \nu)$, $\mathcal{X}_1 := (T_1, r_1, \rho_1, \nu)$, $\mathcal{X}_2 := (T_2, r_2, \rho_2, \nu), \dots$ be in $\mathbb{T}$. We say that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ in *pointed Gromov-vague topology* iff there are a pointed metric space (E, d_E, ρ_E) and isometries $\varphi_n : T_n \to E$ with $\varphi_n(\rho_n) = \rho_E$, for all $n \in \mathbb{N}$, as well as an isometry $\varphi : T \to E$ with $\varphi(\rho) = \rho_E$ such that the sequence of image measures $(\varphi_{n*}\nu_n)\restriction_{B(\rho_E, R)}$ restricted to the ball of radius R around the root converges weakly for all but countably many $R > 0$.

Before we are in a position to state our main scaling result, notice that the approximating Markov processes may live on different spaces. We therefore agree on the following.

DEFINITION 1.3 (A notion of convergence in path space). For every $n \in \mathbb{N} \cup \{\infty\}$, let X^n be a càdlàg process with values in a metric space (T_n, r_n).

(i) We say that $(X^n)_{n\in\mathbb{N}}$ converges to X^∞ weakly in path space (resp., f.d.d.) if there exists a metric space (E, d_E) and isometric embeddings $\phi_n: T_n \to E$, $n \in \mathbb{N}\cup\{\infty\}$, such that $(\phi_n \circ X^n)_{n\in\mathbb{N}}$ converges to $\phi_\infty \circ X^\infty$ weakly in Skorohod path space (resp., f.d.d.).

(ii) We say that $(X^n)_{n\in\mathbb{N}}$ converges to X^∞ in the one-point compactification weakly in path-space (resp., f.d.d.) if there exists a locally compact space (E, d_E) and embeddings as in (i) such that we have weak path-space (resp., f.d.d.) convergence in the one-point compactification $E \cup \{\infty\}$, where the processes are defined to take the value ∞ after their lifetimes.

To be in a position to state our invariance principle, we recall the notion of the one-point compactification $\hat{E} := E \cup \{\infty\}$ of a separable, locally compact (but non-compact) metric space E, and the life time ζ of a E-valued strong Markov process, that is,

$$\zeta := \inf\{t \geq 0 : X_t = \infty\}. \tag{1.8}$$

Our main result is the following.

THEOREM 1 (Invariance principle). *Assume that* $\mathcal{X} = (T, r, \rho, \nu)$, $\mathcal{X}_1 = (T_1, r_1, \rho_1, \nu_1)$, $\mathcal{X}_2 = (T_2, r_2, \rho_2, \nu_2), \ldots$ *are in* $\mathbb{T}$. *Let X be the speed-ν motion on (T, r) starting in ρ, and for all $n \in \mathbb{N}$, let X^n be the speed-ν_n motion on (T_n, r_n) started in ρ_n. Assume that the following conditions hold*:

(A0) *For all $R > 0$,*

$$\limsup_{n\to\infty} \sup\{r_n(x, z) : x \in B_n(\rho_n, R), z \in T_n, x \sim z\} < \infty. \tag{1.9}$$

(A1) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ pointed Gromov-vaguely.*

(A2) *The uniform local lower mass-bound property* (1.6) *holds.*

Then the following hold:

(i) *X^n converges in the one-point compactification weakly in path-space to a process Y, such that Y stopped at infinity has the same distribution as the speed-ν motion X. In particular, if X is conservative* (*i.e., does not hit infinity*), *then X^n converges weakly in path-space to X.*

(ii) *If* $\sup_{n\in\mathbb{N}} \mathrm{diam}(T_n, r_n) < \infty$, *where* diam *is the diameter, and we assume* (A1) *but not* (A2), *then X^n converges f.d.d. to X.*

REMARK 1.4 (Entrance law). Let $\mathcal{X} := (T, r, \rho, \nu)$, $\mathcal{X}_1 := (T_1, r_1, \rho_1, \nu_1)$, $\mathcal{X}_2 := (T_2, r_2, \rho_2, \nu_2), \ldots$ in $\mathbb{T}$ be such that $\mathcal{X}_n \underset{n\to\infty}{\longrightarrow} \mathcal{X}$ Gromov–Hausdorff-vaguely. The statement of Theorem 1(i) reflects the fact that it is possible that the approximating speed-ν_n motions on (T_n, r_n), as well as their limit processes on the one-point compactification, are recurrent but the speed-ν motion on (T, r) is not. Note that in such a situation we obtain an entrance law and that the limit processes cannot be a strong Markov processes. We explain this in detail in Example 5.5.

We want to briefly illustrate this invariance principle with a first non-trivial example which was established in [11]. Further examples and the relation of Theorem 1 to the existing literature are discussed in Section 7.

EXAMPLE 1.5 (RWs on GW-trees converge to BM on the CRT). Consider a Galton–Watson process in discrete time whose offspring distribution is critical and has finite (positive) variance σ^2. For each $n \in \mathbb{N}$, let $\mathcal{T}_n$ be the corresponding GW-tree conditioned on having n vertices. Given $\mathcal{T}_n$, whenever $v' \sim_{\mathcal{T}_n} v$, put $r_n(v, v') := \frac{\sigma}{\sqrt{n}}$, and let $\nu_n(\{v\}) := \frac{\deg(v)}{2n}$ for all $v \in \mathcal{T}_n$, where deg denotes the degree of node. Notice that given $\mathcal{T}_n$, the speed-ν_n random walk on $(\mathcal{T}_n, r_n)$ is the symmetric nearest neighbor random walk on $\mathcal{T}_n$ with edge lengths rescaled by a factor $\frac{\sigma}{\sqrt{n}}$ and with exponential jump rates

$$\gamma_n(v) = \frac{1}{2\nu_n(\{v\})} \sum_{v' \sim v} r_n^{-1}(v, v') = \frac{1}{2} \cdot \frac{2n}{\deg(v)} \cdot \deg(v) \frac{\sqrt{n}}{\sigma} = \sigma^{-1} \cdot n^{3/2}. \tag{1.10}$$

Denote by μ_n^{ske} the normalized length-measure (see Section 2.1) on the path-connected tree $\bar{\mathcal{T}}_n$ spanned by $\mathcal{T}_n$. Then it is known that $(\bar{\mathcal{T}}_n, r_n, \mu_n^{\mathrm{ske}})$ converges Gromov-vaguely in distribution to some random, compact, path-connected metric measure tree $(\mathcal{T}, r, \mu)$, where $(\mathcal{T}, r)$ is the so-called Brownian continuum random tree (or shortly, the CRT), and μ the "leaf-measure" (see, e.g., [4], Theorem 23). As the Prohorov distance between ν_n and μ_n^{ske} is not greater than $\frac{\sigma}{2\sqrt{n}}$, $(\mathcal{T}_n, r_n, \nu_n)$ also converges Gromov-vaguely to $(\mathcal{T}, r, \mu)$ by [6], Lemma 2.10. Furthermore, it is known that the family $\{\nu_n; n \in \mathbb{N}\}$ satisfies the uniform local lower mass-bound property (compare [4], Corollary 19, together with Proposition 3.8).

We can therefore conclude from Theorem 1 that given a realization of a sequence $(\mathcal{T}_n)_{n\in\mathbb{N}}$ converging Gromov-weakly to some $\mathcal{T}$, the symmetric random walk with jumps rescaled by $\frac{1}{\sqrt{n}}$ and time speeded up by a factor of $n^{3/2}$ converges to μ-Brownian motion on the CRT. This was first conjectured in [3], Section 5.1 and proved in [11]. A more general result on homogeneous scaling limits of random walks on graph trees towards diffusions on continuum trees was established in [12]. We will discuss in Section 7.3 how this result is covered by our invariance principle.

For the proof of the invariance principle, we use the following approach. We first use techniques from Dirichlet forms to construct the speed-ν motion on (T, r). We continue showing tightness based on a version of Aldous' stopping time criterion (Proposition 4.2), and then identify the limit. As we are working with Dirichlet forms, one might be tempted to show f.d.d.-convergence of the motions by verifying the Mosco-convergence introduced in [28] (compare also [29] for its application to Dirichlet forms). It turns out, however, that this is tedious, and we rather identify the limit via the occupation time formula (0.1). For that, we first restrict ourselves to limit metric (finite) measure trees which are compact, and show that any limit point must be a strong Markov process satisfying (0.1). We then reduce the general case to the case of compact limit trees by showing that there are suitably many hitting times which converge.

The rest of the paper is organized as follows: In Section 2, we construct the speed-ν motion on (T, r) and present occupation time formula (0.1). In Section 3 we introduce all the topological concepts needed to deal with convergence of the underlying metric measure spaces. In Section 4, we prove the tightness of a sequence of speed-ν_n motions on (T_n, r_n) provided that the underlying spaces $(T_n, r_n, \nu_n)_{n\in\mathbb{N}}$ converges. In Section 5, we show that any limit point satisfies the strong Markov property and that its occupation time formula agrees with that of the limit variable speed motion. In Section 6, we collect all the ingredients to present the proof of Theorem 1. Finally, in Section 7 we present examples and relate our result to the existing literature.

2. The speed-ν motion on (T, r) and its Dirichlet form.

In this section, we will use Dirichlet form techniques to construct the variable speed motions. We will follow the lines of [5] where the variable speed motion was constructed on path-connected rooted metric measure trees, or rooted measure $\mathbb{R}$-*trees* for short. The main idea behind the generalization to arbitrary rooted metric measure trees is the presentation of a universal notion of the length measure and the gradient. This will be given in Section 2.1. In Section 2.2, we associate the variable speed motion with a Dirichlet form and establish in Section 2.3 the occupation time formula. We will revise (where necessary) the proofs given in [5] to the larger class of underlying rooted metric measure trees.

2.1. *The set-up.*

In this subsection, we discuss preliminaries that are required to construct the variable speed motions.

Recall rooted metric trees and rooted $\mathbb{R}$-trees from Definition 1.1, and notice that a rooted metric tree (T, r, ρ) can be embedded isometrically into an $\mathbb{R}$-*tree*, that is, a path-connected rooted metric tree (see, e.g., Theorem 3.38 in [14]). Furthermore, there is a unique (up to isometry) smallest rooted $\mathbb{R}$-tree, $(\bar{T}, \bar{r}, \rho)$, which contains (T, r, ρ) (compare, e.g., [27], Remark 2.7). $(\bar{T}, \bar{r})$ is the smallest $\mathbb{R}$-tree in the following sense: if $(\hat{T}, \hat{r})$ is another $\mathbb{R}$-tree with $T \subseteq \hat{T}$, and $\hat{r}$ extends r, then there is a unique isometric embedding $\phi: \bar{T} \to \hat{T}$ such that $\phi\restriction_T$ is the identity on T. Heuristically, $(\bar{T}, \bar{r})$ is obtained from (T, r) by replacing edges with line segments of the appropriate length.

Given a rooted metric tree (T, r, ρ), we can define a partial order (with respect to ρ), $\le_\rho$, on T by saying that $x \le_\rho y$ for all $x, y \in T$ with $x \in [\rho, y]$.

To be in a position to capture that our variable speed motions are processes on "natural scale," we need the notion of a length measure. For $\mathbb{R}$-trees, it was first introduced in [15]. It turns out that this measure can be constructed on any separable 0-hyperbolic metric space provided that we have fixed a reference point, say the root ρ. Let therefore (T, r, ρ) be a rooted metric tree, and $\mathcal{B}(T)$ the Borel-σ-algebra of (T, r). We denote the set of *isolated points* (other than the root) by $\mathrm{Iso}(T, r, \rho)$, and define the *skeleton* of (T, r, ρ) as

$$T^o := \mathrm{Iso}(T, r, \rho) \cup \bigcup_{a\in T} (\rho, a). \tag{2.1}$$

Recall that rooted metric trees are Heine–Borel spaces, and thus separable, and observe that if $T' \subset T$ is a dense countable set, then (2.1) holds with T replaced by T'. In particular, $T^o \in \mathcal{B}(T)$ and $\mathcal{B}(T)\restriction_{T^o} = \sigma(\{(a, b); a, b \in T'\})$, where $\mathcal{B}(T)\restriction_{T^o} := \{A \cap T^o;\ A \in \mathcal{B}(T)\}$. Hence, there exist a unique σ-finite measure $\lambda^{(T,r,\rho)}$ on T, such that $\lambda^{(T,r,\rho)}(T \setminus T^o) = 0$ and for all $a \in T$,

$$\lambda^{(T,r,\rho)}\big((\rho, a]\big) = r(\rho, a). \tag{2.2}$$

DEFINITION 2.1 (Length measure). Let (T, r, ρ) be a rooted metric tree. The unique σ-finite measure $\lambda^{(T,r,\rho)}$ satisfying (2.2) and $\lambda^{(T,r,\rho)}(T \setminus T^o) = 0$ is called the *length measure* of (T, r, ρ).

REMARK 2.2 (Length measure; particular instances). (i) If (T, r) is an $\mathbb{R}$-tree, then $\lambda^{(T,r,\rho)}$ does not depend on the root ρ, and is the trace onto T^o of the 1-dimensional Hausdorff-measure on T.

(ii) If (T, r) is discrete as a topological space, that is, all points in T are isolated, the length measure shifts all the "length" sitting on an edge to the end point which is further away from the root. In this case, it does explicitly depend on the root.

(iii) In general, let $(\bar{T}, \bar{r})$ be the $\mathbb{R}$-tree spanned by (T, r) and $\pi : \bar{T} \to T$ defined by

$$\pi(x) := \inf\{y \in T : x \le_\rho y\}, \tag{2.3}$$

for all $x \in \bar{T}$. Note that π is well defined because T is closed and satisfies (1.2). It is therefore easy to check that

$$\lambda^{(T,r,\rho)} = \pi_* \lambda^{(\bar{T},\bar{r})}. \tag{2.4}$$

In order to characterize the variable speed motion analytically (via Dirichlet forms), we use a concept of weak differentiability. Denote the space of continuous functions $f : T \to \mathbb{R}$ by $\mathcal{C}(T)$. We call a function $f \in \mathcal{C}(T)$ *locally absolutely continuous* if and only if for all $\varepsilon > 0$ and all subsets $S \subseteq T$ with $\lambda^{(T,r,\rho)}(S) < \infty$ there exists a $\delta = \delta(\varepsilon, S)$ such that if $[x_1, y_1], \dots, [x_n, y_n] \subseteq S$ are disjoint arcs with $\sum_{i=1}^n r(x_i, y_i) < \delta$ then $\sum_{i=1}^n |f(x_i) - f(y_i)| < \varepsilon$. Put

$$\mathcal{A} = \mathcal{A}^{(T,r)} := \big\{f \in \mathcal{C}(T) : f \text{ is locally absolutely continuous}\big\}. \tag{2.5}$$

Of course, if (T, r) is discrete, then $\mathcal{A}$ equals the space $\mathcal{C}(T)$ of continuous functions.

The definition of the gradient is then based on the following observation which was proved for $\mathbb{R}$-trees in [5], Proposition 1.1.

PROPOSITION 2.3 (Gradient). *Let $f \in \mathcal{A}$. There exists a unique* (*up to $\lambda = \lambda^{(T,r,\rho)}$-zero sets*) *function $g \in L^1_{\text{loc}}(\lambda^{(T,r,\rho)})$ such that*

$$f(y) - f(x) = \int_{[\rho,y]} \lambda(\mathrm{d}z) g(z) - \int_{[\rho,x]} \lambda(\mathrm{d}z) g(z), \tag{2.6}$$

for all $x, y \in T$. Moreover, g is already uniquely determined (*up to $\lambda^{(T,r,\rho)}$-zero sets*) *if we only require* (2.6) *to hold for all $x \le_\rho y$.*

PROOF. For $f \in \mathcal{A}$, we define the linear extension $\bar{f}: \bar{T} \to \mathbb{R}$ by $\bar{f}\restriction_T := f$ and

$$\bar{f}(v) := \frac{r(v,y)}{r(x,y)} f(x) + \frac{r(v,x)}{r(x,y)} f(y), \tag{2.7}$$

whenever (x, y) is an edge of T and $v \in [x, y] \subseteq \bar{T}$. By [5], Proposition 1.1, there is $\bar{g}: \bar{T} \to \mathbb{R}$ such that (2.6) holds for $x, y \in \bar{T}$ and $\bar{\lambda} := \lambda^{(\bar{T},\bar{r})}$ instead of λ. It is easy to see from the definition of $\bar{f}$ that $\bar{g}$ is constant on edges of T, and hence $g: T \to \mathbb{R}$ is well defined by $g \circ \pi := \bar{g}$, with π defined in Remark 2.2(iii). By (2.4),

$$\begin{aligned} f(x) - f(y) &= \int_{[\rho,y]} \mathrm{d}\bar{\lambda}\bar{g} - \int_{[\rho,x]} \mathrm{d}\bar{\lambda}\bar{g} \\ &= \int_{[\rho,y]} \mathrm{d}\lambda g - \int_{[\rho,x]} \mathrm{d}\lambda g. \end{aligned} \tag{2.8}$$

Uniqueness and integrability of g follow from the corresponding properties of $\bar{g}$. □

The statement of Proposition 2.3 yields a general notion of a gradient.

DEFINITION 2.4 (Gradient). The gradient, $\nabla f = \nabla^{(T,r,\rho)} f$, of $f \in \mathcal{A}$ is the unique up to $\lambda^{(T,r,\rho)}$-zero sets function g which satisfies (2.6) for all $x, y \in T$.

2.2. *The regular Dirichlet form.* In this subsection, we recall the construction of the so-called ν-Brownian motion on an $\mathbb{R}$-tree given in [5], and extend it to arbitrary rooted metric measure trees.

As usual, we denote by $\mathcal{C}(T)$ the space of continuous functions $f: T \to \mathbb{R}$, and the subspace of functions vanishing at infinity by

$$\mathcal{C}_\infty(T) := \{ f \in \mathcal{C}(T) : \forall \varepsilon > 0\ \exists K \text{ compact } \forall x \in T \setminus K : |f(x)| \le \varepsilon \}. \tag{2.9}$$

Consider the bilinear form $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$ where

$$\mathcal{E}(f, g) := \frac{1}{2} \int \mathrm{d}\lambda \nabla f \nabla g, \tag{2.10}$$

and

$$\mathcal{D}(\mathcal{E}) := \{f \in L^2(\nu) \cap \mathcal{A} \cap \mathcal{C}_\infty(T) : \nabla f \in L^2(\lambda)\}. \tag{2.11}$$

For technical purposes, we also introduce for all closed subsets $A \subseteq T$ the domain

$$\mathcal{D}_A(\mathcal{E}) := \{f \in \mathcal{D}(\mathcal{E}) : f|_A \equiv 0\}. \tag{2.12}$$

We first note that the bilinear form $(\mathcal{E}, \mathcal{D}_A(\mathcal{E}))$ is closable for all closed sets $A \subseteq T$. Indeed, let $(f_n)_{n\in\mathbb{N}}$ be an $\mathcal{E}$-Cauchy sequence in $\mathcal{D}_A(\mathcal{E}) \subseteq L^2(\nu)$ with $\|f_n\|_{L^2(\nu)} \to 0$. Then, by passing to a subsequence if necessary, we may assume $\nabla f_n \to 0$, $\lambda^{(T,r,\rho)}$-almost surely and $\mathcal{E}(f_n, f_n)$ is uniformly bounded in $n \in \mathbb{N}$ (see, e.g., [5], (2.15), (2.16)).

Let $(\mathcal{E}, \bar{\mathcal{D}}_A(\mathcal{E}))$ be the closure of $(\mathcal{E}, \mathcal{D}_A(\mathcal{E}))$, that is, $\bar{\mathcal{D}}_A(\mathcal{E})$ is the closure of $\mathcal{D}_A(\mathcal{E})$ with respect to $\mathcal{E}_1 = \mathcal{E} + \langle\cdot,\cdot\rangle_\nu$.

REMARK 2.5 (Closing the form might not be necessary). The procedure of closing the form is unnecessary if the global lower mass-bound property holds on $T \setminus A$, that is, for all $\delta > 0$,

$$\inf_{x\in T\setminus A} \nu\big(B(x,\delta)\big) > 0. \tag{2.13}$$

In this case, $\bar{\mathcal{D}}_A(\mathcal{E}) = \mathcal{D}_A(\mathcal{E})$.

The following lemma is an immediate consequence of Proposition 2.4, Lemmas 2.8, 3.4 and Proposition 4.1 in [5].

LEMMA 2.6 (Regular Dirichlet form). *Let (T, r, ν) be a metric boundedly finite measure tree, and $A \subseteq T$ a closed subset. Then the following hold*:

(i) *The bilinear form $(\mathcal{E}, \bar{\mathcal{D}}_A(\mathcal{E}))$ is a regular Dirichlet form.*

(ii) *Dirac measures are of finite energy integral, there exists a constant $C_x > 0$ such that for all $f \in \mathcal{D}(\mathcal{E}) \cap \mathcal{C}_0(T)$,*

$$f(x)^2 \le C_x \mathcal{E}_1(f, f) \tag{2.14}$$

(*see* (2.2.1) *in* [18]).

(iii) *If A is non-empty the Dirichlet form is transient.*

It follows immediately from [18], Theorem 7.2.1, that there is a unique (up to ν-equivalence) ν-symmetric strong Markov process

$$X = \big((X_t)_{t\ge0}, (\mathbb{P}^x)_{x\in T}\big) \tag{2.15}$$

on (T, r) associated with the regular Dirichlet form $(\mathcal{E}, \bar{\mathcal{D}}(\mathcal{E}))$.

DEFINITION 2.7 (Speed-ν motion on (T,r)). Let (T,r,ν) be a metric boundedly finite measure tree. In the following, we refer to the unique ν-symmetric strong Markov process associated with $(\mathcal{E},\bar{\mathcal{D}}(\mathcal{E}))$ as the speed-ν motion on (T,r).

- If (T,r) is discrete, then the speed-ν motion on (T,r) is referred to as speed-ν random walk on (T,r).
- If (T,r) is an $\mathbb{R}$-tree, then the speed-ν motion on (T,r) agrees with the ν-Brownian motion on (T,r) constructed in [5].

REMARK 2.8 (Variable speed motion does not depend on root). Notice that although the definition of the length measure and the gradient depend on the root, the Dirichlet form does not. Therefore, the variable speed motion is independent on the choice of the root.

REMARK 2.9 (Connectedness and continuous paths). Notice that the Dirichlet form satisfies the local property if and only if the underlying space is connected. Thus, the variable speed motion on (T,r) has continuous paths if and only if (T,r) is an $\mathbb{R}$-tree.

Recall the explosion time ζ from (1.8). Notice that the Dirichlet form need not be conservative, which means that the speed-ν motion might exist only for a (random) finite *life time*. This happens when it explodes in finite time, that is, $\zeta<\infty$.

REMARK 2.10 (Finite versus infinite life time). Let (T,r,ν) be a rooted boundedly finite measure tree, and X the speed-ν motion on (T,r). Whether or not $\zeta=\infty$, almost surely, depends on the tree topology and the measure ν.

(i) The speed-ν motion on (T,r) cannot explode if it is recurrent. Recurrence depends on (T,r,ν) only through (T,r). See [5], Theorem 4, for recurrence criteria.
(ii) An example of a transient variable speed motion with finite life time will be discussed in Example 5.5.

LEMMA 2.11 (Variable speed motion on discrete trees is a Markov chain). *Let (T,r,ν) be a metric boundedly finite measure tree such that (T,r) is discrete. Then the speed-ν random walk on (T,r) is a continuous time nearest neighbor Markov chain with jumps from v to $v'\sim v$ at rate $\gamma_{vv'}:=\frac{1}{2\cdot\nu(\{v\})\cdot r(v,v')}$.*

PROOF. Recall from Definition 1.1 that (T,r) is a Heine–Borel space. Thus, each ball around ρ contains only a finite number of branch points, and in consequence the nearest neighbor random walk with the jump rates $(\gamma_{vv'})_{v\sim v'}$ is a well-defined strong Markov process. Its generator Ω acts on the space $\mathcal{C}_c(T)$ of continuous functions which depend only on finitely many $v \in T$ as follows:

$$\Omega f(v) := \frac{1}{2 \cdot \nu(\{v\})} \sum_{v' \sim v} \frac{1}{r(v, v')} \big(f(v') - f(v)\big). \tag{2.16}$$

Notice that for all $f, g \in \mathcal{C}_c(T)$,

$$\begin{aligned} \mathcal{E}(f, g) &= \frac{1}{2} \int \mathrm{d}\lambda \nabla f \nabla g \\ &= \frac{1}{2} \sum_{v \in T} \frac{1}{2} \sum_{v' \sim v} \frac{1}{r(v, v')} \big(f(v') - f(v)\big)\big(g(v') - g(v)\big) \\ &= -\sum_{v \in T} \nu\big(\{v\}\big) \frac{1}{2\nu(\{v\})} \sum_{v' \sim v} \frac{1}{r(v, v')} \big(f(v') - f(v)\big) g(v) \\ &= -(\Omega f, g)_\nu. \end{aligned} \tag{2.17}$$

The statement therefore follows from Example 1.2.5 together with Exercise 4.4.1 in [18]. □

2.3. *The occupation time formula*. We conclude this section by recalling here the occupation time formula as known from speed-ν motions on $\mathbb{R}$ or the ν-Brownian motion on compact metric trees (see, e.g., [5], Proposition 1.9).

As usual, we denote for each $x \in T$ by

$$\tau_x = \tau_x(X) := \inf\{t \geq 0 : X_t = x\} \tag{2.18}$$

the *first hitting time* of x. A standard calculation shows the following.

PROPOSITION 2.12 (Occupation time formula). *Let X be a speed-ν motion on (T, r). If X is recurrent, then for all $x, z \in T$,*

$$\mathbb{E}^x\left[\int_0^{\tau_z} f(X_t)\,\mathrm{d}t\right] = 2 \int_T f(y) \cdot r\big(z, c(x, z, y)\big) \nu(\mathrm{d}y), \tag{2.19}$$

for all bounded, measurable $f : T \to \mathbb{R}$. Moreover, the process $X_{\cdot \wedge \tau_z}$ is transient for all $z \in T$.

PROOF. Let (T, r, ν) be a metric boundedly finite measure tree, and $z \in T$ fixed. By Lemma 2.6(iii), the Dirichlet form $(\mathcal{E}, \mathcal{D}_{\{z\}}(\mathcal{E}))$ is transient. Therefore by Theorem 4.4.1(ii) in [18], $R_{\{z\}} f(x) := \mathbb{E}^x[\int_0^{\tau_z} \mathrm{d}s f(X_s)]$ is the resolvent of the speed-ν motion killed on hitting z, that is,

$$\mathcal{E}(R_{\{z\}} f, h) = \int \mathrm{d}\nu h \cdot f, \tag{2.20}$$

for all $h \in \bar{\mathcal{D}}_{\{z\}}(\mathcal{E})$ and $f \in \mathcal{D}(\mathcal{E})$ with $(R_{\{z\}}f, f)_\nu < \infty$. The resolvent of a Markov process has the form

$$R_{\{z\}} f(x) = \int_T \nu(\mathrm{d}y) \frac{h^*_{\{z\},y}(x)}{\mathrm{cap}_{\{z\}}(y)} f(y), \tag{2.21}$$

where $\mathrm{cap}_{\{z\}}(y) := \inf\{\mathcal{E}(f, f) : f \in \bar{\mathcal{D}}(\mathcal{E}), f(z) = 0, f(y) = 1\}$ and $h^*_{\{z\},y}$ is the unique minimizer for $\mathrm{cap}_{\{z\}}(y)$. This can be shown by essentially rewriting the argument laid out in [5], Section 3. Moreover, for our particular Dirichlet form we find that $h^*_{\{z\},y}(x) := \frac{r(c(x,y,z),z)}{r(y,z)}$ and $\mathrm{cap}_{\{z\}}(y) = \frac{1}{2r(y,z)}$, and thus that

$$\mathbb{E}^x\left[\int_0^{\tau_z} \mathrm{d}s f(X_s)\right] = 2 \int \nu(\mathrm{d}y) r\big(z, c(x, y, z)\big) f(y). \tag{2.22}$$ □

3. Preliminaries on the Gromov-vague topology. Recall the notion of a rooted metric (boundedly finite) measure space (T, r, ρ, ν) from Definition 1.1. Once more, we call two rooted metric measure trees (T, r, ρ, ν) and (T', r', ρ', ν') equivalent iff there is an isometry φ between $\mathrm{supp}(\nu) \cup \{\rho\}$ and $\mathrm{supp}(\nu') \cup \{\rho'\}$ such that $\varphi(\rho) = \rho'$ and $\nu \circ \varphi^{-1} = \nu'$, and denote by

(3.1) $\mathbb{T} :=$ the space of equivalence classes of rooted metric measure trees.

In this section, we want to equip $\mathbb{T}$ with the so-called *Gromov–Hausdorff-vague topology* on which the convergence of the underlying spaces in our invariance principle is based on. We refer the reader to [6] for many detailed discussions. We recall the definition of the pointed Gromov-weak topology on finite metric measure spaces in Section 3.1 and then extend it to a Gromov-vague topology on $\mathcal{T}$ in Section 3.2. Finally, we compare the notions of Gromov-weak and Gromov-vague convergence in Section 3.3.

3.1. *Gromov-weak and Gromov–Hausdorff-weak topology.* In this subsection, we restrict to compact metric spaces and recall the Gromov-weak topology. This topology originates from the work of Gromov [20] who considers topologies allowing to compare metric spaces who might not be subspaces of a common metric space. The *Gromov-weak topology* on complete and separable metric measure spaces was introduced in [19]. In the same paper, the Gromov-weak topology was metrized by the so-called *Gromov–Prohorov-metric* which is equivalent to Gromov's box metric introduced in [20], as was shown in [26]. The topology is closely related to the so-called *measured Gromov–Hausdorff topology* which was first introduced by [17], and further discussed in [16, 25].

REMARK 3.1 (Full-support assumption). Note that, by our definition, the measure ν of a metric boundedly finite measure tree (T, r, ρ, ν) is required to have full support. This is usually not assumed for metric measure spaces, but it is only a minor restriction, because whenever $\rho \in \mathrm{supp}(\nu)$ we can choose representatives with full support.

Consider also the subspace

$$\mathbb{T}_c := \{(T, r, \rho, \nu) \in \mathbb{T} : (T, r) \text{ is compact}\}. \tag{3.2}$$

We shortly recall the basic definitions of the Gromov-weak and Gromov–Hausdorff-weak topologies on $\mathbb{T}_c$.

DEFINITION 3.2 (Gromov-weak and Gromov–Hausdorff-weak topology). Let for each $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n)$ be in $\mathbb{T}_c$. We say that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ in:

(i) *Pointed Gromov-weak topology* if and only if there exists a complete, separable rooted metric space (E, d_E, ρ_E) and for each $n \in \mathbb{N} \cup \{\infty\}$ isometries $\varphi_n : T_n \to E$ with $\varphi_n(\rho_n) = \rho_E$, and such that

$$(\varphi_n)_*\nu_n \underset{n\to\infty}{\Longrightarrow} (\varphi_\infty)_*\nu_\infty. \tag{3.3}$$

(ii) *Pointed Gromov–Hausdorff-weak topology* if and only if there exists a compact metric space (E, d_E, ρ_E) and for each $n \in \mathbb{N} \cup \{\infty\}$ isometries $\varphi_n : T_n \to E$ with $\varphi_n(\rho_n) = \rho_E$, such that (3.3) holds and

$$\mathrm{supp}\big((\varphi_n)_*\nu_n\big) \xrightarrow[n\to\infty]{\text{Hausdorff}} \mathrm{supp}\big((\varphi_\infty)_*\nu_\infty\big). \tag{3.4}$$

REMARK 3.3 (Supports do not converge under Gromov-weak convergence). Consider, for example, $T_n :\equiv \{\rho, \rho'\}$ and $r_n(\rho, \rho') \equiv 1$, and put $\nu_n := \frac{n-1}{n}\delta_\rho + \frac{1}{n}\delta_{\rho'}$ for all $n \in \mathbb{N}$. Clearly, $((T_n, r_n, \rho, \nu_n))_{n\in\mathbb{N}}$ converges pointed Gromov-weakly to the unit mass pointed singleton $(\{\rho\}, \rho, \delta_\rho)$. The supports, however, do not converge. This shows that Gromov-weak is in general weaker than Gromov–Hausdorff-weak convergence.

In order to close the gap between Gromov-weak and Gromov–Hausdorff-weak convergence, we define for each $\delta > 0$ the *lower mass-bound function* $m_\delta : \mathbb{T} \to \mathbb{R}_+$ as

$$m_\delta\big((T, r, \rho, \nu)\big) := \inf\{\nu\big(\bar{B}_r(x, \delta)\big) : x \in T\}. \tag{3.5}$$

It follows from our full-support assumption, $\mathrm{supp}(\nu) = T$, that $m_\delta(\mathcal{X}) > 0$ for all $\delta > 0$ if $\mathcal{X} \in \mathbb{T}_c$.

DEFINITION 3.4 (Global lower mass-bound property). We say that a family $\Gamma \subseteq \mathbb{T}_c$ satisfies the global lower mass-bound property if and only if the lower mass-bound functions are all bounded away from zero uniformly in Γ, that is, for each $\delta > 0$,

$$m_\delta(\Gamma) := \inf_{\mathcal{X}\in\Gamma} m_\delta(\mathcal{X}) > 0. \tag{3.6}$$

The following is Theorem 6.1 in [6].

PROPOSITION 3.5 (Gromov-weak versus Gromov–Hausdorff-weak topology). *Let for each $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n)$ be in $\mathbb{T}_c$ such that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ pointed Gromov-weakly. Then the following are equivalent*:

(i) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ pointed Gromov–Hausdorff-weakly.*

(ii) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ satisfies the global lower mass-bound property.*

3.2. *Gromov-vague and Gromov–Hausdorff-vague topology.* Recently, in [1], the Gromov–Hausdorff-weak topology on rooted compact length spaces was extended to complete locally compact length spaces equipped with locally finite measures. In this subsection, we want, in similar spirit, extend the Gromov(–Hausdorff)-weak topology on $\mathbb{T}_c$ to the Gromov(–Hausdorff)-vague topology on $\mathbb{T}$.

The restriction of $\mathcal{X} = (X, r, \rho, \nu) \in \mathbb{T}$ to the closed ball $\bar{B}(\rho, R)$ of radius $R > 0$ around the root is denoted by

$$\mathcal{X}\restriction_R := \big(\bar{B}(\rho, R), r, \rho, \nu\restriction_{\bar{B}_r(\rho,R)}\big). \tag{3.7}$$

DEFINITION 3.6 (Gromov-vague topology). Let for each $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n)$ be in $\mathbb{T}$. We say that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ in:

(i) *Pointed Gromov-vague topology* if and only if there exists a complete, separable rooted metric space (E, d_E, ρ_E) and for each $n \in \mathbb{N} \cup \{\infty\}$ isometries $\varphi_n : T_n \to E$ with $\varphi_n(\rho_n) = \rho_E$, and such that

$$\big((\varphi_n)_*\nu_n\big)\restriction_R \underset{n\to\infty}{\Longrightarrow} \big((\varphi_\infty)_*\nu_\infty\big)\restriction_R \tag{3.8}$$

for all but countably many $R > 0$.

(ii) *Pointed Gromov–Hausdorff-vague topology* if and only if there exists a rooted Heine–Borel space (E, d_E, ρ_E) and for each $n \in \mathbb{N} \cup \{\infty\}$ isometries $\varphi_n : T_n \to E$ with $\varphi_n(\rho_n) = \rho_E$, and such that (3.8) and

$$\varphi_n(T_n) \cap \bar{B}_{d_E}(\rho_E, R) \xrightarrow[n\to\infty]{\text{Hausdorff}} \varphi(T) \cap \bar{B}_{d_E}(\rho_E, R) \tag{3.9}$$

hold for all but countably many $R > 0$.

Once more, we want to close the gap between Gromov-vague and Gromov–Hausdorff-vague convergence. Define, therefore, for all $\delta > 0$ and $R > 0$ the *local lower mass-bound function* $m_\delta^R : \mathbb{T} \to \mathbb{R}_+ \cup \{\infty\}$ as

$$m_\delta^R\big((T, r, \rho, \nu)\big) := \inf\big\{\nu\big(\bar{B}_r(x, \delta)\big) : x \in B(\rho, R)\big\}. \tag{3.10}$$

Notice that $m_\delta^R(\mathcal{X}) > 0$ for all $\mathcal{X} \in \mathbb{T}$, and $\delta, R > 0$.

DEFINITION 3.7 (Local lower mass-bound property). We say that a family $\Gamma \subseteq \mathbb{T}$ satisfies the local lower mass-bound property if and only if the lower mass-bound functions are all bounded away from zero uniformly in Γ, that is, for each $\delta > 0$ and $R > 0$,

$$m_\delta^R(\Gamma) := \inf_{\mathcal{X} \in \Gamma} m_\delta^R(\mathcal{X}) > 0. \tag{3.11}$$

The following is Corollary 5.2 in [6].

PROPOSITION 3.8 (Gromov-vague versus Gromov–Hausdorff-vague). *Let for each $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n)$ be in $\mathbb{T}$ such that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ pointed Gromov-vaguely. Then the following are equivalent*:

(i) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ pointed Gromov–Hausdorff-vaguely.*

(ii) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ satisfies the local lower mass-bound property.*

3.3. *Gromov-weak versus Gromov-vague convergence*. Note that the concept of Gromov-vague convergence on $\mathbb{T}$ is not strictly an extension of the concept of Gromov-weak convergence on $\mathbb{T}_c$ because in the limit parts might "vanish at infinity," and hence a non-converging sequence of compact spaces with respect to the Gromov-weak or the Gromov–Hausdorff-weak topology may converge in the "locally compact version" of the corresponding topology.

REMARK 3.9 (Gromov-vague versus Gromov-weak). Consider the subspaces $\mathbb{T}_{\text{finite}}$ and $\mathbb{T}_{\text{probability}}$ of $\mathbb{T}$ consisting of spaces $\mathcal{X} = (T, r, \rho, \nu) \in \mathbb{T}$ where ν is a finite or a probability measure, respectively. Then on $\mathbb{T}_{\text{probability}}$ the induced Gromov-vague topology coincides with the Gromov-weak topology. However, on $\mathbb{T}_{\text{finite}}$ and even on $\mathbb{T}_c$ this is not the case, as the total mass might not be preserved under Gromov-vague convergence. In fact, for $\mathcal{X}, \mathcal{X}_n = (T_n, r_n, \rho_n, \nu_n) \in \mathbb{T}_{\text{finite}}$ the following are equivalent:

(i) $\mathcal{X}_n \to \mathcal{X}$ Gromov-weakly.

(ii) $\mathcal{X}_n \to \mathcal{X}$ Gromov-vaguely and $\nu_n(T_n) \to \nu(T)$.

Moreover, $\mathcal{X}_n \to \mathcal{X} \in \mathbb{T}_c$ Gromov–Hausdorff-weakly if and only if $\mathcal{X}_n \to \mathcal{X}$ Gromov–Hausdorff-vaguely and the diameters of (T_n, r_n) are bounded uniformly in n (except for finitely many n).

4. Tightness. Recall the speed-ν motion on (T, r), $X^{(T,r,\nu)}$, from Definition 2.7. In this section, we prove that the sequence $\{X^{(T_n,r_n,\nu_n)}; n \in \mathbb{N}\}$ is tight provided that assumptions (A0), (A1) and (A2) from Theorem 1 are satisfied. The main result is the following.

PROPOSITION 4.1 (Tightness). *Consider rooted metric boundedly finite measure trees* $\mathcal{X} := (T, r, \rho, \nu)$ *and* $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n)$, $n \in \mathbb{N}$. *Assume that for all* $n \in \mathbb{N}$, $\mathcal{X}_n$ *is discrete, and that the following conditions hold*:

(A0) *For all* $R > 0$,

$$\limsup_{n\to\infty} \sup\{r_n(x, z) : x \in B_n(\rho_n, R), z \in T_n, x \sim z\} < \infty. \tag{4.1}$$

(A1) *The sequence* $(\mathcal{X}_n)_{n\in\mathbb{N}}$ *converges to* $\mathcal{X}$ *in the pointed Gromov-vague topology as* $n \to \infty$.

(A2) *The local lower mass-bound property holds uniformly in* $n \in \mathbb{N}$.

Then there is a Heine Borel space (E, d), *such that* T *and all* T_n, $n \in \mathbb{N}$, *are embedded in* (E, d) *and the sequence* X^n, $n \in \mathbb{N}$, *of speed-*ν_n *random walks on* (T_n, r_n) *is tight in the one-point compactification of* E.

For the proof, we rely on the following version of the Aldous tightness criterion (see, [21], Theorems 16.11 and 16.10).

PROPOSITION 4.2 (Aldous tightness criterion). *Let* $X^n = (X^n_t)_{t\geq 0}$, $n \in \mathbb{N}$, *be a sequence of càdlàg processes on a complete*, *separable metric space* (E, d). *Assume that the one-dimensional marginal distributions are tight*, *and for any bounded sequence of* X^n*-stopping times* τ_n *and any* $\delta_n > 0$ *with* $\delta_n \to 0$ *we have*

$$d\big(X^n_{\tau_n}, X^n_{\tau_n+\delta_n}\big) \overset{n\to\infty}{\longrightarrow} 0 \qquad \textit{in probability}. \tag{4.2}$$

Then the sequence $(X^n)_{n\in\mathbb{N}}$ *is tight*.

To verify Proposition 4.2, we have to show that it is unlikely that the walk has moved more than a certain distance in a sufficiently small amount of time, uniformly in n and the starting point.

COROLLARY 4.3. *Let* (E, d) *be a locally compact*, *separable metric space*. *For each* $n \in \mathbb{N}$, *let* $T_n \subseteq E$ *and* $(X^n, (\mathbb{P}^x)_{x\in T_n})$ *a strong Markov process on* T_n. *Assume that for every* $\varepsilon > 0$

$$\lim_{t\to 0} \lim_{n\to\infty} \sup_{x\in T_n} \mathbb{P}^x\{d(x, X^n_t) > \varepsilon\} = 0. \tag{4.3}$$

Then for every sequence of initial distributions $\mu_n \in \mathcal{M}_1(T_n)$ *the sequence* $(X^n)_{n\in\mathbb{N}}$ *is tight as processes on the one-point compactification of* E.

PROOF. Let $\hat{E} = E \cup \{\infty\}$ be the one-point compactification of E. $\hat{E}$ is metrizable, and we can choose a metric $\hat{d}$ with $\hat{d} \leq d$ on $E \times E$. A possible choice is

$$\hat{d}(x, y) := \inf_{n\in\mathbb{N}} \inf_{z_1,\dots,z_n\in E} \sum_{k=0}^{n} e^{-\inf\{j : z_k\in U_j \text{ or } z_{k+1}\in U_j\}} \big(1 \wedge d(z_k, z_{k+1})\big), \tag{4.4}$$

where $z_0 := x$, $z_{n+1} := y$, and $U_1 \subseteq U_2 \subseteq \cdots$ are (fixed) open, relatively compact subsets of E with $E = \bigcup_{n\in\mathbb{N}} U_n$. By the strong Markov property, (4.3) implies (4.2) for d, and hence also for $\hat{d}$. By Proposition 4.2, $(X^n)_{n\in\mathbb{N}}$ is tight on $\hat{E}$. $\square$

From here, we proceed in several steps. We first give an estimate for the probability to reach a particular point in a small amount of time. We are then seeking an estimate for the probability that the walk has moved more than a given distance away from the starting point. For that, we will need a bound on the number of possible directions the random walk might have taken until reaching that distance.

Recall from (2.18) the first hitting time τ_x of a point $x \in T$.

LEMMA 4.4 (Hitting time bound). *Let (T, r, ρ, ν) be a discrete rooted metric boundedly finite measure tree, $x \in T$, X the speed-ν random walk on (T, r) started at x. Fix $v \in T$ and $\delta \in (0, r(x, v))$. Denote by $S := B(x, \delta)$ the subtree δ-close to x and let $R := r(S, v)$. Then, for all $t \geq 0$,*

$$\mathbb{P}^x\{\tau_v \leq t\} \leq 2\left(1 - \frac{R}{R+2\delta} e^{-t/(R\nu(S))}\right). \tag{4.5}$$

PROOF. Assume w.l.o.g. that X is recurrent and let w be the unique point in S with $r(v, w) = R$. Obviously, if X starts in x then it must pass w before hitting v. Neglect the time until w and assume X starts in w instead of x. For $u \in S$, let t_u be the (random) amount of time spent in u before hitting v, $r_u := r(w, u)$, and $m_u := \nu(\{u\})$. Using that a geometric sum of independent, exponentially distributed random variables is again exponentially distributed, it is easy to see that the law of t_u is

$$\mathcal{L}^w(t_u) = \frac{r_u}{R + r_u}\delta_0 + \frac{R}{R + r_u}\operatorname{Exp}\left(\frac{1}{2(R + r_u)m_u}\right), \tag{4.6}$$

where $\operatorname{Exp}(\lambda)$ denotes an exponential distribution with expectation $\frac{1}{\lambda}$, and δ_0 the Dirac measure in 0.

As $\tau_v \geq \sum_{u\in S} t_u$, we find that for every $a > 0$,

$$\tau_v \geq \sum_{u\in S} a m_u \mathbf{1}_{\{t_u \geq a m_u\}} = a\nu\big(\{u \in S : t_u \geq a m_u\}\big). \tag{4.7}$$

Now we pick $a := \frac{2t}{\nu(S)}$ and obtain

$$\begin{aligned}\mathbb{P}^x\{\tau_v \leq t\} &\leq \mathbb{P}^w\left\{\nu\{u \in S : t_u \geq a m_u\} \leq \frac{1}{2}\nu(S)\right\} \\ &= \mathbb{P}^w\left\{\nu\{u \in S : t_u < a m_u\} \geq \frac{1}{2}\nu(S)\right\}\end{aligned} \tag{4.8}$$

$$\begin{aligned}&\leq \frac{2}{\nu(S)}\mathbb{E}^w\big[\nu\{u \in S : t_u < am_u\}\big]\\&= \frac{2}{\nu(S)}\sum_{u\in S} m_u \mathbb{P}^w\biggl\{t_u < 2t\frac{m_u}{\nu(S)}\biggr\},\end{aligned}$$

which together with (4.6) and the fact that $r_u \leq 2\delta$ gives the claim. □

To get bounds on the probability to move sufficiently far from bounds on the probability to hit a pre-specified point, we need a bound on the number of directions the random walk can take in order to get far away. With ε-degree of a node x we mean the number of edges that intersect the ε-sphere around x and are connected to points at least 2ε away from x.

DEFINITION 4.5 (ε-degree). Let (T, r) be a discrete metric tree. For $\varepsilon > 0$, $x \in T$, let $B := B(x, \varepsilon)$ be the ε-ball around x. The ε-degree of x is

$$\begin{aligned}&\deg_\varepsilon(x) := \deg_\varepsilon^T(x)\\&\qquad := \#\bigl\{v \in T \setminus B : \exists u \in B, w \in T \setminus B(x, 2\varepsilon) : u \sim v, v \in [u, w]\bigr\}.\end{aligned} \tag{4.9}$$

We also define the maximal degree as

$$\deg_\varepsilon(T) := \sup_{x\in T} \deg_\varepsilon^T(x). \tag{4.10}$$

LEMMA 4.6 (Topological bound). *Let $\mathcal{X}_n := (T_n, r_n)$, $n \in \mathbb{N}$, be discrete metric trees, and $\mathcal{X} := (T, r)$ a compact metric tree. If $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ in Gromov–Hausdorff topology, then for every $\varepsilon > 0$,*

$$\limsup_{n\to\infty} \deg_\varepsilon(T_n) < \infty. \tag{4.11}$$

PROOF. Fix $\varepsilon > 0$. As $\mathcal{X}_n \to \mathcal{X}$ in Gromov–Hausdorff topology, there exists a finite ε-net S in T, and ε-nets S_n in T_n, such that for all sufficiently large $n \in \mathbb{N}$, S_n has the same cardinality as S (see, e.g., [10], Proposition 7.4.12). Obviously, this common cardinality is an upper bound for $\{\deg_\varepsilon(T_n); n \in \mathbb{N}\}$. □

With the notion of an ε-degree of a tree, we can immediately conclude the following.

LEMMA 4.7 (Speed bound). *Let (T, r, ρ, ν) be a discrete metric boundedly finite measure tree, $x \in T$, and X the speed-ν random walk on (T, r). Then for every $\varepsilon > 0$, $\delta \in (0, \varepsilon)$ and $t < (\varepsilon - \delta)m$, where $m := \nu(B(x, \delta))$,*

$$\mathbb{P}^x\Bigl\{\sup_{s\in[0,t]} r(X_s, x) > 2\varepsilon\Bigr\} \leq 2\deg_\varepsilon(x)\biggl(1 - \frac{\varepsilon - \delta}{\varepsilon + \delta}\exp\biggl(-\frac{t}{\varepsilon m}\biggr)\biggr). \tag{4.12}$$

PROOF. Let $v_1, \dots, v_N$ be the points outside $B(x, \varepsilon)$ that are neighbours of a point inside $B(x, \varepsilon)$ and on the way from x to a point outside $B(x, 2\varepsilon)$. Then $N \le \deg_\varepsilon(x)$. Under $\mathbb{P}^x$, if $r(X_s, x) > 2\varepsilon$ for some $s \le t$, X must have hit at least one point in $\{v_1, \dots, v_N\}$ before time s. Hence, the claim follows from Lemma 4.4. □

PROOF OF PROPOSITION 4.1. According to Proposition 3.8, $\mathcal{X}_n \to \mathcal{X}$ in Gromov–Hausdorff-vague topology. Hence, we may assume that there is a rooted Heine–Borel space (E, d, ρ_E), such that $T_n, T \subseteq E$, $\rho_E = \rho = \rho_n$ for all $n \in \mathbb{N}$, and, for all but countably many $R > 0$, we have both

$$T_n \cap \bar{B}_d(\rho, R) \to T \cap \bar{B}_d(\rho, R) \tag{4.13}$$

as subsets of E in Hausdorff topology, and

$$\nu_n{\restriction}_R \Rightarrow \nu{\restriction}_R. \tag{4.14}$$

Let $\hat{E} = E \cup \{\infty\}$ be the one-point compactification of E, metrized by a metric $\hat{d}$ with $\hat{d} \le d$ on E^2 [see, e.g., (4.4)]. For each $x \in \hat{E}$ and $N \in \mathbb{N}$, write $B_{\hat{d}}(x, \frac{1}{N}) := \{y \in \hat{E} : \hat{d}(x, y) < \frac{1}{N}\}$ and put

$$K_N := \hat{E} \setminus B_{\hat{d}}\left(\infty, \frac{1}{N}\right) \subseteq E. \tag{4.15}$$

Notice that K_N is compact by definition.

To show tightness, we show that condition (4.3) of Corollary 4.3 is satisfied for the metric $\hat{d}$, that is, for given $\varepsilon, \hat{\varepsilon} > 0$, we can construct $t_0 > 0$ such that

$$\sup_{x \in T_n} \mathbb{P}^x\{\hat{d}(x, X_t^n) > \varepsilon\} \le \hat{\varepsilon}, \tag{4.16}$$

for all $t \in [0, t_0]$ and all $n \in \mathbb{N}$.

Fix $\varepsilon > 0$, and choose $N > \frac{4}{\varepsilon}$. Then the diameter of $\hat{E} \setminus K_N$ with respect to $\hat{d}$ is at most $\frac{1}{2}\varepsilon$. Let

$$e_N := \sup_{n \in \mathbb{N}} \sup_{x \in T_n \cap K_N} \sup_{y \sim x} d(x, y) \tag{4.17}$$

be the supremum of edge-lengths emanating from points in $T_n \cap K_N$, and note that $e_N < \infty$ by assumption. Now choose $M > N$ such that K_M contains the e_N-neighbourhood of K_N, that is, $\{x' \in E : d(K_N, x') < e_N\} \subseteq K_M$. Then all points of K_M which are connected to a point in $E \setminus K_M$ (within some T_n) are actually in $K_M \setminus K_N$.

Consider the hitting time of K_M, $\tau_{K_M} := \inf\{s \ge 0 : X_s^n \in K_M\}$, and recall that the $\hat{d}$-diameter of $\hat{E} \setminus K_N$ is at most $\frac{\varepsilon}{2}$. Therefore, if X^n starts in $x \in T_n$, then $\hat{d}(x, X_t^n) > \varepsilon$ implies $\tau_{K_M} < t$ and $\hat{d}(x, X^n_{\tau_{K_M}}) \le \frac{\varepsilon}{2}$. Using the strong Markov property at τ_{K_M}, we obtain for all $n \in \mathbb{N}$, $x \in T_n$,

$$\mathbb{P}^x\{\hat{d}(x, X_t^n) > \varepsilon\} \le \sup_{y \in T_n \cap K_M} \sup_{s \in [0,t]} \mathbb{P}^y\left\{\hat{d}(y, X_s^n) > \frac{1}{2}\varepsilon\right\}. \tag{4.18}$$

Applying Lemma 4.7, we conclude for all $\delta \in (0, \varepsilon)$ and $t < \frac{1}{4}(\varepsilon - \delta)m_\delta$, where $m_\delta := \inf_{n\in\mathbb{N}} \inf_{y\in T_n\cap K_M} \nu_n(B(y, \frac{\delta}{4}))$,

$$\mathbb{P}^x\{\hat{d}(x, X_t^n) > \varepsilon\} \le 2\deg_{\varepsilon/4}(T_n \cap K_M)\left(1 - \frac{\varepsilon - \delta}{\varepsilon + \delta}\exp\left(-\frac{4t}{\varepsilon m_\delta}\right)\right). \tag{4.19}$$

As $D := \sup_{n\in\mathbb{N}} \deg_{\varepsilon/4}(T_n \cap K_M) < \infty$ by Lemma 4.6, and $m_\delta > 0$ by the local lower mass-bound property (A2), we can choose $\delta > 0$ small enough such that $\frac{\varepsilon-\delta}{\varepsilon+\delta} > 1 - \frac{\hat{\varepsilon}}{4D}$, and subsequently $t_0 < \frac{1}{4}(\varepsilon - \delta)m_\delta$ such that $\exp(-\frac{4t_0}{\varepsilon m_\delta}) > 1 - \frac{\hat{\varepsilon}}{4D}$. Inserting this into (4.19), we obtain (4.16) and tightness follows from Corollary 4.3. □

5. Identifying the limit. In this section, we identify the limit process. For this purpose, we use a characterization from [3], Section 5, where the existence of a diffusion process on a particular non-trivial continuum tree, the so-called Brownian CRT (T, r, ν) from Example 1.5, was shown. Aldous defines this diffusion as a strong Markov process on T with continuous path such that ν is the reversible equilibrium and it satisfies the following two properties:

(i) For all $a, b, x \in T$ with $x \in [a, b]$, $\mathbb{P}^x\{\tau_a < \tau_b\} = \frac{r(x,b)}{r(a,b)}$.
(ii) The occupation time formula (0.1) holds.

While (i) reflects the fact that this diffusion is on "natural scale," (ii) recovers ν as the "speed" measure. At several places in the literature, constructions of diffusions on the CRT and more general continuum random trees rely on Aldous' characterisation (see, e.g., [11, 12, 24]). Albeit the diffusions can be indeed characterised by (i) and (ii) uniquely, a formal proof for this fact has to the best of our knowledge never been given anywhere. We want to close this gap, and even show that the requirement (i) is redundant.

The following result will be proven in Section 6.1.

PROPOSITION 5.1 (Characterization via occupation time formula). *Assume that (T, r) is a compact metric (finite) measure tree, and that we are given two T-valued strong Markov processes X and Y such that for all $x, y \in T$, and bounded measurable $f : T \to \mathbb{R}_+$,*

$$\mathbb{E}^x\left[\int_0^{\tau_y} \mathrm{d}t\, f(X_t)\right] = \mathbb{E}^x\left[\int_0^{\tau_y} \mathrm{d}t\, f(Y_t)\right]. \tag{5.1}$$

Assume further that $X_{\cdot\wedge\tau_y}$ is transient for all $y \in T$. Then the laws of X and Y agree.

We will rely on Proposition 5.1 and show for compact limiting trees that any limit point satisfies the strong Markov property in Section 5.1 and the occupation time formula (0.1) in Section 5.2. Note that, if $\mathcal{X}_n = (T_n, r_n, \rho_n, \nu_n)$ converges to $\mathcal{X} = (T, r, \rho, \nu)$ pointed Gromov–Hausdorff-vaguely [i.e., we assume (A1) and (A2) of Theorem 1], then compactness of $\mathcal{X}$ together with assumption (A0) of Theorem 1 is equivalent to the uniform diameter bound $\sup_{n\in\mathbb{N}} \operatorname{diam}(T_n, r_n) < \infty$.

5.1. *The strong Markov property of the limit.* In this subsection, we show that any limit point has the strong Markov property. To be more precise, the main result is the following.

PROPOSITION 5.2 (Strong Markov property). *Let* $\mathcal{X} := (T, r, \nu)$ *and* $\mathcal{X}_n := (T_n, r_n, \nu_n)$, $n \in \mathbb{N}$, *be metric boundedly finite measure trees. Assume that all* $\mathcal{X}_n$, $n \in \mathbb{N}$, *are discrete with* $\sup_{n\in\mathbb{N}} \operatorname{diam}(T_n, r_n) < \infty$, *and that the sequence* $(\mathcal{X}_n)_{n\in\mathbb{N}}$ *converges to* $\mathcal{X}$ *Gromov–Hausdorff-vaguely as* $n \to \infty$. *If* X^n *is the speed-*ν_n *random walk on* (T_n, r_n) *and* $X^n \underset{n\to\infty}{\Longrightarrow} \tilde{X}$ *in path space, then* $\tilde{X}$ *is a (strong Markov) Feller process.*

In order to prove Proposition 5.2, we will first show that under its assumptions the family of functions $\{P_n : n \in \mathbb{N}\}$, where for each $n \in \mathbb{N}$

$$P_n\colon \begin{cases} T_n \times \mathbb{R}_+ \to \mathcal{M}_1(E), \\ (x, t) \mapsto \mathcal{L}^x\big(X_t^n\big) =: P_{n,t}^x, \end{cases} \tag{5.2}$$

is uniformly equicontinuous. Here, $\mathcal{L}^x(X_t^n)$ denotes the law of X_t^n, where X^n is started in $x \in T_n$, E is a metric space containing all T_n, and $\mathcal{M}_1(E)$ is equipped with the Prohorov metric.

LEMMA 5.3 (Equicontinuity). *Let* $\mathcal{X} := (T, r, \nu)$ *and* $\mathcal{X}_n := (T_n, r_n, \nu_n)$, $n \in \mathbb{N}$, *be metric boundedly finite measure trees. Assume that all* $\mathcal{X}_n$, $n \in \mathbb{N}$, *are discrete with* $\sup_{n\in\mathbb{N}} \operatorname{diam}(T_n, r_n) < \infty$, *and that* $\mathcal{X}_n \to \mathcal{X}$ *Gromov–Hausdorff-vaguely. If for each* $n \in \mathbb{N}$, X^n *is the speed-*ν_n *random walk on* (T_n, r_n), *and* $P_n\colon \mathbb{R}_+ \times T_n \to \mathcal{M}_1(E)$ *is defined as in* (5.2), *then the family* $\{P_n : n \in \mathbb{N}\}$ *is uniformly equicontinuous.*

PROOF. Fix $\varepsilon > 0$. We construct a $\delta > 0$, independent of n, such that $P_{n,s}^x$ and $P_{n,t}^y$ are ε-close whenever $x, y \in T_n$, $s, t \in \mathbb{R}_+$ are such that $r_n(x, y) < \delta$ and $s \le t \le s + \delta$.

Fix $n \in \mathbb{N}$, and denote for any two $x, y \in T_n$ by X^x and X^y speed-ν_n random walks on (T_n, r_n) starting in x and y, respectively, which are coupled as follows: let the random walks X^x, X^y run independently until X^x hits y for the first time, that is, until $\tau := \inf\{t \ge 0 : X_t^x = y\}$, and put $X_{\tau+\cdot}^x = X^y$. In particular, whenever $s \ge \tau$, we obtain $X_s^x = X_{t-u}^y$ for $u = \tau + t - s$.

Using the strong Markov property of X^y, we can estimate for any $c \in [t - s, t]$

$$\mathbb{P}\big\{r_n\big(X_s^x, X_t^y\big) > \varepsilon\big\} \le \mathbb{P}\{\tau > c - t + s\} + \sup_{z\in T_n} \mathbb{P}\Big\{\sup_{u\in[0,c]} r_n\big(z, X_u^z\big) > \varepsilon\Big\}. \tag{5.3}$$

For small t, we need another estimate, namely for $r_n(x, y) \le \frac{1}{3}\varepsilon$ we have

$$\mathbb{P}\big\{r_n\big(X_s^x, X_t^y\big) > \varepsilon\big\} \le 2 \sup_{z\in T_n} \mathbb{P}\bigg\{\sup_{u\in[0,t]} r_n\big(z, X_u^z\big) > \frac{\varepsilon}{3}\bigg\} =: 2q_t. \tag{5.4}$$

Combining (5.3) and (5.4), we obtain, under the condition $r_n(x,y) \le \frac{1}{3}\varepsilon$, for any $c \ge t-s$

$$\mathbb{P}\{r_n(X_s^x, X_t^y) > \varepsilon\} \le q_c + q_c \vee \mathbb{P}\{\tau > c-(t-s)\}. \tag{5.5}$$

Note that this estimate depends on x, y, s, t only through $r_n(x,y)$ and $t-s$.

The Gromov–Hausdorff-vague convergence together with the uniform diameter bound on (T_n, r_n) implies that (T, r) is compact and (T_n, r_n) converges to (T, r) in Gromov–Hausdorff topology. Hence, $\sup_{n\in\mathbb{N}} \deg_{\varepsilon/6}(T_n) < \infty$, by Lemma 4.6. Furthermore, the global lower mass-bound property is satisfied, that is, for every $\varepsilon' > 0$, $m_{\varepsilon'} := \inf_{n\in\mathbb{N}, x\in T_n} \nu_n(B_n(x,\varepsilon')) > 0$. We can thus apply Lemma 4.7 to obtain a sufficiently small $c = c(\varepsilon) > 0$, independent of n, such that $q_c \le \frac{\varepsilon}{2}$. To estimate (for this c) $\mathbb{P}\{\tau > c-(t-s)\}$, we note that $M := \sup_{n\in\mathbb{N}} \nu_n(T_n) < \infty$ because of the diameter bound, and obtain for $t-s \le \frac{1}{2}c$

$$\mathbb{P}\{\tau > c-(t-s)\} \le \frac{2}{c}\mathbb{E}[\tau] \le \frac{4}{c}M \cdot r_n(x,y). \tag{5.6}$$

Choose therefore $\delta := \frac{\varepsilon}{8M}c \wedge \frac{\varepsilon}{3} \wedge \frac{1}{2}c$. Then for all $x, y \in T_n$ with $r_n(x,y) < \delta$, and $0 \le s \le t < s+\delta$, (5.5) implies $\mathbb{P}\{r_n(X_s^x, X_t^y) > \varepsilon\} \le \varepsilon$, and hence $d_{\mathrm{Pr}}(P_{n,s}^x, P_{n,t}^y) \le \varepsilon$, which is the claimed equicontinuity. $\square$

The proof of Proposition 5.2 relies on the following modification of the Arzelà–Ascoli theorem, which is proven in the same way as the classical theorem.

LEMMA 5.4 (Arzelà–Ascoli). *Let (E, d) be a compact metric space, and (F, d_F) a metric space. Consider closed subsets $T, T_n \subseteq E$ and functions $f_n : T_n \to F$ for $n \in \mathbb{N}$. Further assume that the family $\{f_n; n \in \mathbb{N}\}$ is uniformly equicontinuous with modulus of continuity h, and that for all $x \in T$ there exists $x_n \in T_n$ such that $x_n \to x$ and $\{f_n(x_n) : n \in \mathbb{N}\}$ is relatively compact in F. Then there is a function $f : T \to F$, a subsequence of $(f_n)_{n\in\mathbb{N}}$, again denoted by (f_n), and $\varepsilon_n > 0$ with $\varepsilon_n \to 0$ such that for all $n \in \mathbb{N}$, for all $x \in T$ and $y \in T_n$,*

$$d_F(f(x), f_n(y)) \le h(d(x,y)) + \varepsilon_n. \tag{5.7}$$

Note that (5.7) in particular implies that f is continuous with the same modulus of continuity h, and that $f_n(x_n) \to f(x)$ whenever $x_n \to x$.

PROOF OF PROPOSITION 5.2. By assumption there is a compact metric space (E, d) such that $T, T_1, T_2, \ldots \subseteq E$, $d\restriction_T = r$, $d\restriction_{T_n} = r_n$ for all $n \in \mathbb{N}$, and $(T_n, r_n, \nu_n)_{n\in\mathbb{N}}$ converges Hausdorff-weakly to (T, r, ν).

According to Proposition 4.1 and Lemma 5.3, the assumptions of Arzelà–Ascoli are satisfied for the family of functions P_n, $n \in \mathbb{N}$, defined in (5.2). Thus, we obtain a continuous subsequential limit $P : T \times \mathbb{R}_+ \to \mathcal{M}_1(E)$, $(x,t) \mapsto P_t^x$. Let $S = (S_t)_{t\ge 0}$ and $S^n = (S_t^n)_{t\ge 0}$ be the corresponding operators on $\mathcal{C}(T)$ and $\mathcal{C}(T_n)$, respectively. That is $S_t f(x) := \int_T f \, \mathrm{d}P_t^x$ and $S_t^n f(x) := \int_{T_n} f \, \mathrm{d}P_{n,t}^x$, $n \in \mathbb{N}$. We show that S is indeed a strongly continuous semi-group.

To this end, it is enough to show $\lim_{t\to 0} \|S_t f - f\|_\infty = 0$ and $S_{t+s} f = S_t(S_s f)$, $s, t > 0$, for Lipschitz continuous $f \in \mathcal{C}(T)$ with Lipschitz constant (at most) 1 and $\|f\|_\infty \le 1$. We can extend every such f to a function on E with the same properties. Let $\mathrm{Lip}_1 = \mathrm{Lip}_1(E)$ be the space of such (extended) f and recall that the Kantorovich–Rubinshtein metric between two measures $\mu, \hat{\mu} \in \mathcal{M}_1(E)$,

$$d_{\mathrm{KR}}(\mu, \hat{\mu}) := \sup_{f \in \mathrm{Lip}_1} \int f \, \mathrm{d}(\mu - \hat{\mu}), \tag{5.8}$$

is uniformly equivalent to the Prohorov metric (see [9], Theorem 8.10.43). For the rest of the proof, $\mathcal{M}_1(E)$ is equipped with d_{KR}. Let h be a common modulus of continuity for all P_n, $n \in \mathbb{N}$, which exists according to Lemma 5.3. Due to Lemma 5.4, P has the same modulus of continuity and hence, for all $f \in \mathrm{Lip}_1$,

$$\|S_t f - f\|_\infty \le \sup_{x \in T} d_{\mathrm{KR}}\big(P_t^x, P_0^x\big) \le h(t) \underset{t\to 0}{\longrightarrow} 0, \tag{5.9}$$

that is, S is strongly continuous.

Because T_n converges to T in the Hausdorff metric, we find $g_n \colon T_n \to T$ such that

$$\alpha_n := \sup_{y \in T_n} d\big(y, g_n(y)\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{5.10}$$

W.l.o.g. we may also assume that $T_1, T_2, \dots,$ are disjoint. As the spaces (T_n, r_n), $n \in \mathbb{N}$, are discrete, the map

$$g \colon T \cup \bigcup_{n \in \mathbb{N}} T_n \to T, \qquad x \mapsto \begin{cases} x, & x \in T, \\ g_n(x), & x \in T_n \end{cases} \tag{5.11}$$

is continuous. Now we apply (5.7) to P_n and P and obtain for all $n \in \mathbb{N}$, $f \in \mathrm{Lip}_1$ and $s > 0$

$$\begin{aligned} \sup_{y \in T_n} \big|S_s^n f(y) - (S_s f)\big(g(y)\big)\big| &\le \sup_{y \in T_n} d_{\mathrm{KR}}\big(P_{n,s}^y, P_s^{g(y)}\big) \\ &\le h(\alpha_n) + \varepsilon_n \underset{n\to\infty}{\longrightarrow} 0, \end{aligned} \tag{5.12}$$

where ε_n is obtained in Lemma 5.4. For $x \in T$, there exists $x_n \in T_n$ with $x_n \to x$, and thus, using (5.12) and the semi-group property of S^n,

$$\begin{aligned} S_{t+s} f(x) &= \lim_{n\to\infty} S_{t+s}^n f(x_n) = \lim_{n\to\infty} S_t^n\big(S_s^n f\big)(x_n) \\ &= \lim_{n\to\infty} S_t^n(S_s f \circ g)(x_n) = S_t(S_s f \circ g)(x) \\ &= S_t(S_s f)(x). \end{aligned} \tag{5.13}$$

Now it is standard to see that S comes from a Feller process, and this process has to be $\tilde{X}$. $\square$

We can conclude immediately from Proposition 5.2 that in the general locally compact case any limit process has the strong Markov property, at least up to the first time it hits the boundary at infinity.

The following example shows that in general we loose the strong Markov property once we hit infinity.

EXAMPLE 5.5 (Entrance law). Let (T, r, ρ) be the discrete binary tree with unit edge-lengths, that is,

$$T := \bigcup_{n\in\mathbb{N}} \{0,1\}^n \cup \{\rho\}, \tag{5.14}$$

$r(\rho, x) := n$ for all $x \in \{0,1\}^n$, and there is an edge $x \sim y$ if and only if $y = (x, i)$ or $x = (y, i)$ for $i \in \{0, 1\}$.

Put $h(x) := r(\rho, x)$, and consider the speed measure $\nu(\{x\}) := e^{-h(x)}$, $x \in T$. Obviously, the speed-ν random walk on X is transient, as $h(X)$ is a reflected random walk on $\mathbb{N}$ with constant drift to the right.

Now consider (T_n, r, ρ, ν) with $T_n := \{x \in T : h(x) \le n\}$, where the metric and the measure are understood to be restricted to T_n. Because T_n is finite and the speed-ν_n random walk X^n has no absorbing points, it is positive recurrent. We may therefore conclude from Proposition 2.12 that for all $x \in T$, $n \in \mathbb{N}$ suitably large,

$$\mathbb{E}^x\big[\tau_\rho^n\big] = 2\sum_{y\in T_n} h\big(c(\rho, x, y)\big)e^{-h(y)} \le \sum_{k=1}^{n} k2^k e^{-k} < \infty. \tag{5.15}$$

Therefore, in contrast to the transience of the speed-ν random walk on (T, r), any "limiting" process Y of the speed-ν_n random walks on (T_n, r_n) is also positive recurrent. This shows that in Theorem 1 we indeed have to stop limiting processes at infinity in order for them to coincide with the speed-ν motion on (T, r). Consequently, this also means that the speed-ν motion has an entrance law on (T, r) from infinity, which we obtain by considering excursions of Y away from infinity. Finally, the limit Y obviously loses its strong Markov property at hitting infinity, because in the one-point compactification, we are identifying all ends at infinity.

5.2. *The occupation time formula of the limit.* In this section, we assume that the limiting tree is compact and show that all limit points satisfy the occupation time formula (0.1). The main result is the following.

PROPOSITION 5.6 (Occupation time formula). *Let $\mathcal{X} := (T, r, \nu)$ and $\mathcal{X}_n := (T_n, r_n, \nu_n)$, $n \in \mathbb{N}$, be metric boundedly finite measure trees. Assume that all $\mathcal{X}_n$, $n \in \mathbb{N}$, are discrete with $\sup_{n\in\mathbb{N}} \operatorname{diam}(T_n, r_n) < \infty$, and that $\mathcal{X}_n \to \mathcal{X}$ Gromov–Hausdorff-vaguely as $n \to \infty$. If X^n is the speed-ν_n random walk on (T_n, r_n) and $X^n \underset{n\to\infty}{\Longrightarrow} \tilde{X}$ in path space, then $\tilde{X}$ satisfies (0.1).*

To prove this formula, we need a lemma about semi-continuity of hitting times in Skorohod space. This semi-continuity does not hold in general, but we rather have to use that the limiting path satisfies a certain regularity property.

If $\mathrm{supp}(\nu)$ is not connected, the paths of the limit process are obviously not continuous. They satisfy, however, the following weaker *closedness condition*.

DEFINITION 5.7 (Closed-interval property). Let E be a topological space. We say that a function $w: \mathbb{R}_+ \to E$ has the *closed-interval property* if $w([s,t]) \subseteq E$ is closed for all $0 \le s < t$.

LEMMA 5.8 (Speed-ν motions have the closed-interval property). *The path of the limit process $\tilde{X}$ has the closed-interval property, almost surely.*

PROOF. Let $A \subseteq T$ be the set of endpoints of edges of T. Recall from Remark 1.2 that A is at most countable. Jumps of the limit process $\tilde{X}$ can only occur over edges of T, hence $\tilde{X}_{t-} := \lim_{s \nearrow t} \tilde{X}_s \neq \tilde{X}_t$ implies $\tilde{X}_{t-} \in A$.

Fix $a \in A$. We first show that if $\tau_a^- := \inf\{t > 0 : \tilde{X}_{t-} = a\}$ denotes the first time when the left limit of $\tilde{X}$ reaches a, we have $\tilde{X}_{\tau_a^-} = a$ almost surely, that is, $\tilde{X}$ does not jump at time τ_a^- almost surely. Indeed, for every $\varepsilon > 0$ we can use the right-continuity of the paths of $\tilde{X}$ together with Feller-continuity to find $s_0 > 0$ and $\delta > 0$ such that for all $x \in B(a, \delta)$,

$$\mathbb{P}^x\Big\{\sup_{s\in[0,s_0]} r(a, \tilde{X}_s) > \varepsilon\Big\} < \frac{1}{2}\varepsilon. \tag{5.16}$$

Define the stopping times $\tau_n := \inf\{t \ge 0 : r(\tilde{X}_t, a) \le \frac{1}{n}\}$, and note that $\tau_n \uparrow \tau_a^-$. If $n > \frac{1}{\delta}$ is such that $\mathbb{P}^x\{\tau_a^- - \tau_n > s_0\} < \frac{1}{2}\varepsilon$, then by Proposition 5.2,

$$\mathbb{P}^x\big\{r(\tilde{X}_{\tau_a^-}, a) > \varepsilon\big\} \le \frac{1}{2}\varepsilon + \mathbb{E}^x\Big[\mathbb{P}^{\tilde{X}_{\tau_n}}\Big\{\sup_{s\in[0,s_0]} r(a, \tilde{X}_s) > \varepsilon\Big\}\Big] \le \varepsilon. \tag{5.17}$$

Since ε is arbitrary, this proves $\tilde{X}_{\tau_a^-} = a$ almost surely.

Because A is countable, this implies that $\{\tilde{X}_u : u \in [0,t]\}$ is closed for all $t \ge 0$, almost surely. Again using the Markov property, we also obtain almost surely closedness of $\{\tilde{X}_u : u \in [s,t]\}$ for all $t \ge 0$, $s \in \mathbb{Q}_+$, which implies closedness for all $s \ge 0$ by right-continuity. $\square$

We omit the proof of the following lemma, because it is straight-forward.

LEMMA 5.9 (Semi-continuity of the hitting time functional). *Let E be a Polish space and $\mathcal{D}_E = \mathcal{D}_E(\mathbb{R}_+)$ the corresponding Skorohod space. For a set $A \subseteq E$, define*

$$\sigma_A: \mathcal{D}_E \to \mathbb{R}_+ \cup \{\infty\}, \qquad w \mapsto \inf\big\{t \in \mathbb{R}_+ : w(t) \in A\big\}. \tag{5.18}$$

Then if A is open, σ_A is upper semi-continuous, and if A is closed, the set of lower semi-continuity points of σ_A contains the set of paths with the closed-interval property.

REMARK 5.10. For $A \subseteq E$ closed, σ_A is in general not lower semi-continuous.

PROOF OF PROPOSITION 5.6. Fix $x, y \in T$ and let τ_y be the first time when $\tilde{X}$ hits y. It is enough to show (0.1) for non-negative $f \in \mathcal{C}_b(T)$. Because T is closed in E, we can extend f to a bounded continuous function on E, again denoted by f. For $A \subseteq E$, recall the definition of σ_A from (5.18) and consider the function

$$F_A : \mathcal{D}_E \to \mathbb{R}_+ \cup \{\infty\}, \qquad w \mapsto \int_0^{\sigma_A(w)} f\bigl(w(t)\bigr)\,\mathrm{d}t. \tag{5.19}$$

Note that the left-hand side of (0.1) coincides with $\mathbb{E}^x[F_y(\tilde{X})]$, where we abbreviate $F_y := F_{\{y\}}$. The strategy is to approximate F_y by F_A for small neighbourhoods A of y and then use semi-continuity properties of F_A and the occupation time formula of the approximating X^n.

Denote for each $\varepsilon > 0$ the closed ε-ball in E around y by A_ε. We claim that almost surely

$$\tau := \sup_{\varepsilon>0} \sigma_{A_\varepsilon}(\tilde{X}) = \sigma_{\{y\}}(\tilde{X}) = \tau_y. \tag{5.20}$$

Indeed, $\tau \leq \tau_y$ is obvious. For the converse inequality, recall that the path of $\tilde{X}$ almost surely has the closed-interval property by Lemma 5.8, which means that $\{\tilde{X}_t : t \in [0, \tau]\}$ is almost surely a closed set containing points in every A_ε, $\varepsilon > 0$, hence also y. Therefore, $\tau_y \leq \tau$ almost surely.

Because f is non-negative, (5.20) implies that

$$\sup_{\varepsilon>0} F_{A_\varepsilon}(\tilde{X}) = F_y(\tilde{X}), \tag{5.21}$$

almost surely. Furthermore, it follows from the definition of the Skorohod topology that whenever w is a lower- or upper semi-continuity point of σ_A, the same is true for F_A. Hence, Lemma 5.9 together with Lemma 5.8 implies that the path of $\tilde{X}$ is almost surely a lower semi-continuity point of F_A for closed sets A, and an upper semi-continuity point for open sets A.

Choose $x_n, y_n \in T_n$ with $y_n \to y$ and $x_n \to x$, and note that $y_n \in A_\varepsilon$ for all sufficiently large n. Since $X^n \underset{n\to\infty}{\Longrightarrow} \tilde{X}$, and $\tilde{X}$ is almost surely a lower semi-continuity point of F_A,

$$\begin{aligned} \mathbb{E}^x\bigl[F_y(\tilde{X})\bigr] &= \sup_{\varepsilon>0} \mathbb{E}^x\bigl[F_{A_\varepsilon}(\tilde{X})\bigr] \\ &\leq \sup_{\varepsilon>0} \liminf_{n\to\infty} \mathbb{E}^{x_n}\bigl[F_{A_\varepsilon}\bigl(X^n\bigr)\bigr] \\ &\leq \liminf_{n\to\infty} \mathbb{E}^{x_n}\bigl[F_{y_n}\bigl(X^n\bigr)\bigr]. \end{aligned} \tag{5.22}$$

Note that the functions $(x_n, y_n, z_n) \mapsto 2r_n(y_n, c_n(x_n, y_n, z_n))$ on T_n^3 and $(x, y, z) \mapsto 2r(y, c(x, y, z))$ on T^3 have a common Lipschitz continuous extension to E given by

$$\xi(x, y, z) := d(y, x) + d(y, z) - d(z, x). \tag{5.23}$$

Therefore, we obtain from (5.22) and the occupation time formula for X^n (Proposition 2.12) that

$$\begin{aligned} \mathbb{E}^x\big[F_y(\tilde{X})\big] &\leq \liminf_{n\to\infty} \int \nu_n(\mathrm{d}z)\xi(x_n, y_n, z) f(z) \\ &= 2\int \nu(\mathrm{d}z) r\big(y, c(x, y, z)\big) f(z). \end{aligned} \tag{5.24}$$

On the other hand, for every sufficiently small $\varepsilon > 0$ and large $n \in \mathbb{N}$, there is a unique point $y_n' \in B(y_n, 2\varepsilon) \cap T_n$ closest to x_n, and using that $\tilde{X}$ is almost surely an upper semi-continuity point of $F_{B(y,\varepsilon)}$, we obtain

$$\begin{aligned} \mathbb{E}^x\big[F_y(\tilde{X})\big] &\geq \limsup_{n\to\infty} \mathbb{E}^{x_n}\big[F_{B(y,\varepsilon)}\big(X^n\big)\big] \\ &\geq \limsup_{n\to\infty} \mathbb{E}^{x_n}\big[F_{B(y_n,2\varepsilon)}\big(X^n\big)\big] \\ &= \limsup_{n\to\infty} \mathbb{E}^{x_n}\big[F_{y_n'}\big(X^n\big)\big] \\ &\geq 2\int \nu(\mathrm{d}z)\big(r\big(y, c(x, y, z)\big) - 2\varepsilon\big) f(z). \end{aligned} \tag{5.25}$$

The claim follows with $\varepsilon \to 0$. $\square$

6. Proof of Theorem 1. In this section, we collect all the pieces we have proven so far and present the proof of our invariance principle.

As we have stated all the results which characterize the limiting process for approximating rooted metric measure trees $(T_n, r_n, \rho_n, \nu_n)$ where (T_n, r_n) was assumed to be discrete, we start with a lemma which states that each rooted metric boundedly finite measure tree can be approximated by discrete trees.

LEMMA 6.1 (Approximation by discrete trees). *Let (T, r, ρ, ν) be a rooted metric boundedly finite measure tree $\mathcal{X}$. Then we can find a sequence $\mathcal{X}_n := (T_n, r_n, \rho, \nu_n)$ of rooted discrete metric boundedly finite measure trees such that $\mathcal{X}_n \to \mathcal{X}$ pointed Gromov–Hausdorff-vaguely.*

PROOF. Let (T, r, ρ, ν) be a rooted metric boundedly finite measure tree, and for each $n \in \mathbb{N}$, S_n a finite $\frac{1}{n}$-net of $B(\rho, n)$ containing $\{\rho\}$. Let $T_n \subseteq T$ be the smallest metric tree containing S_n, that is, the union of S_n and all branching points $x \in T$ with

$$r(x, s_1) = \tfrac{1}{2}\big(r(s_1, s_2) + r(s_1, s_3) - r(s_2, s_3)\big), \tag{6.1}$$

for some $s_1, s_2, s_3 \in S_n$. As usual, let r_n be the restriction of r to T_n, and note that T_n is a finite set, hence (T_n, r_n) is a discrete metric tree.

Consider for each $n \in \mathbb{N}$ the map $\psi_n : T \to T_n$ which sends a point in T to the nearest point on the way from x to ρ which belongs to T_n, that is,

$$\psi_n(x) := \sup\{y \in T_n : y \in [\rho, x]\}. \tag{6.2}$$

Finally, put

$$\nu_n := (\psi_n)_* \nu\restriction_{B(\rho,n)}. \tag{6.3}$$

Then obviously the Prohorov distance between $\nu\restriction_{B(\rho,n)}$ and ν_n is not larger than $\frac{1}{n}$. Thus, (T_n, r_n, ρ, ν_n) converges pointed Gromov-vaguely and also pointed Gromov–Hausdorff-vaguely to (T, r, ρ, ν). $\square$

6.1. *Compact limit trees.* In this subsection, we restrict to the case where the limiting tree is compact. We start with the proof of Proposition 5.1, on which we shall rely the characterization of the limit process.

PROOF OF PROPOSITION 5.1. Consider X and Y satisfying the assumption on Proposition 5.1. In particular, assume that $X_{\cdot\wedge\tau_y}$ is transient for all $y \in T$. Consider for each $y \in T$ the family of resolvent operators $\{G_\alpha^{X,y}; \alpha > 0\}$ and $\{G_\alpha^{Y,y}\alpha > 0\}$ associated with $\{X_{\cdot\wedge\tau_y}; y \in T\}$ and $Y_{\cdot\wedge\tau_y}$, and put $G_X^y := \lim_{N\to\infty} G_{1/N}^{X,y}$ and $G_Y^y := \lim_{N\to\infty} G_{1/N}^{Y,y}$, respectively. By transience, $G_X^y < \infty$ for all $y \in T$. Moreover, for all $x \in T$, and bounded, measurable $f : T \to \mathbb{R}_+$,

$$G_X^y f(x) = \mathbb{E}^x\left[\int_0^{\tau_y} \mathrm{d}s f(X_s)\right]. \tag{6.4}$$

By (5.1), $G_Y^y f(x) < \infty$ as well.

As X is a strong Markov processes, the resolvent identity holds, that is,

$$G_\alpha^{X,y} = G_\beta^{X,y} + (\alpha - \beta) G_\alpha^{X,y} G_\beta^{X,y}. \tag{6.5}$$

Iterating the latter with $\alpha > \beta > 0$ and $|\alpha - \beta| \le \frac{1}{2\|G_\beta^{X,y}\|}$, we have

$$G_\alpha^{X,y} = G_\beta^{X,y} + (\alpha - \beta)\big(G_\beta^{X,y}\big)^2 + (\alpha - \beta)^2\big(G_\beta^{X,y}\big)^3 + \cdots. \tag{6.6}$$

We note that $\|G_\beta^{X,y}\| \le \|G^{X,y}\|$ for all $\beta \ge 0$. So it is bounded above independent of β. Hence, (6.6) holds for $\beta = 0$ by taking limits. Further, by the same arguments, (6.6) also holds for Y instead of X, and by (5.1) $G_0^{Y,y} := G_Y^y = G_X^y$. Therefore, for all small enough $\alpha > 0$, $G_\alpha^{X,y} = G_\alpha^{Y,y}$. Thus, for all small enough $\alpha > 0$,

$$\mathbb{E}^x\left[\int_0^{\tau_y} \mathrm{d}t e^{-\alpha t} \cdot f(X_t)\right] = \mathbb{E}^x\left[\int_0^{\tau_y} \mathrm{d}t e^{-\alpha t} \cdot f(Y_t)\right]. \tag{6.7}$$

Therefore, by uniqueness of the Laplace transform,

$$\mathbb{E}^x\big[f(X_t);\{t<\tau_y\}\big]=\mathbb{E}^x\big[f(Y_t);\{t<\tau_y\}\big] \tag{6.8}$$

for all $y\in T$ and for all $t>0$. Therefore, the one-dimensional distributions of $X_{\cdot\wedge\tau_y}$ and $Y_{\cdot\wedge\tau_y}$ are the same for all $y\in T$. By the strong Markov property, this implies that the laws of X and Y agree. □

To show f.d.d. convergence, we need to control the probability that X_t is in an "exceptional" set of small ν-measure. To this end, we use the following simple heat-kernel bound. We will see in Corollary 6.4 below that the technical assumption $\nu(\{x\})>0$ can be dropped.

LEMMA 6.2. *Let $\mathcal{X}:=(T,r,\nu)$ be a compact metric finite measure tree, $x\in T$ with $\nu(\{x\})>0$, and X the speed-ν motion on (T,r) started in x. Then the law of X_t has for every $t>0$ a density $q_t(x,\cdot)\in L^2(\nu)$ w.r.t. ν, and*

$$\|q_t(x,\cdot)\|_2^2\le\nu(T)^{-1}+\operatorname{diam}(T)\cdot t^{-1}\qquad\forall t>0, \tag{6.9}$$

where $\|\cdot\|_2$ is the norm in $L^2(\nu)$. In particular, for any $A\subseteq T$, we have

$$\mathbb{P}^x\{X_t\in A\}\le\gamma_t\sqrt{\nu(A)}\qquad\forall t>0, \tag{6.10}$$

where the constant $\gamma_t:=1+\nu(T)^{-1}+\operatorname{diam}(T)\cdot t^{-1}$ is independent of x and depends on (T,r,ν) only through $\nu(T)$ and $\operatorname{diam}(T)$.

PROOF. 1. Let $f:=\nu(\{x\})^{-1}\mathbf{1}_{\{x\}}$ be the density of δ_x w.r.t. ν, and

$$f_t:=P_tf,\qquad g(t):=\|f_t\|_2^2, \tag{6.11}$$

where $(P_t)_{t\ge0}$ is the semi-group of the speed-ν motion. Due to reversibility of ν it is easy to see that $f_t=q_t(x,\cdot)$ is the density of X_t w.r.t. ν. Furthermore,

$$g'(t)=2\langle Gf_t,f_t\rangle_\nu=-2\mathcal{E}(f_t,f_t), \tag{6.12}$$

where G is the generator of $(P_t)_{t\ge0}$. Let $a:=\operatorname{diam}(T)^{-1}$. Because $\|f_t\|_1=1$, we find a point $y\in T$ with $f_t(y)\le b:=\nu(T)^{-1}$. For every $z\in T$ with $f_t(z)\ge b$, we have

$$\mathcal{E}(f_t,f_t)\ge\big(f_t(z)-f_t(y)\big)^2\cdot\big(2r(z,y)\big)^{-1}\ge\tfrac12a\big(f_t(z)-b\big)^2. \tag{6.13}$$

Combining (6.13) and (6.12), and using $g(t)=\|f_t\|_2^2\le\|f_t\|_\infty\|f_t\|_1=\|f_t\|_\infty$, we obtain the differential inequality

$$g'(t)\le-a\big(\|f_t\|_\infty-b\big)^2\le-a\big(g(t)-b\big)^2. \tag{6.14}$$

In the above, we have used that $g(t)\ge b$. Solving $h_u'(t)=-a(h_u(t)-b)^2$, $h_u(0)=u$, and using monotonicity of the solution in u, we conclude

$$g(t)\le\lim_{u\to\infty}h_u(t)=\lim_{u\to\infty}\frac{u(1+abt)-ab^2t}{uat-bat+1}=b+(at)^{-1}, \tag{6.15}$$

which is the desired bound (6.9).

2. For $u := \nu(A)^{-1/2}$, we obtain

$$\mathbb{P}^x\{X_t \in A\} \le u\nu(A) + \int_{\{f_t > u\}} \frac{f_t^2}{u}\,\mathrm{d}\nu \le \sqrt{\nu(A)}\big(1 + \|f_t\|_2^2\big). \tag{6.16}$$

Together with (6.9) this implies the desired bound (6.10). $\square$

PROPOSITION 6.3 (Theorem 1 holds for compact limit trees). *Let $\mathcal{X} := (T, r, \rho, \nu)$, $\mathcal{X}_1 := (T_1, r_1, \rho_1, \nu_1)$, $\mathcal{X}_2 := (T_2, r_2, \rho_2, \nu_2), \ldots$ be rooted metric boundedly finite measure trees with $\sup_{n\in\mathbb{N}} \operatorname{diam}(T_n, r_n) < \infty$. Let X be the speed-ν motion on (T, r) starting in ρ, and for all $n \in \mathbb{N}$, X^n the speed-ν_n motion on (T_n, r_n) started in ρ_n. Assume that the following conditions hold*:

(A1) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ pointed Gromov-vaguely.*
(A2) *The uniform local lower mass-bound property* (1.6) *holds.*

Then the following hold:

(i) *X^n converges weakly in path-space to X.*
(ii) *If we assume only* (A1) *but not* (A2), *then X^n converges in finite dimensional distributions to X.*

PROOF. Assume w.l.o.g. that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ are discrete trees (the general result is then obtained by Lemma 6.1 and a diagonal argument). Let X^n be a sequence of ν_n-random walks on (T_n, r_n) starting in ρ_n.

(i) By Proposition 4.1, we know that the sequence is tight. Let $\tilde{X}$ be a weak subsequential limit on (T, r). Then in particular, $\tilde{X}_0 = \rho$ almost surely. From Proposition 5.2 together with Proposition 5.6, we know that $\tilde{X}$ is a strong Markov process and $\mathbb{E}^x[\int_0^{\tau_z} \mathrm{d}s f(\tilde{X}_s)] = 2\int \nu(\mathrm{d}y) r(z, c(x, y, z)) f(y)$.

Let X be the speed-ν motion on (T, r) starting in ρ. Then X is the strong Markov process associated with the Dirichlet form $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$. X is recurrent as clearly $\mathbf{1} \in \mathcal{D}(\mathcal{E})$ and $\mathcal{E}(\mathbf{1}, \mathbf{1}) = 0$. Thus, X satisfies (0.1) by Proposition 2.12. Moreover, it follows from Lemma 2.6 that $X_{\cdot\wedge\tau_y}$ is transient for all $y \in T$. Therefore, the laws of $\tilde{X}$ and X agree by Proposition 5.1.

(ii) Using that $\mathcal{X}_n$ converges Gromov-weakly to $\mathcal{X}$, and $\mathcal{X}$ is compact, we can construct subsets $A_n \subseteq T_n$ with $\nu_n(A_n) \to 0$, $\rho_n \notin A_n$ and the following property. The measure trees $\tilde{\mathcal{X}}_n := (\tilde{T}_n, r_n, \rho_n, \nu_n)$, where $\tilde{T}_n := T_n \setminus A_n$, satisfy the lower mass-bound (1.6) and still converge Gromov-weakly to $\mathcal{X}$. Let $\tilde{X}^n$ be the ν_n-random walk on $(\tilde{T}_n, r_n)$. Then $\tilde{X}^n$ converges in distribution to X by part (i). We show that every finite-dimensional marginal of $\tilde{X}^n$ is weakly merging with the corresponding marginal of X^n. For this it is enough to show for all $t \ge 0$ the uniform merging of one-dimensional marginals, that is,

$$\lim_{n\to\infty} \sup_{x\in\tilde{T}_n} d_{\mathrm{Pr}}^{(T_n, r_n)}\big(\mathcal{L}^x(X_t^n), \mathcal{L}^x(\tilde{X}_t^n)\big) = 0, \tag{6.17}$$

where $d_{\mathrm{Pr}}^{(T_n,r_n)}$ is the Prohorov metric associated to r_n. The finite-dimensional statement then follows from the Markov property of the speed-ν motions together with the Feller continuity of the limiting process (proven in Proposition 5.2).

Recall that (T_n, r_n) is discrete and thus $\nu_n(\{x\}) > 0$ for all $x \in T_n$. Using Lemma 6.2, and the fact that $\mathrm{diam}(T_n)$ and $\nu_n(T_n)^{-1}$ are bounded uniformly in n, we obtain $\gamma_t > 0$, independent of n, such that

$$\sup_{x\in T_n} \mathbb{P}^x\{X_t^n \in A_n\} \le \gamma_t \sqrt{\nu_n(A_n)}. \tag{6.18}$$

We can couple X^n and $\tilde{X}^n$ by a time transformation such that $\tilde{X}_t^n = X^n_{L_n^{-1}(t)}$, where $L_n^{-1}(t) = \inf\{s \ge 0 : \int_0^s \mathbf{1}_{\tilde{T}_n}(X_u^n)\,\mathrm{d}u > t\}$. For (6.17), it is enough to show for every fixed $t, \varepsilon > 0$ that

$$\sup_{x\in \tilde{T}_n} \mathbb{P}^x\{r_n(X_t^n, \tilde{X}_t^n) > \varepsilon\} \le 4\varepsilon, \tag{6.19}$$

for all sufficiently large $n \in \mathbb{N}$. The idea is that X_t^n and $\tilde{X}_t^n$ do not differ too much, because $\tilde{X}_t^n$ cannot move far in a short amount of time and will be ahead of X_t^n only a small amount of time, controlled via the occupation time formula by the (small) ν_n-measure of $A_n = T_n \setminus \tilde{T}_n$.

Because $\tilde{\mathcal{X}}_n$ converges Gromov–Hausdorff weakly, we can use the speed bound, Lemma 4.7, to find $c > 0$ such that the probability that $\tilde{X}^n$ moves ε within time c is bounded by ε, that is,

$$\sup_{x\in \tilde{T}_n} \mathbb{P}^x\Big\{\sup_{s\in[0,c]} r_n(\tilde{X}_s^n, x) > \varepsilon\Big\} \le \varepsilon. \tag{6.20}$$

In order to use the occupation time formula, we fix two points $y_n, z_n \in \tilde{T}_n$ with $r_n(y_n, z_n) > \varepsilon$ and define recursively the times where X^n hits y_n and z_n in alternation, that is, $\tau_n^0 := 0$, $\tau_n^k := \inf\{t > \tau_n^{k-1} : X_t^n = y_n\}$ for k odd and $\tau_n^k := \inf\{t > \tau_n^{k-1} : X_t^n = z_n\}$ for k even. Let $\tilde{\tau}_n^k$, $k \in \mathbb{N}$, be the analogous stopping times for $\tilde{X}^n$ instead of X^n. Because the lower bound for the distance of y_n and z_n is independent of n, we can use Lemma 4.7 again to find $k \in \mathbb{N}$, independent of n, such that $\mathbb{P}\{\tilde{\tau}_n^k < t\} < \varepsilon$. Because $\tau_n^k \ge \tilde{\tau}_n^k$, we also obtain

$$\sup_{x\in \tilde{T}_n} \mathbb{P}^x\{\tau_n^k < t\} < \varepsilon. \tag{6.21}$$

Now consider the accumulated time difference between X^n and $\tilde{X}^n$ until τ_n^k, that is,

$$\delta_n := \int_0^{\tau_n^k} \mathbf{1}_{A_n}(X_t^n)\,\mathrm{d}t. \tag{6.22}$$

Then, by the occupation time formula,

$$\sup_{x\in \tilde{T}_n} \mathbb{E}^x[\delta_n] \le k \cdot 2\,\mathrm{diam}(T_n)\nu_n(A_n). \tag{6.23}$$

The right-hand side tends to zero as n tends to infinity, because $\operatorname{diam}(T_n)$ is uniformly bounded by assumption and k is independent of n. Therefore, for sufficiently large n depending on c chosen in (6.20),

$$\sup_{x \in \tilde{T}_n} \mathbb{P}^x\{\delta_n > c\} < \varepsilon. \tag{6.24}$$

On the event $\{X_t^n \notin A_n\}$, we have $X_t^n = \tilde{X}^n_{L_n(t)}$, and on the event $\{\tau_n^k \geq t\}$, we have $t - L_n(t) < \delta_n$. Hence, using (6.18) and (6.21), we obtain for all $x \in \tilde{T}_n$,

$$\begin{aligned} &\mathbb{P}^x\{r_n(X_t^n, \tilde{X}_t^n) > \varepsilon\} \\ &\quad \leq \mathbb{P}^x\{X_t^n \in A_n\} + \mathbb{P}^x\{\tau_n^k < t\} + \mathbb{P}^x\{t - L_n(t) < \delta_n, r_n(\tilde{X}^n_{L_n(t)}, \tilde{X}_t^n) > \varepsilon\} \\ &\quad \leq \gamma_t \sqrt{\nu_n(A_n)} + \varepsilon + \mathbb{P}^x\{\delta_n > c\} + \mathbb{P}^x\Big\{\sup_{s \in [t-c,t]} r_n(\tilde{X}_s^n, \tilde{X}_t^n)\Big\}, \end{aligned} \tag{6.25}$$

which is bounded by 4ε for large n due to $\nu_n(A_n) \to 0$, (6.24) and (6.20) together with the Markov property of $\tilde{X}^n$. This proves (6.19), and hence the claimed f.d.d. convergence. $\square$

COROLLARY 6.4 (Pointwise L^2-heat-kernel bound). *Lemma 6.2 remains correct if we drop the assumption $\nu(\{x\}) > 0$. In particular, for every compact metric finite measure tree $\mathcal{X} := (T, r, \nu)$, the following bound on the $L^2(\nu)$-norm of the heat-kernel q_t (defined in Lemma 6.2) holds*:

$$\|q_t(x, \cdot)\|_2^2 \leq \nu(T)^{-1} + \operatorname{diam}(T) \cdot t^{-1} \qquad \forall x \in T, t > 0. \tag{6.26}$$

PROOF. Fix $x \in T$, $t > 0$ and let $\nu_n := \nu + \frac{1}{n}\delta_x$. Let X^n and X be the speed-ν_n and speed-ν motion on (T, r), respectively, all started in x. According to Proposition 6.3 for $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n) := (T, r, x, \nu + \frac{1}{n}\delta_x)$, the law $\mu_{n,t}$ of X_t^n converges weakly to the law μ_t of X_t. According to Lemma 6.2, there is $f_{n,t} \in L^2(\nu)$ with $\mu_{n,t} = f_{n,t} \cdot \nu$, and $\|f_{n,t}\|_2$ is bounded uniformly in n. Therefore, the weak limit μ_t also admits a density with the same bound on its $L^2(\nu)$-norm. $\square$

We conclude this subsection with examples showing how the violation of the tightness condition (A2) destroys convergence in path space, while f.d.d. convergence still holds.

EXAMPLE 6.5 (F.d.d. convergence but not path-wise). Let $r, r_1, r_2, \ldots$ be the Euclidean metric on $[0, 1]$.

(i) Let $T_n = \{0, 1\}$, and $\nu_n = \delta_0 + \frac{1}{n}\delta_1$ for $n \in \mathbb{N}$. Then $\mathcal{X}_n := (T_n, r, 0, \nu_n)$ converges pointed Gromov-vaguely to $\mathcal{X} := (\{0\}, r, 0, \delta_0)$. The speed-$\nu_n$ motion X^n is a two-state Markov chain that jumps from 0 to 1 at rate $\frac{1}{2}$ and from 1 to 0 at rate $\frac{n}{2}$. It obviously converges f.d.d. to the constant process, but not in path-space.

(ii) Let $T_n = [0, 1]$, and $\nu_n = \delta_0 + \delta_1 + \frac{1}{n}\lambda_{[0,1]}$, where $\lambda_{[0,1]}$ is Lebesgue measure on $[0, 1]$. Then $(T_n, r, 0, \nu_n)$ converges pointed Gromov-vaguely to $(\{0, 1\}, r, 0, \nu)$ with $\nu = \delta_0 + \delta_1$. The speed-$\nu$ motion X is the symmetric Markov chain on $\{0, 1\}$ with jump-rate $\frac{1}{2}$, and the speed-ν_n motions X^n are sticky Brownian motions on $[0, 1]$ with diverging speed on $(0, 1)$, as n tends to ∞. As X^n has continuous paths for each $n \in \mathbb{N}$ but X has discontinuous paths, the convergence cannot be in path space. The finite dimensional distributions of X^n, however, converge to those of X, as the processes X^n spend less and less times in discontinuity points.

6.2. *From compact to locally compact limit trees.* In this subsection, we extend the proof of Theorem 1 to locally compact trees equipped with boundedly finite speed measures. In order to reduce this to the compact case, we stop the processes upon reaching a height R. For that purpose, we need the following lemma whose proof is straight-forward and will therefore be omitted.

Recall the closed interval property from Definition 5.7.

LEMMA 6.6 (Continuity points). *Let (E, d) be a Polish space, $\rho \in E$, and $R > 0$. Define the function*

$$\psi_R : \mathcal{D}_E \to \mathcal{D}_E, \qquad \psi_R(w)(t) := w\big(t \wedge \inf\{s : d\big(\rho, w(s)\big) \geq R\}\big). \tag{6.27}$$

Assume that $w \in \mathcal{D}_E$ has the closed-interval property, and that the map $t \mapsto d(\rho, w(t))$ does not have a local maximum at height R. Then w is a continuity point of ψ_R.

PROOF OF THEOREM 1. (ii) has already been shown in Proposition 6.3.

(i) We call a point $v \in T$ *extremal leaf* of T if the height function $h: T \to \mathbb{R}_+$, $x \mapsto r(\rho, x)$ has a local maximum at v. Note that, although there can be uncountably many extremal leaves, the set of heights of extremal leaves is at most countable due to separability of T. Now choose $R_k > 0$, $k \in \mathbb{N}$, with $R_k \to \infty$ such that there is no extremal leaf of T at height R_k and $\nu\{x' \in T : r(\rho, x') = R_k\} = 0$.

Let X be the speed-ν motion on (T, r) started in ρ, and recall that $X = X_{\cdot\wedge\zeta}$, where $\zeta := \inf\{t \geq 0 : r(\rho, X_t) = \infty\}$. We show that the law of X coincides with the law of $\tilde{X}_{\cdot\wedge\zeta} := \psi_\infty(\tilde{X})$, where $\tilde{X}$ is any limit process. Using that there is no extremal leaf of T at height R_k and that $\tilde{X}$ and X have the closed-interval property, we obtain from Lemma 6.6 that (the paths of) $\tilde{X}$ and X are almost surely continuity points of ψ_{R_k}.

Let X_k^n be the speed-ν_n motion on the compact metric measure tree $T_n\restriction_{B(\rho_n, R_k)}$ and X_k the speed-ν motion on the compact metric measure tree $T\restriction_{B(\rho, R_k)}$. Then, for every $k \in \mathbb{N}$, $X_k^n \underset{n\to\infty}{\Longrightarrow} X_k$, as $n \to \infty$, by Proposition 6.3. Furthermore, for every k there is an $\ell = \ell_k$, such that the laws of $\psi_{R_k}(X^n)$ and $\psi_{R_k}(X_\ell^n)$ coincide; and the same is true for $\psi_{R_k}(X)$ and $\psi_{R_k}(X_\ell)$.

By continuity of ψ_{R_k} in $\tilde{X}$ and X, we obtain

$$\psi_{R_k}(X_\ell^n) \stackrel{\mathcal{L}}{=} \psi_{R_k}(X^n) \underset{n\to\infty}{\overset{\mathcal{L}}{\Longrightarrow}} \psi_{R_k}(\tilde{X}), \tag{6.28}$$

and on the other hand

$$\psi_{R_k}(X_\ell^n) \underset{n\to\infty}{\overset{\mathcal{L}}{\Longrightarrow}} \psi_{R_k}(X_\ell) \stackrel{\mathcal{L}}{=} \psi_{R_k}(X). \tag{6.29}$$

Hence, $\psi_{R_k}(\tilde{X}) \stackrel{\mathcal{L}}{=} \psi_{R_k}(X)$ for all $k \in \mathbb{N}$ and, therefore, $\psi_\infty(\tilde{X}) \stackrel{\mathcal{L}}{=} \psi_\infty(X) = X$ as claimed. $\square$

7. Examples and related work. We conclude the paper with a discussion on how our invariance principle relates to results from the existing literature. These results have often been proven via quite different techniques but they all follow in a unified way from Theorem 1.

In Section 7.1, we revisit [31] which (including a killing part) proves the invariance principle in the particular situation when the underlying metric trees are closed subsets of $\mathbb{R}$, or equivalently, linear trees. In Section 7.2, we connect our invariance principle with the construction of diffusions on so-called dendrites, or equivalently, $\mathbb{R}$-trees, which is given in [23]. We continue in Section 7.3 with [12], where the classical convergence of rescaled simple random walks on $\mathbb{Z}$ to Brownian motion on $\mathbb{R}$ is generalized in a different direction than in [31]. Namely, simple random walks on discrete trees with uniform edge-lengths are proven to converge to Brownian motion on a limiting rooted compact $\mathbb{R}$-tree which additionally has to satisfy some conditions. Finally, in Section 7.4 we consider the nearest neighbor random walk on a size-biased branching tree for which the suitably rescaled height process averaged over all realizations is tight according to [22], while for almost every fixed realization it is not tight by [7].

7.1. *Invariance principle on* $\mathbb{R}$. In this subsection, we consider the special case of linear trees, that is, closed subsets of $\mathbb{R}$.

Let ν, ν_n, $n \in \mathbb{N}$, be locally finite measures on $\mathbb{R}$, $T := \operatorname{supp}(\nu)$ and $T_n := \operatorname{supp}(\nu_n)$. Denote the Euclidean metric on $\mathbb{R}$ by r. Then $(T, r, 0, \nu)$ and $(T_n, r, 0, \nu_n)$ are obviously rooted metric boundedly finite measure trees in the sense of Definition 1.1. Also note that the speed-ν motion is conservative (i.e., does not hit infinity), because the tree (T, r) is recurrent (see, e.g., [5], Theorem 4). Now if ν_n converges vaguely to ν, and the uniform local lower mass-bound (1.6) holds, Theorem 1 implies that the speed-ν_n motions converge in path-space to the speed-ν motion. This (essentially) is Theorem 1(i) obtained in [31] in the special case, where the killing measures are not present.

The methods used in [31] are quite different from ours. In that paper, all processes are represented as time-changes of standard Brownian motion and a jointly continuous version of local times is used.

EXAMPLE 7.1 (Standard motion on disconnected sets). A particular instance of Stone's invariance principle was studied in detail in [8]. Put for each $q > 1$, $T_q := \{\pm q^k; k \in \mathbb{Z}\} \cup \{0\}$ and $\rho_q = 0$. Then $(T_q)_{q>1}$ converges, as $q \downarrow 1$, to $\mathbb{R}$ with respect to the localized Hausdorff distance. Recall the length measure from (2.2). Obviously, as the length measure is always boundedly finite on linear trees, the embedding which sends a rooted tree (T, ρ) with $T \subseteq \mathbb{R}$ to the measure tree $(T, \rho, \lambda^{(T,\rho)})$ is a homeomorphism onto its image. Thus, $(T_q, 0, \lambda^{(T_q,0)})$ converges Hausdorff-vaguely to $(\mathbb{R}, 0, \lambda)$, as $q \downarrow 1$, where λ is the Lebesgue measure. It therefore follows that the speed-$\lambda^{(T_q,0)}$ motion on T_q converges in path space to the standard Brownian motion on $\mathbb{R}$ by Theorem 1. The latter is Proposition 5.1 in [8].

7.2. *Diffusions on dendrites.* In [23], diffusions on dendrites (which are $\mathbb{R}$-trees) are constructed via approximating Dirichlet forms rather than processes. In this subsection, we relate our invariance principle to this construction.

Let (T, r, ρ, ν) be a complete, locally compact, rooted boundedly finite measure $\mathbb{R}$-tree. Let furthermore $(T_m)_{m\in\mathbb{N}}$ be an increasing family of finite subsets of T. Put for all $f, g : T_m \to \mathbb{R}$

$$\mathcal{E}_m(f, g) := \frac{1}{2}\int_{T_m} \lambda^{(T_m, r_m, \rho)}(\mathrm{d}y)\nabla f(y)\nabla g(y). \tag{7.1}$$

Assume for each $m \in \mathbb{N}$ that T_m contains all the branch points of the subtree spanned by T_m [see our condition (1.2)]. Then for all $m \le m'$, and for all $f : T_m \to \mathbb{R}$,

$$\mathcal{E}_m(f, f) = \min\{\mathcal{E}_{m'}(g, g) : g : T_{m'} \to \mathbb{R}, g\restriction_{T_m} = f\}. \tag{7.2}$$

That is, the sequence $(T_m, \mathcal{E}_m)_{m\in\mathbb{N}}$ is compatible in the sense of Definition 0.2 (and the following paragraph) in [23]. Assume further that $T^* := \bigcup_{m\in\mathbb{N}} T_m$ is dense in T, and consider the bilinear form:

$$\mathcal{E}^{\mathrm{Kigami}}(f, g) := \lim_{m\to\infty} \mathcal{E}_m(f\restriction_{T_m}, g\restriction_{T_m}) \tag{7.3}$$

with domain

$$\mathcal{F}^{\mathrm{Kigami}} := \{f : T^* \to \mathbb{R} : \text{limit on RHS of (7.3) exists}\}. \tag{7.4}$$

Let $\mathcal{D}(\mathcal{E}^{\mathrm{Kigami}})$ be the completion of $\mathcal{F}^{\mathrm{Kigami}} \cap \mathcal{C}_c(T)$ with respect to the $\mathcal{E}^{\mathrm{Kigami}} + (\cdot, \cdot)_\nu$-norm. By Theorem 5.4 in [23], $(\mathcal{E}^{\mathrm{Kigami}}, \bar{\mathcal{D}}(\mathcal{E}^{\mathrm{Kigami}}))$ is a regular Dirichlet form.

It was shown in Remark 3.1 in [5] that the unique ν-symmetric strong Markov process associated with $(\mathcal{E}^{\mathrm{Kigami}}, \bar{\mathcal{D}}(\mathcal{E}^{\mathrm{Kigami}}))$ is the speed-ν motion on (T, r).

The bilinear form $\mathcal{E}^{\mathrm{Kigami}}$ describes the discrete time embedded Markov chains evaluated at T_n, $n \in \mathbb{N}$. The fact that it is a resistance form means that the projective limit diffusion is on "natural scale," which we additionally equip with speed measure ν. We can, of course, also approximate the speed-ν motion on (T, r) by continuous time Markov chains evaluated at T_n, $n \in \mathbb{N}$. Similarly, as in the proof of Lemma 6.1, consider for each $n \in \mathbb{N}$ the map $\psi_n : T \to T_n$ which sends a point in T to the nearest point on the way from x to ρ which belongs to T_n, that is,

$$\psi_n(x) := \sup\{y \in T_n : y \in [\rho, x]\}, \tag{7.5}$$

and equip T_n with

$$\nu_n := (\psi_n)_* \nu. \tag{7.6}$$

As T^* is dense, $(\nu_n)_{n\in\mathbb{N}}$ converges vaguely to ν, and thus $(T_n, r, \nu_n)_{n\in\mathbb{N}}$ converges Gromov–Hausdorff-vaguely to (T, r, ν). It therefore follows from our invariance principle that the continuous time Markov chains which jump from $v \in T_n$ to a neighboring $v \sim v'$ at rate $(2\nu_n(\{v\})r(v, v'))^{-1}$ converges weakly in path space to the speed-ν motion on (T, r).

7.3. *Invariance principle with homogeneous rescaling*. In this subsection, we relate our invariance principle to the one obtain earlier in [12]. We first recall the excursion representation of a rooted compact measure $\mathbb{R}$-tree. We denote by

$$\mathcal{E} := \{e: [0, 1] \to \mathbb{R}_+ | e \text{ is continuous}, e(0) = e(1) = 0\} \tag{7.7}$$

the set of continuous excursions on $[0, 1]$. From each excursion $e \in \mathcal{E}$, we can define a measure $\mathbb{R}$-tree in the following way:

- Define a pseudo-distance on $[0, 1]$ by $r_e(x, y) := e(x) + e(y) - 2 \inf_{[x,y]} e$.
- Call $x, y \in [0, 1]$ equivalent, $x \sim_e y$, if $r_e(x, y) = 0$.
- Endow the image of the canonical projection $\pi_e: [0, 1] \to [0, 1]/\sim_e$ with the push forward of r_e (again denoted r_e). Then $T_e := (T_e, r_e, \rho_e) := (\pi_e([0, 1]), r_e, \pi_e(0))$ is a rooted compact $\mathbb{R}$-tree.
- Endow this space with the probability measure $\mu_e := \pi_{e*}\lambda_{[0,1]}$ which is the push forward of the Lebesgue measure on $[0, 1]$.

We denote by $g : \mathcal{E} \to \mathbb{T}_c$ the resulting "glue function,"

$$g(e) := (T_e, r_e, \rho_e, \mu_e), \tag{7.8}$$

which sends an excursion to a rooted probability measure $\mathbb{R}$-tree.

Recall $\mathbb{T}_c$ from (3.2). Given $\mathcal{X} := (T, r, \rho, \nu) \in \mathbb{T}_c$, we say that $\mathcal{X}$ satisfies a polynomial lower bound for the volume of balls, or short a *polynomial lower bound* if there is a $\kappa > 0$ such that

$$\liminf_{\delta \downarrow 0} \inf_{x \in T} \delta^{-\kappa} \nu\big(B_r(x, \delta)\big) > 0. \tag{7.9}$$

In [12], the following subspace of $\mathbb{T}_c$ is considered:

$$\begin{aligned} \mathbb{T}^* := \{\mathcal{X} = (T, r, \rho, \nu) \in \mathbb{T}_c :& \\ & (a)\ \nu \text{ is non-atomic, (b) } \nu \text{ is supported on the leaves, and} \\ & (c)\ \nu \text{ satisfies a polynomial lower bound}\}. \end{aligned} \tag{7.10}$$

Let $((T_n, \rho_n))_{n\in\mathbb{N}}$, be a sequence of rooted graph trees with $\#T_n = n$, whose search-depth functions e_n in $\mathcal{E}$ with uniform topology satisfy

$$\frac{1}{a_n} e_n \underset{n\to\infty}{\longrightarrow} e \tag{7.11}$$

for a sequence $(a_n)_{n\in\mathbb{N}}$ and some $e \in \mathcal{E}$ with $(T_e, r_e, 0, \mu_e) \in \mathbb{T}^*$. In Theorem 1.1 of [12], it is shown that the discrete-time simple random walks on T_n starting in ρ_n with jump sizes rescaled by $1/a_n$ and speeded up by a factor of $n \cdot a_n$ converge to the μ_e-Brownian motion on T_e starting in 0.

To connect the above construction with Theorem 1 notice that the map g from (7.8) is continuous if $\mathbb{T}_c$ is endowed with the rooted Gromov–Hausdorff-weak topology, and $\mathcal{E}$ with the uniform topology (see [2], Proposition 2.9; compare also [26], Theorem 4.8, for a generalization to lower semi-continuous excursions). Thus, it follows from (7.11) that if we put $\nu_n := \mu_{a_n^{-1} e_n}$, then (T_n, ν_n) converges to (T_e, μ_e) rooted Gromov–Hausdorff-weakly. Analogously to Example 1.5, we obtain that $d_{\mathrm{Pr}}^{(T_n, r_n)}(\nu_n, \tilde{\nu}_n) \le a_n^{-1}$, where

$$\tilde{\nu}_n(\{v\}) := \frac{\deg(v)}{2n}, \tag{7.12}$$

and that thus also $(T_n, \tilde{\nu}_n)$ converges to (T_e, μ_e) rooted Gromov–Hausdorff-weakly by [6], Lemma 2.10. Theorem 1 then implies that unit rate simple random walks with edge lengths rescaled by a_n^{-1} and speeded up by $n \cdot a_n$ converge to the speed-μ_e motion on (T_e, r_e). As μ_e always has full support, the requirement that μ_e is supported on the leaves already implies that (T_e, r_e) is an $\mathbb{R}$-tree and thus the speed-μ_e motion on (T_e, r_e) has continuous paths.

Note that in contrast to [12] our Theorem 1 does not require any additional assumptions on the limiting tree, which also does not have to be an $\mathbb{R}$-tree. The polynomial lower bound or that ν is non-atomic and supported on the leaves are not required. Also note that Theorem 1.1 of [12] does only allow for homogeneous (non-state-dependent) rescaling. This means, for example, that in the particular case where the trees (T_n, r_n) are subsets of $\mathbb{R}$, only the case $T_n = a_n^{-1}\mathbb{Z} \cap [0, na_n^{-1}]$ and $\nu_n(\{x\}) = n^{-1}$, $x \in T_n$, is covered.

7.4. *Random walk on the size-biased branching tree.* Theorem 1 applies to trees that are complete and locally compact. The extension from compact to complete, locally compact trees is relatively straight-forward. However, this extension helps us to cover the random walk on the size-biased Galton–Watson tree studied in [22] in the annealed regime and in [7] in the quenched regime. In this subsection, we want to illuminate these results and put them in the context of our invariance principle.

Consider a random graph theoretical tree $\mathcal{T}_{\mathrm{Kesten}}$ which is distributed like the rooted Galton–Watson process with finite variance mean 1 offspring distribution conditioned to never die out. Let X be the (discrete-time) nearest neighbor random walk on $\mathcal{T}_{\text{Kesten}}$ and d the graph distance on $\mathcal{T}_{\text{Kesten}}$. Consider the rescaled height process

$$Z_t^{(n)} := n^{-1/3} \cdot d(\rho, X_{\lfloor nt \rfloor}), \qquad t \geq 0. \tag{7.13}$$

In [22], it is shown that if $\tau_{B^c(\rho,N)} := \inf\{n \geq 0 : d(\rho, X_n) = N\}$, then for all $\varepsilon > 0$ there exists λ_1, λ_2 such that under the annealed law $\mathbb{P}^*$,

$$\mathbb{P}^*\{\lambda_1 \leq N^{-3}\tau_{B^c(\rho,N)} \leq \lambda_2\} \geq 1 - \varepsilon,$$

for all $N \geq 1$. Moreover, under $\mathbb{P}^*$, the process $Z^{(n)}$ converges weakly in path space to a non-trivial process Z with continuous paths.

In contrast to this annealed regime, in [7] (in the continuous time setting) it is shown that for almost all realizations of $\mathcal{T}_{\text{Kesten}}$, the family $\{Z^{(n)}; n \in \mathbb{N}\}$ is not tight.

These two statements relate to our invariance principle as follows. Recall from (7.7) the space of continuous excursions on [0, 1] and from (7.8) the glue map g which sends an excursion $e \in \mathcal{E}$ to a rooted metric tree $([0, 1]/{\sim_e}, r_e, 0)$ as well the map π_e, which given $e \in \mathcal{E}$, sends a point from the excursion interval [0, 1] to T_e. We can easily extend the maps g and π_e to the space

$$\mathcal{E}_\infty := \Big\{e : \mathbb{R} \to \mathbb{R}_+ \big| e \text{ is continuous}, e(0) = 0, \lim_{x \to \pm\infty} e(x) = \infty\Big\} \tag{7.14}$$

of continuous, two-sided, transient excursions on $\mathbb{R}$. To this end, we use the semi-metric defined by

$$r_e(x, y) := \begin{cases} e(x) + e(y) - 2 \inf\limits_{z \in [x,y]} e(z), & xy \geq 0, \\ e(x) + e(y) - 2 \inf\limits_{z \in \mathbb{R}\setminus[x,y]} e(z), & xy < 0, \end{cases} \tag{7.15}$$

for $x \leq y$ (see [13]). Then $g(e)$ is a rooted locally compact metric measure tree with a boundedly finite measure, for all $e \in \mathcal{E}_\infty$. It is not hard to show that the map g from (7.8) is continuous if $\mathbb{T}$ is endowed with the rooted Gromov–Hausdorff-vague topology, and $\mathcal{E}_\infty$ with the uniform topology on compact sets (see [6], Proposition 7.5).

In the particular case of a geometric offspring distribution, $\mathcal{T}_{\text{Kesten}}$ can be associated with the (two-sided) random excursion $\tilde{W}$, where for all $t \in \mathbb{R}$,

$$\tilde{W}_t := \begin{cases} W_t - 2 \inf\limits_{s \in [0,t]} W_s, & t \geq 0, \\ W_t - 2 \inf\limits_{s \in [t,0]} W_s, & t < 0, \end{cases} \tag{7.16}$$

with a simple two-sided random walk path $(W_n)_{n \in \mathbb{Z}}$, $W_0 = 0$, linearly interpolated. As W converges, after Brownian rescaling, weakly in path space towards (two-sided) standard Brownian motion $(B_t)_{t \in \mathbb{R}}$, we have

$$\big(n^{-1/3} \tilde{W}_{n^{2/3}t}\big)_{t \in \mathbb{R}} \underset{n \to \infty}{\Longrightarrow} (\tilde{B}_t)_{t \in \mathbb{R}}, \tag{7.17}$$

where $\tilde{B}_t := B_t - 2 \inf_{s \in [0 \wedge t, t \vee 0]} B_s$.

Given a realization e of $\tilde{W}$, define $e_n := n^{-1/3}e(n^{2/3}\cdot) \in \mathcal{E}_\infty$ and denote by ν_n the rescaled degree measure on T_{e_n}, that is, for all $A \subseteq T_{e_n}$,

$$\nu_n(A) := n^{-2/3} \sum_{v \in A} \frac{1}{2} \deg(v). \tag{7.18}$$

By Proposition 2.8 in [7], for almost all realizations e of $\tilde{W}$,

$$\liminf_{n\to\infty} \nu_n\big(B(\rho, R)\big) = 0 \quad \text{and} \quad \limsup_{n\to\infty} \nu_n\big(B(\rho, R)\big) = \infty, \tag{7.19}$$

and thus the sequence $\{\nu_n; n \in \mathbb{N}\}$ does not converge. Consider once more the map which sends all points of a half edge to its end point, and notice that the image measure of $\mu_{e_n} = (\pi_{e_n})_*\lambda_{\mathbb{R}_+}$ under this map equals ν_n. Thus, the Prohorov distance between μ_{e_n} and ν_n is at most $n^{-1/3}$, and thus for almost all realizations e of $\tilde{W}$, also the sequence $\{\mu_{e_n}; n \in \mathbb{N}\}$ does not converge. Hence, the assumptions on our invariance principle fail for almost all realizations of $\mathcal{T}_{\text{Kesten}}$.

Notice that we can choose for each $n \in \mathbb{N}$ a realization e_n of $n^{-1/3}\tilde{W}_{n^{2/3}\cdot}$, and a realization e of $\tilde{B}$, such that $e_n \underset{n\to\infty}{\longrightarrow} e$, almost surely. To understand why the quenched rescaling failed, notice that $e_n \underset{n\to\infty}{\longrightarrow} e$ CANNOT be realized via a coupling such that all the e_n come from the same realization of $\tilde{W}$. As now $g(e_n)$ clearly converges to $g(e)$ by continuity of g, Theorem 1 implies that the speed-μ_{e_n} random walk X^n on (T_{e_n}, r_{e_n}) starting in ρ_{e_n} converges weakly in path space to the μ_e-Brownian motion $X = (X_t)_{t\geq 0}$ on (T_e, r_e) started in ρ_e for almost all realizations. We can interpret this as *annealed convergence* in law of X^n to X, which we define—in analogy to Definition 1.3 and in view of Skorohod's representation theorem—as follows. There exists a coupling of the underlying random spaces $\mathcal{X} = (T_e, r_e, \mu_e)$, $\mathcal{X}_n = (T_{e_n}, r_{e_n}, \mu_{e_n})$, $n \in \mathbb{N}$, such that almost surely, conditioned on these spaces, X^n converges weakly in path space to X in the sense of Definition 1.3. In particular, the rescaled height processes $Z^{(n)}$, defined in (7.13), converge under the annealed law to the height process $Z = (Z_t)_{t\geq 0}$ defined by $Z_t := r_e(\rho_e, X_t)$. As X is recurrent by Theorem 4 in [5], its life time is infinite, and Z is non-trivial.

7.5. *Motions on Λ-coalescent measure trees.* We conclude the example section with the example of speed-ν motions on the Λ-coalescent measure trees for appropriate measures ν. These have not been considered in the literature so far.

Let Λ be a finite measure on $([0, 1], \mathcal{B}([0, 1]))$ which satisfies

$$\sum_{n=2}^{\infty} \left(\int_0^1 \sum_{k=2}^{n} \binom{n}{k} (k-1)x^{k-2}(1-x)^{n-k} \Lambda(\mathrm{d}x) \right)^{-1} < \infty. \tag{7.20}$$

Denote by $\mathbb{S}$ the set of all partitions of $\mathbb{N}$, and for each $n \in \mathbb{N}$ by $\mathbb{S}_n$ the set of all partitions of $\{1, \dots, n\}$. Write ρ_n for the restriction map from $\mathbb{S}$ to $\mathbb{S}_n$.

The Λ-coalescent is the unique $\mathbb{S}$-valued strong Markov process ζ, such that for each $n \in \mathbb{N}$ the restricted process $\rho_n(\zeta)$ is the following $\mathbb{S}_n$-valued continuous time Markov chain. Given the current partition $\mathcal{P} \in \mathbb{S}_n$, every k-tuple of its partition elements merges independently at rate

$$\lambda_{k,\#\mathcal{P}} := \int \Lambda(\mathrm{d}x)x^{k-2}(1-x)^{\#\mathcal{P}-k} \tag{7.21}$$

into one partition element, thereby forming a new partition. It is known that condition (7.20) is equivalent to the Λ-coalescent coming down from infinity, that is, under (7.20), $\#\zeta_t < \infty$ for each $t > 0$, almost surely [30]. Furthermore, (7.20) implies the so-called *dust-free property*, that is, $\int_0^1 \Lambda(\mathrm{d}x)x^{-1} = \infty$.

Equip for each realization of the Λ-coalescent started in $\mathcal{P}_0 := \{\{i\} : i \in \mathbb{N}\}$ the set $\mathbb{N}$ with the genealogical distances, that is, $r(i, j)$ is for all $i, j \in \mathbb{N}$ the first time when i and j belong to the same partition element. Denote the completion of $(\mathbb{N}, r)$ by $(\mathcal{T}_\Lambda, r)$. Obviously, coming down from infinity implies (and is in fact equivalent to) the compactness of $\mathcal{T}_\Lambda$. Further, equip for each $n \in \mathbb{N}$, $\mathcal{T}_\Lambda$ with the sampling measure $\mu^n := \frac{1}{n}\sum_{i=1}^n \delta_i$. By Theorem 4 in [19] the sequence $((\mathcal{T}_\Lambda, r, \mu^n))_{n\in\mathbb{N}}$ converges weakly in Gromov-weak topology towards the so-called Λ-*coalescent measure tree*, $(\mathcal{T}_\Lambda, r, \mu)$.

Consider next the $\mathbb{R}$-tree $(\bar{\mathcal{T}}_\Lambda, \bar{r})$ spanned by $(\mathcal{T}_\Lambda, r)$, and notice that $\mathcal{T}_\Lambda$ is ultra-metric. We therefore find a unique point $\rho \in \bar{\mathcal{T}}_\Lambda$ whose distance to $\mathcal{T}_\Lambda$ equals $\mathrm{diam}(\bar{\mathcal{T}}_\Lambda)/2$, which we choose as the root. For each point $x \in \bar{\mathcal{T}}_\Lambda$ denote by

$$S^x := \{z \in \mathcal{T}_\Lambda : x \in [\rho, z]\} \tag{7.22}$$

the (leaves of the) subtree above x, and recall from (2.2) the notion of the length measure $\lambda^{(T,r,\rho)}$ of a rooted compact metric tree (T, r, ρ).

Define the speed measures ν^n, $n \in \mathbb{N}$, and ν on $\bar{\mathcal{T}}_\Lambda$ as being absolutely continuous with respect to the length measure with densities

$$\frac{\mathrm{d}\nu^n}{\mathrm{d}\lambda^{\bar{\mathcal{T}}_\Lambda}}(x) := \mu^n(S^x) \quad \text{and} \quad \frac{\mathrm{d}\nu}{\mathrm{d}\lambda^{\bar{\mathcal{T}}_\Lambda}}(x) := \mu(S^x), \tag{7.23}$$

for all $x \in \bar{\mathcal{T}}_\Lambda$. Obviously, ν^n, $n \in \mathbb{N}$, and ν are finite measures with total masses at most (and in fact due to the dust-free property equal to) $\mathrm{diam}(\bar{\mathcal{T}}_\Lambda)/2$. Note that for every ultra-metric space (T, r), the map $\xi^{(T,r)}$ which sends a pair $(t, x) \in [0, \infty) \times T$ to the unique "ancestor" of x a time t back, that is, the unique $y \in \bar{T}$ ($\bar{T}$ denoting the span of T) with $\bar{r}(y, x) = t \wedge \frac{1}{2}\mathrm{diam}(\bar{T})$ is continuous. Hence, using the convergence alluded to earlier (Theorem 4 in [19]) the sequence $((\bar{\mathcal{T}}_\Lambda, \nu^n))_{n\in\mathbb{N}}$ converges weakly in Gromov-weak topology toward $(\bar{\mathcal{T}}_\Lambda, \nu)$. Our invariance principle therefore implies that the ν_n-Brownian motion on $(\mathrm{supp}(\nu^n), \bar{r})$ converges weakly to the ν-Brownian motion on $(\bar{\mathcal{T}}_\Lambda, \bar{r})$ in the sense of finite dimensional marginals (provided all Brownian motions start at the same point). Applying once more the dust-free property implies that the global lower mass-bound holds, and thus the convergence holds even in path space.

We can modify the example such that we obtain path-wise convergence of a continuous time Markov chain to a motion on a totally disconnected (limiting) tree. For that purpose, denote by $\mathrm{Br}(\bar{\mathcal{T}}_\Lambda)$ the set of branch points of $\bar{\mathcal{T}}_\Lambda$, that is, the set of those $x \in \bar{\mathcal{T}}_\Lambda$ such that either $x = \rho$ or $\bar{\mathcal{T}}_\Lambda \setminus \{x\}$ consists of at least 3 connected components. Consider now the (atomic) length measure on $\mathrm{Br}(\bar{\mathcal{T}}_\Lambda)$ and the Dirac measure δ_ρ, and define

$$\hat{\lambda} := \lambda^{(\mathrm{Br}(\bar{\mathcal{T}}_\Lambda), \bar{r}, \rho)} + \delta_\rho. \tag{7.24}$$

We use the speed measures $\tilde{\nu}^n$, $n \in \mathbb{N}$, and $\tilde{\nu}$ on $\bar{\mathcal{T}}_\Lambda$ which are absolutely continuous with respect to $\hat{\lambda}$ with densities

$$\frac{\mathrm{d}\tilde{\nu}^n}{\mathrm{d}\hat{\lambda}}(x) := \mu^n(S^x) \quad \text{and} \quad \frac{\mathrm{d}\tilde{\nu}}{\mathrm{d}\hat{\lambda}}(x) := \mu(S^x) \tag{7.25}$$

for all $x \in \mathrm{Br}(\bar{\mathcal{T}}_\Lambda)$. For each $\varepsilon \in (0, \frac{1}{2}\operatorname{diam}(\bar{\mathcal{T}}_\Lambda))$ and for all suitably large $n \in \mathbb{N}$, we have $\operatorname{supp}(\tilde{\nu}^n) \cap \{x \in \mathrm{Br}(\bar{\mathcal{T}}_\Lambda) : \bar{r}(x, \mathcal{T}_\Lambda) \geq \varepsilon\} = \{x \in \mathrm{Br}(\bar{\mathcal{T}}_\Lambda) : \bar{r}(x, \mathcal{T}_\Lambda) \geq \varepsilon\}$. Therefore, the sequence $((\bar{\mathcal{T}}_\Lambda, \tilde{\nu}^n))_{n\in\mathbb{N}}$ also converges weakly in Gromov-weak topology toward $(\bar{\mathcal{T}}_\Lambda, \tilde{\nu})$. Thus our invariance principle applies to the speed-$\tilde{\nu}^n$ random walk on $\operatorname{supp}(\tilde{\nu}^n)$ and the speed-$\tilde{\nu}$ motion on $\operatorname{supp}(\tilde{\nu}) = \mathrm{Br}(\bar{\mathcal{T}}_\Lambda) \cup \mathcal{T}_\Lambda$.

Acknowledgments. We would like to thank Steve Evans and Fabian Gerle for discussions, and the anonymous referees for their detailed reports, which enabled us to improve the paper.

REFERENCES

[1] ABRAHAM, R., DELMAS, J.-F. and HOSCHEIT, P. (2013). A note on the Gromov–Hausdorff–Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.* **18** no. 14, 21. MR3035742

[2] ABRAHAM, R., DELMAS, J.-F. and HOSCHEIT, P. (2014). Exit times for an increasing Lévy tree-valued process. *Probab. Theory Related Fields* **159** 357–403. MR3201925

[3] ALDOUS, D. (1991). The continuum random tree. II. An overview. In *Stochastic Analysis* (*Durham*, 1990). *London Mathematical Society Lecture Note Series* **167** 23–70. Cambridge Univ. Press, Cambridge. MR1166406

[4] ALDOUS, D. (1993). The continuum random tree. III. *Ann. Probab.* **21** 248–289. MR1207226

[5] ATHREYA, S., ECKHOFF, M. and WINTER, A. (2013). Brownian motion on $\mathbb{R}$-trees. *Trans. Amer. Math. Soc.* **365** 3115–3150. MR3034461

[6] ATHREYA, S., LÖHR, W. and WINTER, A. (2016). The gap between Gromov-vague and Gromov–Hausdorff-vague topology. *Stochastic Process. Appl.* **126** 2527–2553. MR3522292

[7] BARLOW, M. T. and KUMAGAI, T. (2006). Random walk on the incipient infinite cluster on trees. *Illinois J. Math.* **50** 33–65 (electronic). MR2247823

[8] BHAMIDI, S., EVANS, S. N., PELED, R. and RALPH, P. (2008). Brownian motion on disconnected sets, basic hypergeometric functions, and some continued fractions of Ramanujan. In *Probability and Statistics*: *Essays in Honor of David A. Freedman*. *Inst. Math. Stat. Collect.* **2** 42–75. IMS, Beachwood, OH. MR2459949

[9] BOGACHEV, V. I. (2007). *Measure Theory. Vol. I, II*. Springer, Berlin. MR2267655

[10] BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Amer. Math. Soc., Providence, RI. MR1835418

[11] CROYDON, D. (2008). Convergence of simple random walks on random discrete trees to Brownian motion on the continuum random tree. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 987–1019. MR2469332

[12] CROYDON, D. A. (2010). Scaling limits for simple random walks on random ordered graph trees. *Adv. in Appl. Probab.* **42** 528–558. MR2675115

[13] DUQUESNE, T. (2009). Continuum random trees and branching processes with immigration. *Stochastic Process. Appl.* **119** 99–129. MR2485021

[14] EVANS, S. N. (2008). *Probability and Real Trees*. *Lecture Notes in Math.* **1920**. Springer, Berlin. MR2351587

[15] EVANS, S. N., PITMAN, J. and WINTER, A. (2006). Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields* **134** 81–126. MR2221786

[16] EVANS, S. N. and WINTER, A. (2006). Subtree prune and regraft: A reversible real tree-valued Markov process. *Ann. Probab.* **34** 918–961. MR2243874

[17] FUKAYA, K. (1987). Collapsing of Riemannian manifolds and eigenvalues of Laplace operator. *Invent. Math.* **87** 517–547. MR0874035

[18] FUKUSHIMA, M., OSHIMA, Y. and TAKEDA, M. (2011). *Dirichlet Forms and Symmetric Markov Processes*, extended ed. *de Gruyter Studies in Mathematics* **19**. de Gruyter, Berlin. MR2778606

[19] GREVEN, A., PFAFFELHUBER, P. and WINTER, A. (2009). Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probab. Theory Related Fields* **145** 285–322. MR2520129

[20] GROMOV, M. (2007). *Metric Structures for Riemannian and Non-Riemannian Spaces*, English ed. *Modern Birkhäuser Classics*. Birkhäuser, Boston, MA. MR2307192

[21] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. MR1876169

[22] KESTEN, H. (1986). Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. Henri Poincaré Probab. Stat.* **22** 425–487. MR0871905

[23] KIGAMI, J. (1995). Harmonic calculus on limits of networks and its application to dendrites. *J. Funct. Anal.* **128** 48–86. MR1317710

[24] KREBS, W. B. (1995). Brownian motion on the continuum tree. *Probab. Theory Related Fields* **101** 421–433. MR1324094

[25] KUWAE, K. and SHIOYA, T. (2003). Convergence of spectral structures: A functional analytic theory and its applications to spectral geometry. *Comm. Anal. Geom.* **11** 599–673. MR2015170

[26] LÖHR, W. (2013). Equivalence of Gromov–Prohorov- and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces. *Electron. Commun. Probab.* **18** no. 17, 10. MR3037215

[27] LÖHR, W., VOISIN, G. and WINTER, A. (2015). Convergence of bi-measure $\mathbb{R}$-trees and the pruning process. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 1342–1368. MR3414450

[28] MOSCO, U. (1969). Convergence of convex sets and of solutions of variational inequalities. *Adv. Math.* **3** 510–585. MR0298508

[29] MOSCO, U. (1994). Composite media and asymptotic Dirichlet forms. *J. Funct. Anal.* **123** 368–421. MR1283033

[30] SCHWEINSBERG, J. (2000). A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Commun. Probab.* **5** 1–11 (electronic). MR1736720

[31] STONE, C. (1963). Limit theorems for random walks, birth and death processes, and diffusion processes. *Illinois J. Math.* **7** 638–660. MR0158440

S. ATHREYA
STATISTICS AND MATHEMATICS UNIT
INDIAN STATISTICAL INSTITUTE
8TH MILE MYSORE ROAD
BANGALORE 560059
INDIA
E-MAIL: athreya@isibang.ac.in

W. LÖHR
A. WINTER
FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT DUISBURG-ESSEN
THEA-LEYMANN-STR. 9
45117 ESSEN
GERMANY
E-MAIL: wolfgang.loehr@uni-due.de
anita.winter@uni-due.de