

# ON THE COVER TIME OF BROWNIAN MOTION ON THE BROWNIAN CONTINUUM RANDOM TREE

BY G. ANDRIOPOULOS<sup>1,a</sup>, D. A. CROYDON<sup>2,b</sup>, V. MARGARINT<sup>3,c</sup>, AND L. MÉNARD<sup>4,d</sup>

<sup>1</sup>*Department of Computer Science, New York University Abu Dhabi, [ga73@nyu.edu](mailto:ga73@nyu.edu)*

<sup>2</sup>*Research Institute for Mathematical Sciences, Kyoto University, Japan, [croydon@kurims.kyoto-u.ac.jp](mailto:croydon@kurims.kyoto-u.ac.jp)*

<sup>3</sup>*Department of Mathematics and Statistics, University of North Carolina at Charlotte, [vmargari@charlotte.edu](mailto:vmargari@charlotte.edu)*

<sup>4</sup>*Modal'X, UMR CNRS 9023, UPL, Univ. Paris-Nanterre, F92000 Nanterre, France, [laurent.menard@normalesup.org](mailto:laurent.menard@normalesup.org)*

Upon almost-every realisation of the Brownian continuum random tree (CRT), it is possible to define a canonical diffusion process or ‘Brownian motion’. The main result of this article establishes that the cover time of the Brownian motion on the Brownian CRT (i.e. the time taken by the process in question to visit the entire state space) is equal to the infimum over the times at which the associated local times are strictly positive everywhere. The proof of this result depends on the recursive self-similarity of the Brownian CRT and a novel version of the first Ray-Knight theorem for trees, which is of independent interest. As a consequence, we obtain that the suitably-rescaled cover times of simple random walks on critical, finite variance Galton-Watson trees converge in distribution with respect to their annealed laws to the cover time of Brownian motion on the Brownian CRT. Other families of graphs that have the Brownian CRT as a scaling limit are also covered. Additionally, we partially confirm a 1991 conjecture of David Aldous regarding related cover-and-return times.

## CONTENTS

1	Introduction . . . . .	2
2	Brownian motion on compact real trees . . . . .	9
	2.1 Definition and basic properties . . . . .	9
	2.2 Approximations for the cover time . . . . .	10
	2.3 A Ray-Knight theorem for trees . . . . .	16
3	Height decomposition for the Brownian CRT . . . . .	26
4	A path property of BESQ-Brownian snakes . . . . .	30
	4.1 Definition and statement of the path property . . . . .	30
	4.2 The special Markov property . . . . .	32
	4.3 Proof of Theorem 4.2 . . . . .	34
5	Proof of Theorem 1.1 . . . . .	36
	5.1 An extended Gromov-Hausdorff-type topology . . . . .	36
	5.2 Checking the assumptions and the result with $X^T$ started from $\rho$ . . . . .	39
	5.3 Extending to arbitrary starting points . . . . .	40
6	Convergence to the cover time of the Brownian CRT . . . . .	42
7	Moments of cover times for critical Galton-Watson trees . . . . .	44
8	Cover-and-return times . . . . .	49
	Funding . . . . .	52
	References . . . . .	52

---

*MSC2020 subject classifications:* Primary 60K37; secondary 05C81, 60G17, 60J55, 60J68.

*Keywords and phrases:* continuum random tree, Brownian snake, Ray-Knight theorem, random walk, local times, cover time, scaling limit.

**1. Introduction.** The cover time, that is, the time taken to explore the entire state space, is a natural quantity to study for a stochastic process. In this direction, a question that has been widely considered in the probability and computer science literature is how the cover time behaves asymptotically for a sequence of random walks on graphs of increasing sizes. However, the cover time is not always an easy quantity to pin down. Indeed, even when one has a functional scaling limit for a sequence of random walks, the scaling of the cover time can not typically be read off from this, since such results do not take into account the fine structure of the paths of the processes. For instance, to see a non-trivial functional limit theorem for the random walk on the discrete torus  $(\mathbb{Z}/n\mathbb{Z})^d$ , equipped with nearest neighbor edges, the time-scaling is  $n^2$ , but the cover time scales like  $n^2(\log n)^2$  for  $d = 2$  and  $n^d(\log n)$  for  $d \geq 3$  (see [32] for a more detailed result in two dimensions, and [56, Chapter 11] for an argument that yields the order of growth in higher dimensions). Moreover, although the random walk on the largest supercritical percolation cluster contained within the box  $\{-n, -n + 1, \dots, n\}^d$  satisfies the same global scaling (up to constants, and modulo the boundary condition), the cover time in this case is of order  $n^d(\log n)^2$  for all  $d \geq 2$ , see [2]. (It is reasonable to expect the same result for supercritical percolation on the discrete torus.) Thus we see that the cover time can be highly sensitive to the detailed geometry of the underlying space.

The goal of this article is to demonstrate convergence of the rescaled cover times of random walks on graphs that converge to the Brownian continuum random tree (CRT). The latter object has, since its introduction by Aldous in [9, 10, 13], come to be a central object in modern probability theory. It is known to appear as a scaling limit of critical Galton-Watson trees [10], uniform spanning trees on high-dimensional tori [17], critical percolation clusters on certain random hyperbolic maps [16, 24], amongst many other examples, and is also conjectured to appear in the description of the scaling limit of critical percolation clusters on the integer lattice in high dimensions [42]. The limiting cover time in our main convergence result (see Theorem 1.2 below) will be expressed in terms of the cover time of the natural diffusion, or ‘Brownian motion’, on the Brownian CRT, which is a process whose definition was suggested by Aldous in [10] and which was first rigorously defined by Krebs in [51]. In the course of our arguments, we establish various new results for this process.

In fact, cover times of random walks on graph trees, and more general graphs, was something of a research focus for Aldous. For example, in [11], he derived some scaling results concerning the cover time of random walks on some particular examples of trees, including giving precise asymptotics in the case of the balanced  $b$ -ary tree. In the same paper, he discussed the cover times of random walks on critical Galton-Watson trees, including giving a specific conjecture on the scaling limit of the expected ‘cover-and-return’ time (i.e. the time taken to hit all vertices and then return to the starting point) for a tree with a Poisson(1) offspring distribution. As a corollary of our main result, we will partially confirm his conjecture, in the sense that we confirm a limit of the relevant sequence exists (see Corollary 1.5 below). Moreover, since our limit is written in terms of the expectation of the cover-and-return time of the Brownian motion on the Brownian CRT, this also supports his conjecture on the exact value of this constant. We highlight that understanding the cover time of the latter process was listed in a section of Aldous’ paper [10] entitled ‘hard distributional properties [of Brownian motion on continuum trees]’.

Part of the difficulty of studying the cover time of Brownian motion on the Brownian CRT is that it does not concentrate on its mean. In [12], Aldous gave a characterisation of when concentration of the cover time occurs; namely, the phenomenon arises if the maximal expected hitting time of vertices is asymptotically of a smaller order than the expected cover time, but there exist starting states for which concentration is not seen if this is not the case. (See [1] for work that helps explain which kinds of graph fall into each of these categories. Also, a recent result in [43] gives a related characterisation of cover time concentration for

a class of graphs that includes transitive ones in terms of the spectral gaps of the graphs in question.) In the case when concentration occurs, the first order behaviour of the cover time can be understood from its expectation, which was shown in the seminal work of [34] to be strongly connected to the expectation of the maximum of an associated Gaussian field; see also [33] for tighter estimates in the case of bounded degree graphs and trees (in the concentration regime). However, in ‘low-dimensional’ settings, it is easy to find examples where the cover time does not concentrate on its mean, and so its expectation is not sufficient to describe the asymptotic distribution to first order. Consider, for example, one-dimensional Brownian motion on the unit interval  $[0, 1]$  and the simple random walks on  $(n^{-1}\mathbb{Z}) \cap [0, 1]$  approximating this. Clearly the cover times of the graphs, when rescaled by  $n^{-2}$ , have a random limit, and it is a simple exercise to show that this can be described in terms of the cover time of the limiting diffusion. We will show a corresponding result when the limiting space is the Brownian CRT.

In order to state our main results, let us introduce some of the main objects of the subsequent discussion. In this introduction, we denote the Brownian CRT, equipped with its natural metric, by  $(\mathcal{T}, d_{\mathcal{T}})$  (see Section 3 for a precise definition of this space in terms of a Brownian excursion, including our choice of normalisation). We also write  $\rho$  for the root vertex of  $\mathcal{T}$ ,  $\mu_{\mathcal{T}}$  for its canonical measure, and  $\mathbf{P}$  for the probability measure on the underlying probability space upon which  $\mathcal{T}$  is built. Given a typical realisation of  $(\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}})$ , one has an associated, canonical Brownian motion  $((X_t^{\mathcal{T}})_{t \geq 0}, (P_x^{\mathcal{T}})_{x \in \mathcal{T}})$  (see Section 2 for the definition of such a process on a compact real tree more generally). Setting  $X_{[0,t]}^{\mathcal{T}} := \{X_s^{\mathcal{T}} : s \in [0, t]\}$  for the range of  $X^{\mathcal{T}}$  up to time  $t$ , the cover time of  $X^{\mathcal{T}}$  is defined as follows:

$$(1) \quad \tau_{\text{cov}} := \inf \left\{ t \geq 0 : X_{[0,t]}^{\mathcal{T}} = \mathcal{T} \right\};$$

as we will note in Section 2, this is a finite random variable, almost-surely. Additionally, the Brownian motion on the Brownian CRT is known to admit jointly continuous local times  $(L_t^{\mathcal{T}}(x))_{x \in \mathcal{T}, t \geq 0}$ , and a study of these will be central to our work. Whilst it is obviously the case for random walks on graphs that the cover time is the first time that the local times are strictly positive everywhere (possibly with a difference of one time-step in the case of discrete-time random walks), and the same property is readily checked for one-dimensional Brownian motion on  $[0, 1]$ , for example, it is by no means clear for diffusions on more general spaces, especially ones with a fractal structure such as the Brownian CRT. Yet, as was discussed in [28, Remark 7.4] (cf. [15]) and will be used in our proof below, checking this property of local times yields a potential route for establishing convergence of the cover times of the random walks approximating the relevant diffusion. Thus it is a key target of ours to prove the following result. Indeed, the majority of the article is devoted to its proof.

**THEOREM 1.1.** *For  $\mathbf{P}$ -a.e. realisation of  $\mathcal{T}$ , for every  $x \in \mathcal{T}$ , it  $P_x^{\mathcal{T}}$ -a.s. holds that*

$$(2) \quad \tau_{\text{cov}} = \inf \left\{ t \geq 0 : L_t^{\mathcal{T}}(y) > 0, \forall y \in \mathcal{T} \right\}.$$

Before proceeding to the application of the above theorem to scaling limits of cover times, let us briefly outline our strategy for verifying it. As is at the heart of [34], we will appeal to the strong connection between local times and Gaussian fields via a Ray-Knight theorem. More specifically, we give a version of the classical first Ray-Knight theorem in which we describe the distribution of the local time at the hitting time of a specified vertex in terms of certain squared Bessel (BESQ) processes (tree-indexed in our case), see Lemma 2.12. We will apply this at a certain approximation to the cover time that occurs (just after) the time  $X^{\mathcal{T}}$  covers a subtree  $\mathcal{T}^{(\varepsilon)} \subseteq \mathcal{T}$  with finitely many branches that approximates  $\mathcal{T}$  to within  $d_{\mathcal{T}}$ -Hausdorff distance  $\varepsilon$ . When  $\mathcal{T}$  is selected according to the Itô excursion measure and

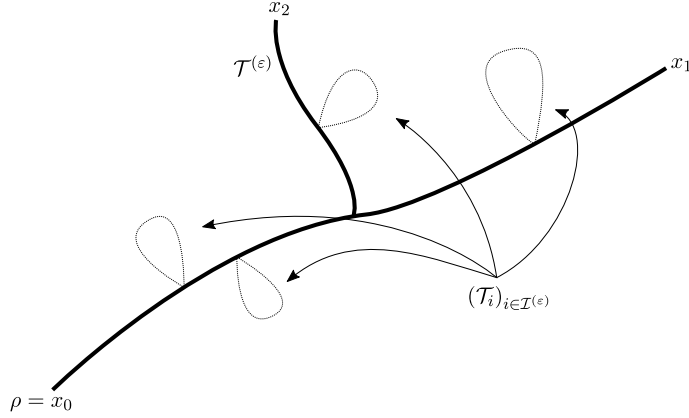


FIG 1. The above illustration shows the decomposition of the Brownian CRT that we will apply. The tree  $\mathcal{T}^{(\epsilon)}$ , shown with bold lines, spans a collection of leaves  $(x_i)_{i=1}^{M(\epsilon)}$  such that the Hausdorff distance between  $\mathcal{T}^{(\epsilon)}$  and  $\mathcal{T}$  is smaller than  $\epsilon$ . Moreover, our particular choice of leaves means that the closures of the connected components  $\mathcal{T} \setminus \mathcal{T}^{(\epsilon)}$ , which are denoted  $(\mathcal{T}_i)_{i \in \mathcal{I}(\epsilon)}$  are distributed as rescaled copies of the Brownian CRT. On  $\mathcal{T}^{(\epsilon)}$ , the local times are easily checked to take strictly positive values just after the cover time. To check the same is true on the remaining parts of  $\mathcal{T} \setminus \mathcal{T}^{(\epsilon)}$ , we use a Ray-Knight theorem, which enables us to understand the local times within these in terms of a tree-indexed zero-dimensional squared Bessel process, and thereby apply results from the literature for generalised Brownian snakes.

$\mathcal{T}^{(\epsilon)}$  is chosen according to a particular ‘height decomposition’ (see Section 3 for details), we have that the (closures of) the components of  $\mathcal{T} \setminus \mathcal{T}^{(\epsilon)}$  form a Poisson process of copies of rescaled Brownian CRTs, see Figure 1. Together with our Ray-Knight theorem, this implies that the local times of  $X^{\mathcal{T}}$  at our cover time approximation behave on these components as independent tree-indexed zero-dimensional BESQ processes. With this perspective, we are able to obtain the result of interest by appealing to a certain path property of BESQ-Brownian snakes, which we check by applying an argument that involves expressing the measures of certain events as solutions of differential equations; this part of the work is based on techniques developed by Le Gall in [53]. We emphasise this means that, in our approach, rather than sampling the entire tree, and then the processes  $X^{\mathcal{T}}$  and  $L^{\mathcal{T}}$  upon this, we first sample  $\mathcal{T}^{(\epsilon)}$ , then the local times upon this set at a particular stopping time, before finally sampling the remaining parts of the tree and local times together. In particular, we interleave the selection of the random environment and the stochastic process that has this as its state space.

We now come to our main convergence result. Although our motivation is principally to understand cover times of random walks, we will not need to restrict to such. Rather, we will work in a more general setting of stochastic processes associated with resistance forms, as developed in [29, 31, 47, 48, 61], for example. In particular, we consider objects of the form

$$(3) \quad \mathcal{K} = (K, R_K, \mu_K, \rho_K, P^K),$$

where  $(K, R_K)$  is a compact resistance metric space that satisfies a certain metric entropy condition (see Section 5.1 below for a definition of this and the other components that we now introduce),  $\mu_K$  is a finite Borel measure on  $K$  of full support,  $\rho_K$  is a marked point of  $K$ , and  $P^K$  is the law of the associated stochastic process  $X^K = (X_t^K)_{t \geq 0}$  and its jointly continuous local times  $L^K = (L_t^K(x))_{x \in K, t \geq 0}$ , where we will always assume that  $X^K$  is started from the root  $\rho_K$ . We note that the process  $X^K$  takes values in  $D(\mathbb{R}_+, K)$ , the space of càdlàg paths in  $K$ , equipped with the usual Skorohod  $J_1$  topology, and  $L^K$  takes values in  $C(K \times \mathbb{R}_+, \mathbb{R}_+)$ . A convenient topology for describing convergence of such objects was introduced in [61, 62],

and this will be recalled below in Section 5.1. Essentially the convergence  $\mathcal{K}_n \rightarrow \mathcal{K}$  means that all the spaces can be isometrically embedded into a common metric space in such a way that the sets converge with respect to Hausdorff distance, the measures converge weakly, the roots are identified, and the laws of the stochastic processes converge weakly, with the local times  $L^{K_n}$  close to those of  $L^K$  for nearby points (see Lemma 6.1 for details). Following [61], we write the collection of  $(K, R_K, \mu_K, \rho_K)$ , identified if they are root-preserving and measure-preserving isometric, as  $\check{\mathbb{F}}_c$ , and the space containing elements of the form  $\mathcal{K}$  as  $\mathbb{M}_{L,c}$ . (In fact,  $\mathbb{M}_{L,c}$  is a larger space that contains elements with fewer restrictions, but this is unimportant here.) In this setting, we have the following consequence of Theorem 1.1, where we write  $\tau_{\text{cov}}(K)$  for the cover time of a process  $X^K$ , defined analogously to (1).

**THEOREM 1.2.** *For  $\mathbf{P}$ -a.e. realisation of  $\mathcal{T}$ , if  $(K_n, R_{K_n}, \mu_{K_n}, \rho_{K_n}) \in \check{\mathbb{F}}_c$  are such that*

$$(4) \quad (K_n, R_{K_n}, \mu_{K_n}, \rho_{K_n}, P^{K_n}) \rightarrow (\mathcal{T}, 2d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho, P_{\rho}^{\mathcal{T}})$$

*in  $\mathbb{M}_{L,c}$ , and moreover, for each  $n$ , either  $K_n$  is at most a countable set or  $X^{K_n}$  is  $P^{K_n}$ -a.s. continuous, then the law of  $\tau_{\text{cov}}(K_n)$  under  $P^{K_n}$  converges to that of  $\tau_{\text{cov}}(\mathcal{T})$  under  $P_{\rho}^{\mathcal{T}}$ .*

For random spaces, it is also natural to deal with the annealed law of the cover time, i.e. the law averaged over the randomness of the space. Specifically, we write

$$(5) \quad \mathbb{P}^K(\tau_{\text{cov}}(K) \in \cdot) = \int P^K(\tau_{\text{cov}}(K) \in \cdot) \mathbf{P}(d(K, R_K, \mu_K, \rho_K)),$$

where we suppose that the underlying probability space for  $(K, R_K, \mu_K, \rho_K)$  is also denoted  $\mathbf{P}$ . We note that the measurability of the map  $(K, R_K, \mu_K, \rho_K) \mapsto \mathcal{K}$  was checked in [61, Proposition 6.1], which means the above integral is well-defined. From the previous result, it is straightforward to deduce the following.

**COROLLARY 1.3.** *If  $(K_n, R_{K_n}, \mu_{K_n}, \rho_{K_n})$  are random elements of  $\check{\mathbb{F}}_c$  such that (4) holds in distribution, and moreover, for each  $n$ , either  $K_n$  is at most a countable set or  $X^{K_n}$  is  $P^{K_n}$ -a.s. continuous,  $\mathbf{P}$ -a.s., then the law of  $\tau_{\text{cov}}(K_n)$  under  $\mathbb{P}^{K_n}$  converges to that of  $\tau_{\text{cov}}(\mathcal{T})$  under its annealed law  $\mathbb{P}^{\mathcal{T}}$  (defined similarly to (5), with the measure within the integrand being given by  $P_{\rho}^{\mathcal{T}}$ ).*

Whilst the statements of Theorem 1.2 and Corollary 1.3 are broad, they are maybe a little abstract for those not familiar with resistance forms. So, let us outline how they apply in the case of random walks on graphs. In particular, suppose  $K$  is the vertex set of a connected, finite graph containing no loops or multiple edges. Equip the graph with conductances  $(c(x, y))_{x, y \in K}$ , i.e.  $c(x, y)$  are non-negative numbers such that  $c(x, y) = c(y, x)$  and  $c(x, y) > 0$  if and only if  $\{x, y\}$  is an edge of the graph. Write  $R_K = (R_K(x, y))_{x, y \in K}$  for the associated effective resistance (see, for example, [56, Section 9.4] for a definition). Moreover, let  $\mu_K$  be a finite measure on  $K$  placing a non-zero mass on each vertex. In this case,  $(K, R_K, \mu_K, \rho_K) \in \check{\mathbb{F}}_c$  for any vertex  $\rho_K \in K$ , and the associated stochastic process  $(X_t^K)_{t \geq 0}$  is the continuous time Markov chain with generator given by

$$\Delta_K f(x) := \frac{1}{\mu_K(\{x\})} \sum_{y \in K} c(x, y) (f(y) - f(x)),$$

for  $f : K \rightarrow \mathbb{R}$ . One standard choice for  $\mu_K$  is the counting measure, i.e.  $\mu_K(\{x\}) = 1$  for all  $x \in K$ , in which case  $X^K$  is called the *variable-speed random walk (VSRW)*. Another is the conductance measure, i.e.  $\mu_K(\{x\}) = \sum_{y \in K} c(x, y)$  for all  $x \in K$ , in which case the process is called the *constant-speed random walk (CSRW)* since it has unit mean holding times. The

latter property means the CSRW closely resembles the discrete-time simple random walk with transition probabilities given by  $P(x, y) = c(x, y) / \sum_{z \in K} c(x, z)$ . Although we will not discuss the discrete-time process to any great extent later, where we have the conclusion of Theorem 1.2 or Corollary 1.3 for the CSRW, we will typically be able to transfer this to the discrete-time random walk with no problem. (For moments of the cover time, one should be slightly careful in moving between discrete and continuous time, see Section 7.) As for the local times of  $X^K$ , these are given by

$$L_t^K(x) = \frac{1}{\mu_K(\{x\})} \int_0^t \mathbf{1}_{\{X_s^K=x\}} ds.$$

In examples, we will generally discuss a sequence of graphs  $(K_n)_{n \geq 1}$  for which the effective resistance and measure have been rescaled. That is, writing  $P^{K_n}$  for the law of  $X^{K_n}$  started from  $\rho_{K_n}$ , we are able to check

$$(6) \quad \left( K_n, a_n^{-1} R_{K_n}, b_n^{-1} \mu_{K_n}, \rho_{K_n}, P^{K_n} \left( \left( \left( X_{ta_n b_n}^{K_n} \right)_{t \geq 0}, \left( a_n^{-1} L_{ta_n b_n}^{K_n}(x) \right)_{x \in K_n, t \geq 0} \right) \in \cdot \right) \right)$$

converges to the right-hand side of (4), either for a typical realisation of the Brownian CRT (as in Theorem 1.2) or in distribution (as in Corollary 1.3), where  $(a_n)_{n \geq 1}, (b_n)_{n \geq 1}$  are some divergent sequences of non-negative numbers. We then obtain that

$$(7) \quad (a_n b_n)^{-1} \tau_{\text{cov}}(K_n) \rightarrow \tau_{\text{cov}}(\mathcal{T})$$

in the appropriate sense. Various examples for which the convergence of objects as at (6) can be checked are presented in [61] (see also [15, 63]), and these include the following. We stress that, although these two examples are graph trees, our result should be applicable beyond such. See Problem 2 below for details of some non-tree models for which we expect similar results to hold.

**Example 1:** Critical, finite variance Galton-Watson trees. Let  $T_n$  be the tree generated by a Galton-Watson process whose offspring distribution is non-trivial, critical (mean one) and has finite variance  $\sigma^2$ , conditioned to have  $n$  vertices. (One may consider only those  $n$  for which the probability of the latter event is strictly positive.) Let  $c_{T_n}(x, y) = 1$  if  $\{x, y\}$  is a parent-child bond in  $T_n$ , and set  $c_{T_n}(x, y) = 0$  otherwise. Also, let  $\mu_{T_n}^V$  be the counting measure on  $T_n$ , and  $\mu_{T_n}^C$  be the conductance (or degree) measure. Since the spaces are trees, one has the associated effective resistance  $R_n$  is the usual shortest path graph distance on  $T_n$ , and we have from [15, 61] that the convergence described at (6) holds in distribution, with scaling factors given by

$$a_n := \frac{\sqrt{n}}{2\sigma},$$

$$b_n^V := n, \quad b_n^C := 2n,$$

where the superscripts  $V$  and  $C$  distinguish between the cases of the variable- and constant-speed random walks. (The convergence of the random walks was initially established in [25].) Moreover, we highlight that in results concerning the scaling of graph trees to the Brownian CRT, the most common measure considered is the counting measure. However, for a tree with unit conductances, it is easy to check the discrepancy between this and half the conductance measure is small, and so there is no problem in checking the result for both the VSRW and the CSRW.) Thus we have that (7) holds in distribution under the relevant annealed laws. In particular, for the CSRW (or discrete-time random walk) on  $T_n$ , the time scaling factor is given by  $a_n b_n^C = \sigma^{-1} n^{3/2}$ .

**Example 2:** Uniform spanning trees on high-dimensional tori. Consider the uniform spanning tree  $U_n$  on a  $d$ -dimensional torus  $(\mathbb{Z}/n\mathbb{Z})^d$  with  $d \geq 5$ . Equip edges of the graph with unit conductances, and suppose  $\mu_{U_n}$  is either the counting measure or the conductance measure on the vertices of  $U_n$ . Based on the work of [17], it is then shown in [61] that the convergence at (6) holds in distribution, with scaling factors given by

$$a_n := \alpha(d)n^{d/2},$$

$$b_n^V := n^d, \quad b_n^C := 2n^d,$$

where  $\alpha(d)$  is some dimension-dependent constant and again the superscripts  $V$  and  $C$  distinguish between the cases of the variable- and constant-speed random walks. (In [17, 61], only the counting measure was considered, but, similarly to the comment of the previous example, it is straightforward to replace this with the conductance measure, and thereby obtain the result for both the VSRW and the CSRW.) In particular, for the CSRW (or discrete-time random walk) on  $T_n$ , the time scaling factor is given by  $a_n b_n^C = 2\alpha(d)n^{3d/2}$ . We note that [17, 61] also consider a more general model of a ‘high-dimensional’ uniform spanning tree and our results would cover this; we restrict to the case of tori simply for brevity.

Whilst the above discussion covers the main contribution of this paper, we continue this introduction by presenting a couple of further consequences of our results that we believe are also of interest. Firstly, for any  $p \geq 1$ , we check the uniform  $L^p$ -integrability of the cover times of random walks on critical Galton-Watson trees whose offspring distribution admits finite exponential moments. In conjunction with our earlier results, this allows us to conclude that the  $p$ th moments of the cover times of the discrete models converge, with again the limit being expressed in terms of the cover time of Brownian motion on the Brownian CRT. As part of the result we note that the latter quantity also admits finite moments of all orders. (Our argument could also be used to yield a corresponding result for certain stretched exponential moments, see Remark 7.2 below.) Note that we write  $\mathbb{E}^K$  for the expectation under  $\mathbb{P}^K$  (and similarly for other spaces).

**COROLLARY 1.4.** *As in Example 1 above, let  $T_n$  be the tree generated by a Galton-Watson process whose offspring distribution is non-trivial and critical (mean one), conditioned to have size  $n$ . Additionally, suppose that the offspring distribution  $(p_k)_{k \geq 0}$  has finite exponential moments of some order, i.e.  $\sum_{k \geq 0} e^{\lambda k} p_k < \infty$  for some  $\lambda > 0$ . Writing  $\sigma^2$  for the variance of the offspring distribution and  $\tau_{\text{cov}}(T_n)$  for the cover time of the discrete-time or constant-speed random walk on  $T_n$  started from the root, it then holds that, for all  $p \geq 1$ ,*

$$\sigma^p n^{-3p/2} \mathbb{E}^{T_n} (\tau_{\text{cov}}(T_n)^p) \rightarrow \mathbb{E}^{\mathcal{T}} (\tau_{\text{cov}}(\mathcal{T})^p),$$

where the right-hand side above is finite. In particular, the cover time of Brownian motion on the Brownian CRT admits finite polynomial moments of all orders.

Secondly, again motivated by the work of Aldous in [10, 11], we consider the cover-and-return time. In particular, for Brownian motion on the Brownian CRT started from  $\rho$ , this is defined as

$$\tau_{\text{cov}}^+ := \inf \{ t \geq \tau_{\text{cov}} : X_t^{\mathcal{T}} = \rho \},$$

i.e. the first time  $X^{\mathcal{T}}$  returns to its starting point  $\rho$  after the cover time  $\tau_{\text{cov}}$ . We define the corresponding stopping time similarly for other processes. It is not difficult to incorporate the additional time from  $\tau_{\text{cov}}$  to  $\tau_{\text{cov}}^+$  into our arguments and thereby obtain the following result.

**COROLLARY 1.5.** (a) Under the assumptions of Theorem 1.2, the law of  $\tau_{\text{cov}}^+(K_n)$  under  $P^{K_n}$  converges to that of  $\tau_{\text{cov}}^+(\mathcal{T})$  under  $P_\rho^\mathcal{T}$ .

(b) Under the assumptions of Corollary 1.3, the law of  $\tau_{\text{cov}}^+(K_n)$  under  $\mathbb{P}^{K_n}$  converges to that of  $\tau_{\text{cov}}^+(\mathcal{T})$  under its annealed law  $\mathbb{P}^\mathcal{T}$ .

(c) In the setting of Corollary 1.4,

$$\sigma^p n^{-3p/2} \mathbb{E}^{T_n} (\tau_{\text{cov}}^+(T_n)^p) \rightarrow \mathbb{E}^\mathcal{T} (\tau_{\text{cov}}^+(\mathcal{T})^p),$$

where the right-hand side above is finite. In particular, the cover-and-return time of Brownian motion on the Brownian CRT admits finite polynomial moments of all orders.

Finally, we set out some problems of varying difficulty that are left open by this work.

**Problem 1.** Whilst the argument we give for the Brownian CRT has some appealing aspects, we believe that it should be possible to generalise the conclusion of Theorem 1.1 significantly. Somewhat obvious modifications should allow us to cover tree-like spaces with similarly nice recursive decompositions, such as  $\alpha$ -stable trees (see [3]) or the scaling limit of critical Erdős-Rényi graphs (see [5, 6] for a recursive decomposition of this space and [20] for a study of the associated cover time asymptotics); indeed, essentially all our technical arguments for Brownian motion on real trees have been written in a general way so that they could easily be applied directly or adapted to such spaces. More ambitiously, though, it is not unreasonable to conjecture that the identity at (2) holds for any of the stochastic processes associated with elements of  $\check{\mathbb{F}}_c$ , such as nested fractals (as introduced in [58]) or the two-dimensional Sierpiński carpet. For all these spaces, the processes in question admit jointly continuous local times, and so it is harder to conceive that they admit exceptional points, as would be needed for the conclusion of Theorem 1.1 to fail. At the moment, however, we have no idea of a strategy to tackle this issue.

**Problem 2.** Other examples for which a Brownian CRT scaling limit is known include the high-dimensional critical branching random walk [21, 22] and lattice trees [23], as well as the critical percolation cluster on the random hyperbolic half-planar triangulation [16]. To apply Corollary 1.3 in these cases, it will suffice to verify the metric entropy condition of [61], as described in Remark 6.2 below. At least for the branching random walk model, this should be an easy consequence of the corresponding result for Galton-Watson trees. Another model for which we would also expect the result to be applicable is the incipient infinite cluster of percolation on the integer lattice in high dimensions. However, deriving the scaling limit in this case is still an open question. (See [42] for rigorous work in this direction, and [29] for a detailed conjecture on the resistance scaling of this model.)

**Problem 3.** We do not believe exponential moments are essential for the conclusion of Corollary 1.4 to be applicable, rather we expect that a finite variance assumption is sufficient. The one part of the argument for which we apply the exponential moment assumption is in deducing a suitable quantitative bound for the metric cover size of the graphs in question (see the comment above Lemma 7.6 for further detail). As with the bounds in [7], however, one might hope that second moments of the offspring distribution are enough to yield exponential tail estimates for the distribution of the metric cover size associated with the conditioned Galton-Watson tree.

**Problem 4.** Concerning cover-and-return times, in [10], Aldous conjectured that

$$\mathbb{E} (\tau_{\text{cov}}^+(\mathcal{T})) = 6\sqrt{2\pi}.$$

This assertion was based on a consideration of the asymptotics of cover-and-return times for random walks on conditioned Poisson(1) trees, as studied in [11]. (We note that such trees can also be interpreted as uniform random trees on  $n$  labelled vertices  $\{1, 2, \dots, n\}$ ,

with root vertex 1, see [11, Lemma 12] or [50, Section 2.2], for example.) In partial support of this conjecture, Corollaries 1.4 and 1.5 confirm that indeed the cover-and-return times of the relevant discrete models converge to the corresponding quantity for the Brownian CRT. To complete the proof of the above identity, one also needs to make rigorous various heuristic arguments in [11]. We believe that the local time scaling limits now known to hold in this setting will contribute to this program, but it seems more detailed estimates will be needed to complete it.

The remainder of the article is organised as follows. In Section 2, we derive basic results for Brownian motions on compact real trees, including a version of the first Ray-Knight theorem for trees. It is here that we reduce the problem of proving Theorem 1.1 to a technical condition (see Assumption 2) that will later be checked using BESQ-Brownian snakes. Following this, our recursive height decomposition of the Brownian CRT is described in Section 3, and then the desired property of BESQ-Brownian snakes is checked in Section 4. Putting these pieces together, we prove Theorem 1.1 in Section 5. Finally, we derive the convergence results of Theorem 1.2 and Corollary 1.3 in Section 6, the moment convergence statement of Corollary 1.4 in Section 7, and our result on cover-and-return times, i.e. Corollary 1.5, in Section 8. We note that, in various estimates, we will write  $C$  and  $c$  for constants whose particular values are unimportant, and these might change from line to line.

**2. Brownian motion on compact real trees.** The aim of this section is to develop the fundamental results concerning Brownian motions on compact real trees that will provide the basis for the proof of Theorem 1.1. In particular, after introducing the process of interest in Section 2.1 and setting out an assumption that ensures the existence of jointly continuous local times (see Assumption 1 and Lemma 2.1), in Section 2.2, we give a version of Theorem 1.1 (see Proposition 2.7) under an additional assumption (see Assumption 2). The latter assumption is rather technical, involving a specific property of the local times at a certain approximation of the cover time. Towards checking this, in Section 2.3, we give a Ray-Knight theorem for local times that holds at the relevant approximation of the cover time. Together with a key property of a BESQ-Brownian snake that we will check in Section 4, such a Ray-Knight theorem will enable us to check Assumption 2 in our setting, and thereby establish Theorem 1.1.

*2.1. Definition and basic properties.* We start by introducing the setting of this section. Since the construction and basic properties of Brownian motions on real trees are by now rather well understood, we will be brief with the details, and refer the interested reader to [18] for comprehensive background. (See also [46] for earlier work on this topic, and [10, 26, 51] for an introduction to Brownian motion on the Brownian CRT in particular.) Specifically, for now, we assume that  $(\mathcal{T}, d_{\mathcal{T}})$  is a fixed compact real tree with root (distinguished vertex)  $\rho$ . We also suppose that  $\mu_{\mathcal{T}}$  is a Borel probability measure on  $\mathcal{T}$  of full support. As is described in [18, 46] (see [18, Theorem 1] in particular), there is a canonical local, regular Dirichlet form  $(\mathcal{E}^{\mathcal{T}}, \mathcal{F}^{\mathcal{T}})$  on  $L^2(\mathcal{T}, \mu_{\mathcal{T}})$  that satisfies

$$(8) \quad 2d_{\mathcal{T}}(x, y) = (\inf \{ \mathcal{E}^{\mathcal{T}}(f, f) : f \in \mathcal{F}^{\mathcal{T}}, f(x) = 0, f(y) = 1 \})^{-1}, \quad \forall x, y \in \mathcal{T}, x \neq y,$$

and the  $\mu_{\mathcal{T}}$ -symmetric strong Markov process associated with this through the standard correspondence (of [38, Theorem 7.2.1], for example), which we will denote by  $((X_t^{\mathcal{T}})_{t \geq 0}, (P_x^{\mathcal{T}})_{x \in \mathcal{T}})$ , has a law that is uniquely defined from every starting point. Moreover,  $X^{\mathcal{T}}$  has continuous sample paths,  $P_x^{\mathcal{T}}$ -a.s. for every  $x \in \mathcal{T}$ . This process is called the Brownian motion associated with  $(\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}})$ . (In various works, the constant 2 is not included in the characterisation of the form at (8), but we choose to do so here to align the

definition better with standard one-dimensional Brownian motion, and also the convention in works such as [10, 18, 19].) Since points have positive capacity in this setting (see [18, Lemma 3.4]), it is known that  $X^\mathcal{T}$  admits jointly measurable local times  $(L_t^\mathcal{T}(x))_{x \in \mathcal{T}, t \geq 0}$  that satisfy the occupation density formula, i.e.

$$(9) \quad \int_0^t f(X_s^\mathcal{T}) ds = \int_{\mathcal{T}} f(y) L_t^\mathcal{T}(y) \mu_{\mathcal{T}}(dy),$$

for all continuous  $f : \mathcal{T} \rightarrow \mathbb{R}$  and  $t \geq 0$ ,  $P_x^\mathcal{T}$ -a.s. for every  $x \in \mathcal{T}$ , see [39, Section 2]. (Precisely, when we say that  $X^\mathcal{T}$  or  $L^\mathcal{T}$  admits a property  $P_x^\mathcal{T}$ -a.s., we mean that there exists a version of the relevant process that admits the property in question.)

We finish this subsection by giving a simple sufficient condition for the local time process  $L^\mathcal{T}$  to admit a jointly continuous version. In particular, we will generally suppose the following, which can be verified for  $\mathbf{P}$ -a.e. realisation of the Brownian CRT (equipped with its canonical measure) by applying the volume estimates of [26, Theorem 1.2], for example.

ASSUMPTION 1. There exist constants  $C$  and  $\eta$  such that

$$\inf_{x \in \mathcal{T}} \mu_{\mathcal{T}}(B_{\mathcal{T}}(x, r)) \geq Cr^\eta, \quad \forall r \in (0, 1),$$

where  $B_{\mathcal{T}}(x, r) := \{y \in \mathcal{T} : d_{\mathcal{T}}(x, y) < r\}$ .

LEMMA 2.1. *Suppose Assumption 1 holds. It is then the case that the local times  $(L_t^\mathcal{T}(x))_{x \in \mathcal{T}, t \geq 0}$  are continuous,  $P_y^\mathcal{T}$ -a.s. for every  $y \in \mathcal{T}$ .*

PROOF. See [27, Lemma 2.2]. □

REMARK 2.2. A weaker sufficient condition for the continuity of local times is the metric entropy condition of [61], as described at (40). In fact, for the results of Section 2.2, we only need local time continuity, and so it would be sufficient to make this itself the assumption. However, for our Ray-Knight theorem of Section 2.3, we require continuity of squared Bessel processes on various subtrees, and Assumption 1 is convenient for checking this.

2.2. *Approximations for the cover time.* In order to prove the main result of this section (Proposition 2.7), we will consider a sequence of approximations to  $\mathcal{T}$  with finitely many branches. The precise choice of approximations that we will introduce for the Brownian CRT will be presented in Section 3. For now, we merely need to suppose  $(x_i)_{i \geq 0}$  is a dense sequence of points in  $\mathcal{T}$ . It will be consistent with our decomposition for the Brownian CRT and result in no loss of generality to assume that  $x_0 = \rho$  and that the elements of the sequence  $(x_i)_{i \geq 0}$  are distinct. We also let  $M = M(\varepsilon)$ ,  $\varepsilon > 0$ , be natural numbers that are non-increasing as a function of  $\varepsilon$  and diverge as  $\varepsilon \rightarrow 0$ . For each  $\varepsilon > 0$ , we define

$$(10) \quad \mathcal{T}^{(\varepsilon)} := \bigcup_{i=1}^{M(\varepsilon)} [[\rho, x_i]]_{\mathcal{T}},$$

where for points  $x, y \in \mathcal{T}$ , we write  $[[x, y]]_{\mathcal{T}}$  for the geodesic arc between  $x$  and  $y$ . We equip  $\mathcal{T}^{(\varepsilon)}$  with the restriction of  $d_{\mathcal{T}}$  to  $\mathcal{T}^{(\varepsilon)}$  (we will not introduce special notation for this restriction), so that  $(\mathcal{T}^{(\varepsilon)}, d_{\mathcal{T}})$  is a compact real tree with finitely many branches. The natural projection from  $\mathcal{T}$  to  $\mathcal{T}^{(\varepsilon)}$  that maps  $x \in \mathcal{T}$  to the (unique) closest point in  $\mathcal{T}^{(\varepsilon)}$  will be written  $\pi^{(\varepsilon)}$ . It is straightforward to check that

$$d_{\mathcal{T}}^H(\mathcal{T}^{(\varepsilon)}, \mathcal{T}) = \sup_{x \in \mathcal{T}} d_{\mathcal{T}}(x, \pi^{(\varepsilon)}(x)) \rightarrow 0,$$

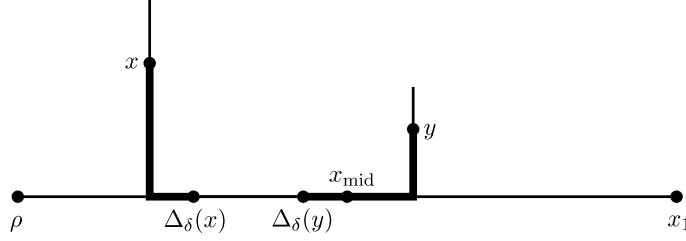


FIG 2. An illustration of how the map  $\Delta_\delta$  operates. The bold arcs represent  $[[x, \Delta_\delta(x)]]_{\mathcal{T}}$  and  $[[y, \Delta_\delta(y)]]_{\mathcal{T}}$  for points  $x$  and  $y$  on opposite sides of the point  $x_{\text{mid}}$ , which is the midpoint of  $[[\rho, x_1]]_{\mathcal{T}}$ .

as  $\varepsilon \rightarrow 0$ , where  $d_{\mathcal{T}}^H$  is the Hausdorff distance between compact subsets of  $\mathcal{T}$ .

We are now able to proceed with our discussion of cover times. The principal quantity of interest will be  $\tau_{\text{cov}}$ , defined from  $X^{\mathcal{T}}$  as at (1). To show that  $\tau_{\text{cov}}$  can be written in terms of local times, as per the conclusion of Theorem 1.1, we will introduce various approximations of  $\tau_{\text{cov}}$ . Firstly, for  $\varepsilon > 0$ , we set

$$\tau_{\text{cov}}^{(\varepsilon)} := \inf \left\{ t \geq 0 : X_{[0,t]}^{\mathcal{T}} \supseteq \mathcal{T}^{(\varepsilon)} \right\},$$

i.e.  $\tau_{\text{cov}}^{(\varepsilon)}$  is the first time that  $X^{\mathcal{T}}$  has visited every point of  $\mathcal{T}^{(\varepsilon)}$ . We will actually run the process  $X^{\mathcal{T}}$  for a slightly longer time than this, until it hits a specified point in  $\mathcal{T}^{(\varepsilon)}$  at a distance  $\delta$  away from  $X_{\tau_{\text{cov}}^{(\varepsilon)}}^{\mathcal{T}}$ . To this end, for each  $\delta \in (0, d_{\mathcal{T}}(\rho, x_1)/2)$ , we define a map  $\Delta_\delta : \mathcal{T} \rightarrow \mathcal{T}$  such that, writing  $b^{\mathcal{T}}(x, y, z)$  for the branch point of  $x, y, z \in \mathcal{T}$  (in particular, this is the unique point that lies on each of the arcs  $[[x, y]]_{\mathcal{T}}$ ,  $[[y, z]]_{\mathcal{T}}$  and  $[[z, x]]_{\mathcal{T}}$ ):

- if  $d_{\mathcal{T}}(\rho, b^{\mathcal{T}}(\rho, x, x_1)) \geq d_{\mathcal{T}}(\rho, x_1)/2$ , then  $\Delta_\delta(x)$  is the point on  $[[\rho, x]]_{\mathcal{T}}$  at a distance  $\delta$  from  $x$ ;
- if  $d_{\mathcal{T}}(\rho, b^{\mathcal{T}}(\rho, x, x_1)) < d_{\mathcal{T}}(\rho, x_1)/2$ , then  $\Delta_\delta(x)$  is the point on  $[[x_1, x]]_{\mathcal{T}}$  at a distance  $\delta$  from  $x$ .

See Figure 2. We observe that, for each  $\delta$ , the map  $\Delta_\delta$  is continuous at every point of  $\mathcal{T} \setminus \{x_{\text{mid}}\}$ , where  $x_{\text{mid}}$  is the mid-point of  $[[\rho, x_1]]_{\mathcal{T}}$ , i.e. the unique point of  $\mathcal{T}$  such that

$$d_{\mathcal{T}}(\rho, x_{\text{mid}}) = d_{\mathcal{T}}(x_{\text{mid}}, x_1) = d_{\mathcal{T}}(\rho, x_1)/2.$$

Moreover, we highlight that  $\Delta_\delta(\mathcal{T}^{(\varepsilon)}) \subseteq \mathcal{T}^{(\varepsilon)}$ . With these preparations in place, we define, for  $\varepsilon > 0$  and  $\delta \in (0, d_{\mathcal{T}}(\rho, x_1)/2)$ ,

$$\tau_{\text{cov}}^{(\varepsilon)}(\delta) := \inf \left\{ t \geq \tau_{\text{cov}}^{(\varepsilon)} : X_t^{\mathcal{T}} = \Delta_\delta \left( X_{\tau_{\text{cov}}^{(\varepsilon)}}^{\mathcal{T}} \right) \right\}.$$

We similarly set

$$\tau_{\text{cov}}(\delta) := \inf \left\{ t \geq \tau_{\text{cov}} : X_t^{\mathcal{T}} = \Delta_\delta \left( X_{\tau_{\text{cov}}}^{\mathcal{T}} \right) \right\}.$$

Before continuing, we note [27, Lemma 2.3] yields that, under Assumption 1, the local times of  $X^{\mathcal{T}}$  diverge uniformly (i.e.  $\inf_{x \in \mathcal{T}} L_t^{\mathcal{T}}(x) \rightarrow \infty$  as  $t \rightarrow \infty$ ,  $P_\rho^{\mathcal{T}}$ -a.s.), which implies in turn that  $\tau_{\text{cov}}$  and  $\tau_{\text{cov}}^{(\varepsilon)}$  are finite,  $P_\rho^{\mathcal{T}}$ -a.s. Moreover, since  $X^{\mathcal{T}}$  satisfies the commute time identity, i.e.

$$(11) \quad E_x^{\mathcal{T}}(\tau_y) + E_y^{\mathcal{T}}(\tau_x) = 2d_{\mathcal{T}}(x, y), \quad \forall x, y \in \mathcal{T},$$

where  $E_x^{\mathcal{T}}$  is the expectation under  $P_x^{\mathcal{T}}$  and  $\tau_x := \inf \{ t \geq 0 : X_t^{\mathcal{T}} = x \}$  is the hitting time of  $x$  by  $X^{\mathcal{T}}$  (see [31, equation (2.17)], for example), it further holds that, under Assumption 1, both  $\tau_{\text{cov}}^{(\varepsilon)}(\delta)$  and  $\tau_{\text{cov}}(\delta)$  are also finite,  $P_\rho^{\mathcal{T}}$ -a.s.

With  $\tau_{\text{cov}}^{(\varepsilon)}(\delta)$  and  $\tau_{\text{cov}}(\delta)$  defined, we are now able to introduce our second main assumption. This excludes the possibility that, at time  $\tau_{\text{cov}}^{(\varepsilon)}(\delta)$ , the local time  $L^{\mathcal{T}}$  takes the value zero on a small set.

ASSUMPTION 2. For each  $\varepsilon, 2\delta \in (0, d_{\mathcal{T}}(\rho, x_1)) \cap \mathbb{Q}$ ,

$$P_{\rho}^{\mathcal{T}}(E_1(\varepsilon, \delta) \cup E_2(\varepsilon, \delta)) = 1,$$

where

$$(12) \quad E_1(\varepsilon, \delta) := \left\{ L_{\tau_{\text{cov}}^{(\varepsilon)}(\delta)}^{\mathcal{T}}(x) > 0, \forall x \in \mathcal{T} \right\},$$

$$(13) \quad E_2(\varepsilon, \delta) := \left\{ \exists x \in \mathcal{T}, r > 0 : L_{\tau_{\text{cov}}^{(\varepsilon)}(\delta)}^{\mathcal{T}}(y) = 0, \forall y \in B_{\mathcal{T}}(x, r) \right\}.$$

In order to explain how we apply this assumption, which we will do in Proposition 2.7, we need several preparatory lemmas (none of which require Assumption 2 itself). The first of these lemmas establishes that  $\tau_{\text{cov}}(\delta)$  is indeed an approximation of  $\tau_{\text{cov}}$ . This is a general result in the sense that we need not make use of either of our general assumptions.

LEMMA 2.3. *It  $P_{\rho}^{\mathcal{T}}$ -a.s. holds that*

$$\tau_{\text{cov}}(\delta) \rightarrow \tau_{\text{cov}},$$

as  $\delta \rightarrow 0$ .

PROOF. Clearly  $\tau_{\text{cov}}(\delta)$  decreases monotonically as  $\delta \rightarrow 0$ ,  $P_{\rho}^{\mathcal{T}}$ -a.s., and so it will suffice to show the desired convergence statement holds in  $P_{\rho}^{\mathcal{T}}$ -probability. Conditioning on the value of  $X_{\tau_{\text{cov}}}^{\mathcal{T}}$ , from the strong Markov property we obtain: for any  $\eta > 0$ ,

$$P_{\rho}^{\mathcal{T}}(|\tau_{\text{cov}}(\delta) - \tau_{\text{cov}}| > \eta) \leq \sup_{x \in \mathcal{T}} P_x^{\mathcal{T}}(\tau_{\Delta_{\delta}(x)} > \eta).$$

Now, by applying Markov's inequality and the commute time identity (11), we have that the right-hand side here is bounded above by  $2\delta\eta^{-1}$ . Hence, it converges to zero as  $\delta \rightarrow 0$ . This is enough to complete the proof.  $\square$

Next, we rule out the possibility that  $X^{\mathcal{T}}$  is located at the point  $x_{\text{mid}}$ , which we recall is the midpoint of  $[[\rho, x_1]]_{\mathcal{T}}$ , at time  $\tau_{\text{cov}}$ .

LEMMA 2.4. *If Assumption 1 holds, then*

$$P_{\rho}^{\mathcal{T}}(X_{\tau_{\text{cov}}}^{\mathcal{T}} = x_{\text{mid}}) = 0.$$

PROOF. We may assume that, for each fixed point  $x \in \mathcal{T}$ , the local time  $(L_t^{\mathcal{T}}(x))_{t \geq 0}$  is a continuous additive functional satisfying  $P_x^{\mathcal{T}}$ -a.s.,

$$(14) \quad \inf \{t \geq 0 : L_t^{\mathcal{T}}(x) > 0\} = 0,$$

see [59, Section 3.6], for example. It is also straightforward to check from (9) and the continuity of the various processes that,  $P_{\rho}^{\mathcal{T}}$ -a.s.,  $L_{\tau_x}(x) = 0$ . Hence, by applying the strong Markov property at time  $\tau_{x_{\text{mid}}}$ , we deduce that,  $P_{\rho}^{\mathcal{T}}$ -a.s.,

$$(15) \quad \inf \{t \geq 0 : L_t^{\mathcal{T}}(x_{\text{mid}}) > 0\} = \tau_{x_{\text{mid}}}.$$

Applying the continuity of  $X^\mathcal{T}$ , it moreover  $P_\rho^\mathcal{T}$ -a.s. holds that  $\tau_{x_{\text{mid}}} < \tau_{x_1} \leq \tau_{\text{cov}}$ . Hence, writing  $\tau$  for  $(\tau_{x_{\text{mid}}} + \tau_{\text{cov}})/2$ , we can conclude that,  $P_\rho^\mathcal{T}$ -a.s.,  $L_\tau^\mathcal{T}(x_{\text{mid}}) > 0$ . Since the local times are jointly continuous, it follows that,  $P_\rho^\mathcal{T}$ -a.s., for some  $r > 0$ ,

$$L_\tau^\mathcal{T}(x) > 0, \quad \forall x \in B_{\mathcal{T}}(x_{\text{mid}}, r).$$

From this, it is an elementary application of the occupation density formula (9) to deduce that,  $P_\rho^\mathcal{T}$ -a.s., for some  $r > 0$ ,  $\tau_x \leq \tau$  for all  $x \in B_{\mathcal{T}}(x_{\text{mid}}, r)$ . The result is a straightforward consequence of this (and the continuity of  $X^\mathcal{T}$ ).  $\square$

We now present a useful property concerning the uniformity of the distribution of the time to hit  $\Delta_\delta(x)$  from  $x$ . Again, this requires neither Assumption 1 nor 2.

LEMMA 2.5. *For each  $\delta \in (0, d_{\mathcal{T}}(\rho, x_1)/2)$ , it holds that*

$$\sup_{x \in \mathcal{T}} P_x^\mathcal{T}(\eta > \tau_{\Delta_\delta(x)}) \rightarrow 0,$$

as  $\eta \rightarrow 0$ .

PROOF. We first claim that if  $x_n \rightarrow x$  and  $y_n \rightarrow y$  in  $(\mathcal{T}, d_{\mathcal{T}})$ , then

$$(16) \quad P_{x_n}^\mathcal{T}(\tau_{y_n} \in \cdot) \rightarrow P_x^\mathcal{T}(\tau_y \in \cdot)$$

weakly as probability measures on  $\mathbb{R}$ . To prove this, note that, under  $P_{x_n}^\mathcal{T}$ , it holds that  $\tau_{y_n} \leq \tau_1^n + \tau_2^n + \tau_3^n$ , where  $\tau_1^n$  is the time taken by  $X^\mathcal{T}$  to hit  $x$ ,  $\tau_2^n$  is the time taken from then to hit  $y$  and  $\tau_3^n$  is the subsequent time to hit  $y_n$ . Clearly  $\tau_2^n$  under  $P_{x_n}^\mathcal{T}$  has the distribution of  $\tau_y$  under  $P_x^\mathcal{T}$ . Moreover, applying (11), we find that, for any  $\eta > 0$ ,

$$P_{x_n}^\mathcal{T}(\tau_1^n + \tau_3^n \geq \eta) \leq P_{x_n}^\mathcal{T}(\tau_x \geq \eta/2) + P_{y_n}^\mathcal{T}(\tau_{y_n} \geq \eta/2) \leq \frac{4(d_{\mathcal{T}}(x_n, x) + d_{\mathcal{T}}(y_n, y))}{\eta} \rightarrow 0,$$

as  $n \rightarrow \infty$ . Hence we obtain that, for any  $\eta > 0$ ,

$$\limsup_{n \rightarrow \infty} P_{x_n}^\mathcal{T}(\tau_{y_n} \geq \eta) \leq P_x^\mathcal{T}(\tau_y \geq \eta).$$

By considering a similar upper bound for  $\tau_y$  under  $P_x^\mathcal{T}$ , we are also able to deduce that, for any  $\eta > 0$ ,

$$\liminf_{n \rightarrow \infty} P_{x_n}^\mathcal{T}(\tau_{y_n} > \eta) \geq P_x^\mathcal{T}(\tau_y > \eta).$$

Combining the two previous displays yields the claim.

Now, for each  $\eta > 0$ , define  $\phi_\eta(u) := \max\{\min\{1, 2 - u/\eta\}, 0\}$  for  $u \geq 0$ ; this is a continuous bounded function that strictly dominates  $\mathbf{1}_{\{u < \eta\}}$ . Hence to prove the lemma, it will suffice to show that

$$\sup_{x \in \mathcal{T}} E_x^\mathcal{T}(\phi_\eta(\tau_{\Delta_\delta(x)})) \rightarrow 0,$$

as  $\eta \rightarrow 0$ . To simplify notation, we define  $f_\eta(x) := E_x^\mathcal{T}(\phi_\eta(\tau_{\Delta_\delta(x)}))$ . Since for any  $x \in \mathcal{T}$ ,  $\tau_{\Delta_\delta(x)}$  is strictly positive,  $P_x^\mathcal{T}$ -a.s., we have from the monotone convergence theorem that  $f_\eta$  decreases monotonically to zero pointwise as  $\eta \rightarrow 0$ . To extend this to uniform convergence, we will appeal to Dini's theorem, which requires continuity. Whilst we can not apply this on the whole space  $\mathcal{T}$ , we can do on certain subsets. In particular, define

$$A := \{x \in \mathcal{T} : d_{\mathcal{T}}(\rho, b^\mathcal{T}(\rho, x, x_1)) \geq d_{\mathcal{T}}(\rho, x_1)/2\}.$$

By construction, the function  $\Delta_\delta$  is continuous on  $A$ . Thus from the claim that we proved in the first paragraph of the proof, i.e. (16), it follows that, for each  $\eta$ ,  $f_\eta$  is also continuous on  $A$ . Since  $A$  is clearly a compact subset of  $\mathcal{T}$ , we obtain as a consequence that

$$\sup_{x \in A} f_\eta(x) \rightarrow 0,$$

as  $\eta \rightarrow 0$ . Similarly, set

$$\tilde{A} := \{x \in \mathcal{T} : d_{\mathcal{T}}(\rho, b^{\mathcal{T}}(\rho, x, x_1)) \leq d_{\mathcal{T}}(\rho, x_1)/2\},$$

and consider the function  $\tilde{\Delta}_\delta$  given by changing  $\geq$  and  $<$  to  $>$  and  $\leq$ , respectively, in the definition of  $\Delta_\delta$ . Note that  $\tilde{\Delta}_\delta$  is continuous on  $\tilde{A}$  and  $\tilde{\Delta}_\delta = \Delta_\delta$  on  $A^c$ . Thus, proceeding as in the argument for the set  $A$ , we find that

$$\sup_{x \in A^c} f_\eta(x) \leq \sup_{x \in \tilde{A}} E_x^{\mathcal{T}} \left( \phi_\eta \left( \tau_{\tilde{\Delta}_\delta}(x) \right) \right) \rightarrow 0,$$

as  $\eta \rightarrow 0$ . This completes the proof.  $\square$

Applying the previous two lemmas, we are able to prove the following result, which, in conjunction with Lemma 2.3, gives that  $\tau_{\text{cov}}^{(\varepsilon)}(\delta)$  is also a good approximation for  $\tau_{\text{cov}}$  when both  $\varepsilon$  and  $\delta$  are small.

LEMMA 2.6. *If Assumption 1 holds, then, for each  $\delta \in (0, d_{\mathcal{T}}(\rho, x_1)/2)$ ,*

$$\tau_{\text{cov}}^{(\varepsilon)}(\delta) \rightarrow \tau_{\text{cov}}(\delta)$$

in  $P_\rho^{\mathcal{T}}$ -probability as  $\varepsilon \rightarrow 0$ .

PROOF. Clearly  $\tau_{\text{cov}}^{(\varepsilon)} \leq \tau_{\text{cov}}$  for every  $\varepsilon > 0$ ,  $P_\rho^{\mathcal{T}}$ -a.s. Moreover,  $\tau_{\text{cov}}^{(\varepsilon)}$  monotonically increases as  $\varepsilon$  decreases, and so has a  $P_\rho^{\mathcal{T}}$ -a.s. limit as  $\varepsilon \rightarrow 0$ ,  $\tau_{\text{cov}}^{(0)}$ , say. Suppose  $\tau_{\text{cov}}^{(0)} < \tau_{\text{cov}}$ . If  $t \in [\tau_{\text{cov}}^{(0)}, \tau_{\text{cov}})$ , then  $X_{[0,t]}^{\mathcal{T}} \supseteq \cup_{\varepsilon > 0} \mathcal{T}^{(\varepsilon)}$ , and so, by the denseness of  $(x_i)_{i \geq 0}$  and the continuity of  $X^{\mathcal{T}}$ ,  $X_{[0,t]}^{\mathcal{T}} = \mathcal{T}$ , but this contradicts that  $t < \tau_{\text{cov}}$ . Hence, it must  $P_\rho^{\mathcal{T}}$ -a.s. hold that

$$(17) \quad \tau_{\text{cov}}^{(\varepsilon)} \rightarrow \tau_{\text{cov}},$$

as  $\varepsilon \rightarrow 0$ . Applying the continuity of  $X^{\mathcal{T}}$ , it is thus also the case that  $X_{\tau_{\text{cov}}^{(\varepsilon)}}^{\mathcal{T}} \rightarrow X_{\tau_{\text{cov}}}^{\mathcal{T}}$ , and moreover, from Lemma 2.4 and the continuity of  $\Delta_\delta$  on  $\mathcal{T} \setminus \{x_{\text{mid}}\}$ , we deduce that, for every  $\delta \in (0, d_{\mathcal{T}}(\rho, x_1)/2)$ ,  $P_\rho^{\mathcal{T}}$ -a.s.,

$$(18) \quad \Delta_\delta \left( X_{\tau_{\text{cov}}^{(\varepsilon)}}^{\mathcal{T}} \right) \rightarrow \Delta_\delta \left( X_{\tau_{\text{cov}}}^{\mathcal{T}} \right),$$

as  $\varepsilon \rightarrow 0$ .

We will next apply Lemma 2.5 and (17) to check that, for every  $\delta \in (0, d_{\mathcal{T}}(\rho, x_1)/2)$ ,

$$(19) \quad P_\rho^{\mathcal{T}} \left( \tau_{\text{cov}} \leq \tau_{\text{cov}}^{(\varepsilon)}(\delta) \right) \rightarrow 1,$$

as  $\varepsilon \rightarrow 0$ . In particular, we have that, for every  $\eta > 0$ ,

$$\begin{aligned} P_\rho^{\mathcal{T}} \left( \tau_{\text{cov}} > \tau_{\text{cov}}^{(\varepsilon)}(\delta) \right) &= P_\rho^{\mathcal{T}} \left( \tau_{\text{cov}} - \tau_{\text{cov}}^{(\varepsilon)} > \tau_{\text{cov}}^{(\varepsilon)}(\delta) - \tau_{\text{cov}}^{(\varepsilon)} \right) \\ &\leq P_\rho^{\mathcal{T}} \left( \tau_{\text{cov}} - \tau_{\text{cov}}^{(\varepsilon)} > \eta \right) + P_\rho^{\mathcal{T}} \left( \tau_{\text{cov}}^{(\varepsilon)}(\delta) - \tau_{\text{cov}}^{(\varepsilon)} < \eta \right). \end{aligned}$$

The first term here converges to zero as  $\varepsilon \rightarrow 0$  by (17). As for the second term, this is bounded above by  $\sup_{x \in \mathcal{T}} P_x^\mathcal{T}(\eta > \tau_{\Delta_\delta(x)})$ , which, by Lemma 2.5, can be made arbitrarily small by choosing  $\eta$  appropriately. This establishes (19), as desired.

Now, consider the following two stopping times:

$$\begin{aligned}\tau_1 &= \inf \left\{ t \geq \tau_{\text{cov}} : X_t^\mathcal{T} \in \left\{ \Delta_\delta \left( X_{\tau_{\text{cov}}}^\mathcal{T} \right), \Delta_\delta \left( X_{\tau_{\text{cov}}}^\mathcal{T} \right) \right\} \right\}, \\ \tau_2 &= \inf \left\{ t \geq \tau_{\text{cov}} : X_{[\tau_{\text{cov}}, t]}^\mathcal{T} \supseteq \left\{ \Delta_\delta \left( X_{\tau_{\text{cov}}}^\mathcal{T} \right), \Delta_\delta \left( X_{\tau_{\text{cov}}}^\mathcal{T} \right) \right\} \right\}.\end{aligned}$$

Clearly, on the event  $\{\tau_{\text{cov}} \leq \tau_{\text{cov}}^{(\varepsilon)}(\delta)\}$ , it holds that

$$\tau_2 - \tau_1 = \left| \tau_{\text{cov}}^{(\varepsilon)}(\delta) - \tau_{\text{cov}}(\delta) \right|.$$

From (19) and the conclusion of the previous paragraph, to complete the proof, it will consequently suffice to show that  $\tau_2 - \tau_1 \rightarrow 0$  as  $\varepsilon \rightarrow 0$  in  $P_\rho^\mathcal{T}$ -probability. To this end, we define

$$d(\varepsilon, \delta) = d_\mathcal{T} \left( \Delta_\delta \left( X_{\tau_{\text{cov}}}^\mathcal{T} \right), \Delta_\delta \left( X_{\tau_{\text{cov}}}^\mathcal{T} \right) \right),$$

and then apply the strong Markov property at  $\tau_1$  (noting that both  $\Delta_\delta(X_{\tau_{\text{cov}}}^\mathcal{T})$  and  $\Delta_\delta(X_{\tau_{\text{cov}}}^\mathcal{T})$  are  $(X_t^\mathcal{T})_{t=0}^{\tau_{\text{cov}}}$ -measurable) to deduce that, for every  $\eta > 0$ ,

$$P_\rho^\mathcal{T}(\tau_2 - \tau_1 > \eta) \leq E_\rho^\mathcal{T} \left( \sup_{\substack{x, y \in \mathcal{T}: \\ d_\mathcal{T}(x, y) \leq d(\varepsilon, \delta)}} P_x^\mathcal{T}(\tau_y > \eta) \right) \leq E_\rho^\mathcal{T}(\min\{1, 2\eta^{-1}d(\varepsilon, \delta)\}),$$

where the second inequality follows from (11). In conjunction with (18), this implies the result.  $\square$

Putting the above pieces together, we arrive at the main result of this subsection.

**PROPOSITION 2.7.** *If Assumptions 1 and 2 hold, then,  $P_\rho^\mathcal{T}$ -a.s.,*

$$\tau_{\text{cov}} = \inf \left\{ t \geq 0 : L_t^\mathcal{T}(x) > 0, \forall x \in \mathcal{T} \right\}.$$

**PROOF.** Fix rational  $\delta \in (0, d_\mathcal{T}(\rho, x_1)/2)$  and suppose there exists a rational sequence  $(\varepsilon_i)$  with  $\varepsilon_i \rightarrow 0$  such that

$$P_\rho^\mathcal{T}(E_1(\varepsilon_i, \delta) \text{ holds for finitely many } i) > 0.$$

Then, from Assumption 2, it follows that

$$(20) \quad P_\rho^\mathcal{T}(E_2(\varepsilon_i, \delta) \text{ holds for all large } i) > 0.$$

Clearly, on  $E_2(\varepsilon_i, \delta)$ , the occupation density formula (9) and the continuity of  $X^\mathcal{T}$  imply that

$$\tau_{\text{cov}} \geq \tau_{\text{cov}}^{(\varepsilon_i)}(\delta).$$

Moreover, by Lemma 2.6 and [44, Lemma 4.2], it is possible to find a (deterministic) subsequence  $(\varepsilon_{i_j})$  such that  $\tau_{\text{cov}}^{(\varepsilon_{i_j})}(\delta)$  converges to  $\tau_{\text{cov}}(\delta)$  as  $j \rightarrow \infty$ ,  $P_\rho^\mathcal{T}$ -a.s. Consequently, from (20), we obtain that

$$P_\rho^\mathcal{T}(\tau_{\text{cov}} \geq \tau_{\text{cov}}(\delta)) > 0.$$

However, from the continuity of  $X^\mathcal{T}$ , it is clear that  $\tau_{\text{cov}}(\delta) > \tau_{\text{cov}}$ ,  $P_\rho^\mathcal{T}$ -a.s., and so the above inequality is a contradiction. Thus we have established that, for any rational sequence  $(\varepsilon_i)$  with  $\varepsilon_i \rightarrow 0$ ,

$$P_\rho^\mathcal{T}(E_1(\varepsilon_i, \delta) \text{ holds for infinitely many } i) = 1.$$

Now, again appealing to Lemma 2.6 and [44, Lemma 4.2], we can find a (deterministic) sequence of rationals  $(\varepsilon_i)$  such that  $\tau_{\text{cov}}^{(\varepsilon_i)}(\delta)$  converges to  $\tau_{\text{cov}}(\delta)$  as  $i \rightarrow \infty$ ,  $P_\rho^\mathcal{T}$ -a.s. On the event that  $E_1(\varepsilon_i, \delta)$  holds, we clearly have that

$$\inf \{t \geq 0 : L_t^\mathcal{T}(x) > 0, \forall x \in \mathcal{T}\} \leq \tau_{\text{cov}}^{(\varepsilon_i)}(\delta).$$

Since this holds infinitely often along the sequence with probability one, it follows that,  $P_\rho^\mathcal{T}$ -a.s.,

$$\inf \{t \geq 0 : L_t^\mathcal{T}(x) > 0, \forall x \in \mathcal{T}\} \leq \limsup_{i \rightarrow \infty} \tau_{\text{cov}}^{(\varepsilon_i)}(\delta) = \tau_{\text{cov}}(\delta).$$

As  $\delta \in (0, d_{\mathcal{T}}(\rho, x_1)/2)$  can be chosen as an arbitrarily small rational number, it follows from this and Lemma 2.3 that,  $P_\rho^\mathcal{T}$ -a.s.,

$$(21) \quad \inf \{t \geq 0 : L_t^\mathcal{T}(x) > 0, \forall x \in \mathcal{T}\} \leq \tau_{\text{cov}}.$$

On the other hand, if  $t$  is such that  $L_t^\mathcal{T}(x) > 0$  for all  $x \in \mathcal{T}$ , then it readily follows from the occupation density formula (9) and the continuity of  $X^\mathcal{T}$  that  $t \geq \tau_{\text{cov}}$ . Hence, it  $P_\rho^\mathcal{T}$ -a.s. holds that

$$(22) \quad \inf \{t \geq 0 : L_t^\mathcal{T}(x) > 0, \forall x \in \mathcal{T}\} \geq \tau_{\text{cov}}.$$

Combining (21) and (22) yields the desired result.  $\square$

**2.3. A Ray-Knight theorem for trees.** The aim of this section is to give a probabilistic characterisation of the local time process  $L^\mathcal{T}$  at time  $\tau_{\text{cov}}^{(\varepsilon)}(\delta)$  that will facilitate the proof that Assumption 2 holds in our setting. The main result, see Proposition 2.9 below, is in the spirit of the classical Ray-Knight theorem, which relates the local times of one-dimensional Brownian motion at a hitting time and a BESQ process. (For background on Ray-Knight theorems, see [65, Section XI.2] or [59, Chapter 2], for example; of the results that appear there, the statement closest to ours is probably [59, Theorem 2.6.3]. We also note that a related version of the second Ray-Knight theorem on tree-like spaces is studied in [57].) To state our conclusion, we first need to introduce several additional objects: a certain Brownian motion on  $\mathcal{T}^{(\varepsilon)}$ , a description of the connected components of  $\mathcal{T} \setminus \mathcal{T}^{(\varepsilon)}$ , and the notion of a tree-indexed Bessel process.

Recall the definition of  $\mathcal{T}^{(\varepsilon)}$  from (10). Clearly, for each  $\varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))$ , this is a real tree of finite and strictly positive length. Hence the associated one-dimensional Hausdorff measure  $\lambda^{(\varepsilon)}$  is a finite Borel measure of full support on  $\mathcal{T}^{(\varepsilon)}$ . Consequently, similarly to the discussion at the start of Section 2.1, it is possible to define a canonical Brownian motion  $((X_t^{(\varepsilon)})_{t \geq 0}, (P_x^{(\varepsilon)})_{x \in \mathcal{T}^{(\varepsilon)}})$  from  $(\mathcal{T}^{(\varepsilon)}, d_{\mathcal{T}}, \lambda^{(\varepsilon)})$ . Since  $\lambda^{(\varepsilon)}$  is readily checked to satisfy Assumption 1, we may assume that  $X^{(\varepsilon)}$  admits jointly continuous local times  $(L_t^{(\varepsilon)}(x))_{x \in \mathcal{T}^{(\varepsilon)}, t \geq 0}$ . As one further piece of notation concerning the process  $X^{(\varepsilon)}$ , we define  $\tau(\varepsilon, \delta)$  in analogous way to how  $\tau_{\text{cov}}^{(\varepsilon)}(\delta)$  was defined from  $X^\mathcal{T}$ . In particular, if  $\tau_{\text{cov}}(X^{(\varepsilon)})$  is the cover time of  $X^{(\varepsilon)}$  (defined as at (1)), then

$$\tau(\varepsilon, \delta) := \inf \left\{ t \geq \tau_{\text{cov}}(X^{(\varepsilon)}) : X_t^{(\varepsilon)} = \Delta_\delta \left( X_{\tau_{\text{cov}}(X^{(\varepsilon)})}^{(\varepsilon)} \right) \right\}.$$

Note we require  $\delta \in (0, d_{\mathcal{T}}(\rho, x_1)/2)$  to ensure this is well-defined.

Next, we observe that because  $(\mathcal{T}, d_{\mathcal{T}})$  is a compact real tree, it has at most a countable number of branch points (see [37, Lemma 3.1], for example). This means that, for each  $\varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))$ , the set  $\mathcal{T} \setminus \mathcal{T}^{(\varepsilon)}$  consists of at most a countable number of connected components,  $(\mathcal{T}_i^o)_{i \in \mathcal{I}^{(\varepsilon)}}$  say. We write  $(\mathcal{T}_i)_{i \in \mathcal{I}^{(\varepsilon)}}$  for the closures of these components in  $\mathcal{T}$ , which means adding one additional point to each of the sets. Namely, for each  $i$ , there exists a  $\rho_i \in \mathcal{T}^{(\varepsilon)}$  such that  $\mathcal{T}_i \setminus \mathcal{T}_i^o = \{\rho_i\}$ . We then have that each  $(\mathcal{T}_i, d_{\mathcal{T}})$  is a compact real tree (again, we do not introduce specific notation for the restriction of  $d_{\mathcal{T}}$ ), and we will consider the root of each to be the relevant  $\rho_i$ .

As the final ingredient for stating Proposition 2.9, we need the notion of a tree-indexed (squared) Bessel process. To introduce this, we start by recalling that a BESQ process of dimension  $d \geq 0$ , started from  $z_0 \geq 0$ , can be characterised as the unique positive strong solution  $(\xi_t)_{t \geq 0}$  of the stochastic differential equation

$$\xi_t := z_0 + 2 \int_0^t \sqrt{\xi_s} dB_s + dt,$$

where  $(B_t)_{t \geq 0}$  is a standard one-dimensional Brownian motion. (See [59, Section 14.2] for a brief introduction to such processes with a focus on aspects that are useful for the study of local times, and also [65, Chapter XI] for a more in-depth discussion.) We will denote the law of such a process by  $\text{BESQ}^d(z_0)$ . Heuristically, a  $\mathcal{T}$ -indexed  $\text{BESQ}^d(z_0)$  process is a random function  $(\xi_x^{\mathcal{T}})_{x \in \mathcal{T}}$  such that, along each arc  $[[\rho, x]]_{\mathcal{T}}$ ,  $(\xi_y^{\mathcal{T}})_{y \in [[\rho, x]]_{\mathcal{T}}}$  is a one-dimensional  $\text{BESQ}^d(z_0)$  process (where the arc is parameterised using the distance from the root), and, if  $x$  and  $x'$  are two distinct vertices, then conditional on  $\xi_{b_{\mathcal{T}}(\rho, x, x')}$ , the values of  $(\xi_y^{\mathcal{T}})_{y \in [[b_{\mathcal{T}}(\rho, x, x'), x]]_{\mathcal{T}}}$  and  $(\xi_y^{\mathcal{T}})_{y \in [[b_{\mathcal{T}}(\rho, x, x'), x']]_{\mathcal{T}}}$  are independent. When the number of leaves in the tree is finite, such as in each  $\mathcal{T}^{(\varepsilon)}$ , there is no problem in making this definition rigorous by applying the Markov property for one-dimensional Bessel processes (see [65, Chapter XI]). However, for a general compact real tree, some care is needed. More generally, we will say that  $(\xi_x^{\mathcal{T}})_{x \in \mathcal{T}}$  is a  $\mathcal{T}$ -indexed  $\text{BESQ}^d(z_0)$  process if it is a continuous stochastic process such that, for each subtree  $\mathcal{T}' \subseteq \mathcal{T}$  that spans a finite set of vertices including the root,  $(\xi_x^{\mathcal{T}})_{x \in \mathcal{T}'}$  satisfies the properties just described. Following the work of Le Gall [53], we will later give a construction of such a process in our setting (i.e. where the underlying tree is the Brownian CRT) using the notion of a Brownian snake. For now, though, it is convenient to give a preliminary result that ensures certain 0-dimensional tree-indexed BESQ processes are well-defined.

**LEMMA 2.8.** *Suppose Assumption 1 holds. It is then possible to construct, for any  $z_0 \geq 0$ , a  $\mathcal{T}$ -indexed  $\text{BESQ}^0(z_0)$  process. The law of such a process is unique.*

**PROOF.** Since for  $z_0 = 0$ , the one-dimensional  $\text{BESQ}^0(z_0)$  process remains at zero (see [65, Theorem XI.1.5]), the generalisation to trees is trivial in this case. So, we fix  $z_0 > 0$ . The key input to the proof in this case will be the following estimate for a one-dimensional  $\text{BESQ}^0(z_0)$  process,  $(\xi_t)_{t \geq 0}$ : for any  $T \geq 0$ ,

$$(23) \quad \mathbf{P}(|\xi_s - \xi_t| > \varepsilon) \leq C \left(1 + \frac{T}{t-s}\right) \exp\left(-\frac{c\varepsilon}{\sqrt{t-s}}\right), \quad \forall \varepsilon > 0, 0 \leq s < t \leq T,$$

where the constants  $C$  and  $c$  depend only upon  $z_0$  and  $T$ , and here we write  $\mathbf{P}$  for the probability measure on the space upon which  $\xi$  is defined. To prove this, we use the fact that the  $\text{BESQ}^0(z_0)$  process can, up to the hitting time of 0, be written as a time-change of one-dimensional Brownian motion  $(B_t)_{t \geq 0}$  started from  $z_0$ , with speed measure given by  $\frac{1}{4x} dx$ , see [66, Section V.47.27 and V.48]. In particular, if

$$A_t := \begin{cases} \int_0^t \frac{1}{4B_s} ds, & \text{for } t < \tau_0^B, \\ \infty, & \text{for } t \geq \tau_0^B, \end{cases}$$

where  $\tau_0^B$  is the hitting time of 0 by  $B$ , and  $\alpha(t) := \inf\{s : A_s > t\}$  is its right-continuous inverse, then it is possible to construct a version of  $Z$  by setting  $\xi_t = B_{\alpha(t)}$ . Now, on the event that  $\sup_{s \leq T_0} B_s \leq \lambda$ , we clearly have that  $A_t - A_s \geq (t - s)/4\lambda$  for any  $0 \leq s < t \leq \min\{T_0, \tau_0^B\}$ . Hence  $\alpha(t) - \alpha(s) \leq 4\lambda(t - s)$  for any  $0 \leq s < t \leq T_0/4\lambda$ . Consequently, taking  $T_0 := 4\lambda T$ , we deduce that: for any  $0 \leq s < t \leq T$  and  $\lambda > z_0$ ,

$$\begin{aligned} & \mathbf{P}(|\xi_s - \xi_t| > \varepsilon) \\ & \leq \mathbf{P}\left(\sup_{s \leq T_0} B_s > \lambda\right) + \mathbf{P}\left(|B_{\alpha(s)} - B_{\alpha(t)}| > \varepsilon, \sup_{s \leq T_0} B_s \leq \lambda\right) \\ & \leq 2\mathbf{P}(B_{T_0} > \lambda) + \mathbf{P}\left(\sup_{q \leq 4\lambda s} \sup_{r \leq 4\lambda(t-s)} |B_{q+r} - B_q| > \varepsilon\right) \\ & \leq 2\mathbf{P}\left(B_1 - B_0 > \frac{\lambda - z_0}{2\sqrt{\lambda T}}\right) + \left(1 + \frac{T}{t-s}\right) \mathbf{P}\left(\sup_{r \leq 8\lambda(t-s)} |B_r - B_0| > \varepsilon\right) \\ & \leq 2\mathbf{P}\left(B_1 - B_0 > \frac{\lambda - z_0}{2\sqrt{\lambda T}}\right) + 4\left(1 + \frac{T}{t-s}\right) \mathbf{P}\left(B_1 - B_0 > \frac{\varepsilon}{\sqrt{8\lambda(t-s)}}\right), \end{aligned}$$

where we apply a union bound to deduce the third inequality, and the reflection principle to bound the terms involving the supremum of Brownian motion. Now, since  $B_1 - B_0$  is simply a standard normal random variable, we have that  $\mathbf{P}(B_1 - B_0 > x) \leq C e^{-x^2/2}$ , and so applying the above bound with  $\lambda = \varepsilon/\sqrt{t-s}$  gives the result at (23) when  $\varepsilon/\sqrt{t-s} > z_0$ . We can extend this to  $\varepsilon/\sqrt{t-s} \leq z_0$  by adjusting  $C$  as appropriate.

To apply (23) to deduce the existence of a  $\mathcal{T}$ -indexed BESQ $^0(z_0)$  process, we will apply a version of the Kolmogorov-Chentsov continuity theorem, as appears in [61, Section 3]. To this end, for each  $n \geq 0$ , let  $D_n''$  be a minimal  $2^{-n}$  covering of  $\mathcal{T}$  (i.e., a collection  $(x_i)_{i=1}^{N_n}$  of minimal size such that  $\cup_{i=1}^{N_n} B_{\mathcal{T}}(x_i, 2^{-n}) = \mathcal{T}$ ). From Assumption 1, it is an elementary exercise to check that  $N_n \leq C 2^{\eta n}$ , where  $\eta$  is the constant of that assumption. We also let  $D_n' = \cup_{m \leq n} D_m''$ , and then set  $D_n$  to be the union of  $D_n'$  and all those points at a distance that is an integer multiple of  $2^{-n}$  from the root on one of the arcs  $[[\rho, x]]_{\mathcal{T}}$  for some  $x \in D_n'$ . Note that  $D_n$  forms an increasing sequence and each has size bounded by  $C 2^{(1+\eta)n}$ . Finally, we set  $D = \cup_{n \geq 0} D_n$ , which is countable. If  $\mathcal{T}_n := \cup_{x \in D_n} [[\rho, x]]_{\mathcal{T}}$ , then it is straightforward to construct a  $\mathcal{T}_n$ -indexed BESQ $^0(z_0)$  process directly, by piecing together one-dimensional BESQ processes on each of the arcs. Moreover, by Kolmogorov's extension theorem, we can readily extend this construction to give a BESQ $^0(z_0)$  on  $\mathcal{T}_D := \cup_{x \in D} [[\rho, x]]_{\mathcal{T}}$ . Let us denote the latter process by  $(Z_x^{\mathcal{T}_D})_{x \in \mathcal{T}_D}$ . Now, since the diameter of  $\mathcal{T}$  is finite, we obtain from (23) with  $\varepsilon = (t-s)^{\frac{1}{2}-\delta}$  for some  $\delta > 0$  that:

$$\mathbf{P}\left(|\xi_x^{\mathcal{T}_D} - \xi_y^{\mathcal{T}_D}| > d_{\mathcal{T}}(x, y)^{\frac{1}{2}-\delta}\right) \leq C \exp\left(-c d_{\mathcal{T}}(x, y)^{-\delta}\right), \quad \forall x, y \in \mathcal{T}_D,$$

where the constants here depend only upon the diameter of  $\mathcal{T}$  and  $z_0$ . Hence we can apply [61, Corollary 3.2] to deduce that

$$\mathbf{P}\left(\sup_{\substack{x, y \in D: \\ d_{\mathcal{T}}(x, y) \leq 2^{-(n-1)}}} |\xi_x^{\mathcal{T}_D} - \xi_y^{\mathcal{T}_D}| > C 2^{-(\frac{1}{2}-\delta)n}\right) \leq C e^{-2^{\delta n}}.$$

Since the upper bound here is summable in  $n$ , we deduce from a Borel-Cantelli argument that  $(\xi_x^{\mathcal{T}_D})_{x \in D}$  is (almost-surely) uniformly continuous on  $D$ , and thus can be extended uniquely to a continuous function  $(\xi_x^{\mathcal{T}})_{x \in \mathcal{T}}$ . We further note that the inclusion of the additional points

in  $D_n$  as compared to  $D'_n$  means that  $D \cap [[\rho, x_i]]_{\mathcal{T}}$  is dense in  $[[\rho, x_i]]_{\mathcal{T}}$  for each  $x_i \in D$ , and from this it is straightforward to check that  $\xi_x^{\mathcal{T}} = \xi_x^{\mathcal{T}_D}$  on the whole of  $\mathcal{T}_D$ .

To complete the proof and obtain uniqueness of the law of  $(\xi_x^{\mathcal{T}})_{x \in \mathcal{T}}$ , it will suffice to show that for any subtree  $\mathcal{T}' \subseteq \mathcal{T}$  spanning a finite set of vertices including the root, the process  $(\xi_x^{\mathcal{T}})_{x \in \mathcal{T}'}$  is a  $\mathcal{T}'$ -indexed BESQ $^0(z_0)$  process. Suppose  $\mathcal{T}' = \cup_{i=1}^M [[\rho, x_i]]_{\mathcal{T}}$ . By the denseness of  $D$ , for each  $i$ , there exist sequences  $(x_i^n)_{n \geq 0}$  in  $D$  such that  $d_{\mathcal{T}}(x_i^n, x_i) \rightarrow 0$  as  $n \rightarrow \infty$ . It follows that,  $b^{\mathcal{T}}(\rho, x_i, x_i^n) \rightarrow 0$ . Now, from the construction of  $(\xi_x^{\mathcal{T}_D})_{x \in \mathcal{T}_D}$ , we obtain that the law of  $(\xi_x^{\mathcal{T}})_{x \in \mathcal{T}'}$ , where  $\mathcal{T}'_n := \cup_{i=1}^M [[\rho, b^{\mathcal{T}}(\rho, x_i, x_i^n)]]_{\mathcal{T}}$  is a  $\mathcal{T}'_n$ -indexed BESQ $^0(z_0)$  process, for each  $n$ . From this and the continuity of  $(\xi_x^{\mathcal{T}})_{x \in \mathcal{T}}$ , one can deduce the corresponding result holds for  $\mathcal{T}'$ , as required.  $\square$

Often we will apply Lemma 2.8 to subsets of  $\mathcal{T}$ . Specifically, for each  $x, y \in \mathcal{T}$ ,  $x \neq y$ , write  $(\mathcal{T}_i)_{i \in \mathcal{I}_{x,y}}$  for the closures of the connected components of  $\mathcal{T} \setminus [[x, y]]_{\mathcal{T}}$ , defined similarly to  $(\mathcal{T}_i)_{i \in \mathcal{I}(\varepsilon)}$ . If Assumption 1 holds for  $\mathcal{T}$ , then it also holds for the restriction of  $\mu_{\mathcal{T}}$  to each component  $\mathcal{T}_i$ ,  $i \in \mathcal{I}_{x,y}$  (with possibly different constants for each). Hence, for any  $z_0 \geq 0$ , it is possible to construct a  $\mathcal{T}_i$ -indexed BESQ $^0(z_0)$  process for any  $i \in \mathcal{I}_{x,y}$ . Obviously, the same result also holds for each component of  $(\mathcal{T}_i)_{i \in \mathcal{I}(\varepsilon)}$ .

We are now ready to state the main result of this section, which relates local times and certain BESQ processes.

**PROPOSITION 2.9.** *Suppose Assumption 1 holds, and that  $2\delta, \varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))$ . Under  $P_{\rho}^{\mathcal{T}}$ , the law of the local time process  $L^{\mathcal{T}}$  at time  $\tau_{\text{cov}}^{(\varepsilon)}(\delta)$  is characterised as follows.*

- (a) *The law of  $(L_{\tau_{\text{cov}}^{(\varepsilon)}(\delta)}^{\mathcal{T}}(x))_{x \in \mathcal{T}^{(\varepsilon)}}$  is equal to that of  $(L_{\tau_{(\varepsilon, \delta)}^{(\varepsilon)}}(x))_{x \in \mathcal{T}^{(\varepsilon)}}$  under  $P_{\rho}^{(\varepsilon)}$ .*
- (b) *Conditional on  $(L_{\tau_{\text{cov}}^{(\varepsilon)}(\delta)}^{\mathcal{T}}(x))_{x \in \mathcal{T}^{(\varepsilon)}}$ , the local times on the various components of  $\mathcal{T} \setminus \mathcal{T}^{(\varepsilon)}$ , that is,  $(L_{\tau_{\text{cov}}^{(\varepsilon)}(\delta)}^{\mathcal{T}}(x))_{x \in \mathcal{T}_i}$ ,  $i \in \mathcal{I}^{(\varepsilon)}$ , are independent, with each distributed as a  $\mathcal{T}_i$ -indexed 0-dimensional BESQ process started from  $L_{\tau_{\text{cov}}^{(\varepsilon)}(\delta)}^{\mathcal{T}}(\rho_i)$ .*

Towards checking part (a) of the above result, we will use the fact that  $X^{(\varepsilon)}$  can be represented as the time change of  $X^{\mathcal{T}}$ . In particular, we introduce a continuous additive functional  $A^{(\varepsilon)}$  by setting

$$A_t^{(\varepsilon)} := \int_{\mathcal{T}^{(\varepsilon)}} L_t^{\mathcal{T}}(x) \lambda^{(\varepsilon)}(dx),$$

and write

$$\alpha^{(\varepsilon)}(t) := \inf \left\{ s \geq 0 : A_s^{(\varepsilon)} > t \right\}$$

for its right-continuous inverse. We then have the following connection.

**LEMMA 2.10.** *Suppose Assumption 1 holds. For any  $\varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))$  and  $x \in \mathcal{T}^{(\varepsilon)}$ , it holds that the joint law of  $(X^{(\varepsilon)}, L^{(\varepsilon)})$  under  $P_x^{(\varepsilon)}$  is equal to that of*

$$\left( \left( X_{\alpha^{(\varepsilon)}(t)}^{\mathcal{T}} \right)_{t \geq 0}, \left( L_{\alpha^{(\varepsilon)}(t)}^{\mathcal{T}}(x) \right)_{x \in \mathcal{T}^{(\varepsilon)}, t \geq 0} \right)$$

under  $P_x^{\mathcal{T}}$ .

**PROOF.** The proof is identical to that of [26, Lemmas 2.6 and 3.4], which dealt with realisations of the Brownian CRT, but applies to more general compact real trees. The one

point that should be expanded upon is that the starting points of the two processes are the same, i.e., that  $X_{\alpha^{(\varepsilon)}(0)}^{\mathcal{T}} = x$  almost-surely under  $P_x^{\mathcal{T}}$ . For this, it is sufficient to show that  $\alpha^{(\varepsilon)}(0) = 0$  almost-surely under  $P_x^{\mathcal{T}}$ . However, it is an elementary exercise to check this fact by applying that  $\lambda^{(\varepsilon)}$  has full support, (14) holds and the local times  $L^{\mathcal{T}}$  are almost-surely jointly continuous.  $\square$

In what follows, we will often suppose that  $X^{(\varepsilon)}$  (and its local times) are coupled with  $X^{\mathcal{T}}$  (and its local times) in the way given by the previous result. Namely, we will suppose that both processes are built on the same probability space in such a way that we have equality between  $X_t^{(\varepsilon)}$  and  $X_{\alpha^{(\varepsilon)}(t)}^{\mathcal{T}}$  for all  $t \geq 0$  (and similarly for the associated local times), almost-surely. When this is the case, we will simply say the processes are coupled, and write the joint law as  $P_x^{\mathcal{T}}$  if the starting point is  $x \in \mathcal{T}^{(\varepsilon)}$ . Under such a coupling, we also have that various stopping times of interest for the processes align with respect to the time change. Recall  $\tau_x$  is the hitting time of  $x$  by  $X^{\mathcal{T}}$ . We will write  $\tau_x^{(\varepsilon)}$  for the corresponding quantity for  $X^{(\varepsilon)}$ .

LEMMA 2.11. *Suppose Assumption 1 holds, that  $2\delta, \varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))$ ,  $x, y \in \mathcal{T}^{(\varepsilon)}$ , and that  $X^{(\varepsilon)}$  and  $X^{\mathcal{T}}$  are coupled, each with starting point  $x$ . It then  $P_x^{\mathcal{T}}$ -a.s. holds that*

$$\tau_y = \alpha^{(\varepsilon)}\left(\tau_y^{(\varepsilon)}\right), \quad \tau_{\text{cov}}^{(\varepsilon)}(\delta) = \alpha^{(\varepsilon)}\left(\tau(\varepsilon, \delta)\right).$$

PROOF. Under the coupling, we  $P_x^{\mathcal{T}}$ -a.s. have that

$$X_{\alpha^{(\varepsilon)}(\tau_y^{(\varepsilon)})}^{\mathcal{T}} = X_{\tau_y^{(\varepsilon)}}^{(\varepsilon)} = y.$$

Hence  $\alpha^{(\varepsilon)}(\tau_y^{(\varepsilon)}) \geq \tau_y$ . On the other hand, from the version of (15) with  $x_{\text{mid}}$  replaced by  $y$  (and the continuity of the local times, as well as the full support of  $\lambda^{(\varepsilon)}$ ), we have that  $A_{\tau_y+t}^{(\varepsilon)} > A_{\tau_y}^{(\varepsilon)} = 0$  for all  $t > 0$ ,  $P_x^{\mathcal{T}}$ -a.s. Hence, it  $P_x^{\mathcal{T}}$ -a.s. holds that  $\alpha^{(\varepsilon)}\left(A_{\tau_y}^{(\varepsilon)}\right) = \tau_y$ , and so

$$X_{A_{\tau_y}^{(\varepsilon)}}^{(\varepsilon)} = X_{\alpha^{(\varepsilon)}(A_{\tau_y}^{(\varepsilon)})}^{\mathcal{T}} = X_{\tau_y}^{\mathcal{T}} = y,$$

which implies  $A_{\tau_y}^{(\varepsilon)} \geq \tau_y^{(\varepsilon)}$ , and in turn that  $\tau_y \geq \alpha^{(\varepsilon)}(\tau_y^{(\varepsilon)})$ . Thus we have completed the proof of the first equality.

To verify the second equality, we start by applying the first one to deduce that,  $P_x^{\mathcal{T}}$ -a.s.,

$$\tau_{\text{cov}}^{(\varepsilon)} = \max_{i=1, \dots, M(\varepsilon)} \tau_{x_i} = \max_{i=1, \dots, M(\varepsilon)} \alpha^{(\varepsilon)}\left(\tau_{x_i}^{(\varepsilon)}\right) = \alpha^{(\varepsilon)}\left(\max_{i=1, \dots, M(\varepsilon)} \tau_{x_i}^{(\varepsilon)}\right) = \alpha^{(\varepsilon)}\left(\tau_{\text{cov}}\left(X^{(\varepsilon)}\right)\right).$$

We then apply the strong Markov property at this time, together with another application of the first equality to deduce the result.  $\square$

With these preparations, we can prove the first part of Proposition 2.9.

PROOF OF PROPOSITION 2.9(A). Supposing that  $X^{(\varepsilon)}$  and  $X^{\mathcal{T}}$  are coupled, each with starting point  $\rho$ , we have from Lemma 2.11 that,  $P_{\rho}^{\mathcal{T}}$ -a.s.,

$$L_{\tau(\varepsilon, \delta)}^{(\varepsilon)}(x) = L_{\alpha^{(\varepsilon)}(\tau(\varepsilon, \delta))}^{\mathcal{T}}(x) = L_{\tau_{\text{cov}}^{(\varepsilon)}(\delta)}^{\mathcal{T}}(x), \quad \forall x \in \mathcal{T}^{(\varepsilon)}.$$

Hence the claim follows from Lemma 2.10.  $\square$

For the proof of part (b) of Proposition 2.9, the additional ingredients we need are the following Ray-Knight-type result at a single hitting time (Lemma 2.12) and certain additivity and Markov properties for tree-indexed BESQ processes (Lemma 2.13).

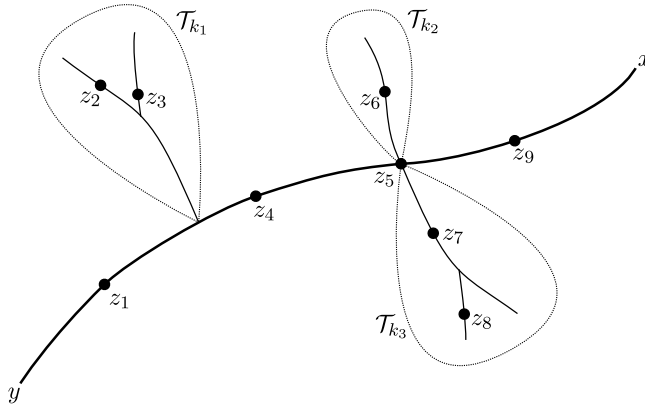


FIG 3. An example of the labelling of the points  $(z_i)_{i=1}^M$  and subtrees  $(\mathcal{T}_{k_n})_{n=1}^N$  from the proof of Lemma 2.12. The arc  $[[x, y]]_{\mathcal{T}}$  is shown in bold.

LEMMA 2.12. Suppose Assumption 1 holds, and that  $x, y \in \mathcal{T}$ ,  $x \neq y$ . Under  $P_x^{\mathcal{T}}$ , the law of the local time process  $L^{\mathcal{T}}$  at time  $\tau_y$  is characterised as follows.

- (a) The law of  $(L_{\tau_y}^{\mathcal{T}}(z))_{z \in [[x, y]]_{\mathcal{T}}}$  is given by a BESQ $^2(0)$  distribution, with time parameterised by distance from  $y$ .
- (b) Conditional on  $(L_{\tau_y}^{\mathcal{T}}(z))_{z \in [[x, y]]_{\mathcal{T}}}$ , the local times on the various components of  $\mathcal{T} \setminus [[x, y]]_{\mathcal{T}}$ , that is,  $(L_{\tau_y}^{\mathcal{T}}(z))_{z \in \mathcal{T}_i}$ ,  $i \in \mathcal{I}_{x, y}$ , are independent, with each distributed as a  $\mathcal{T}_i$ -indexed 0-dimensional BESQ process started from  $L_{\tau_y}^{\mathcal{T}}(\rho_i)$ .

PROOF. In this proof, we consider the Gaussian process  $(G_z)_{z \in \mathcal{T}}$  whose distribution is characterised by

$$\mathbf{E}(G_z) = 0,$$

$$\text{Cov}(G_w, G_z) = d_{\mathcal{T}}(y, b^{\mathcal{T}}(y, w, z)).$$

As per the discussion in [36, Section 6], it is possible to check that, under Assumption 1, this process has a continuous version. (Alternatively, one could follow the proof of Lemma 2.8, using a Gaussian estimate in place of (23).) We will show that

$$(24) \quad \left( L_{\tau_y}^{\mathcal{T}}(z) + (G_z - G_{b^{\mathcal{T}}(x, y, z)})^2 + (\bar{G}_z - \bar{G}_{b^{\mathcal{T}}(x, y, z)})^2 \right)_{z \in \mathcal{T}} \stackrel{(d)}{=} (G_z^2 + \bar{G}_z^2)_{z \in \mathcal{T}},$$

where  $(\bar{G}_z)_{z \in \mathcal{T}}$  is an independent copy of  $(G_z)_{z \in \mathcal{T}}$ , and both  $G$  and  $\bar{G}$  are independent of  $L_{\tau_y}^{\mathcal{T}}$ . From this, the result follows exactly as in the one-dimensional case; see the proof of [59, Theorem 2.6.3].

Since all the processes in question are continuous, it will suffice to check that the finite-dimensional distributions of the various objects align. In particular, for the remainder of the proof, we fix vertices  $(z_i)_{i=1}^M$  in  $\mathcal{T}$ . Without loss of generality, we will suppose these are distinct. Moreover, it will be convenient to suppose these are ordered so that  $d_{\mathcal{T}}(y, b^{\mathcal{T}}(x, y, z_i))$  is non-decreasing, and if  $d_{\mathcal{T}}(y, b^{\mathcal{T}}(x, y, z_i)) = d_{\mathcal{T}}(y, b^{\mathcal{T}}(x, y, z_j))$  with  $z_i \in [[x, y]]_{\mathcal{T}}$ , then  $i \leq j$  (i.e., vertices on the arc  $[[x, y]]_{\mathcal{T}}$  appear in the order before those in the subtrees that emanate from them). Moreover, vertices within the same subtree off  $[[x, y]]_{\mathcal{T}}$  should be consecutive in the sequence, though the order within these is unimportant. See Figure 3 for an example of this labelling. To understand the distribution of  $(L_{\tau_y}^{\mathcal{T}}(z_i))_{i=1}^M$ , we consider

its moment generating function. In particular, from [59, Lemma 3.10.2], we have that, for  $\lambda_1, \dots, \lambda_M$  sufficiently small,

$$(25) \quad E_x^T \left( \exp \left( \sum_{i=1}^M \lambda_i L_{\tau_y}^T(z_i) \right) \right) = \frac{\det(I - \hat{\Sigma}\Lambda)}{\det(I - \Sigma\Lambda)},$$

where  $I = (I_{i,j})_{i,j=1}^M$  is the  $M \times M$  identity matrix,  $\Lambda$  is the  $M \times M$  diagonal matrix with entries given by  $\Lambda_{i,j} = \lambda_i I_{i,j}$ ,  $\Sigma = (\Sigma_{i,j})_{i,j=1}^M$  is given by

$$\Sigma_{i,j} = 2d_{\mathcal{T}}(y, b^T(y, z_i, z_j)),$$

and  $\hat{\Sigma} = (\hat{\Sigma}_{i,j})_{i,j=1}^M$  is given by

$$\hat{\Sigma}_{i,j} = 2d_{\mathcal{T}}(y, b^T(y, z_i, z_j)) - 2d_{\mathcal{T}}(y, b^T(x, y, z_j)).$$

We highlight that, in [59, Lemma 3.10.2],  $\Sigma$  and  $\hat{\Sigma}$  are expressed in terms of a certain potential density, namely  $\Sigma_{i,j} = u_{\tau_y}(z_i, z_j)$  and  $\hat{\Sigma}_{i,j} = u_{\tau_y}(z_i, z_j) - u_{\tau_y}(x, z_j)$ , where  $u_{\tau_y}(z, \cdot)$  is the occupation density for  $X^T$ , started from  $z$  and run until hitting  $y$ . In our case, it is known that  $u_{\tau_y}(z, w) = 2d_{\mathcal{T}}(y, b^T(y, z, w))$ , see [18, Corollary 3.10] or [19, Proposition 2.12], for example. Moreover, [59, Lemma 3.10.2] requires a continuous  $\alpha$ -potential density, but this readily follows from the existence of a jointly continuous transition density for  $X^T$ ,  $(p_t^T(w, z))_{w,z \in \mathcal{T}, t > 0}$  say (see [48, Theorem 10.4], for example), which, under Assumption 1, satisfies  $p_t^T(z, z) \leq Ct^{-\eta/(1+\eta)}$  for  $t \in (0, 1)$  (as follows from the argument of [52, Proposition 4.1], for example; applying such ideas, for the Brownian CRT in particular, detailed transition density estimates were presented in [26]).

Thus it remains to show that the right-hand side of (25) can be expressed in terms of the Laplace transforms of the relevant Gaussian processes. Firstly, we observe that  $\text{Cov}(G_{z_i}, G_{z_j}) = \frac{1}{2}\Sigma_{i,j}$  by definition. Hence, writing  $g^T$  for the transpose of a vector, we can directly compute that

$$\begin{aligned} \mathbf{E} \left( \exp \left( \sum_{i=1}^M \lambda_i G_{z_i}^2 \right) \right) &= \frac{1}{(2\pi)^{M/2} \det(\frac{1}{2}\Sigma)^{1/2}} \int_{\mathbb{R}^M} \exp \left( g^T \Lambda g - \frac{1}{2} g^T \left( \frac{1}{2}\Sigma \right)^{-1} g \right) dg \\ &= \frac{1}{(2\pi)^{M/2} \det(\frac{1}{2}\Sigma)^{1/2}} \int_{\mathbb{R}^M} \exp \left( -\frac{1}{2} g^T (2\Sigma^{-1} - 2\Lambda) g \right) dg \\ &= \frac{\det \left( (2\Sigma^{-1} - 2\Lambda)^{-1} \right)^{1/2}}{\det(\frac{1}{2}\Sigma)^{1/2}} \\ &= \det(1 - \Sigma\Lambda)^{-1/2}, \end{aligned}$$

which of course implies

$$(26) \quad \mathbf{E} \left( \exp \left( \sum_{i=1}^M \lambda_i (G_{z_i}^2 + \bar{G}_{z_i}^2) \right) \right) = \det(1 - \Sigma\Lambda)^{-1}.$$

To handle the numerator of (25), we will consider the form of  $\hat{\Sigma}$  carefully. To do this, we first note that every  $z_i$  is either on the arc  $[[x, y]]_{\mathcal{T}}$  or in a set of the form  $\mathcal{T}_k \setminus \{\rho_k\}$ , where  $\mathcal{T}_k$  is one of the trees in the collection  $(\mathcal{T}_l)_{l \in \mathcal{I}_{x,y}}$ . Write  $(\mathcal{T}_{k_n})_{n=1}^N$  for the set of trees that contain such a vertex, ordered in a way that is consistent with the ordering of the points from



Thus, by following the calculation used to deduce (26), we find that

$$(28) \quad \mathbf{E} \left( \exp \left( \sum_{i=1}^M \lambda_i \left( (G_{z_i} - G_{b^{\mathcal{T}}(x,y,z_i)})^2 + (\bar{G}_{z_i} - \bar{G}_{b^{\mathcal{T}}(x,y,z_i)})^2 \right) \right) \right) = \det \left( 1 - \tilde{\Sigma} \Lambda \right)^{-1}.$$

Combining (25), (26), (27) and (28) yields (24), as desired.  $\square$

LEMMA 2.13. (a) Suppose  $(\xi_x^i)_{x \in \mathcal{T}}$ ,  $i = 1, 2$ , are independent  $\mathcal{T}$ -indexed BESQ $^{d_i}(z_i)$  processes. It is then the case that  $(\xi_x^1 + \xi_x^2)_{x \in \mathcal{T}}$  is a  $\mathcal{T}$ -indexed BESQ $^{d_1+d_2}(z_1 + z_2)$  process. (b) Suppose  $(\xi_x)_{x \in \mathcal{T}}$  is a  $\mathcal{T}$ -indexed BESQ $^d(z_0)$  process and  $\mathcal{T}' \subseteq \mathcal{T}$  is a subtree spanning a finite set of vertices including the root. Write  $(\mathcal{T}_i)_{i \in \mathcal{I}'}$  for the closures of the components of  $\mathcal{T} \setminus \mathcal{T}'$ . It is then the case that, conditional on  $(\xi_x)_{x \in \mathcal{T}'}$ ,  $(\xi_x)_{x \in \mathcal{T}_i}$ ,  $i \in \mathcal{I}'$ , are independent, with each distributed as a  $\mathcal{T}_i$ -indexed BESQ $^d(\xi_{\rho_i})$  process.

PROOF. Given the continuity of the processes in question, for both parts (a) and (b), it suffices to check the claims for the finite-dimensional distributions of the processes in question. This is straightforward from the construction of the tree-indexed BESQ processes and the analogous claims that hold for the one-dimensional process. In particular, for the additivity property, see [65, Theorem XI.1.2]. The Markovianity of the processes is also discussed in [65, Chapter XI].  $\square$

PROOF OF PROPOSITION 2.9(B). In this proof, we write

$$\tau_{(i)} := \max_{j=1, \dots, i} \tau_{x_j},$$

so that  $\tau_{\text{cov}}^{(\varepsilon)} = \tau_{(M(\varepsilon))}$ .

By Lemma 2.12, we have that, conditional on  $(L_{\tau_{(1)}}^{\mathcal{T}}(z))_{z \in [[\rho, x_1]]_{\mathcal{T}}}$ ,  $(L_{\tau_{(1)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}_i}$ ,  $i \in \mathcal{I}_{\rho, x_1}$ , are independent, with each distributed as a  $\mathcal{T}_i$ -indexed 0-dimensional BESQ process started from  $L_{\tau_{(1)}}^{\mathcal{T}}(\rho_i)$ . Using the Markov property for tree-indexed Bessel processes of Lemma 2.13(b), we obtain from this that, conditional on  $(L_{\tau_{(1)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ ,  $(L_{\tau_{(1)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}_i}$ ,  $i \in \mathcal{I}^{(\varepsilon)}$ , are independent, with each distributed as a  $\mathcal{T}_i$ -indexed 0-dimensional BESQ process started from  $L_{\tau_{(1)}}^{\mathcal{T}}(\rho_i)$ .

We proceed from the previous observation inductively. Suppose for some  $i < M(\varepsilon)$  we have that, conditional on  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ ,  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}_j}$ ,  $j \in \mathcal{I}^{(\varepsilon)}$ , are independent, with each distributed as a  $\mathcal{T}_j$ -indexed 0-dimensional BESQ process started from  $L_{\tau_{(i)}}^{\mathcal{T}}(\rho_j)$ . We have two situations to consider: either  $\tau_{x_{i+1}} < \tau_{(i)}$  or  $\tau_{x_{i+1}} > \tau_{(i)}$ . (The times  $\tau_{x_{i+1}}$  and  $\tau_{(i)}$  can not be equal, because we assumed the vertices  $(x_j)_{j \geq 0}$  to be distinct.) Moreover, we observe that, since it  $P_{\rho}^{\mathcal{T}}$ -a.s. holds that  $\tau_{x_{i+1}} = \inf\{t \geq 0 : L_t^{\mathcal{T}}(x_{i+1}) > 0\}$  (cf. (14)), we have that  $L_{\tau_{(i)}}^{\mathcal{T}}(x_{i+1}) > 0$  implies  $\tau_{x_{i+1}} < \tau_{(i)}$  and  $L_{\tau_{(i)}}^{\mathcal{T}}(x_{i+1}) = 0$  implies  $\tau_{x_{i+1}} > \tau_{(i)}$ . In particular, which of the events  $\{\tau_{x_{i+1}} < \tau_{(i)}\}$  or  $\{\tau_{x_{i+1}} > \tau_{(i)}\}$  occurs is determined by  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ . Similarly,  $X_{\tau_{(i)}}^{\mathcal{T}}$  is determined by  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ , being the unique one of the vertices  $x_1, \dots, x_i$  with zero local time at time  $\tau_{(i)}$ .

Now, conditional on  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ , if  $\tau_{x_{i+1}} > \tau_{(i)}$ , then

$$\left( L_{\tau_{(i+1)}}^{\mathcal{T}}(z) - L_{\tau_{(i)}}^{\mathcal{T}}(z) \right)_{z \in \mathcal{T}},$$

is distributed as the local time  $(L_{\tau_{x_{i+1}}}^{\mathcal{T}}(z))_{z \in \mathcal{T}}$  under  $P_{X_{\tau_{(i)}}^{\mathcal{T}}}^{\mathcal{T}}$ , independently of  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}_j}$ ,  $j \in \mathcal{I}^{(\varepsilon)}$ . In particular, applying Lemma 2.12 and the Markov property for tree-indexed

Bessel processes (i.e. Lemma 2.13(b)), conditional on  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$  and  $(L_{\tau_{(i+1)}}^{\mathcal{T}}(z) - L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ , if  $\tau_{x_{i+1}} > \tau_{(i)}$ , then

$$\left( L_{\tau_{(i+1)}}^{\mathcal{T}}(z) - L_{\tau_{(i)}}^{\mathcal{T}}(z) \right)_{z \in \mathcal{T}_j}, \quad j \in \mathcal{I}^{(\varepsilon)},$$

are independent, with each distributed as a  $\mathcal{T}_j$ -indexed 0-dimensional BESQ process started from  $L_{\tau_{(i+1)}}^{\mathcal{T}}(\rho_j) - L_{\tau_{(i)}}^{\mathcal{T}}(\rho_j)$  (again, independently of  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}_j}$ ,  $j \in \mathcal{I}^{(\varepsilon)}$ ). On the other hand, conditional on  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ , if  $\tau_{x_{i+1}} < \tau_{(i)}$ , then  $\tau_{(i+1)} = \tau_{(i)}$ , and so

$$L_{\tau_{(i+1)}}^{\mathcal{T}}(z) - L_{\tau_{(i)}}^{\mathcal{T}}(z) = 0, \quad \forall z \in \mathcal{T}.$$

Since the 0-dimensional BESQ process started from 0 remains at 0, then we can express this in an analogous way to the conclusion in the first case. Namely, conditional on  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$  and  $(L_{\tau_{(i+1)}}^{\mathcal{T}}(z) - L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ ,

$$\left( L_{\tau_{(i+1)}}^{\mathcal{T}}(z) - L_{\tau_{(i)}}^{\mathcal{T}}(z) \right)_{z \in \mathcal{T}_j}, \quad j \in \mathcal{I}^{(\varepsilon)},$$

are independent, with each distributed as a  $\mathcal{T}_j$ -indexed 0-dimensional BESQ process started from  $L_{\tau_{(i+1)}}^{\mathcal{T}}(\rho_j) - L_{\tau_{(i)}}^{\mathcal{T}}(\rho_j)$  (again, independently of  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}_j}$ ,  $j \in \mathcal{I}^{(\varepsilon)}$ ). Consequently, applying the inductive hypothesis and the additivity property of Lemma 2.13(a), we conclude that, conditional on  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$  and  $(L_{\tau_{(i+1)}}^{\mathcal{T}}(z) - L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ ,

$$\left( L_{\tau_{(i+1)}}^{\mathcal{T}}(z) \right)_{z \in \mathcal{T}_j}, \quad j \in \mathcal{I}^{(\varepsilon)},$$

are independent, with each distributed as a  $\mathcal{T}_j$ -indexed 0-dimensional BESQ process started from  $L_{\tau_{(i+1)}}^{\mathcal{T}}(\rho_j)$ . Since the law only depends on the pointwise sum of  $(L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$  and  $(L_{\tau_{(i+1)}}^{\mathcal{T}}(z) - L_{\tau_{(i)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ , it follows that we have the same result when we condition simply upon  $(L_{\tau_{(i+1)}}^{\mathcal{T}}(z))_{z \in \mathcal{T}^{(\varepsilon)}}$ .

The above argument gives the result corresponding to Proposition 2.9(b) at the time  $\tau_{\text{cov}}^{(\varepsilon)}$ . To get from here to the time  $\tau_{\text{cov}}^{(\varepsilon)}(\delta)$ , we need to add the local time accumulated between these two times, but this can be dealt with in an identical fashion to the inductive step of the proof, and so we omit the details. (In fact, the argument is easier, because we certainly have that  $\tau_{\text{cov}}^{(\varepsilon)}(\delta) > \tau_{\text{cov}}^{(\varepsilon)}$ , and so we only have one case to deal with.)  $\square$

Before concluding the section, we give one other simple consequence of the Ray-Knight theorem for hitting times.

**COROLLARY 2.14.** *Suppose Assumption 1 holds, and that  $2\delta, \varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))$ . It then  $P_{\rho}^{\mathcal{T}}$ -a.s. holds that*

$$\inf_{z \in \mathcal{T}^{(\varepsilon)}} L_{\tau_{\text{cov}}^{(\varepsilon)}(\delta)}^{\mathcal{T}}(z) > 0.$$

**PROOF.** Since a  $\text{BESQ}^2(0)$  process is simply the square of the modulus of a two-dimensional Brownian motion started from 0 (see [59, Section 14.2], for example), which is a process that almost-surely never returns to 0, it follows from Lemma 2.12(a) that, for each  $i$ ,  $P_{\rho}^{\mathcal{T}}$ -a.s.,

$$L_{\tau_{x_i}}^{\mathcal{T}}(z) > 0, \quad \forall z \in [[\rho, x_i]]_{\mathcal{T}} \setminus \{x_i\}.$$

As we also know that  $\tau_{x_i} := \inf\{t \geq 0 : L_t^{\mathcal{T}}(x_i) > 0\}$  (cf. (15)), it follows that

$$\tau_{x_i} := \inf\{t \geq 0 : L_t^{\mathcal{T}}(z) > 0, \forall z \in [[\rho, x_i]]_{\mathcal{T}}\}.$$

Putting the results together for each of the arcs forming  $\mathcal{T}^{(\varepsilon)}$ , we consequently obtain,  $P_{\rho}^{\mathcal{T}}$ -a.s.,

$$\tau_{\text{cov}}^{(\varepsilon)} := \inf\left\{t \geq 0 : L_t^{\mathcal{T}}(z) > 0, \forall z \in \mathcal{T}^{(\varepsilon)}\right\}.$$

Since  $\tau_{\text{cov}}^{(\varepsilon)}(\delta) > \tau_{\text{cov}}^{(\varepsilon)}$  and the local times are continuous and non-decreasing, the result follows.  $\square$

**3. Height decomposition for the Brownian CRT.** The aim of this section is to give an approximation procedure for the Brownian CRT by recursively adding branches of maximal height. This procedure relies on William's decomposition of Brownian excursions under the Itô measure [67]. The following description of this decomposition is based on [65, Chapter XII.4]. We note that there are other recursive decompositions of the Brownian CRT, such as the stick-breaking construction of [9, Process 3] or the self-similarity seen when splitting at the branch point of the root and two uniformly chosen vertices, as described in [14, Theorem 2] (see also [8, 30]), but we adopt the current approach due to the convenience of the description of the remainder of the tree in terms of a Poisson process of smaller copies of the Brownian CRT. See Proposition 3.2 below for details.

We denote by  $\mathbf{n}$  the Itô measure on excursions  $\zeta = (\zeta_s)_{0 \leq s \leq \sigma}$  normalised such that

$$(29) \quad \mathbf{n}\left(\sup_{s \geq 0} \zeta_s > \varepsilon\right) = \frac{1}{2\varepsilon} \quad \text{and} \quad \mathbf{n}(\sigma > \varepsilon) = \frac{1}{\sqrt{2\pi\varepsilon}}.$$

(See [65, Chapter XII] for background on the excursion measure of Brownian motion.) Each excursion in the support of  $\mathbf{n}$  corresponds to a compact real tree via contour functions. In particular, to define  $\mathcal{T} = \mathcal{T}_{\zeta}$ , we first introduce

$$d_{\zeta}(s, t) := \zeta(s) + \zeta(t) - 2 \min_{r \in [s \wedge t, s \vee t]} \zeta(r), \quad s, t \geq 0,$$

and then set  $\mathcal{T} := [0, \sigma]/\sim$ , where we suppose  $s \sim t$  if and only if  $d_{\zeta}(s, t) = 0$ . (Here and below, we use the notation  $s \wedge t := \min\{s, t\}$  and  $s \vee t := \max\{s, t\}$ .) We equip  $\mathcal{T}$  with the quotient metric induced by  $2d_{\zeta}$  (the scaling factor of 2 is included to align our definition with that of Aldous, see [13, Corollary 22]), the measure  $\mu_{\mathcal{T}}$  obtained as the push-forward of Lebesgue measure on  $[0, \sigma]$  by the quotient map, and a root  $\rho$ , as given by the equivalence class of 0 with respect to  $\sim$ . (See [54, Section 2] for an introduction to this coding procedure.) It is known that the push-forward map from  $(\zeta, \sigma)$  to  $(\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho)$  is continuous when the image space is equipped with the marked Gromov-Hausdorff-Prohorov topology (cf. [4, Proposition 3.3]), and thus the map  $\zeta \mapsto (\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho)$  is measurable. Hence,  $(\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho)$  is a well-defined random object, and we denote its law by  $\mathbf{N}$ . With a slight abuse of notation, the law  $\mathbf{P}$  of the Brownian CRT is given by

$$\mathbf{P} = \mathbf{N}(\cdot | \sigma = 1).$$

We remark that, from the scaling property of Brownian motion, it is possible to check that

$$(30) \quad \mathbf{N}((\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho) \in \cdot | \sigma = s) = \mathbf{P}\left(\left(\mathcal{T}, s^{1/2}d_{\mathcal{T}}, s\mu_{\mathcal{T}}, \rho\right) \in \cdot\right).$$

Since, by (29), it holds that  $\mathbf{N} = \int_0^{\infty} \frac{1}{2\sqrt{2\pi s^3}} \mathbf{N}(\cdot | \sigma = s) ds$ , if a property of  $(\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho)$  that does not depend on the particular scaling of the tree holds  $\mathbf{N}$ -a.s., then it also holds  $\mathbf{P}$ -a.s. (and vice versa).

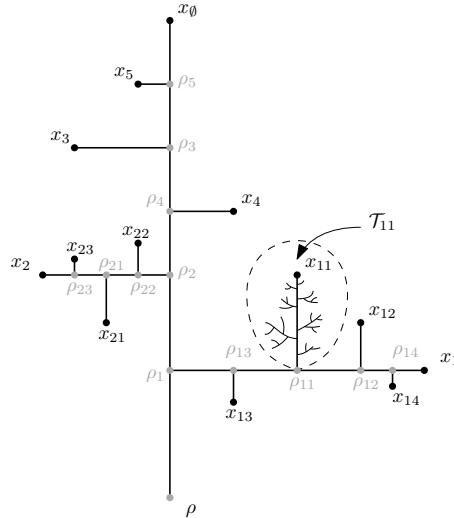


FIG 4. Height decomposition of the Brownian CRT. A subset of leaves  $x_i$  with  $|i| \leq 2$  is represented in black; the corresponding roots  $\rho_i$  are represented in gray.

The Brownian CRT  $\mathcal{T}$  has a unique vertex  $x_\emptyset$  of maximal height (i.e. distance from the root),  $\mathbf{P}$ -a.s. Indeed, this vertex corresponds to the maximum of a Brownian excursion, which is unique by William's decomposition, for example. (See the proof of Lemma 3.1 below for further details.) In the same spirit as in previous sections, we denote by  $(\mathcal{T}_i)_{i \geq 1}$  the subtrees of  $\mathcal{T}$  corresponding to the closures of the connected components of  $\mathcal{T} \setminus [[\rho, x_\emptyset]]_{\mathcal{T}}$  ordered by decreasing heights. (Our subsequent decomposition will make clear that, almost-surely, these heights are distinct.) For every  $i$ , we denote by  $\rho_i$  the root of  $\mathcal{T}_i$  (i.e. the point of its intersection with  $[[\rho, x_\emptyset]]_{\mathcal{T}}$ ), and by  $x_i$  its  $\mathbf{P}$ -a.s. unique vertex of maximal height (again, our decomposition will make clear such points are almost-surely well-defined). We will also denote the height of  $\mathcal{T}_i$ , which is the length of the segment  $[[\rho_i, x_i]]_{\mathcal{T}}$ , by  $H(\mathcal{T}_i)$ . We highlight that, unlike the assumption of the previous section, the points  $(x_i)_{i \geq 1}$  will not be dense in  $(\mathcal{T}, d_{\mathcal{T}})$ .

We can extend this procedure by induction; the following makes sense  $\mathbf{P}$ -a.s. As an index set, we introduce  $\Sigma_* = \bigcup_{n \geq 0} \Sigma_n$ , where  $\Sigma_0 := \{\emptyset\}$  and, for every  $n \geq 1$ ,  $\Sigma_n := \mathbb{N}^n$ . Suppose that subtrees  $\mathcal{T}_i$  of  $\mathcal{T}$  with respective roots  $\rho_i$  and vertices of maximal height  $x_i$  have been constructed for  $i \in \Sigma_*$  such that  $|i| \leq n$ . Fix  $i \in \Sigma_n$ . We define the trees  $(\mathcal{T}_{ij})_{j \geq 1}$  as the subtrees of  $\mathcal{T}_i$  corresponding to the closures of the connected components of  $\mathcal{T}_i \setminus [[\rho_i, x_i]]_{\mathcal{T}}$  ordered by decreasing height. Each  $\mathcal{T}_{ij}$  has root  $\rho_{ij} \in [[\rho_i, x_i]]_{\mathcal{T}}$  and we denote by  $x_{ij}$  its  $\mathbf{P}$ -a.s. unique vertex of maximal height. As before, we also denote the height of  $\mathcal{T}_{ij}$  by  $H(\mathcal{T}_{ij})$ . See Figure 4 for an illustration. The next lemma will ensure that, for any  $\varepsilon > 0$ , there will only be a finite number of arcs  $[[\rho_i, x_i]]_{\mathcal{T}}$ ,  $i \in \Sigma_*$ , with a length that exceeds  $\varepsilon$ .

LEMMA 3.1. *It  $\mathbf{P}$ -a.s. holds that*

$$\sup_{i \in \Sigma_n} H(\mathcal{T}_i) \xrightarrow{n \rightarrow \infty} 0.$$

PROOF. We start by presenting William's decomposition of the Brownian excursion [65, Chapter XII.4]. In the context of random trees, this description can also be found in [3, Theorem 3.3]. Denoting  $H(\mathcal{T})$  the height of a rooted real tree  $\mathcal{T}$ , our normalisation of the

Brownian excursion gives

$$\mathbf{N}(H > \varepsilon) = \mathbf{n} \left( \sup_{s \geq 0} \zeta_s > \varepsilon/2 \right) = \frac{1}{\varepsilon},$$

and, as follows from [3, Theorem 3.3],  $\mathbf{N}$ -a.e. realisation of  $\mathcal{T}$  has a unique vertex of height  $H$ . For  $h > 0$ , we set  $\mathbf{N}^{H=h}(\cdot) := \mathbf{N}(\cdot \mid H = h)$ , so that

$$\mathbf{N} = \int_0^\infty \frac{dh}{h^2} \mathbf{N}^{H=h}.$$

Fix  $h > 0$ . Under  $\mathbf{N}^{H=h}$ , a tree has a unique vertex  $x_h$  at height  $h$  almost-surely, and we denote the (possibly empty) subtrees grafted on the left-hand side and right-hand side of the segment  $[[\rho, x_h]]_{\mathcal{T}}$  at distance  $x$  from  $x_h$  and  $h - x$  from  $\rho$  by  $\overleftarrow{\mathcal{T}}_x^h$  and  $\overrightarrow{\mathcal{T}}_x^h$ . (The left- and right- sides of the tree are determined by the whether the part of the excursion  $\zeta$  that describes them appears before or after the unique index that corresponds to the vertex  $x_h$ .) Williams' decomposition states that, under  $\mathbf{N}^{H=h}$ , the collection  $(x, \overleftarrow{\mathcal{T}}_x^h, \overrightarrow{\mathcal{T}}_x^h)$  with  $x \in [0, h]$  and at least one of  $\overleftarrow{\mathcal{T}}_x^h$  and  $\overrightarrow{\mathcal{T}}_x^h$  being non-empty is countable, and that the point measure over such elements

$$\sum_{x \in [0, h]: \overleftarrow{\mathcal{T}}_x^h \cup \overrightarrow{\mathcal{T}}_x^h \neq \emptyset} \delta_{(x, \overleftarrow{\mathcal{T}}_x^h, \overrightarrow{\mathcal{T}}_x^h)}$$

is a Poisson random measure with intensity

$$(31) \quad \frac{1}{4} \mathbf{1}_{[0, h]}(x) dx \left( \delta_\emptyset(d\overleftarrow{\mathcal{T}}_x^h) \mathbf{N} \left( \mathbf{1}_{\{H \leq x\}} d\overrightarrow{\mathcal{T}}_x^h \right) + \delta_\emptyset(d\overrightarrow{\mathcal{T}}_x^h) \mathbf{N} \left( \mathbf{1}_{\{H \leq x\}} d\overleftarrow{\mathcal{T}}_x^h \right) \right).$$

(The factor  $\frac{1}{4}$  differs from that in [3] due to our choice of normalisation of the Brownian CRT.) As a consequence, conditional on their branching points  $x \in [0, h]$  and on their heights  $(h_x)_x$ , the atoms of the point process, the trees  $\overrightarrow{\mathcal{T}}_x^h$  or  $\overleftarrow{\mathcal{T}}_x^h$  are independent and have respective laws  $\mathbf{N}^{H=h_x}$ . It is straightforward to track this decomposition in our recursive construction of  $(\mathcal{T}_i)_{i \in \Sigma_*}$ . Indeed, for every  $n \geq 1$ , conditional on  $(h_i, \rho_i, x_i)_{|i| \leq n}$ , the random trees  $(\mathcal{T}_i)_{|i| \leq n}$  are independent and have respective laws  $\mathbf{N}^{H=h_i}$ .

Now fix  $\theta > 1$ . From the intensity measure (31), we calculate

$$\begin{aligned} \mathbf{N}^{H=h} \left( \sup_{x \in [0, h]} \left\{ H(\overrightarrow{\mathcal{T}}_x^h)^\theta \vee H(\overleftarrow{\mathcal{T}}_x^h)^\theta \right\} \right) &\leq \mathbf{N}^{H=h} \left( \sum_{x \in [0, h]} H(\overrightarrow{\mathcal{T}}_x^h)^\theta + H(\overleftarrow{\mathcal{T}}_x^h)^\theta \right) \\ &= \frac{1}{2} \int_0^h dx \mathbf{N} \left( H^\theta \mathbf{1}_{H \leq x} \right) \\ &= \frac{1}{2} \int_0^h dx \int_0^x \frac{u^\theta}{u^2} du \\ &= \frac{1}{2\theta(\theta - 1)} h^\theta. \end{aligned}$$

This immediately yields, for every  $i \in \Sigma_*$ :

$$\mathbf{N}^{H=h} \left( \sup_{j \geq 1} H(\mathcal{T}_{ij})^\theta \mid H(\mathcal{T}_i) \right) \leq \frac{1}{2\theta(\theta - 1)} H(\mathcal{T}_i)^\theta.$$

By induction, for any  $k \geq 1$ , we therefore have

$$\mathbf{N}^{H=h} \left( \sup_{i \in \Sigma_k} H(\mathcal{T}_i)^\theta \right) \leq \left( \frac{1}{2\theta(\theta - 1)} \right)^k h^\theta.$$

Applying Borel-Cantelli with  $\theta$  suitably large gives the result.  $\square$

We can now define our approximations  $\mathcal{T}^{(\varepsilon)}$  for  $\mathcal{T}$  that fit with the setting of Section 2. Supposing  $\mathcal{T}$  is chosen according to  $\mathbf{N}^{H=h}$ , for every  $\varepsilon > 0$ , we set

$$(32) \quad \mathcal{T}^{(\varepsilon)} := [[\rho, x_\emptyset]]_{\mathcal{T}} \cup \bigcup_{i \in \Sigma_* : h_i > \varepsilon} [[\rho_i, x_i]]_{\mathcal{T}}.$$

Lemma 3.1 ensures that, almost-surely, the trees  $\mathcal{T}^{(\varepsilon)}$  are finite unions of lines and that  $d_{\mathcal{T}}^H(\mathcal{T}^{(\varepsilon)}, \mathcal{T}) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . In addition, from our definition of the  $\mathcal{T}_i$ 's, the  $\mathcal{T}^{(\varepsilon)}$ 's are subtrees of  $\mathcal{T}$  and are nested when  $\varepsilon$  decreases. In particular, by relabelling, it is possible to suppose that the  $x_i$ 's that appear in the above description can be written as  $(x_i)_{i=0}^{M(\varepsilon)}$  for some sequences  $(x_i)_{i=0}^\infty$  and  $(M(\varepsilon))_{\varepsilon > 0}$  that satisfy the assumptions set out at the start of Section 2.2. Finally, we have the following Poisson description.

**PROPOSITION 3.2.** *Fix  $\varepsilon > 0$  and denote by  $\lambda^{(\varepsilon)}$  the one-dimensional Hausdorff measure on  $\mathcal{T}^{(\varepsilon)}$ . Denote by  $(\mathcal{T}_x)_{x \in \mathcal{T}^{(\varepsilon)}}$  the countable collection of closures of the connected components of  $\mathcal{T} \setminus \mathcal{T}^{(\varepsilon)}$ . For every  $x \in \mathcal{T}^{(\varepsilon)}$ , we denote by  $h_x$  the distance between  $x$  and the leaf  $x_i$  such that  $x$  belongs to the branch  $[[\rho_i, x_i]]_{\mathcal{T}}$  (which is unique apart from when  $x$  is a branch point of  $\mathcal{T}^{(\varepsilon)}$ ). Conditional on  $\mathcal{T}^{(\varepsilon)}$ , the point process*

$$\sum_{x \in \mathcal{T}^{(\varepsilon)}} \delta_{(x, \mathcal{T}_x)}$$

*is an inhomogeneous Poisson point process with intensity measure*

$$\frac{1}{2} \lambda^{(\varepsilon)}(dx) \mathbf{N}^{H \leq h_x \wedge \varepsilon}(d\mathcal{T}_x),$$

where  $\mathbf{N}^{H \leq h} := \int_0^h \frac{dh'}{(h')^2} \mathbf{N}^{H=h'} = \mathbf{N}(\cdot \cap \{H \leq h\})$ .

**REMARK 3.3.** Note that, for a given  $\varepsilon > 0$ , the intensity measure of Proposition 3.2 is dominated by the homogeneous measure

$$\frac{1}{2} \lambda^{(\varepsilon)}(dx) \mathbf{N}^{H \leq \varepsilon}(d\mathcal{T}_x).$$

Moreover, although in the statement of the above result we only explicitly keep track of the subsets of the form  $\mathcal{T}_x$ , we have implicitly included the restriction of the metric  $d_{\mathcal{T}}$  on these and the marked point that represents their root. Additionally, we could have included the restriction of the measure  $\mu_{\mathcal{T}}$ , and thus considered each  $\mathcal{T}_x$  as a marked compact metric space equipped with a finite Borel measure. We will later need this more refined viewpoint.

**PROOF.** Fix  $\varepsilon > 0$  and set  $I_\varepsilon = \{i \in \Sigma_* : h_i > \varepsilon\}$ . Under  $\mathbf{P}$ , Lemma 3.1 ensures that the random set  $I_\varepsilon$  is finite almost-surely. In addition,  $\mathcal{T}^{(\varepsilon)} = [[\rho, x_\emptyset]] \cup (\bigcup_{i \in I_\varepsilon} [[\rho_i, x_i]])$  by construction. Therefore, the collection  $(\mathcal{T}_x)_{x \in \mathcal{T}^{(\varepsilon)}}$  of Proposition 3.2 consists of all the trees  $\mathcal{T}_j$  with  $j \notin I_\varepsilon$ . Alternatively,

$$(\mathcal{T}_x)_{x \in \mathcal{T}^{(\varepsilon)}} = \bigcup_{i \in I_\varepsilon} \left( \bigcup_{j \geq 1 : ij \notin I_\varepsilon} \mathcal{T}_{ij} \right) = \bigcup_{i \in I_\varepsilon} \left( \bigcup_{j \geq 1 : h_{ij} \leq \varepsilon} \mathcal{T}_{ij} \right),$$

where the unions are disjoint. Given  $\mathcal{T}^{(\varepsilon)}$  and thanks to (31), for every  $i \in I_\varepsilon$ , the point measure

$$\mathcal{P}_i := \sum_{j \geq 1} \delta_{(\rho_{ij}, \mathcal{T}_{ij})}$$

is a Poisson point process with intensity  $\frac{1}{2}\mathbf{1}_{[[\rho_i, x_i]]_{\mathcal{T}}}(x)dx \mathbf{N}(\mathbf{1}_{H \leq h_x} d\mathcal{T}_x)$ . Therefore the point measure

$$\mathcal{P}_i^{(\varepsilon)} := \sum_{j \geq 1 : h_{ij} \leq \varepsilon} \delta_{(\rho_{ij}, \mathcal{T}_{ij})}$$

is a Poisson point process with intensity

$$\frac{1}{2}\mathbf{1}_{[[\rho_i, x_i]]_{\mathcal{T}}}(x)dx \mathbf{N}(\mathbf{1}_{H \leq h_x \wedge \varepsilon} d\mathcal{T}_x) = \frac{1}{2}\mathbf{1}_{[[\rho_i, x_i]]_{\mathcal{T}}}(x)dx \mathbf{N}^{h_x \wedge \varepsilon}(d\mathcal{T}_x).$$

Given  $\mathcal{T}^{(\varepsilon)}$ , the point measures  $\mathcal{P}_i^{(\varepsilon)}$  have disjoint supports and are independent, and the result follows.  $\square$

**4. A path property of BESQ-Brownian snakes.** In this section we to prove a path property of snakes driven by a Brownian motion with spatial displacements following a 0-dimensional BESQ process. In Section 4.1, we first give a quick overview of snakes and then formally state our path property as Theorem 4.2. The proof of this result involves the special Markov property of snakes that we state in Section 4.2. These two sections cover well-known properties of Brownian snakes and we refer the reader to Le Gall's book [53, Chapter IV] for more details. (An interested reader might also want to refer to [35], where more general branching mechanisms are treated.) Finally, in Section 4.3, we prove Theorem 4.2.

4.1. *Definition and statement of the path property.* We start by giving a description of the snakes we will work with, which are path-valued processes. A finite path is a continuous mapping  $w : [0, \zeta] \rightarrow \mathbb{R}$ , and  $\zeta$  is called the lifetime of  $w$ . Let  $\mathcal{W}$  be the set of all finite paths in  $\mathbb{R}$ . For  $w \in \mathcal{W}$ , we denote by  $\zeta_{(w)}$  the lifetime of  $w$ . The tip of the path  $w$  is denoted by  $\widehat{w} = w(\zeta_{(w)})$ . For every  $x \in \mathbb{R}$ , we set  $\mathcal{W}_x = \{w \in \mathcal{W} : w(0) = x\}$ , and we identify the trivial path of  $\mathcal{W}_x$  with lifetime 0 with the point  $x$ . The space  $\mathcal{W}$  is equipped with the following distance:

$$(33) \quad d(w_1, w_2) := |\zeta_{(w_1)} - \zeta_{(w_2)}| + \sup_{t \geq 0} |w_1(t \wedge \zeta_{(w_1)}) - w_2(t \wedge \zeta_{(w_2)})|.$$

A path-valued process is a random element of the space  $\mathcal{C}(\mathbb{R}_+, \mathcal{W})$  of continuous functions from  $\mathbb{R}_+$  to  $\mathcal{W}$  equipped with the topology of uniform convergence on every compact subset of  $\mathbb{R}_+$ .

For  $w \in \mathcal{W}$ , the law of the *zero-dimensional BESQ-Brownian snake* started from  $w$ , which we will denote by  $\mathbb{P}_w$ , is characterised as follows. Note that we will write  $(W_s)_{s \geq 0}$  for a random element of  $\mathcal{C}(\mathbb{R}_+, \mathcal{W})$  sampled according to  $\mathbb{P}_w$ . First, under  $\mathbb{P}_w$ , the lifetime process  $(\zeta_{(W_s)})_{s \geq 0}$  is a reflected Brownian motion started at  $\zeta_{(w)}$ . To simplify notation we write  $\zeta_s := \zeta_{(W_s)}$  for every  $s \geq 0$ . Second, conditionally on the lifetime process  $(\zeta_s)_{s \geq 0}$ , the law of process  $(W_s)_{s \geq 0}$ , denoted by  $\Theta_w^\zeta$ , satisfies

1. The initial value  $W_0$  is  $\Theta_w^\zeta$ -almost-surely equal to  $w$ .
2. The process  $(W_s)_{s \geq 0}$  is time-homogeneous Markov under  $\Theta_w^\zeta$ . Its transition kernel is specified as follows. For  $0 \leq s \leq s'$ :
  - $W_{s'}(t) = W_s(t)$  for all  $t \leq m_\zeta(s, s') := \min\{\zeta_r : s \leq r \leq s'\}$ ;
  - The random path  $(W_{s'}(m_\zeta(s, s') + t))_{0 \leq t \leq \zeta_{s'} - m_\zeta(s, s')}$  only depends on  $W_s$  through  $W_s(m_\zeta(s, s'))$  and is distributed as a zero-dimensional squared Bessel process started at  $W_s(m_\zeta(s, s'))$  and stopped at time  $\zeta_{s'} - m_\zeta(s, s')$ .

The full construction of this process is covered in [53, Chapter IV] when the zero-dimensional BESQ process is replaced by a generic Markov process with uniformly Hölder continuous sample paths (see Assumption (C) in Section IV.4 of [53]). In particular, [53, Chapter IV] proves that the process has continuous sample paths (Proposition 5) and is strong Markov (Theorem 6). Inconveniently for our purposes, the zero-dimensional BESQ process does not have uniformly Hölder continuous sample paths and a (very minor) tweak is needed to prove that our snake process is still continuous. We include this proof (which replaces the proof of Proposition 5 in [53, Chapter IV]) for the sake of completeness. The proof of the strong Markov property (Theorem 6 in [53, Chapter IV]) is identical as it does not use the Hölder property.

**PROPOSITION 4.1** (cf. [53, Proposition IV.5]). *For every  $w \in \mathcal{W}$ , the process  $(W_s)_{s \geq 0}$  has a continuous modification under  $\mathbb{P}_w$ .*

**PROOF.** The proof is almost identical to that of Le Gall and we only sketch the main difference. In particular, we only prove the statement for  $w \equiv x$  for some  $x > 0$ . The extension to any starting path is then identical to Le Gall's proof. We denote by  $\xi = (\xi_t)_{t \geq 0}$  a zero-dimensional squared Bessel process and by  $\Pi_x$  the law of  $\xi$  started from  $x$ . Fix a deterministic function  $f \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}_+)$  such that  $f(0) = 0$ , as well as constants  $p > 2$  and  $T > 0$ . By the definition of the distance at (33), for every  $0 \leq s \leq s' \leq T$ , one has

$$\Theta_x^f(d(W_s, W_{s'})^p) \leq c_p \left( |f(s) - f(s')|^p + \Pi_x \left( \sup_{t \in [m_f(s, s'), f(s)]} |\xi_t - \xi_{m_f(s, s')}|^p \right) + \Pi_x \left( \sup_{t \in [m_f(s, s'), f(s')] } |\xi_t - \xi_{m_f(s, s')}|^p \right) \right).$$

Applying the Markov property of  $\xi$ , we get

$$(34) \quad \Theta_x^f(d(W_s, W_{s'})^p) \leq c_p \left( |f(s) - f(s')|^p + 2\Pi_x \left( \Pi_{\xi_{m_f(s, s')}} \left( \sup_{t \in [0, f(s) \vee f(s') - m_f(s, s')] } |\xi_t - \xi_0|^p \right) \right) \right).$$

Now, since we are we will be supposing that  $f$  is a typical realisation of a Brownian path, and such paths are  $(2 - \eta)$ -Hölder continuous, for  $\eta$  arbitrarily small, we can suppose that there exists a constant  $C$  such that  $f$  satisfies

$$|f(s) - f(s')| \leq C|s - s'|^{\frac{1}{2} - \eta}.$$

for  $0 \leq s \leq s' \leq T$ . This is where we need to tweak Le Gall's proof, since the Hölder continuity of  $\xi$  is not uniform with respect to its starting point. However, by the scaling property of Bessel processes, one can check that, for every  $\varepsilon > 0$ ,  $x \geq 0$  and  $T \geq 0$ , there exists  $q > 0$  and a constant  $C$  that does not depend on  $x$  or  $T$  such that

$$\Pi_x \left( \sup_{0 \leq t \leq T} |\xi_t - x|^p \right) \leq C x^q T^{p/2}.$$

Plugging this into (34) yields

$$\begin{aligned} \Theta_x^f(d(W_s, W_{s'})^p) &\leq C' \left( |s - s'|^{p(\frac{1}{2} - \eta)} + (f(s) \vee f(s') - m_f(s, s'))^{p/2} \Pi_x \left( \xi_{m_f(s, s')}^q \right) \right), \\ &\leq C' \left( |s - s'|^{p(\frac{1}{2} - \eta)} + |s - s'|^{\frac{p}{2}(\frac{1}{2} - \eta)} \right), \end{aligned}$$

where the constant  $C'$ , which can depend on  $p, q, x, f, T$ , is finite, since Bessel processes have finite polynomial moments. Choosing  $p > 4$  and then  $\eta$  accordingly, we get

$$\Theta_x^f(d(W_s, W_{s'})^p) \leq C |s - s'|^{p'},$$

with  $p' > 1$ , and  $C$  a finite positive constant that depends on  $x, f$  and  $T$  (and  $p$  and  $p'$ ). Kolmogorov's lemma ensures that there is a continuous modification of  $(W_s)_{s \geq 0}$  under  $\Theta_x^f$ , and the proposition follows.  $\square$

We will denote by  $\mathbb{P}_x$  the law of the snake started at  $x > 0$ . Since 0 is an absorbing point for the zero-dimensional Bessel process, we can extend the definition to  $x = 0$  by simply setting  $W_s(t) = 0$  for all  $s \geq 0$  and  $0 \leq t \leq \zeta_s$ . Every point  $x \geq 0$  is then regular for  $W$  in the sense that  $\mathbb{P}_x(\inf\{s > 0 : W_s = x\} = 0) = 1$ . The general excursion theory for Markov processes can then be used to construct the excursion measure of  $W$  away from  $x$ , which we denote by  $\mathbb{N}_x$ .

Under the (infinite) measure  $\mathbb{N}_x$ , the lifetime process  $(\zeta_s)_{s \geq 0}$  is distributed according to the Itô excursion measure  $\mathbf{n}$ , normalised as in (29). As before, we will denote by  $\sigma$  the duration of an excursion under  $\mathbb{N}_x$ , that is  $\sigma = \sup\{s > 0 : \zeta_s > 0\}$ . To connect with the discussion of Section 2, we highlight that if  $\mathcal{T}$  is the real tree associated with  $\zeta$  through the procedure described at the start of Section 3, then the map  $s \mapsto \widehat{W}_s = W_s(\zeta_s)$  to the tip of the path induces a well-defined map  $y \mapsto \xi_y^{\mathcal{T}}$  on  $\mathcal{T}$  given by setting  $\xi_y^{\mathcal{T}} = \widehat{W}_s$  for any  $s$  in the equivalence class that corresponds to  $y \in \mathcal{T}$ . It is straightforward to check from the construction that, conditional on  $\mathcal{T}$ ,  $(\xi_y^{\mathcal{T}})_{y \in \mathcal{T}}$  is a  $\mathcal{T}$ -indexed 0-dimensional BESQ process started from  $x$ . (Cf. the discussion of [36], which presented a similar discussion for tree-indexed Brownian motion.)

The main conclusion of this section is the following. For its statement, we again recall that  $\widehat{W}_s = W_s(\zeta_s)$  is the tip of the path  $W_s$ . Specifically, the result gives that either  $\widehat{W}$  does not hit zero, or it takes the value zero on an open set; this will be a key input into checking Assumption 2 for the Brownian CRT.

**THEOREM 4.2.** *We have, for every  $x > 0$ :*

$$\mathbb{N}_x \left( \left\{ \widehat{W}_s = 0 \text{ on an open interval} \right\} \cup \left\{ \forall s \in [0, 1], \widehat{W}_s \neq 0 \right\} \right) = 1.$$

This result is a consequence of the special Markov property of the Brownian snake, first stated in Le Gall's book [53]. A more general version is available in the Appendix of [55]. In the next section, we state the relevant property in the setting that we will use to prove Theorem 4.2.

**4.2. The special Markov property.** In this section,  $x > 0$  is fixed. For any path  $w \in \mathcal{W}_x$ , we write

$$\tau(w) = \inf\{t \geq 0 : w(t) = 0\}$$

for the hitting time of 0 by  $w$  (or equivalently, since  $x > 0$ , the first exit time of the domain  $(0, +\infty)$ ), with the convention  $\inf \emptyset = +\infty$ . We define the first hitting time of zero by the snake to be

$$T := \inf \left\{ s \geq 0 : \widehat{W}_s = 0 \right\} = \inf \left\{ s \geq 0 : \tau(W_s) < +\infty \right\}.$$

Since the spatial displacement of our Brownian snake is a zero-dimensional squared Bessel process, we have

$$\mathbb{P}_x(T = +\infty) = 0 \quad \text{and} \quad 0 < \mathbb{N}_x(T < +\infty) < \infty.$$

Indeed, the only non-straightforward inequality is the last one, namely

$$(35) \quad \mathbb{N}_x(T < +\infty) = \mathbb{N}_x\left(\exists s \in [0, \sigma] : \widehat{W}_s = 0\right) < \infty.$$

If (35) was not true, then by excursion theory, under  $\mathbb{P}_x$ , infinitely many excursions of the snake would exit  $[x/2, \infty)$  before any given positive time. This contradicts the fact that the snake is continuous under  $\mathbb{P}_x$ .

We will work under the probability measure  $\mathbb{N}_x$  conditioned on the event  $\{T < +\infty\}$ :

$$\mathbb{N}_x^T := \mathbb{N}_x(\cdot | T < +\infty).$$

This probability measure can also be interpreted as the law under  $\mathbb{P}_x$  of the first Brownian snake excursion away from  $x$  that hits 0. The exit local time process from  $(0, +\infty)$  under  $\mathbb{N}_x^T$  is the process  $(\Lambda_s)_{0 \leq s \leq \sigma}$  defined by

$$\Lambda_s := \lim_{\varepsilon \rightarrow 0} \int_0^s \mathbf{1}_{\{\tau(W_r) < \zeta_r < \tau(W_r) + \varepsilon\}} dr,$$

and the measure  $d\Lambda_s$  is supported on  $\{s \in [0, \sigma] : \tau(W_s) = \zeta_s\}$  (see [53, Section V.1] and [55, Appendix]).

In our setting, the special Markov property describes the law of the excursions of the zero-dimensional BESQ-Brownian snake at 0, conditionally on the information provided by the paths of the snake before they hit 0. Let us next describe the filtration  $\mathcal{F}$  containing this information. For every  $s \geq 0$ , we set

$$\eta_s := \inf \left\{ t \geq 0 : \int_0^t \mathbf{1}_{\{\zeta_r \leq \tau(W_r)\}} dr > s \right\}.$$

This definition makes sense  $\mathbb{P}_x$  almost-surely, and we define  $\mathcal{F}$  to be the  $\sigma$ -field generated by the process  $(W_{\eta_s})_{s \geq 0}$  and the  $\mathbb{P}_x$ -negligible sets. We note that  $(\Lambda_{\eta_s})_{s=0}^\sigma$  is  $\mathcal{F}$ -measurable (see [55, Lemma 19]).

Now let us define the excursions at 0 of the snake. For every  $s \geq 0$ , we set  $\gamma_s := (\zeta_s - \tau(W_s))^+$ . Under  $\mathbb{P}_x$ , the set  $\{s \geq 0 : \gamma_s > 0\}$  is almost-surely a countable union of disjoint open intervals:

$$\{s \geq 0 : \gamma_s > 0\} = \bigcup_{i \in \mathbb{N}} (a_i, b_i),$$

and the enumeration can be fixed in such a way that the variables  $a_i$  and  $b_i$  are measurable with respect to the  $\sigma$ -field generated by  $(\gamma_s)_{s \geq 0}$ .

From the properties of the Brownian snake and of the zero-dimensional squared Bessel process, we have that  $\mathbb{P}_x$ -a.s., for every  $i$  and every  $s \in [a_i, b_i]$ ,

$$\tau(W_s) = \tau(W_{a_i}) = \zeta_{a_i},$$

and that all the paths  $W_s$  for  $s \in [a_i, b_i]$  coincide up to their hitting time of 0, from which time they remain at 0 up to their lifetime. For every  $i$ , we define  $W^{(i)}$  by setting

$$W_s^{(i)}(t) := W_{(a_i+s) \wedge b_i}(\zeta_{a_i} + t), \quad \text{for } s \geq 0 \text{ and } 0 \leq t \leq \zeta_s^{(i)} := \zeta_{(a_i+s) \wedge b_i} - \zeta_{a_i}.$$

These variables are the excursions of the zero-dimensional BESQ-Brownian snake at 0, and their structure is fully described by the special Markov property. The variant of this property that we will use is the following.

**THEOREM 4.3** (Special Markov property, [55, Corollary 22]). *Under  $\mathbb{N}_x^T$ , conditionally on the  $\sigma$ -field  $\mathcal{F}$ , the point measure*

$$\sum_{i \in \mathbb{N}} \delta_{W^{(i)}}(d\omega)$$

*is Poisson with intensity*

$$\int \Lambda_\sigma \mathbb{N}_0(d\omega) = \int \Lambda_\sigma \mathbf{n}(d\zeta),$$

*with a slight abuse of notation in the last equality (the associated spatial process is understood to be the constant 0).*

**4.3. Proof of Theorem 4.2.** Theorem 4.2 states that  $\mathbb{N}_x$ -almost-surely, on the event when the snake hits 0, it has at least one excursion of positive length at 0. A direct reformulation of the statement we want to prove is:  $\mathbb{N}_x$ -a.s.,

$$\left\{ \widehat{W}_s = 0 \text{ on an open interval} \right\} = \left\{ \exists s \in [0, \sigma] : \widehat{W}_s = 0 \right\}.$$

For  $\varepsilon \geq 0$ , set

$$A_\varepsilon := \left\{ \widehat{W}_s = 0 \text{ on an open interval of length } > \varepsilon \right\}.$$

The special Markov property stated in Theorem 4.3 and our normalisation for Itô excursion measure (see (29)) gives, for  $\varepsilon > 0$ :

$$\mathbb{N}_x^T(A_\varepsilon) = \mathbb{N}_x^T \left( 1 - \exp \left( - \frac{\Lambda_\sigma}{\sqrt{2\pi\varepsilon}} \right) \right).$$

Taking the limit when  $\varepsilon > 0$  we obtain

$$\lim_{\varepsilon \rightarrow 0} \mathbb{N}_x^T(A_\varepsilon) = \mathbb{N}_x^T(\Lambda_\sigma > 0),$$

or equivalently,  $\mathbb{N}_x$ -a.s.,

$$(36) \quad \{\Lambda_\sigma > 0\} = \left\{ \widehat{W}_s = 0 \text{ on an open interval} \right\}.$$

We will now prove,  $\mathbb{N}_x$ -a.s.,

$$\{\Lambda_\sigma > 0\} = \left\{ \exists s \in [0, \sigma], \widehat{W}_s = 0 \right\},$$

which will then immediately give Theorem 4.2.

For every  $x > 0$ , set

$$u(x) := \mathbb{N}_x \left( \exists s \in [0, \sigma], \widehat{W}_s = 0 \right).$$

Writing  $\mathbb{N}_x^{(u)}$  for  $\mathbb{N}_x(\cdot | \sigma = u)$ , the scaling properties of Brownian excursions (cf. (30)) and squared Bessel processes yield, for  $\lambda > 0$ :

$$\begin{aligned} u(x) &= \mathbb{N}_x \left( \exists s \in [0, \sigma], \widehat{W}_s = 0 \right) \\ &= \int_0^\infty \frac{du}{2\sqrt{2\pi}u^3} \mathbb{N}_x^{(u)} \left( \exists s \in [0, \sigma], \widehat{W}_s = 0 \right) \\ &= \frac{1}{\lambda} \int_0^\infty \frac{du}{2\sqrt{2\pi}u^3} \mathbb{N}_x^{(\lambda^2 u)} \left( \exists s \in [0, \sigma/\lambda^2], \frac{1}{\lambda} \widehat{W}_{\lambda^2 s} = 0 \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\lambda} \int_0^\infty \frac{du}{2\sqrt{2\pi u^3}} \mathbb{N}_{x/\lambda}^{(u)} \left( \exists s \in [0, \sigma], \widehat{W}_s = 0 \right) \\
&= \lambda^{-1} \mathbb{N}_{x/\lambda} \left( \exists s \in [0, \sigma], \widehat{W}_s = 0 \right) \\
&= \lambda^{-1} u(x/\lambda).
\end{aligned}$$

Since we know from (35) that the function  $u$  is not infinite, there is a finite constant  $C = u(1) \geq 0$  such that, for every  $x > 0$ ,

$$u(x) = C x^{-1}.$$

In addition, Proposition VI.2 of Le Gall's book [53] states that the function  $u$  is the maximal non-negative solution of the equation

$$(37) \quad \mathcal{A}u(x) = 2u^2(x), \quad \forall x > 0,$$

where  $\mathcal{A}$  is the Markov generator of the zero-dimensional BESQ process:

$$\mathcal{A}f(x) = 2xf''(x),$$

see [65, Section XI.1]. Precisely, [53, Proposition VI.2] treats a Brownian snake whose spatial displacements follow a Brownian motion. However, it is possible to check that, with (35) in place of [53, Proposition V.9(i)], the same argument applies. Applying (37), a quick calculation gives that either  $C = 0$  or  $C = 2$ , and since  $u$  is maximal, we have  $u(x) = 2x^{-1}$  for every  $x > 0$ . (Note that, apart from the difference in the Markov process describing the spatial displacements, this argument is the same as that given by Le Gall in the example appearing after [53, Proposition VI.2].)

Similarly, we can calculate  $v(x) := \mathbb{N}_x(\Lambda_\sigma > 0)$ . Since  $v(x) \leq \mathbb{N}_x(\exists s \in [0, \sigma], \widehat{W}_s = 0)$ , we obtain from (35) that  $v(x) < \infty$  for any  $x > 0$ . Moreover, the same scaling arguments as for  $u$  give

$$v(x) = \mathbb{N}_x(\Lambda_\sigma > 0) = \lambda^{-1} \mathbb{N}_{x/\lambda}(\Lambda_\sigma > 0) = \lambda^{-1} v(x/\lambda).$$

Therefore, there is another finite constant  $C' \geq 0$  such that, for every  $x > 0$ ,

$$v(x) = C' x^{-1}.$$

For  $n > 0$ , set

$$v_n(x) := \mathbb{N}_x(1 - \exp(-n \Lambda_\sigma)).$$

A straightforward adaptation to the case of a 0-dimensional BESQ-Brownian snake of [53, Theorem V.6] gives that  $v_n$  satisfies

$$\begin{cases} \mathcal{A}v_n = 2v_n^2 & \text{on } (0, +\infty), \\ \lim_{x \rightarrow 0^+} v_n(x) = n. \end{cases}$$

We highlight that, as for the result of [53] that was cited in the previous part of the proof, [53, Theorem V.6] is written for a Brownian snake whose spatial displacements follow a Brownian motion. However, the proof is readily adapted by using the fact that a 0-dimensional BESQ process can be written as the time-change of a Brownian motion. (See the proof of Lemma 2.8 for details.) Indeed, the only sensitive part of Le Gall's argument is an appeal to [64, Theorem 6.6] to check that, for a suitable function  $f$ , it holds that  $\frac{1}{2} \Delta \int G_B^{\text{BM}}(x, y) f(y) dy = -f(x)$ , where  $G_B^{\text{BM}}(x, y)$  is the Green's function of Brownian motion killed on exiting a ball  $B$  contained in  $(0, \infty)$ . In our case, we are required to check  $\mathcal{A} \int G_B^{\text{BESQ}^0}(x, y) f(y) dy = -f(x)$  for

suitable  $f$ , where  $G_B^{\text{BESQ}^0}(x, y)$  is the corresponding Green's function of the 0-dimensional BESQ process. However, since  $G_B^{\text{BESQ}^0}(x, y) = \frac{1}{4y} G_B^{\text{BM}}(x, y)$  and the density of the speed measure,  $\frac{1}{4y}$ , is bounded away from 0 and  $\infty$  on compact subsets of  $(0, \infty)$ , there is no problem in applying the result of [64] in our case. We next note that, since  $v$  is the pointwise limit of the increasing sequence of functions  $(v_n)$  and the set of non-negative solutions on  $(0, +\infty)$  of the equation  $\mathcal{A}f = 2f^2$  is closed under pointwise convergence (to check this, one may follow the argument of [53, Proposition V.9(iii)], for example), the function  $v$  also satisfies  $\mathcal{A}v = 2v^2$  on  $(0, +\infty)$  and is not identically equal to 0. Therefore we have  $C' = 2$ , that is, for every  $x > 0$ :

$$\mathbb{N}_x(\Lambda_\sigma > 0) = \mathbb{N}_x\left(\exists s \in [0, \sigma], \widehat{W}_s = 0\right) = \frac{2}{x}.$$

Since it clearly follows from (36) that  $\{\Lambda_\sigma > 0\} \subseteq \{\exists s \in [0, \sigma], \widehat{W}_s = 0\}$ ,  $\mathbb{N}_x$ -a.s., the two events are therefore  $\mathbb{N}_x$ -a.s. equal, and Theorem 4.2 follows.

**5. Proof of Theorem 1.1.** The aim of this section is to prove Theorem 1.1. Clearly, by the comment below the scaling property at (30), establishing that the desired property of the Brownian CRT holds  $\mathbf{P}$ -a.s. is equivalent to showing the result almost-surely for trees under the push-forward of the Itô excursion measure (i.e.,  $\mathbf{N}$ ), as introduced in Section 3. Our strategy is to first prove that Assumptions 1 and 2 hold for  $\mathbf{N}$ -typical trees, which readily yields the conclusion of Theorem 1.1 when  $X^T$  is started from the root (see Section 5.2). To complete the proof, we give a small additional argument involving random re-rooting and a continuity property to verify that the result can be extended to arbitrary starting points (see Section 5.3). Before this, in Section 5.1, we formalise the setting so as to ensure all the objects of the discussion are well-defined.

5.1. *An extended Gromov-Hausdorff-type topology.* In this section, we introduce (a simplification of) the extended Gromov-Hausdorff-type topology used in [61] for the purposes of studying random metric-measure spaces equipped with stochastic processes and corresponding local times. (A similar viewpoint was also set out in [15].) In order to keep the presentation concise, we will be brief with the details; for a more in-depth presentation, we refer the reader to [62] (cf. [45]), which also provides a discussion of earlier works in this area.

We will consider two basic spaces:  $\check{\mathbb{F}}_c$  and  $\mathbb{M}_{L,c}$ . For the first of these, we consider objects of the form

$$(38) \quad (K, R_K, \mu_K, \rho_K),$$

where, as in the introduction,  $(K, R_K)$  is a (non-empty) compact resistance metric space,  $\mu_K$  is a finite Borel measure on  $K$  of full support, and  $\rho_K$  is a marked point (or root) of  $K$ . Here, we recall that a resistance metric, which was introduced by Kigami in the context of analysis on fractals [47, 48], is defined as follows.

**DEFINITION 5.1** ([47, Definition 2.3.2]). Let  $K$  be a set. A function  $R : K \times K \rightarrow \mathbb{R}$  is a resistance metric if, for every finite  $K' \subseteq K$ , one can find a weighted (i.e. equipped with conductances) graph with vertex set  $K'$  for which  $R|_{K' \times K'}$  is the associated effective resistance.

Similarly to [29, 31, 61], the collection of spaces of the form at (38), identified if they are measure- and root-preserving isometric, will be written  $\mathbb{F}_c$ . (The subscript  $c$  here refers to the restriction to compact spaces.) It is known that such spaces can be associated with a

canonical stochastic process  $((X_t^K)_{t \geq 0}, (P_x^K)_{x \in K})$  that admits jointly measurable local times  $(L_t^K(x))_{x \in K, t \geq 0}$  satisfying an occupation density formula as at (9). (See the discussion in [29, 31] for the existence of the process, and [31, Lemma 2.4], in particular, for the existence of local times. In these papers, a certain regularity condition is imposed on the resistance metric, but this is immediate in the compact case, see [48, Corollary 6.4].) To ensure that the local times are jointly continuous, we assume a further condition involving the metric entropy, as introduced by Noda in [61], which we now describe. In particular, for a compact metric space  $(K, d_K)$  and  $\varepsilon > 0$ ,  $K' \subseteq K$  is called an  $\varepsilon$ -covering of  $(K, d_K)$  if for each point  $x \in K$ , there exists an  $x' \in K'$  such that  $d_K(x, x') \leq \varepsilon$ . Let

$$(39) \quad N(K, d_K, \varepsilon) := \min \{|K'| : K' \text{ is an } \varepsilon\text{-covering of } (K, d_K)\}.$$

It is standard in the literature to call the family (over  $\varepsilon$ ) of the logarithm of  $N(K, d_K, \varepsilon)$ , the metric entropy of  $(K, d_K)$ , though in [61] (and some other places), this term is used to denote  $(N(K, d_K, \varepsilon))_{\varepsilon > 0}$  instead. It is shown in [61, Corollary 4.12] that if  $(K, R_K, \mu_K, \rho_K) \in \mathbb{F}_c$  and there exists an  $\alpha \in (0, 1/2)$  such that

$$(40) \quad \sum_{k \geq 1}^{\infty} N(K, R_K, 2^{-k})^2 \exp(-2^{\alpha k}) = 0,$$

then the local times  $L^K$  can be constructed to be jointly continuous. We denote the subset of  $\mathbb{F}_c$  containing elements for which (40) is satisfied by  $\check{\mathbb{F}}_c$ ; this is the space appearing in the statements of Theorem 1.2 and Corollary 1.3.

The space  $\mathbb{F}_c$  may be equipped with a metric yielding the marked Gromov-Hausdorff-Prohorov metric, as given by setting

$$(41) \quad \begin{aligned} & d_{GHP}((K_1, R_{K_1}, \mu_{K_1}, \rho_{K_1}), (K_2, R_{K_2}, \mu_{K_2}, \rho_{K_2})) \\ &= \inf_{(M, d_M), \phi_1, \phi_2} \max \{d_M^H(\phi_1(K_1), \phi_2(K_2)), d_M^P(\mu_{K_1} \circ \phi_1^{-1}, \mu_{K_2} \circ \phi_2^{-1})\}, \end{aligned}$$

where the infimum is taken over compact rooted metric spaces  $(M, d_M, \rho_M)$  and root-preserving isometric embeddings  $\phi_i : K_i \rightarrow M$ ,  $d_M^H$  is the Hausdorff metric between compact subsets of  $(M, d_M)$ , and  $d_M^P$  is the Prohorov metric between Borel measures on  $(M, d_M)$ . Note that there are various slightly different versions of the above metric used in the literature, but, as discussed in [61, Section 2.1], they all give rise to the same topology. On the larger collection of marked compact metric spaces equipped with a finite Borel measure, it is known that this topology is complete and separable, as is convenient for various probabilistic applications. To say  $(K, R_K, \mu_K, \rho_K)$  is a random element of  $\check{\mathbb{F}}_c$ , as we do in Corollary 1.3, we mean that it is chosen according to a probability measure on this larger space whose support is contained in  $\check{\mathbb{F}}_c$ .

Secondly, we consider objects of the form at (3), i.e. to each element  $(K, R_K, \mu_K, \rho_K) \in \check{\mathbb{F}}_c$ , we add  $P^K$ , which is the law of  $((X_t^K)_{t \geq 0}, (L_t^K(x))_{x \in K, t \geq 0})$  under  $P_{\rho_K}^K$ . Now, for such spaces, we have that  $X^K$  is  $P^K$ -a.s. an element of  $D(\mathbb{R}_+, K)$ , the space of càdlàg paths in  $K$ , which we equip with the usual Skorohod  $J_1$  topology, and  $L^K$  takes values in  $C(K \times \mathbb{R}_+, \mathbb{R}_+)$ . The larger space, where we drop the constraints that  $R_K$  is a resistance metric satisfying (40) and  $\mu_K$  has full support, and allow  $P^K$  to be any probability measure on  $D(\mathbb{R}_+, K) \times C(K \times \mathbb{R}_+, \mathbb{R}_+)$ , will be denoted  $\mathbb{M}_{L,c}$ . (This is the restriction of the space  $\mathbb{M}_L$  of [61] to those elements for which the metric space is compact; we do not need the generality of non-compact spaces considered in that article. The subscript  $L$  refers to the fact elements of the space are equipped with local time-type functions, and  $c$  again refers to the restriction to compact spaces.) To measure distances between two objects  $(K_i, R_{K_i}, \mu_{K_i}, \rho_{K_i}, P^{K_i})$ ,  $i = 1, 2$ , we use a similar metric to (41), but incorporate an additional distance

$$d_M^L(P^{K_1} \circ \phi_1^{-1}, P^{K_2} \circ \phi_2^{-1})$$

which is the Prohorov distance between the laws of  $((\phi_i(X_t^{K_i}))_{t \geq 0}, (L_t^{K_i}(\phi_i^{-1}(x)))_{x \in \phi_i(K_i), t \geq 0})$ ,  $i = 1, 2$ . Of course, to give full definition of this Prohorov metric, one needs to define a distance between objects of the latter form. To this end, in [61], such a distance is defined in terms of a standard metrisation of the Skorohod  $J_1$  topology for the components  $(\phi_i(X_t^{K_i}))_{t \geq 0}$ ,  $i = 1, 2$ , and a notion of uniform convergence for functions with variable domains for the components  $(L_t^{K_i}(\phi_i^{-1}(x)))_{x \in \phi_i(K_i), t \geq 0}$ ,  $i = 1, 2$ . See [61, Section 2.2] for details. Whilst we omit the precise definition, the most important consequence of it for our purposes is set out in Lemma 6.1, which describes how convergence in  $\mathbb{M}_{L,c}$  implies the existence of isometric embeddings of the spaces in question with various nice properties.

Before proceeding, let us finish this section with a few remarks about measurability issues that are relevant here. Firstly, [61, Proposition 6.1] ensures that the map  $(K, R_K, \mu_K, \rho_K) \mapsto (K, R_K, \mu_K, \rho_K, P^K)$  is a measurable mapping from  $\check{\mathbb{F}}_c$  to  $\mathbb{M}_{L,c}$ . This ensures the integrals defining annealed measures, cf. (5), are well-defined. We will also need in the proof of Theorem 1.2 a similar result for a map

$$(42) \quad (\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}}, \ell) \mapsto \left( \mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}}, \ell, P^{\mathcal{T}\text{-BESQ}^0(\ell)} \right),$$

which takes a compact, measured (with a probability measure of full support), rooted real tree  $(\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}})$  satisfying Assumption 1 (or, more generally, (40)) and an value  $\ell \geq 0$ , and returns in addition the law of the associated tree-indexed 0-dimensional BESQ process started at the root from  $\ell$ ,  $P^{\mathcal{T}\text{-BESQ}^0(\ell)}$ ; we recall from Lemma 2.8 that, under Assumption 1, a continuous version of such a process can be constructed. (The topology on the domain above can be assumed to be given by the product of the marked Gromov-Hausdorff-Prohorov topology on  $(\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}})$  and the usual Euclidean topology for the final component. As for the codomain, a topology here can be constructed as for local-time-type functions, using the notion of uniform convergence of functions with variable domains of [61].) To check the measurability in this case, one can follow the approach of [25, Lemma 8.1], which gave a similar result for the laws of Brownian motions on real trees. In particular, it is easy to check that convergence of the form

$$(\mathcal{T}_n, d_{\mathcal{T}_n}, \mu_{\mathcal{T}_n}, \rho_{\mathcal{T}_n}, \ell_n) \rightarrow (\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}}, \ell)$$

implies, for each  $k \geq 1$ , convergence in distribution of

$$(\mathcal{T}_n, d_{\mathcal{T}_n}, \mu_{\mathcal{T}_n}, \rho_{\mathcal{T}_n}, \ell_n, x_1^n, \dots, x_k^n) \rightarrow (\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}}, \ell, x_1, \dots, x_k),$$

where we add additional marks to the spaces by selecting  $k$  points according to the measure on the space in question. (For a suitable topology for such convergence, see [29, Section 8.3] or [45, Section 3.7.2] or [62, Section 4.2].) Using the description of the moment generating functions of the associated BESQ process at the points  $(x_1, \dots, x_k)$ , cf. the proof of Lemma 2.12, it is clear that the laws of the finite-dimensional marginals of these random variables simultaneously converge in distribution (considered as probability measures on the space of continuous functions with variable domains). Summarising so far, this gives that, for each  $k$ , the map from  $(\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}}, \ell)$  to the law of

$$\left( \mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}}, \ell, P^{\mathcal{T}\text{-BESQ}^0(\ell)}|_{x_1, \dots, x_k} \right),$$

is continuous. Finally, since the uniformly selected points become increasingly dense in  $\mathcal{T}$ , and the BESQ processes are continuous, it readily follows that

$$\left( \mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}}, \ell, P^{\mathcal{T}\text{-BESQ}^0(\ell)}|_{x_1, \dots, x_k} \right) \rightarrow \left( \mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}}, \ell, P^{\mathcal{T}\text{-BESQ}^0(\ell)} \right),$$

almost-surely, which yields that the map from the element  $(\mathcal{T}, d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho_{\mathcal{T}}, \ell)$  to the law of the right-hand side above is measurable. However, the latter is deterministic, and so we get the measurability of the map at (42), as desired.

5.2. *Checking the assumptions and the result with  $X^{\mathcal{T}}$  started from  $\rho$ .* We are now ready to start putting the pieces together and check the main result when  $X^{\mathcal{T}}$  is started from the root  $\rho$  (see Lemma 5.4). We first make the following observation, which readily follows from the measure estimates of [26] (and the scaling property of  $\mathbf{N}$ , as described at (30)).

LEMMA 5.2 ([26, Theorem 1.2]). *For  $\mathbf{N}$ -a.e. realisation of  $\mathcal{T}$ , Assumption 1 is satisfied.*

More challenging to check is Assumption 2, but we are able to do this by combining the Ray-Knight theorem of Section 2, the recursive decomposition of Section 3 and the path property of snakes proved in Section 4. In what follows in this section, we suppose that  $\mathcal{T}^{(\varepsilon)}$  is the subtree defined in Section 3, see (32) in particular. We note that the map  $\mathcal{T} \mapsto (\mathcal{T}^{(\varepsilon)}, (\rho_i, \mathcal{T}_i)_{i \in \mathcal{I}^{(\varepsilon)}})$  can be defined measurably with respect to a suitable product Gromov-Hausdorff-type topology on the image space (and, indeed, continuously at typical realisations of the Brownian CRT); since it is straightforward to check this, we omit the details.

LEMMA 5.3. *For  $\mathbf{N}$ -a.e. realisation of  $\mathcal{T}$ , Assumption 2 is satisfied.*

PROOF. Since Assumption 2 only requires us to check the desired property holds for a countable number of  $\varepsilon$  and  $\delta$ , it will suffice to check it for a single pair. In particular, in this proof, we fix  $\varepsilon, \delta > 0$  and write  $E := E_1(\varepsilon, \delta) \cup E_2(\varepsilon, \delta)$ , where the latter events were defined at (12) and (13). We will show that  $\mathcal{N}(\varepsilon, \delta) = 0$ , where

$$\mathcal{N}(\varepsilon, \delta) := \int \mathbf{N}(d\mathcal{T}) \mathbf{1}_{\{2\delta, \varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))\}} P_{\rho}^{\mathcal{T}}(E^c).$$

Now, by Corollary 2.14 and Lemma 5.2, we readily deduce that

$$\begin{aligned} \mathcal{N}(\varepsilon, \delta) &= \int \mathbf{N}(d\mathcal{T}) \mathbf{1}_{\{2\delta, \varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))\}} P_{\rho}^{\mathcal{T}} \left( E^c, \inf_{z \in \mathcal{T}^{(\varepsilon)}} L_{\tau_{\text{cov}}^{(\varepsilon)}(\delta)}^{\mathcal{T}}(z) > 0 \right) \\ &\leq \int \mathbf{N}(d\mathcal{T}) \mathbf{1}_{\{2\delta, \varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))\}} P_{\rho}^{\mathcal{T}} \left( E(\mathcal{T}_i)^c \text{ holds for some } i \in \mathcal{I}^{(\varepsilon)} \right) \\ &\leq \int \mathbf{N}(d\mathcal{T}) \mathbf{1}_{\{2\delta, \varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))\}} \sum_{i \in \mathcal{I}^{(\varepsilon)}} P_{\rho}^{\mathcal{T}}(E(\mathcal{T}_i)^c), \end{aligned}$$

where  $E(\mathcal{T}_i)$  is defined exactly as  $E$  is, but with  $\mathcal{T}$  replaced by  $\mathcal{T}_i$ . Applying the tower property of conditional expectation and the Ray-Knight result of Proposition 2.9 to the inner probability, we deduce that

$$\begin{aligned} P_{\rho}^{\mathcal{T}}(E(\mathcal{T}_i)^c) &= E_{\rho}^{\mathcal{T}} \left( P_{\rho}^{\mathcal{T}} \left( E(\mathcal{T}_i)^c \mid L_{\tau_{\text{cov}}^{(\varepsilon)}(\delta)}^{\mathcal{T}}(z), z \in \mathcal{T}^{(\varepsilon)} \right) \right) \\ &= \int P_{\rho}^{(\varepsilon)} \left( dL_{\tau(\varepsilon, \delta)}^{(\varepsilon)} \right) p \left( \mathcal{T}_i, L_{\tau(\varepsilon, \delta)}^{(\varepsilon)}(\rho_i) \right), \end{aligned}$$

where  $p(\mathcal{T}, \ell)$  is the probability a  $\mathcal{T}$ -indexed BESQ<sup>0</sup>( $\ell$ ) process satisfies the claim equivalent to  $E^c$ , i.e. it hits zero, but is not zero on any open ball. (The measurability of this function, as is required for the above integral to be well-defined, was discussed at the end of Section 5.1; technically, to proceed as described there, one also needs to incorporate suitable measures on the trees, but there is no problem in equipping each tree  $\mathcal{T}_i$  with the probability measure  $\mu(\cdot \cap \mathcal{T}_i) / \mu(\mathcal{T}_i)$ , and checking that this measure has the desired properties, almost-surely.) Next,

we insert this into the above upper bound for  $\mathcal{N}(\varepsilon, \delta)$ , decompose  $\mathbf{N}(d\mathcal{T})$  by conditioning on  $\mathcal{T}^{(\varepsilon)}$  and apply Fubini to yield the following:

$$\begin{aligned} \mathcal{N}(\varepsilon, \delta) &\leq \int \mathbf{N}(d\mathcal{T}) \mathbf{1}_{\{2\delta, \varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))\}} \int P_{\rho}^{(\varepsilon)}(dL_{\tau(\varepsilon, \delta)}^{(\varepsilon)}) \sum_{i \in \mathcal{I}^{(\varepsilon)}} p(\mathcal{T}_i, L_{\tau(\varepsilon, \delta)}^{(\varepsilon)}(\rho_i)) \\ &= \int \mathbf{N}(d\mathcal{T}^{(\varepsilon)}) \mathbf{1}_{\{2\delta, \varepsilon \in (0, d_{\mathcal{T}}(\rho, x_1))\}} \int P_{\rho}^{(\varepsilon)}(dL_{\tau(\varepsilon, \delta)}^{(\varepsilon)}) \\ &\quad \int \mathbf{N}(d((\mathcal{T}_i, \rho_i)_{i \in \mathcal{I}^{(\varepsilon)}}) \mid \mathcal{T}^{(\varepsilon)}) \sum_{i \in \mathcal{I}^{(\varepsilon)}} p(\mathcal{T}_i, L_{\tau(\varepsilon, \delta)}^{(\varepsilon)}(\rho_i)). \end{aligned}$$

Under the conditional law  $\mathbf{N}(\cdot \mid \mathcal{T}^{(\varepsilon)})$ , we have from Proposition 3.2 (see also Remark 3.3) that  $(\mathcal{T}_i, \rho_i)_{i \in \mathcal{I}^{(\varepsilon)}}$  are the atoms of a Poisson process with intensity measure bounded above by  $\lambda^{(\varepsilon)}(d\tilde{\rho})\mathbf{N}^{H \leq \varepsilon}(d\tilde{\mathcal{T}})$ . Hence Campbell's theorem (see [49, Section 3.2], for example) implies that

$$\begin{aligned} &\int \mathbf{N}(d((\mathcal{T}_i, \rho_i)_{i \in \mathcal{I}^{(\varepsilon)}}) \mid \mathcal{T}^{(\varepsilon)}) \sum_{i \in \mathcal{I}^{(\varepsilon)}} p(\mathcal{T}_i, L_{\tau(\varepsilon, \delta)}^{(\varepsilon)}(\rho_i)) \\ &\leq \int \lambda^{(\varepsilon)}(d\tilde{\rho}) \int \mathbf{N}^{H \leq \varepsilon}(d\tilde{\mathcal{T}}) p(\tilde{\mathcal{T}}, L_{\tau(\varepsilon, \delta)}^{(\varepsilon)}(\tilde{\rho})). \end{aligned}$$

Finally, we know from Theorem 4.2 that

$$\int \mathbf{N}^{H \leq \varepsilon}(d\tilde{\mathcal{T}}) p(\tilde{\mathcal{T}}, \ell) \leq \int \mathbf{N}(d\tilde{\mathcal{T}}) p(\tilde{\mathcal{T}}, \ell) = 0,$$

for any  $\ell \geq 0$ , and substituting this into our earlier bound gives that  $\mathcal{N}(\varepsilon, \delta) = 0$ , as desired.  $\square$

LEMMA 5.4. *For  $\mathbf{P}$ -a.e. realisation of  $\mathcal{T}$ , the equality at (2) holds  $P_{\rho}^{\mathcal{T}}$ -a.s.*

PROOF. By Lemmas 5.2 and 5.3,  $\mathbf{N}$ -a.e. realisation of  $\mathcal{T}$  satisfies Assumptions 1 and 2. From the scaling property of  $\mathbf{N}$ , as set out at (30), the same is true under  $\mathbf{P}$ . Hence we obtain the result from Proposition 2.7.  $\square$

5.3. *Extending to arbitrary starting points.* The proof of Theorem 1.1 is almost complete. It remains to explain how to generalise to arbitrary starting points. For this, the following root invariance property is useful.

LEMMA 5.5 ([10, (20)], [36, Proposition 4.8]). *The law of the rooted compact real tree  $(\mathcal{T}, \rho)$  under  $\mathbf{N}$  is the same as that of  $(\mathcal{T}, x)$  under  $\mathbf{N}(d\mathcal{T})\mu_{\mathcal{T}}(dx)/\mu_{\mathcal{T}}(\mathcal{T})$ .*

From the scaling property of  $\mathbf{N}$  (see (30)), it readily follows that the conclusion of Lemma 5.5 holds when  $\mathbf{N}$  is replaced by  $\mathbf{P}$ . Combining this observation with Lemma 5.4 yields the following.

LEMMA 5.6. *For  $\mathbf{P}$ -a.e. realisation of  $\mathcal{T}$ , for  $\mu_{\mathcal{T}}$ -almost-every  $x$ , the equality at (2) holds  $P_x^{\mathcal{T}}$ -a.s.*

We can now conclude our proof.

PROOF OF THEOREM 1.1 (FINAL STEP). Let us consider the general situation for a compact real tree  $(\mathcal{T}, d_{\mathcal{T}})$  equipped with a finite Borel measure of full support  $\mu_{\mathcal{T}}$  such that Assumption 1 holds (so that we have a continuous stochastic process  $X^{\mathcal{T}}$  with jointly continuous local times  $L^{\mathcal{T}}$ ). Moreover, suppose that the conclusion of Lemma 5.6 holds. In particular, there exists a dense subset  $D \subseteq \mathcal{T}$  for which (2) holds,  $P_x^{\mathcal{T}}$ -a.s., for all  $x \in D$ .

Now, let  $x \in \mathcal{T}$  and  $(x_n)_{n \geq 0}$  be a sequence in  $D$  such that  $d_{\mathcal{T}}(x_n, x) \rightarrow 0$ . Clearly, from the commute time identity (11), we have for any  $\varepsilon > 0$  that

$$P_x^{\mathcal{T}}(\tau_{x_n} > \varepsilon) \leq 2\varepsilon^{-1}d_{\mathcal{T}}(x_n, x) \rightarrow 0.$$

Hence, by the continuity of  $X^{\mathcal{T}}$ , there exists a deterministic sequence  $(\varepsilon_n)_{n \geq 0}$  with  $\varepsilon_n \downarrow 0$  such that

$$(43) \quad P_x^{\mathcal{T}}\left(X_{[0, \tau_{x_n}]}^{\mathcal{T}} \subseteq B_{\mathcal{T}}(x, \varepsilon_n)\right) \rightarrow 1.$$

For the next part of the proof, we define  $\theta_t$  to be the temporal shift (by time  $t$ ), so that  $\tau_x \circ \theta_{\tau_{x_n}}$  is the time taken to return to  $x$  after hitting  $x_n$ . Then, for  $\varepsilon, \eta > 0$ , we have that

$$\begin{aligned} & P_x^{\mathcal{T}}\left(B_{\mathcal{T}}(x, \varepsilon) \not\subseteq X_{[\tau_{x_n}, \tau_{\text{cov}}]}^{\mathcal{T}}\right) \\ & \leq P_x^{\mathcal{T}}\left(B_{\mathcal{T}}(x, \varepsilon) \not\subseteq X_{[\tau_{x_n} + \tau_x \circ \theta_{\tau_{x_n}}, \eta + \tau_{x_n} + \tau_x \circ \theta_{\tau_{x_n}}]}^{\mathcal{T}}\right) + P_x^{\mathcal{T}}(\eta + \tau_{x_n} + \tau_x \circ \theta_{\tau_{x_n}} > \tau_{\text{cov}}) \\ & \leq P_x^{\mathcal{T}}\left(B_{\mathcal{T}}(x, \varepsilon) \not\subseteq X_{[0, \eta]}^{\mathcal{T}}\right) + P_x^{\mathcal{T}}(2\eta > \tau_{\text{cov}}) + 2\eta^{-1}d_{\mathcal{T}}(x_n, x), \end{aligned}$$

where the first term has been simplified using the strong Markov property, and we have used the commute time identity again to control the probability of the event  $\tau_{x_n} + \tau_x \circ \theta_{\tau_{x_n}} \geq \eta$ . Next, since local times increase at the starting point immediately (i.e., (14)) holds, it is readily deduced from the continuity of the local times that,  $P_x^{\mathcal{T}}$ -a.s.,  $\forall \eta > 0$ ,  $X_{[0, \eta]}^{\mathcal{T}} \supseteq B_{\mathcal{T}}(0, \varepsilon)$  for some (random)  $\varepsilon > 0$  (cf. the argument of Lemma 2.4). Thus, for any  $\delta, \eta > 0$ , by choosing (deterministic)  $\varepsilon_{\eta}$  suitably small, it is possible to suppose that

$$(44) \quad P_x^{\mathcal{T}}\left(X_{[0, \eta]}^{\mathcal{T}} \not\supseteq B_{\mathcal{T}}(0, \varepsilon_{\eta})\right) \leq \delta.$$

Moreover, since  $\tau_{\text{cov}} > 0$ ,  $P_x^{\mathcal{T}}$ -a.s. (for which we assume that  $\mathcal{T}$  has at least two points to avoid the trivial case that arises otherwise), we have that  $P_x^{\mathcal{T}}(2\eta > \tau_{\text{cov}}) < \delta$  for suitably small  $\eta > 0$ . With this choice of  $\eta$  and  $\varepsilon_{\eta}$  satisfying (44), and taking  $n$  suitably large, it follows that

$$(45) \quad P_x^{\mathcal{T}}\left(B_{\mathcal{T}}(x, \varepsilon_{\eta}) \subseteq X_{[\tau_{x_n}, \tau_{\text{cov}}]}^{\mathcal{T}}\right) \geq 1 - 3\delta.$$

On the intersection of the events in the probabilities shown at (43) and (45), and with  $n$  suitably large, we have that

$$X_{[0, \tau_{x_n}]}^{\mathcal{T}} \subseteq B_{\mathcal{T}}(x, \varepsilon_n) \subseteq B_{\mathcal{T}}(x, \varepsilon_{\eta}) \subseteq X_{[\tau_{x_n}, \tau_{\text{cov}}]}^{\mathcal{T}},$$

which implies that  $\tau_{\text{cov}} = \tau_{x_n} + \tau_{\text{cov}} \circ \theta_{\tau_{x_n}}$ . Consequently, applying the conclusion of Lemma 5.6, we have that  $P_x^{\mathcal{T}}$ -a.s. on the same event,

$$\begin{aligned} \inf\{t : L_t^{\mathcal{T}}(y) > 0, \forall y \in \mathcal{T}\} & \leq \inf\{t \geq \tau_{x_n} : L_t^{\mathcal{T}}(y) - L_{\tau_{x_n}}^{\mathcal{T}}(y) > 0, \forall y \in \mathcal{T}\} \\ & = \tau_{x_n} + \tau_{\text{cov}} \circ \theta_{\tau_{x_n}} \\ & = \tau_{\text{cov}}. \end{aligned}$$

Since this inequality holds with probability greater than  $1 - 4\delta$ , and  $\delta$  was arbitrary, it therefore holds  $P_x^{\mathcal{T}}$ -a.s. Since the reverse inequality (i.e. (22)) is also true (as can be shown by the same simple continuity argument), we obtain that in fact (2) holds  $P_x^{\mathcal{T}}$ -a.s.

Finally, by Lemma 5.2 and Lemma 5.6 (and the scaling property of (30)), we know that the assumptions on  $\mathcal{T}$  hold  $\mathbf{P}$ -a.s. for the Brownian CRT, and so we are done.  $\square$

**6. Convergence to the cover time of the Brownian CRT.** The aim of this section is to prove Theorem 1.2 and Corollary 1.3. The basic strategy is as set out for the Sierpiński gasket in [28], see [28, Corollary 7.3 and Remark 7.4] in particular, though in that paper the equivalent to the conclusion of Theorem 1.1 was not known.

We start by presenting a convenient embedding result, which characterises convergence in the space  $\mathbb{M}_{L,c}$ . The result is essentially a simplification of [61, Theorem 2.28], which also covered the non-compact case. The notation  $d_M^H$ ,  $d_M^P$  and  $d_M^L$  was introduced in Section 5.1.

LEMMA 6.1 ([61]). *Let  $(K_n, d_{K_n}, \mu_{K_n}, \rho_{K_n}, P^{K_n})$ ,  $n \geq 1$ , and  $(K, d_K, \mu_K, \rho_K, P^K)$  be elements of  $\mathbb{M}_{L,c}$  such that*

$$(46) \quad (K_n, d_{K_n}, \mu_{K_n}, \rho_{K_n}, P^{K_n}) \rightarrow (K, d_K, \mu_K, \rho_K, P^K)$$

*in  $\mathbb{M}_{L,c}$ . It is then the case that there exists a rooted compact metric space  $(M, d_M, \rho_M)$  and isometric embeddings  $\phi_n : K_n \rightarrow M$ ,  $n \geq 1$ , and  $\phi : K \rightarrow M$  such that*

$$(47) \quad \phi_n(\rho_{K_n}) = \rho_M = \phi(\rho_K),$$

$$(48) \quad d_M^H(\phi_n(K_n), \phi(K)) \rightarrow 0,$$

$$(49) \quad d_M^P(\mu_{K_n} \circ \phi_n^{-1}, \mu_K \circ \phi^{-1}) \rightarrow 0,$$

$$(50) \quad d_M^L(P^{K_n} \circ \phi_n^{-1}, P^K \circ \phi^{-1}) \rightarrow 0.$$

*Moreover, given (48), the final condition is equivalent to the existence of a coupling of  $P^{K_n} \circ \phi_n^{-1}$ ,  $n \geq 1$ , and  $P^K \circ \phi^{-1}$  under which, almost-surely,*

$$(51) \quad \left( \phi_n \left( X_t^{K_n} \right) \right)_{t \geq 0} \rightarrow \left( \phi \left( X_t^K \right) \right)_{t \geq 0}$$

*in  $D(\mathbb{R}_+, M)$  and also, for any  $T > 0$ ,*

$$(52) \quad \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\substack{x_n \in K_n, x \in K: \\ d_M(\phi_n(x_n), \phi(x)) \leq \delta}} \sup_{t \in [0, T]} \left| L_t^{K_n}(x_n) - L_t^K(x) \right| = 0.$$

PROOF. See [61, Theorem 2.28] for an embedding result that covers (47), (48), (49) and (50). The coupling at (51) and (52) follows from [61, Theorem 2.19] and Skorohod's representation theorem.  $\square$

REMARK 6.2. The main result of [61] provides conditions under which (46) holds. In addition to the convergence of spaces, this involves the metric entropy, as introduced in Section 5.1. In [61], it is shown that if  $(K_n, d_{K_n}, \mu_{K_n}, \rho_{K_n})$ ,  $n \geq 1$ , are elements of  $\check{\mathbb{F}}_c$  satisfying

$$(K_n, d_{K_n}, \mu_{K_n}, \rho_{K_n}) \rightarrow (K, d_K, \mu_K, \rho_K)$$

for some element of  $\mathbb{F}_c$  and there exists an  $\alpha \in (0, 1/2)$  such that

$$(53) \quad \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{k=m}^{\infty} N \left( K_n, d_n, 2^{-k} \right)^2 \exp \left( -2^{\alpha k} \right) = 0,$$

then  $(K, d_K, \mu_K, \rho_K) \in \check{\mathbb{F}}_c$  and (46) holds (with the laws  $P^{K_n}$ ,  $n \geq 1$ , and  $P^K$  being those naturally associated with the spaces in question). This is a (simplification to compact spaces of) [61, Theorem 1.7]. (See also [61, Theorem 1.9] for a version of this result concerning random spaces.) As discussed in [61], the metric entropy condition at (53) is somewhat natural,

and can be checked using lower volume estimates (see [61, Section 7]). It is similar to conditions used to deduce continuity of Gaussian processes, and ensures the tightness of the laws of local times in the appropriate space. We note that, in [61], for non-compact spaces, this metric entropy condition is imposed on balls of finite radius, and an additional non-explosion condition is assumed.

Applying the embedding result of Lemma 6.1, we prove the following adaptation of Theorem 1.2.

**THEOREM 6.3.** *Suppose  $(K_n, d_{K_n}, \mu_{K_n}, \rho_{K_n})$ ,  $n \geq 1$ , and  $(K, d_K, \mu_K, \rho_K)$  are elements of  $\check{\mathbb{F}}_c$  such that*

$$(K_n, d_{K_n}, \mu_{K_n}, \rho_{K_n}, P^{K_n}) \rightarrow (K, d_K, \mu_K, \rho_K, P^K)$$

*in  $\mathbb{M}_{L,c}$ . Moreover, suppose that, for each  $n$ , either  $K_n$  is at most a countable set or  $X^{K_n}$  is  $P^{K_n}$ -a.s. continuous, and that  $P^K$ -a.s.,  $X^K$  is continuous and*

$$(54) \quad \tau_{\text{cov}}(K) = \inf \{t \geq 0 : L_t^K(x) > 0, \forall x \in K\}.$$

*It is then the case that the law of  $\tau_{\text{cov}}(K_n)$  under  $P^{K_n}$  converges to that of  $\tau_{\text{cov}}(K)$  under  $P^K$ .*

**PROOF.** To prove the result, we will, as noted at the start of the section, generalise the argument of [28, Corollary 7.3]. By Lemma 6.1, we may assume that the spaces  $K_n$  and  $K$  have been isometrically embedded into  $(M, d_M)$  so that (47)-(50) hold, and that the measures  $P^{K_n} \circ \phi_n^{-1}$ ,  $n \geq 1$ , and  $P^K \circ \phi^{-1}$  have been coupled so that (51)-(52) hold. From now on in the proof, we will omit the mappings  $\phi_n$  and  $\phi$  from the notation to simplify the presentation.

Suppose  $t < \tau_{\text{cov}}(K)$ , which means that there exists an  $x \in K$  such that  $x$  is not contained in  $X_{[0,t]}^K$ . From the compactness of  $[0, t]$  and the continuity of  $X^K$ , there exists a  $\varepsilon > 0$  such that

$$X_{[0,t]}^K \cap B_K(x, \varepsilon) = \emptyset.$$

Since we have the almost-sure convergence of processes as at (51), it follows that, for large  $n$ ,

$$X_{[0,t]}^{K_n} \cap B_K(x, \varepsilon/2) = \emptyset.$$

However, we also obtain from (48) that, for large  $n$ ,

$$K_n \cap B_K(x, \varepsilon/2) \neq \emptyset.$$

Hence it must be the case that, for large  $n$ ,  $t \leq \tau_{\text{cov}}(K_n)$ . In particular, on the coupled probability space, we have shown that  $\liminf_{n \rightarrow \infty} \tau_{\text{cov}}(K_n) \geq \tau_{\text{cov}}(K)$  almost-surely.

Next we prove the opposite bound. Write  $\tilde{\tau}_{\text{cov}}(K)$  for the right-hand side of (54), i.e. the first time at which the local times  $L^K$  are strictly positive at each point of  $K$ , and suppose that  $t > \tilde{\tau}_{\text{cov}}(K)$ . Since local times are increasing in  $t$ , it must be the case that  $L_t^K(y) > 0$ , for every  $y \in K$ . Together with the joint continuity of local times, this implies that for some  $\varepsilon > 0$ ,  $L_t^K(y) > \varepsilon$  for every  $y \in K$ . Consequently, applying (48) and (52), we deduce that, for all large  $n$ , it is also the case that  $L_t^{K_n}(y) > \varepsilon/2$  for every  $y \in K_n$ . It therefore holds that  $t \geq \tilde{\tau}_{\text{cov}}(K_n)$ , where the latter quantity is defined analogously to  $\tilde{\tau}_{\text{cov}}(K)$ . Hence this argument yields that, on the coupled probability space,  $\limsup_{n \rightarrow \infty} \tilde{\tau}_{\text{cov}}(K_n) \leq \tilde{\tau}_{\text{cov}}(K)$ . Finally, if  $X^{K_n}$  is continuous, we can check that  $\tilde{\tau}_{\text{cov}}(K_n) \geq \tau_{\text{cov}}(K_n)$  as we did for trees at (22), and if  $K_n$  is at most a countable set, we have from the property of local times at (15) that  $\tilde{\tau}_{\text{cov}}(K_n) = \tau_{\text{cov}}(K_n)$ . (Neither of the arguments used in the previous sentence are specific to trees.) Since we also have (54), it follows that  $\limsup_{n \rightarrow \infty} \tau_{\text{cov}}(K_n) \leq \tau_{\text{cov}}(K)$ , and consequently the proof is complete.  $\square$

PROOF OF THEOREM 1.2. It is confirmed in [61, Corollary 8.1] that  $(\mathcal{T}, 2d_{\mathcal{T}}, \mu_{\mathcal{T}}, \rho) \in \tilde{\mathbb{F}}_c$ ,  $\mathbf{P}$ -a.s. Moreover, as discussed in Section 2, for  $\mathbf{P}$ -a.e. realisation of  $\mathcal{T}$ ,  $X^{\mathcal{T}}$  has continuous sample paths,  $\mathbf{P}$ -a.s. Hence the result follows on combining Theorems 1.1 and 6.3.  $\square$

PROOF OF COROLLARY 1.3. Since the space  $\mathbb{M}_{L,c}$  is a separable metric space (when equipped with the metric described in Section 5.1), the result readily follows from Theorem 1.1 on application of Skorohod's representation theorem.  $\square$

**7. Moments of cover times for critical Galton-Watson trees.** The aim of this section is to prove Corollary 1.4. Unless otherwise stated, we will suppose that  $(T_n)_{n \geq 1}$  is a sequence of critical Galton-Watson trees, as described in the statement of that result. Given Corollary 1.3 and the discussion of Example 1, to obtain the desired conclusion, it will suffice to check the uniform boundedness of the  $L^p$  moments of the cover times in question. In particular, the main result of this section is as follows.

PROPOSITION 7.1. *In the setting of Corollary 1.4, it holds that, for  $p \geq 1$ ,*

$$\sup_{n \geq 1} n^{-3p/2} \mathbb{E}^{T_n} (\tau_{\text{cov}}(T_n))^p < \infty.$$

REMARK 7.2. In fact, our proof of the above result gives that

$$\sup_{n \geq 1} \mathbb{E}^{T_n} \left( e^{c\sqrt{n^{-3/2}\tau_{\text{cov}}(T_n)}} \right)$$

is finite for suitably small  $c > 0$ . However, we are not able to extend this to full exponential moments (i.e. we are not able to remove the square root).

In Proposition 7.1,  $\tau_{\text{cov}}(T_n)$  could be taken to be the cover time of the discrete-time or constant-speed random walk. To avoid constantly switching between these, we make the following observation, which will allow us to concentrate solely on the discrete-time version of the process for the remainder of the section. In the statement, we write  $\|\cdot\|_p^{T_n}$  for the  $L^p$ -norm under the measure  $\mathbb{P}^{T_n}$ . In fact the claim is true for random walks on graphs more generally, but we restrict to the specific case to avoid introducing new notation.

LEMMA 7.3. *In the setting of Corollary 1.4, if  $\tau_{\text{cov}}^D(T_n)$  is the cover time of the discrete-time random walk and  $\tau_{\text{cov}}^C(T_n)$  is the cover time of the constant-speed random walk (each started from the root), then, for  $p = q = 1$  or for  $1 < p < q$ , it holds that*

$$\|\tau_{\text{cov}}^C(T_n)\|_p^{T_n} \leq C_{p,q} \|\tau_{\text{cov}}^D(T_n)\|_q^{T_n},$$

where  $C_{p,q}$  is a constant depending only upon  $p$  and  $q$ .

PROOF. Since the CSRW has unit mean exponential holding times, we clearly have that, in distribution,

$$(55) \quad \tau_{\text{cov}}^C(T_n) = \sum_{i=1}^{\tau_{\text{cov}}^D(T_n)} \xi_i,$$

where  $(\xi_i)_{i \geq 1}$  is a sequence of independent and identically distributed unit mean exponential random variables, independent of  $\tau_{\text{cov}}^D(T_n)$ . Thus the result for  $p = q = 1$  is immediate (and

indeed an equality with  $C_{1,1} = 1$ ). For the case  $1 < p < q$ , we first note that, for any  $\lambda > 0$  and  $N \geq 1$ ,

$$\mathbf{P} \left( \sum_{i=1}^N \xi_i \geq (1 + \lambda)N \right) \leq \mathbf{E} \left( e^{\xi_1/2} \right)^N e^{-N(\lambda+1)/2} = 2^N e^{-N(\lambda+1)/2} = e^{-N(\lambda+1-2\log 2)/2},$$

where we have applied the fact that  $\mathbf{E}(e^{\theta\xi_1}) = (1 - \theta)^{-1}$  for  $\theta < 1$ . From this, supposing that  $\tau_{\text{cov}}^C(T_n)$  and  $\tau_{\text{cov}}^D(T_n)$  are coupled so that (55) holds almost-surely, it is straightforward to deduce that

$$\begin{aligned} \mathbb{P}^{T_n} \left( \frac{\tau_{\text{cov}}^C(T_n) - \tau_{\text{cov}}^D(T_n)}{\tau_{\text{cov}}^D(T_n)} \geq \lambda \right) &= \mathbb{P}^{T_n} \left( \tau_{\text{cov}}^C(T_n) \geq (1 + \lambda)\tau_{\text{cov}}^D(T_n) \right) \\ &\leq \mathbb{E}^{T_n} \left( \min \left\{ 1, e^{-\tau_{\text{cov}}^D(T_n)(\lambda+1-2\log 2)/2} \right\} \right) \\ &\leq \min \left\{ 1, e^{-(\lambda+1-2\log 2)/2} \right\}, \end{aligned}$$

where for the final inequality, we use the obvious bound  $\tau_{\text{cov}}^D(T_n) \geq 1$ . Since the right-hand side is independent of  $n$  and decays exponentially in  $\lambda$ , it follows that, for  $r \geq 1$ ,

$$\sup_{n \geq 1} \left\| \frac{\tau_{\text{cov}}^C(T_n) - \tau_{\text{cov}}^D(T_n)}{\tau_{\text{cov}}^D(T_n)} \right\|_r^{T_n} \leq C_r,$$

where  $C_r$  is a constant only depending on  $r$ . Hence, applying the triangle inequality and Hölder's inequality, we find that

$$\begin{aligned} \left\| \tau_{\text{cov}}^C(T_n) \right\|_p^{T_n} &\leq \left\| \tau_{\text{cov}}^C(T_n) - \tau_{\text{cov}}^D(T_n) \right\|_p^{T_n} + \left\| \tau_{\text{cov}}^D(T_n) \right\|_p^{T_n} \\ &\leq \left\| \frac{\tau_{\text{cov}}^C(T_n) - \tau_{\text{cov}}^D(T_n)}{\tau_{\text{cov}}^D(T_n)} \right\|_{\frac{pq}{q-p}}^{T_n} \left\| \tau_{\text{cov}}^D(T_n) \right\|_q^{T_n} + \left\| \tau_{\text{cov}}^D(T_n) \right\|_p^{T_n} \\ &\leq C_{p,q} \left\| \tau_{\text{cov}}^D(T_n) \right\|_q^{T_n}, \end{aligned}$$

as desired.  $\square$

To prove Proposition 7.1 for the discrete-time random walks, we will apply a general estimate for the expected cover time of a random walk on a finite graph from [20]. To state this, we introduce the notation

$$t_{\text{cov}}(T_n) := \sup_{x \in T_n} E_x^{T_n}(\tau_{\text{cov}}(T_n))$$

for the quenched expectation of  $\tau_{\text{cov}}(T_n)$ , starting from the worst possible vertex. Similarly to the comment preceding the previous lemma, the result of [20] is for general graphs, but we summarise it for the graphs  $(T_n)_{n \geq 1}$  for simplicity of presentation. We recall the notation  $N(K, R_K, \varepsilon)$  for the minimal size of an  $\varepsilon$ -cover of a metric space  $(K, R_K)$  from (39).

LEMMA 7.4 ([20, Theorem 1.1]). *For  $i, n \geq 1$ , set*

$$A_i^n := N(T_n, d_n, 2^{-i}D_n),$$

where  $d_n$  is the shortest path graph metric on  $T_n$  (and so also the resistance metric associated with unit conductances) and  $D_n$  is the diameter of  $(T_n, d_n)$ . Then, there exists a universal constant such that

$$t_{\text{cov}}(T_n) \leq C \left( \sum_{i=1}^{\log_2 \log n} \sqrt{2^{-i} \log A_i^n} \right) n D_n.$$

Naturally, to appeal to the above result, we will need to provide bounds on the diameter and cover sizes of  $T_n$ ,  $n \geq 1$ . These are provided by the next two lemmas. Note that we only require a finite variance assumption for the first of these, which was essentially proved in [7].

LEMMA 7.5 ([7]). *Let  $T_n$  be the tree generated by a Galton-Watson process whose offspring distribution is non-trivial, critical (mean one) and has finite variance, conditioned to have size  $n$ . Writing  $D_n$  for the diameter of  $T_n$  with respect to its shortest path metric, it then holds that, for all  $\lambda \geq 1$ ,*

$$\sup_{n \geq 2} \mathbf{P} \left( D_n \notin \left[ \lambda^{-1} n^{1/2}, \lambda n^{1/2} \right] \right) \leq C e^{-c\lambda^2},$$

where  $C$  and  $c$  are constants.

PROOF. A bound of the form  $Ce^{-c\lambda^2}$  on the probability that the height of  $T_n$  is greater than  $\lambda n^{1/2}$  is provided by [7, Theorem 1.2]. Since the diameter is less than twice the height, this gives one half of the result. Moreover, as is explained at the end of [7, Section 1], it follows from [7, Theorem 1.1] that one may also bound by  $Ce^{-c\lambda^2}$  the probability that the height of  $T_n$  is less than  $\lambda^{-1} n^{1/2}$ . Since the diameter is greater than the height, this completes the proof.  $\square$

We now give our estimate on the distribution of  $A_i^n$ . This is the only part of the proof in which exponential moments of the offspring distribution are required; see Remark 7.7 below for further comment on this point. If one could establish the bound that appears at (56) below under weaker conditions, then the extension of Proposition 7.1 would also follow.

LEMMA 7.6. *In the setting of Corollary 1.4, it holds that, for every  $\alpha < 1/2$ ,*

$$\mathbf{P} (A_i^n \geq \lambda) \leq C 2^i \lambda^{-\alpha},$$

for all  $i, n, \lambda \geq 1$ , where  $C$  is a constant independent of the latter three variables.

PROOF. We start by considering the contour function  $C_n = (C_n(t))_{t \in [0, 2(n-1)]}$  of  $T_n$ . In particular,  $C_n$  records the  $d_n$ -distance from the root of a depth-first walker as it traverses the outline of  $T_n$  from left to right at unit speed; notice that the walker visits every vertex apart from the root a number of times given by its degree. (We extend  $C_n$  to be a continuous time process by linear interpolation.) See [54] for background, including a detailed definition of this process. Let  $\tilde{C}_n = (\tilde{C}_n(t))_{t \in [0, 1]}$  be the normalised version of  $C_n$  obtained by setting

$$\tilde{C}_n(t) = n^{-1/2} C_n(2(n-1)t).$$

Now, it is established in [41] (see equation (1) of that article in particular) that, under the assumptions of the lemma,

$$\mathbf{P} \left( \left| \tilde{C}_n(s) - \tilde{C}_n(t) \right| \geq \varepsilon \right) \leq \frac{c_1}{|s-t|} \exp \left( -c_2 \frac{\varepsilon}{\sqrt{|s-t|}} \right),$$

for all  $s, t \in [0, 1]$  and  $\varepsilon > 0$ , where  $c_1$  and  $c_2$  are constants that do not depend on  $s, t, \varepsilon$  or  $n$ . (In [41], the bound is written for  $s, t \in [0, 1] \cap (2(n-1))^{-1}\mathbb{Z}$ , but there is no problem in extending to the full range.) Integrating this gives that, for any  $p \geq 1$ ,

$$\mathbf{E} \left( \left| \tilde{C}_n(s) - \tilde{C}_n(t) \right|^p \right) \leq c_p |s-t|^{p/2-1},$$

where  $c_p$  depends on  $p$ , but not  $s, t$  or  $n$ . This is the generalisation of Kolmogorov's criterion, as considered in [65, Theorem I.2.1], and therein taking  $\gamma = 2p, d = 1, \varepsilon = p - 2$  with  $p$  large, we deduce that

$$(56) \quad \mathbf{E} \left( \sup_{s,t: |s-t| \leq \lambda^{-1}} \left| \tilde{C}_n(s) - \tilde{C}_n(t) \right| \right) \leq c\lambda^{-\alpha},$$

for every  $\alpha < 1/2$ , where the constant  $c$  does not depend on  $n$  or  $\lambda > 0$ .

We next explain how to transfer the bound at (56) to an estimate on the covering sizes for  $T_n$ . For  $1 \leq k \leq 2(n-1)$ , let  $\chi_n^k$  be the set

$$\left\{ i \left\lfloor \frac{2(n-1)}{k} \right\rfloor : i \in \mathbb{Z} \right\} \cap [0, 2(n-1)].$$

Note that  $|\chi_n^k| < 3k$  for  $n \geq k + 1$ . Moreover, let  $T_n^k$  be those vertices in  $T_n$  that the depth-first walker visits at a time in  $\chi_n^k$ . From the definition of the contour process, it is easy to see that every vertex of  $T_n$  is within a distance

$$2n^{1/2} \sup_{s,t: |s-t| \leq \frac{1}{2(n-1)} \lfloor \frac{2(n-1)}{k} \rfloor} \left| \tilde{C}_n(s) - \tilde{C}_n(t) \right|$$

of a vertex in  $T_n^k$ . In particular, if the above quantity is smaller than  $\varepsilon n^{1/2}$  (and  $n \geq k + 1$ ), then

$$N(T_n, d_n, \varepsilon n^{1/2}) < 3k.$$

Finally, if  $D_n \geq \gamma^{-1} n^{1/2}$ , then  $A_i^n \leq N(T_n, d_n, 2^{-i} \gamma^{-1} n^{1/2})$ , and so from the observation of the previous paragraph and Lemma 7.5 we deduce that, for  $n \geq \lambda$ ,

$$(57) \quad \begin{aligned} \mathbf{P}(A_i^n \geq \lambda) &\leq \mathbf{P}(D_n < \gamma^{-1} n^{1/2}) + \mathbf{P}(N(T_n, d_n, 2^{-i} \gamma^{-1} n^{1/2}) \geq \lambda) \\ &\leq C e^{-c\gamma^2} + \mathbf{P} \left( 2 \sup_{s,t: |s-t| \leq \frac{1}{2(n-1)} \lfloor \frac{6(n-1)}{\lambda} \rfloor} \left| \tilde{C}_n(s) - \tilde{C}_n(t) \right| > 2^{-i} \gamma^{-1} \right) \\ &\leq C e^{-c\gamma^2} + C 2^i \gamma \lambda^{-\alpha}, \end{aligned}$$

where we have applied (56) to deduce the final inequality. Obviously the second probability on the right-hand side is zero for  $6n < \lambda$ . Moreover, for  $\lambda \in [n, 6n]$ , we have that, deterministically,

$$\sup_{s,t: |s-t| \leq \frac{1}{2(n-1)} \lfloor \frac{6(n-1)}{\lambda} \rfloor} \left| \tilde{C}_n(s) - \tilde{C}_n(t) \right| \leq 6n^{-1/2} \leq C\lambda^{-\alpha},$$

and so we can obtain a bound of the form  $C 2^i \gamma \lambda^{-\alpha}$  for the probability in question by applying Markov's inequality. In particular, the bound at (57) holds for all  $i, \lambda, \gamma \geq 1$  and  $n \geq 2$ . (The bound on the term involving the diameter is also satisfied for such parameters.) To complete the proof, we choose  $\gamma = \gamma(\lambda) \sim \log \lambda$ , which (after adjusting  $\alpha$ ) yields the desired conclusion.  $\square$

**REMARK 7.7.** In fact, in the reference we cite in proving (56), i.e. [41], it is claimed that a finite variance condition on the offspring distribution is sufficient for the relevant estimate to hold, but in [60, footnote (2)], it is noted that the proof in [41] uses exponential moments. To further elaborate on where the deficiency lies, in [41], on several occasions, results are

used from another paper [40] ([41, reference 3]), which studies the contour function of simply generated random trees. In the latter (especially in Theorem 1.1, as well as in Lemma 3.1), there is an assumption that [41, equation (3.1)] has a minimal positive solution  $\tau < R$ . For conditioned trees, as in our setting, [40, equation (3.1)] designates that the simply generated random tree is associated with a critical Galton-Watson branching process with finite variance. A minimal positive solution  $\tau < R$  is equivalent to the moment generating function of the offspring distribution having a radius of convergence equal to  $R/\tau > 1$ , and hence the distribution having exponential moments.

Putting together the three previous results, we arrive at the following conclusion, which is a refinement of a bound originally proved in [20, Theorem 3.2].

LEMMA 7.8. *In the setting of Corollary 1.4, it holds that*

$$\mathbf{P}\left(n^{-3/2}t_{\text{cov}}(T_n) \geq \lambda\right) \leq Ce^{-c\lambda},$$

for all  $n, \lambda \geq 1$ , where the constants are independent of the latter two variables.

PROOF. The proof is an adaptation of the proof of [20, Theorem 3.2]. From Lemma 7.4, we have that, for  $\lambda, \gamma \geq 1$ ,

$$\begin{aligned} \mathbf{P}\left(n^{-3/2}t_{\text{cov}}(T_n) \geq \lambda\right) &\leq \mathbf{P}\left(C \left(\sum_{i=1}^{\log_2 \log n} \sqrt{2^{-i} \log A_i^n}\right) n^{-1/2} D_n \geq \lambda\right) \\ &\leq \mathbf{P}\left(C \left(\sum_{i=1}^{\log_2 \log n} \sqrt{2^{-i} \log A_i^n}\right) \geq \lambda/\gamma\right) + \mathbf{P}\left(D_n \geq \gamma n^{1/2}\right). \end{aligned}$$

By Lemma 7.5, the second term is bounded above by  $Ce^{-c\gamma^2}$ . Using a union bound and Lemma 7.6, the first term can be estimated from above by

$$\sum_{i=1}^{\infty} \mathbf{P}\left(2^{-i} \log A_i^n \geq \frac{(2^{1/4} - 1)\lambda^2}{2^{i/2}\gamma^2}\right) \leq C \sum_{i=1}^{\infty} 2^i e^{-c\lambda^2\gamma^{-2}2^{i/2}}.$$

Hence, setting  $\gamma = \lambda^{1/2}$ , we have shown that

$$\mathbf{P}\left(n^{-3/2}t_{\text{cov}}(T_n) \geq \lambda\right) \leq Ce^{-c\lambda} + C \sum_{i=1}^{\infty} 2^i e^{-c\lambda 2^{i/2}},$$

from which the result follows  $\square$

To replace  $t_{\text{cov}}(T_n)$  by  $\tau_{\text{cov}}(T_n)$  in the conclusion of the above lemma, we apply the following exponential tail bound for the cover time of discrete-time random walks on general trees. We will write a general finite graph tree as  $T$ , and suppose the associated random walk is started from a root vertex. The result is an adaptation of the concentration result of [33, Corollary 1.3].

LEMMA 7.9. *There exist universal constants  $c_1, c_2$  such that, for any finite rooted graph tree  $T$ ,*

$$P^T(\tau_{\text{cov}}(T) \geq \lambda t_{\text{cov}}(T)) \leq c_1 e^{-c_2 \lambda}, \quad \forall \lambda \geq 1.$$

PROOF. It is proved as [33, Theorem 1.2] that there exist universal constants  $c_1, c_2$  such that, for any finite rooted graph tree  $T$ ,

$$(58) \quad P^T \left( |\tau_{\text{cov}}(T) - E(T)\eta_T^2| \geq \lambda E(T)\eta_T \sqrt{\text{diam}(T)} \right) \leq c_1 e^{-c_2 \lambda}, \quad \forall \lambda \geq 1,$$

where  $\eta_T$  is the expected maximum of the Gaussian free field on  $T$  that takes the value 0 at the root  $\rho$  and has covariance matrix  $(\frac{1}{2}(d_T(\rho, x) + d_T(\rho, y) - d_T(x, y)))_{x, y \in T}$ , with  $d_T$  being the shortest path graph distance on  $T$ . We have also written  $E(T)$  for the number of edges of  $T$  and  $\text{diam}(T)$  for the diameter of  $(T, d_T)$ . (Precisely, [33, Theorem 1.2] is written for the CSRW, but as is explained in [33, Section 1.2], the result also holds for the discrete-time random walk. Moreover, the main result of [34] yields that there exist further universal constants  $c_3, c_4$  such that, for any finite rooted graph tree  $T$ ,

$$(59) \quad c_3 E(T)\eta_T^2 \leq t_{\text{cov}}(T) \leq c_4 E(T)\eta_T^2.$$

(Clearly, there is no distinction between  $t_{\text{cov}}(T)$  for the CSRW and the discrete-time random walk, and so the result is applicable to either.) Combining (58) and (59), we obtain that there exists a universal constant  $c_5$  such that

$$P^T \left( \tau_{\text{cov}}(T) \geq c_5 \left( t_{\text{cov}}(T) + \lambda \sqrt{t_{\text{cov}}(T)E(T)\text{diam}(T)} \right) \right) \leq c_1 e^{-c_2 \lambda}, \quad \forall \lambda \geq 1.$$

By applying that  $t_{\text{cov}}(T) \geq E(T)\text{diam}(T)$  (which is an elementary consequence of the commute time identity for random walks, see [56], for example), the result follows.  $\square$

We are now ready to conclude the proof of the desired result.

PROOF OF PROPOSITION 7.1. We first observe that, for any  $\lambda \geq \gamma \geq 1$ ,

$$\begin{aligned} P^{T_n} \left( n^{-3/2} \tau_{\text{cov}}(T_n) \geq \lambda \right) &\leq \mathbf{1}_{\{n^{-3/2} t_{\text{cov}}(T_n) \geq \lambda/\gamma\}} + P^T \left( \tau_{\text{cov}}(T_n) \geq \gamma t_{\text{cov}}(T_n) \right) \\ &\leq \mathbf{1}_{\{n^{-3/2} t_{\text{cov}}(T_n) \geq \lambda/\gamma\}} + c_1 e^{-c_2 \gamma}, \end{aligned}$$

where  $c_1, c_2$  are the universal constants of Lemma 7.9. Taking expectations with respect to  $\mathbf{P}$ , Lemma 7.8 allows us to deduce from this that, for any  $\varepsilon > 0$ , there exists a  $c_3$  such that

$$\mathbb{P}^{T_n} \left( n^{-3/2} \tau_{\text{cov}}(T) \geq \lambda \right) \leq C e^{-c\lambda/\gamma} + c_1 e^{-c_2 \gamma}.$$

On setting  $\gamma = \lambda^{1/2}$ , this readily implies the result.  $\square$

**8. Cover-and-return times.** In this final section, we consider cover-and-return times, and in particular prove Corollary 1.5. We will again work in the general setting of Section 6. Given our other preparations, the following result will readily yield parts (a) and (b) of the result.

LEMMA 8.1. *In the setting of Theorem 6.3, the law of  $\tau_{\text{cov}}^+(K_n)$  under  $P^{K_n}$  converges to that of  $\tau_{\text{cov}}^+(K)$  under  $P^K$ .*

PROOF. As in the proof of Theorem 6.3, we will assume that the spaces  $K_n$  and  $K$  have been isometrically embedded into  $(M, d_M)$  so that (47)-(50) hold, and that the measures  $P^{K_n} \circ \phi_n^{-1}$ ,  $n \geq 1$ , and  $P^K \circ \phi^{-1}$  have been coupled so that (51)-(52) hold. Also as there, we omit the mappings  $\phi_n$  and  $\phi$  from the notation.

Now, we have from the proof of Theorem 6.3 that, on the coupled space,  $\tau_{\text{cov}}(K_n) \rightarrow \tau_{\text{cov}}(K)$ , almost-surely. Since the processes converge as at (51) and the limit process is continuous, it follows that

$$(60) \quad X_{\tau_{\text{cov}}(K_n)}^{K_n} \rightarrow X_{\tau_{\text{cov}}(K)}^K$$

in  $(M, d_M)$ , almost-surely.

We next claim that if  $x_n \in K_n$ ,  $n \geq 1$ , and  $x \in K$  are such that  $d_M(x_n, x) \rightarrow 0$ , then the law of  $\tau_{\rho_{K_n}}^{K_n}$  under  $P_{x_n}^{K_n}$  converges to that of  $\tau_{\rho_K}^K$  under  $P_x^K$ , where we write  $\tau_{\rho_K}^K$  for the hitting time of  $\rho_K$  by  $X^K$  and  $P_x^K$  for the law of  $X^K$  started from  $x$  (and similarly for the  $n$ -indexed processes). The proof of this is an adaptation of the argument we applied in the first part of the proof of Lemma 2.5. Firstly, it is an easy extension of the conclusion of [29] to deduce that, with the spaces embedded as they are, the law of  $X^{K_n}$  under  $P_{x_n}^{K_n}$  converges to that of  $X^K$  under  $P_x^K$ . (In [29], processes were started from the root, but this is not an essential restriction, and indeed we can just consider the spaces rerooted at  $x_n$  and  $x$ , respectively.) Hence, we can suppose that we have a coupling so that this convergence holds almost-surely. Under such a coupling, we have from (51) that if  $\tau_{\rho_K}^K < t$ , then, for every  $\varepsilon > 0$ ,

$$(61) \quad \limsup_{n \rightarrow \infty} \tau_{\bar{B}_M(\rho_M, \varepsilon)}^{K_n} < t,$$

where  $\tau_{\bar{B}_M(\rho_M, \varepsilon)}^{K_n}$  is the hitting time of the relevant closed ball in  $(M, d_M)$  by  $X^{K_n}$ . Applying the strong Markov property at time  $\tau_{\bar{B}_M(\rho_M, \varepsilon)}^{K_n}$ , it follows that, for all  $\varepsilon, \delta > 0$ ,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} P_{x_n}^{K_n} \left( \tau_{\rho_{K_n}}^{K_n} \geq t \right) \\ & \leq \limsup_{n \rightarrow \infty} \left( P_{x_n}^{K_n} \left( \tau_{\bar{B}_M(\rho_M, \varepsilon)}^{K_n} \geq t - \delta \right) + \sup_{y \in \bar{B}_M(\rho_M, \varepsilon) \cap K_n} P_y^{K_n} \left( \tau_{\rho_{K_n}}^{K_n} \geq \delta \right) \right) \\ & \leq P_x^K \left( \tau_{\rho_K}^K \geq t - \delta \right) + \varepsilon \delta^{-1} \mu_K(K), \end{aligned}$$

where for the final inequality follows from (49), (61) and the commute time identity for stochastic processes associated with resistance forms (see [29, Lemma 2.4], for example). Taking  $\varepsilon \rightarrow 0$  and then  $\delta \rightarrow 0$ , we find that

$$\limsup_{n \rightarrow \infty} P_{x_n}^{K_n} \left( \tau_{\rho_{K_n}}^{K_n} \geq t \right) \leq P_x^K \left( \tau_{\rho_K}^K \geq t \right).$$

Conversely, if  $\tau_{\rho_K}^K > t$ , then, by continuity, there exists an  $\varepsilon > 0$  such that  $X_{[0, t+\varepsilon]}^K \cap \bar{B}_M(\rho_M, \varepsilon) = \emptyset$ . Hence, from (51), we deduce that  $X_{[0, t+\varepsilon]}^{K_n} \cap \bar{B}_M(\rho_M, \varepsilon/2) = \emptyset$  for large  $n$ , and consequently,  $\liminf_{n \rightarrow \infty} \tau_{\rho_{K_n}}^{K_n} > t$ . Hence  $\liminf_{n \rightarrow \infty} \tau_{\rho_{K_n}}^{K_n} \geq \tau_{\rho_K}^K$ , which implies

$$\liminf_{n \rightarrow \infty} P_{x_n}^{K_n} \left( \tau_{\rho_{K_n}}^{K_n} > t \right) \geq \liminf_{n \rightarrow \infty} P_x^K \left( \tau_{\rho_K}^K > t \right).$$

Combining this and the corresponding upper bound confirms the claim.

Finally, together with the convergence at (60), the result of the previous paragraph readily implies that, almost-surely on our coupled probability space, we have the following weak convergence statement:

$$\begin{aligned} & P^{K_n} \left( \tau_{\text{cov}}^+(K_n) - \tau_{\text{cov}}(K_n) \in \cdot \mid \left( X_t^{K_n} \right)_{t \leq \tau_{\text{cov}}(K_n)} \right) \\ & \rightarrow P^K \left( \tau_{\text{cov}}^+(K) - \tau_{\text{cov}}(K) \in \cdot \mid \left( X_t^K \right)_{t \leq \tau_{\text{cov}}(K)} \right). \end{aligned}$$

Since we also know the cover times converge almost-surely on the coupled space, this is enough to complete the proof.  $\square$

To check the remaining part of Corollary 1.5, we strengthen Proposition 7.1 as follows. We stress that, as for Proposition 7.1, this holds for the discrete-time or constant-speed random walk.

LEMMA 8.2. *In the setting of Corollary 1.4, it holds that, for  $p \geq 1$ ,*

$$\sup_{n \geq 1} n^{-3p/2} \mathbb{E}^{T_n} (\tau_{\text{cov}}^+(T_n)^p) < \infty.$$

PROOF. Since

$$\tau_{\text{cov}}^+(T_n)^p \leq 2^{p-1} (\tau_{\text{cov}}(T_n)^p + (\tau_{\text{cov}}^+(T_n) - \tau_{\text{cov}}(T_n))^p),$$

and we already have moment bounds for  $\tau_{\text{cov}}(T_n)$  from Proposition 7.1, it will suffice to check corresponding moment bounds for  $\tau_{\text{cov}}^+(T_n) - \tau_{\text{cov}}(T_n)$ . Specifically, we will show that, for  $p \geq 1$ ,

$$(62) \quad \sup_{n \geq 1} n^{-3p/2} \mathbf{E} \left( \sup_{x, y \in T_n} E_x^{T_n} \left( (\tau_y^{T_n})^p \right) \right) < \infty,$$

from which the desired result will readily follow. To this end, we simply observe that, by applying the Markov property repeatedly, for  $k \in \mathbb{N}$ ,

$$\begin{aligned} \sup_{x, y \in T_n} P_x^{T_n} (\tau_y^{T_n} \geq 4knD_n) &\leq \sup_{x, y \in T_n} P_x^{T_n} (\tau_y^{T_n} \geq 4(k-1)nD_n) \sup_{x, y \in T_n} P_x^{T_n} (\tau_y^{T_n} \geq 4nD_n) \\ &\leq \sup_{x, y \in T_n} P_x^{T_n} (\tau_y^{T_n} \geq 4(k-1)nD_n) \sup_{x, y \in T_n} \frac{E_x^{T_n} (\tau_y^{T_n})}{4nD_n} \\ &\leq \frac{1}{2} \sup_{x, y \in T_n} P_x^{T_n} (\tau_y^{T_n} \geq 4(k-1)nD_n) \\ &\leq 2^{-k}, \end{aligned}$$

where we again write  $D_n$  for the diameter of  $(T_n, d_n)$  and have applied the commute time identity for random walks on graph (which applies to both the CSRW and the discrete-time random walk), see [56, Proposition 10.6], for example, to deduce the third inequality. From this, it is easy to conclude that, for  $\lambda \geq 1$ ,

$$(63) \quad \sup_{n \geq 1} \sup_{x, y \in T_n} P_x^{T_n} (\tau_y^{T_n} \geq \lambda n D_n) \leq C e^{-c\lambda},$$

where  $C$  and  $c$  are deterministic constants. Since we clearly have that

$$\begin{aligned} n^{-3p/2} \mathbf{E} \left( \sup_{x, y \in T_n} E_x^{T_n} \left( (\tau_y^{T_n})^p \right) \right) &= \mathbf{E} \left( \sup_{x, y \in T_n} E_x^{T_n} \left( \left( \frac{\tau_y^{T_n}}{nD_n} \right)^p \right) \times \left( \frac{D_n}{n^{1/2}} \right)^p \right) \\ &\leq \mathbf{E} \left( \sup_{x, y \in T_n} E_x^{T_n} \left( \left( \frac{\tau_y^{T_n}}{nD_n} \right)^{2p} \right) \right)^{1/2} \mathbf{E} \left( \left( \frac{D_n}{n^{1/2}} \right)^{2p} \right)^{1/2}, \end{aligned}$$

the result at (62) follows from (63) and Lemma 7.5.  $\square$

PROOF OF COROLLARY 1.5. With Lemma 8.1 in place of Theorem 6.3, parts (a) and (b) are checked in exactly the same way as Theorem 1.2 and Corollary 1.3 were. As for part (c), since we have Lemma 8.2, again this can be proved in the same way as the corresponding result for the cover time, i.e. Corollary 1.4.  $\square$

**Funding.** DC was supported by JSPS Grant-in-Aid for Scientific Research (C) 19K03540, 24K06758 and the Research Institute for Mathematical Sciences, an International Joint Usage/Research Center located in Kyoto University. LM was supported by the ANR grant ProGraM (ANR-19-CE40-0025).

## REFERENCES

- [1] ABE, Y. (2014). Cover times for sequences of reversible Markov chains on random graphs. *Kyoto J. Math.* **54** 555–576.
- [2] ABE, Y. (2015). Effective resistances for supercritical percolation clusters in boxes. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 935–946.
- [3] ABRAHAM, R. and DELMAS, J. F. (2009). Williams’ decomposition of the Lévy continuum random tree and simultaneous extinction probability for populations with neutral mutations. *Stochastic Process. Appl.* **119** 1124–1143.
- [4] 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.
- [5] ADDARIO-BERRY, L., BROUTIN, N. and GOLDSCHMIDT, C. (2010). Critical random graphs: limiting constructions and distributional properties. *Electron. J. Probab.* **15** no. 25, 741–775.
- [6] ADDARIO-BERRY, L., BROUTIN, N. and GOLDSCHMIDT, C. (2012). The continuum limit of critical random graphs. *Probab. Theory Related Fields* **152** 367–406.
- [7] ADDARIO-BERRY, L., DEVROYE, L. and JANSON, S. (2013). Sub-Gaussian tail bounds for the width and height of conditioned Galton-Watson trees. *Ann. Probab.* **41** 1072–1087.
- [8] ALBENQUE, M. and GOLDSCHMIDT, C. (2015). The Brownian continuum random tree as the unique solution to a fixed point equation. *Electron. Commun. Probab.* **20** no. 61, 14.
- [9] ALDOUS, D. J. (1991). The continuum random tree. I. *Ann. Probab.* **19** 1–28.
- [10] ALDOUS, D. J. (1991). The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*. *London Math. Soc. Lecture Note Ser.* **167** 23–70. Cambridge Univ. Press, Cambridge.
- [11] ALDOUS, D. J. (1991). Random walk covering of some special trees. *J. Math. Anal. Appl.* **157** 271–283.
- [12] ALDOUS, D. J. (1991). Threshold limits for cover times. *J. Theoret. Probab.* **4** 197–211.
- [13] ALDOUS, D. J. (1993). The continuum random tree. III. *Ann. Probab.* **21** 248–289.
- [14] ALDOUS, D. J. (1994). Recursive self-similarity for random trees, random triangulations and Brownian excursion. *Ann. Probab.* **22** 527–545.
- [15] ANDRIPOPOULOS, G. (2023). Convergence of blanket times for sequences of random walks on critical random graphs. *Combin. Probab. Comput.* **32** 478–515. <https://doi.org/10.1017/s0963548322000359MR4574858>
- [16] ARCHER, E. and CROYDON, D. A. (2023). Scaling limit of critical percolation clusters on hyperbolic random half-planar triangulations and the associated random walks. preprint available at arXiv:2311.11993.
- [17] ARCHER, E., NACHMIAS, A. and SHALEV, M. (2024). The GHP scaling limit of uniform spanning trees in high dimensions. *Comm. Math. Phys.* **405** Paper No. 73.
- [18] ATHREYA, S., ECKHOFF, M. and WINTER, A. (2013). Brownian motion on  $\mathbb{R}$ -trees. *Trans. Amer. Math. Soc.* **365** 3115–3150.
- [19] ATHREYA, S., LÖHR, W. and WINTER, A. (2017). Invariance principle for variable speed random walks on trees. *Ann. Probab.* **45** 625–667.
- [20] BARLOW, M. T., DING, J., NACHMIAS, A. and PERES, Y. (2011). The evolution of the cover time. *Combin. Probab. Comput.* **20** 331–345.
- [21] BEN AROUS, G., CABEZAS, M. and FRIBERGH, A. (2019). Scaling limit for the ant in a simple high-dimensional labyrinth. *Probab. Theory Related Fields* **174** 553–646.
- [22] BEN AROUS, G., CABEZAS, M. and FRIBERGH, A. (2019). Scaling limit for the ant in high-dimensional labyrinths. *Comm. Pure Appl. Math.* **72** 669–763.
- [23] CABEZAS, M., FRIBERGH, A., HOLMES, M. and PERKINS, E. (2023). Historical lattice trees. *Comm. Math. Phys.* **401** 435–496.
- [24] CORLIN MARCHAND, D. (2023). Percolation on supercritical causal triangulations. preprint available at arXiv:2307.03746.
- [25] CROYDON, D. A. (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.
- [26] CROYDON, D. A. (2008). Volume growth and heat kernel estimates for the continuum random tree. *Probab. Theory Related Fields* **140** 207–238.

- [27] CROYDON, D. A. (2010). Scaling limits for simple random walks on random ordered graph trees. *Adv. in Appl. Probab.* **42** 528–558.
- [28] CROYDON, D. A. (2015). Moduli of continuity of local times of random walks on graphs in terms of the resistance metric. *Trans. London Math. Soc.* **2** 57–79.
- [29] CROYDON, D. A. (2018). Scaling limits of stochastic processes associated with resistance forms. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** 1939–1968.
- [30] CROYDON, D. A. and HAMBLY, B. M. (2008). Self-similarity and spectral asymptotics for the continuum random tree. *Stochastic Process. Appl.* **118** 730–754.
- [31] CROYDON, D. A., HAMBLY, B. M. and KUMAGAI, T. (2017). Time-changes of stochastic processes associated with resistance forms. *Electron. J. Probab.* **22** Paper No. 82.
- [32] DEMBO, A., PERES, Y., ROSEN, J. and ZEITOUNI, O. (2004). Cover times for Brownian motion and random walks in two dimensions. *Ann. of Math. (2)* **160** 433–464.
- [33] DING, J. (2014). Asymptotics of cover times via Gaussian free fields: bounded-degree graphs and general trees. *Ann. Probab.* **42** 464–496.
- [34] DING, J., LEE, J. R. and PERES, Y. (2012). Cover times, blanket times, and majorizing measures. *Ann. of Math. (2)* **175** 1409–1471.
- [35] DUQUESNE, T. and LE GALL, J. F. (2002). Random trees, Lévy processes and spatial branching processes. *Astérisque* **281** vi+147.
- [36] DUQUESNE, T. and LE GALL, J. F. (2005). Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields* **131** 553–603.
- [37] DUQUESNE, T. and WINKEL, M. (2007). Growth of Lévy trees. *Probab. Theory Related Fields* **139** 313–371.
- [38] FUKUSHIMA, M., OSHIMA, Y. and TAKEDA, M. (2011). *Dirichlet forms and symmetric Markov processes*, extended ed. *De Gruyter Studies in Mathematics* **19**. Walter de Gruyter & Co., Berlin.
- [39] GETTOOR, R. K. and KESTEN, H. (1972). Continuity of local times for Markov processes. *Compositio Math.* **24** 277–303.
- [40] GITTENBERGER, B. (1999). On the contour of random trees. *SIAM J. Discrete Math.* **12** 434–458. <https://doi.org/10.1137/S0895480195289928> MR1720416
- [41] GITTENBERGER, B. (2003). A note on: “State spaces of the snake and its tour—convergence of the discrete snake” [J. Theoret. Probab. **16** (2003), no. 4, 1015–1046; MR2033196] by J.-F. Marckert and A. Mokkadem. *J. Theoret. Probab.* **16** 1063–1067. <https://doi.org/10.1023/B:JOTP.0000012006.91251.e8> MR2033198
- [42] HARA, T. and SLADE, G. (2000). The scaling limit of the incipient infinite cluster in high-dimensional percolation. I. Critical exponents. *J. Statist. Phys.* **99** 1075–1168.
- [43] HERMON, J. (2020). A spectral characterization for concentration of the cover time. *J. Theoret. Probab.* **33** 2167–2184.
- [44] KALLENBERG, O. (2002). *Foundations of modern probability*, second ed. *Probability and its Applications (New York)*. Springer-Verlag, New York.
- [45] KHEZELI, A. (2023). A unified framework for generalizing the Gromov-Hausdorff metric. *Probab. Surv.* **20** 837–896.
- [46] KIGAMI, J. (1995). Harmonic calculus on limits of networks and its application to dendrites. *J. Funct. Anal.* **128** 48–86.
- [47] KIGAMI, J. (2001). *Analysis on fractals. Cambridge Tracts in Mathematics* **143**. Cambridge University Press, Cambridge.
- [48] KIGAMI, J. (2012). Resistance forms, quasisymmetric maps and heat kernel estimates. *Mem. Amer. Math. Soc.* **216** vi+132.
- [49] KINGMAN, J. F. C. (1993). *Poisson processes. Oxford Studies in Probability* **3**. The Clarendon Press, Oxford University Press, New York Oxford Science Publications.
- [50] KOLCHIN, V. F. (1986). *Random mappings. Translation Series in Mathematics and Engineering*. Optimization Software, Inc., Publications Division, New York Translated from the Russian, With a foreword by S. R. S. Varadhan.
- [51] KREBS, W. B. (1995). Brownian motion on the continuum tree. *Probab. Theory Related Fields* **101** 421–433.
- [52] KUMAGAI, T. (2004). Heat kernel estimates and parabolic Harnack inequalities on graphs and resistance forms. *Publ. Res. Inst. Math. Sci.* **40** 793–818.
- [53] LE GALL, J. F. (1999). *Spatial branching processes, random snakes and partial differential equations. Lectures in Mathematics ETH Zürich*. Birkhäuser Verlag, Basel.
- [54] LE GALL, J. F. (2006). Random real trees. *Ann. Fac. Sci. Toulouse Math. (6)* **15** 35–62.
- [55] LE GALL, J. F. (2018). Subordination of trees and the Brownian map. *Probab. Theory Related Fields* **171** 819–864.

- [56] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov chains and mixing times*. American Mathematical Society, Providence, RI With a chapter by James G. Propp and David B. Wilson.
- [57] LI, X. and ZHENG, Y. (2024). Inverting Ray-Knight identities on trees. *Electron. J. Probab.* **29** Paper No. 1.
- [58] LINDSTRØM, T. (1990). Brownian motion on nested fractals. *Mem. Amer. Math. Soc.* **83** iv+128.
- [59] MARCUS, M. B. and ROSEN, J. (2006). *Markov processes, Gaussian processes, and local times*. Cambridge Studies in Advanced Mathematics **100**. Cambridge University Press, Cambridge.
- [60] MARZOUK, C. (2020). Scaling limits of discrete snakes with stable branching. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 502–523.
- [61] NODA, R. (2023). Convergence of local times of stochastic processes associated with resistance forms. preprint available at arXiv:2305.13224.
- [62] NODA, R. (2024). Metrization of Gromov-Hausdorff-type topologies on boundedly-compact spaces. preprint available at arXiv:2404.19681.
- [63] NODA, R. (2024). Scaling limits of discrete-time Markov chains and their local times on electrical networks. preprint available at arXiv:2405.01871.
- [64] PORT, S. C. and STONE, C. J. (1978). *Brownian motion and classical potential theory*. Probability and Mathematical Statistics. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London.
- [65] REVUZ, D. and YOR, M. (1999). *Continuous martingales and Brownian motion*, third ed. *Grund-*lehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]** **293**. Springer-Verlag, Berlin.
- [66] ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions, Markov processes, and martingales*. Vol. 2. *Cambridge Mathematical Library*. Cambridge University Press, Cambridge Itô calculus, Reprint of the second (1994) edition.
- [67] WILLIAMS, D. (1974). Path decomposition and continuity of local time for one-dimensional diffusions. I. *Proc. London Math. Soc. (3)* **28** 738–768.