An invariance principle for branching diffusions in bounded domains

Ellen Powell

Abstract

We study branching diffusions in a bounded domain $D$ of $\mathbb{R}^d$ in which particles are killed upon hitting the boundary $\partial D$. It is known that any such process undergoes a phase transition when the branching rate $\beta$ exceeds a critical value: a multiple of the first eigenvalue of the generator of the diffusion. We investigate the system at criticality and show that the associated genealogical tree, when the process is conditioned to survive for a long time, converges to Aldous’ Continuum Random Tree under appropriate rescaling. The result holds under only a mild assumption on the domain, and is valid for all branching mechanisms with finite variance, and a general class of diffusions.

1 Introduction

This paper concerns branching diffusions in a bounded domain $D$ of $\mathbb{R}^d$. For us, these are processes in which individual particles move according to the law of some diffusion, are killed upon exiting the domain, and branch into a random number of particles (with distribution $A$, independent of position) at constant rate $\beta > 0$. Whenever such a branching event occurs each of the offspring then stochastically repeats the behaviour of its parent, starting from the point of fission, and independently of everything else. The configuration of particles at time $t$ will be represented by the $D$-valued point process $X_t = \{X_u(t) : u \in N_t\}$, where $N_t$ is the set of individuals alive at time $t$ (its size will be denoted by $|N_t|$). We write $\mathbb{P}^x$ for the law of the process initiated from a point $x \in D$. Finally, we will always assume that the offspring distribution $A$ satisfies

$$\mathbb{E}(A) = m > 1 \quad \text{and} \quad \mathbb{E}(A^2) < \infty,$$

and that the generator

$$L = \frac{1}{2} \sum_{i,j} \partial_{x_j}(a^{ij} \partial_{x_i})$$

of the diffusion is uniformly elliptic (see Section 2.1) with coefficients $a^{ij} = a^{ji} \in C^1(\bar{D})$ for $1 \leq i, j \leq d$.

It is known that such a system exhibits a phase transition in the branching rate: for large enough $\beta$ there is a positive probability of survival for all time, but for small $\beta$, including at criticality, there is almost sure extinction. The critical value of $\beta$ is equal to $\frac{\lambda}{m-1}$, where $\lambda$ is the first eigenvalue of $-L$ on $D$ with Dirichlet boundary conditions (see (2.1)). The main goal of this paper will be to study the system at criticality and find a scaling limit for the resulting genealogical tree. This is the continuous plane tree that is generated purely by the birth and death times of particles in the system, and encodes no information about the spatial motion.

More precisely, for $n \in \mathbb{N}$ we condition the critical branching diffusion to survive until time at least $n$, and look at the associated genealogical tree $T$, equipped with its natural distance $d$. Rescaling distances by a factor $n$ gives us a sequence of laws on random compact metric spaces:

$$(T_{n,x}, d_{n,x}) \overset{\text{law}}{=} (T, \frac{1}{n}d) \text{ under } \mathbb{P}^x(\cdot | |N_n| > 0).$$
We will prove that this sequence converges in distribution to a conditioned Brownian continuum random tree as \( n \to \infty \), with respect to the Gromov-Hausdorff topology. Indeed, if we let \( e \) be a Brownian excursion conditioned to reach height at least 1, and write \((T_e, d_e)\) for the real tree whose contour function is given by \( e \), then we obtain the following result.

**Theorem 1.1.** Suppose that \( D \subset \mathbb{R}^d \) is a \( C^1 \) domain\(^1\) and that \( L \) as in (1.2) is uniformly elliptic with coefficients

\[
a^{ij} = a^{ji} \in C^1(D) \quad \text{for} \quad 1 \leq i, j \leq d.\]

Further suppose that \( A \) satisfies (1.1), and that \( \varphi \in C^1(D) \) where \( \varphi \) is the first eigenfunction of \(-L\) on \( D \) (see Section 2.1). Then, at the critical branching rate \( \beta = \lambda/(m - 1) \), and for any starting point \( x \in D \),

\[
(T_{n,x}, d_{n,x}) \xrightarrow{n \to \infty} (T_e, d_e)
\]

in distribution, with respect to the Gromov-Hausdorff topology.

**Remark 1.2.** One sufficient condition to ensure that the hypotheses of Theorem 1.1 are satisfied is to assume that \( D \) is \( C^{2,\alpha} \) for some \( \alpha \in [0, 1] \) (see Lemma 2.3). However, this is also satisfied in many other cases, so we leave the assumptions of Theorem 1.1 as general as possible.

On the way to proving Theorem 1.1 we also obtain new proofs of several other results concerning critical branching diffusions, some of which are already known in various forms. The reason for detailing these proofs here is threefold: firstly, it allows us to pin down the regularity required on the domain \( D \); secondly, it provides a new and somewhat more probabilistic approach to the theory, that we believe is interesting in its own right; and finally, the proofs serve to introduce many concepts and ideas that are crucial for the proof of Theorem 1.1.

Let us first look at the phase transition. This result was originally proved by Sevast’yanov [Sev58] and Watanabe [Wat65], in the case when \( L \) is a constant multiple of \( \Delta \). However, it has also been reworked and generalised since then. In [AH83, Chapter 6], a more general version of the result is given for branching Markov processes whose moment semigroup satisfies a certain criterion. One of the main examples discussed is when the process is a branching diffusion on a manifold with killing at the boundary. This is slightly more general than the set up of the present paper, in that the diffusion is on a manifold and the branching mechanism is allowed to be spatially dependent, however, a (fairly abstract) condition on the moment semigroup is required. In [Her78b] the criterion is shown to be satisfied, for example, if the manifold has \( C^1 \) boundary and the generator of the diffusion is uniformly elliptic with \( C^2 \) coefficients. Here we will prove the result under a weaker assumption on the domain and generator, but in our slightly more specific framework. Related results have also been shown in [EK04], which studies local extinction versus local growth on compactly contained subsets of a (possibly infinite) domain \( D \), and in [HHK12], where \( D \) is taken to be a bounded interval of the real line.

**Theorem 1.3** ([Sev58], [Wat65]). Let \( D \subset \mathbb{R}^d \) be a bounded domain with Lipschitz boundary and suppose that \( L \) and \( A \) are as in Theorem 1.1. Then, for any starting position \( x \in D \), if \( \lambda \) is the first eigenvalue of \(-L\) on \( D \) with Dirichlet boundary conditions then,

- for \( \beta > \frac{\lambda}{m-1} \) the process survives for all time with positive \( \mathbb{P}^x \)-probability.

- for \( \beta \leq \frac{\lambda}{m-1} \) the process becomes extinct \( \mathbb{P}^x \)-almost surely.

Moreover, if \( \beta \leq \frac{\lambda}{m-1} \) then \( \mathbb{P}^x(|N_t| > 0) \to 0 \) uniformly in \( D \).

The rest of the paper will focus on the behaviour of the system at criticality, starting with an asymptotic for the survival probability. The results of Theorems 1.4, Theorem 1.6 and Corollary 1.7 have also been shown in [AH83, Chapter 6] in the same framework discussed above (see also [Her78a, Her78b], and [Her74] for the one-dimensional case). Here we provide new proofs, which hold under fewer assumptions on the domain \( D \), and which we can also modify to give key ingredients for the proof of Theorem 1.1 (see for example Lemma 5.3).

\(^1\)We say that \( D \subset \mathbb{R}^d \) is a \{Lipschitz / \( C^k \) / \( C^{k,\alpha} \)\} domain (for \( k \in \mathbb{Z}_{\geq 0} \) and \( \alpha \in [0, 1] \)) if, at each point \( x_0 \in \partial D \), there exists \( r > 0 \) and a \{Lipschitz / \( C^k \) / \( C^{k,\alpha} \)\} function \( \gamma : \mathbb{R}^{d-1} \to \mathbb{R} \) such that relabelling and reorienting axes as necessary, \( D \cap B(x_0, r) = \{x \in B(x_0, r) : \gamma(x) > \gamma(x^n, \cdots, x^{d-1})\} \).
Theorem 1.4. Suppose that $D$ is $C^1$ and that $L$ and $A$ are as in Theorem 1.3. Then, in the critical case $\beta = \frac{\lambda}{m-1}$, for all $x \in D$ we have

$$
\mathbb{P}^x(|N_t| > 0) \sim \frac{1}{t} \times \frac{2(m-1)\varphi(x)}{\lambda(E[A^2] - E[A])} \int_D \varphi(y)^3 \, dy
$$

(1.4)
as $t \to \infty$. Here $\varphi$ is the first eigenfunction of $L$ on $D$, normalised to have unit $L^2$ norm.

This asymptotic then allows us to study the behaviour of the system when it is conditioned to survive for a long time, which is important for the proof of Theorem 1.1. One tool that we will use is a classical spine change of measure, under which the process has a distinguished particle, the spine, which is conditioned to remain in $D$ forever (as in [Pin85]). Along this spine, families of ordinary critical branching diffusions immigrate at rate $\frac{m}{m-1}\lambda$ according to a biased offspring distribution. Note that there is no extinction under this new measure, which we denote by $Q^x$. We will prove that changing measure in this way is in fact somewhat close to conditioning on survival for all time, in the sense of the following proposition.

Proposition 1.5. Assume the hypotheses of Theorem 1.4. Then for any $T \geq 0$, $x \in D$ and $B \in \mathcal{F}_T$, where $\mathcal{F}$ is the natural filtration of the process, we have

$$
\lim_{t \to \infty} \mathbb{P}^x(B \mid |N_t| > 0) = Q^x(B).
$$

(1.5)

To our knowledge, Proposition 1.5 does not appear in the existing literature.

Finally, we prove a Yaglom-type limit theorem for the positions of the particles in the system at time $t$, given survival.

Theorem 1.6. For any measurable function $f$ on $D$ such that $\int_D f(x)^2 \varphi(x) \, dx < \infty$, we have

$$
\left( t^{-1} \sum_{u \in N_t} f(X_u(t)) \bigg| |N_t| > 0 \right) \to Z
$$
in distribution as $t \to \infty$, where $Z$ is an exponential random variable with mean

$$
\frac{\lambda(E[A^2] - E[A]) (\varphi, f)_{L^2(D)} \int_D \varphi^3}{2(m-1)}.
$$

One consequence of Theorem 1.6 (or rather its proof) is that it allows us to describe the limiting distribution of the particles in the system at time $t$, given survival. It turns out that this is the law with density $\varphi$, normalised to be a probability distribution.

Corollary 1.7. Let

$$
\mu_t := \frac{1}{|N_t|} \sum_{u \in N_t} \delta_{X_u(t)}
$$
be the uniform distribution on all particles alive at time $t$, given survival. Then, for each $f$ as in Theorem 1.6, we have that

$$
\mu_t(f) \to \mu(f)
$$
in distribution, and hence in probability, as $t \to \infty$, where

$$
\mu(f) = \frac{\int_D \varphi(x) f(x) \, dx}{\int_D \varphi(x) \, dx}.
$$
1.1 Context

It is interesting to note the analogy between Theorems 1.3-1.6 and classical results from the theory of Galton–Watson processes. Indeed, for critical Galton–Watson processes, Kolmogorov [Kol38] proved an asymptotic for the probability of survival up to time $n$:

$$P(Z_n > 0) \sim \frac{c}{n}$$

where $Z_n$ is the population size at time $n$, and the constant depends on the variance of the offspring distribution. Moreover, Aldous [Ald91, Ald93] and Duquesne and Le Gall [LGD02] showed that if you condition a critical Galton–Watson process to reach a large generation or have a large total progeny, then you have a scaling limit for the resulting tree. This limit is in the Gromov–Hausdorff topology, after rescaling distances in the tree appropriately, and the limiting object is the Continuum Random Tree, [Ald91]. In fact, this result can be extended to multitype Galton–Watson processes with a finite number of types, as in [Mie08], where the same scaling limit exists. Since a branching diffusion can be thought of as a limit of multitype Galton–Watson processes (considering the types to be positions and discretising the domain appropriately) it is reasonable to conjecture that such a process will have the same limiting genealogy when conditioned to survive for a long time.

On the other hand this result must be non-trivial, given what is known and expected for other types of domain. For example, [Kes78, BBS11, BBS13, BBS14, BBS15] have studied branching Brownian motion on the positive half line (with absorption at the origin) where each particle moves with a drift $-\mu$. In this setup there is a critical value of the branching rate $\beta$, equal to $\beta_c = \mu^2/2$, such that for $\beta \leq \beta_c$ extinction occurs with probability one. In [BBS14], an asymptotic for the survival probability is calculated, which is very different to that of Theorem 1.4. Moreover, results of [BBS13] in the near critical regime suggest that the critical genealogy in this case should be closely related to the Bolthausen–Sznitman coalescent [BS98]. This is also related to non-rigourous physics predictions of Brunet, Derrida, Mueller and Munier, [BDMM06, BDMM]: see [Ber09] for a survey of the area, further references and a general discussion of these predictions. This is of course not just a one-dimensional effect, as the behaviour will trivially be the same when the domain $D$ is a half-space with a drift in the direction of the hyperplane. Hence Theorem 1.1 is not likely to hold if we drop the boundedness assumption on the domain $D$.

More generally, this raises the following question:

**Question 1.8.** If the domain $D$ is allowed to be unbounded, or very irregular, what other behaviours do we see appearing at criticality (whenever we can make sense of this notion)?

In general it is an open, and we believe extremely interesting problem, to try and classify all the possible different behaviours that can occur at criticality depending on the geometry of the domain.

1.2 Organisation of the Paper and Main Ideas

After setting up the relevant notations and preliminary theory, we begin with the proof of Theorem 1.3. The main idea, which differs from the more analytic proofs given in [Sev58, Wat65, AH83], is to exploit the existence of the martingale

$$M_t = e^{(\lambda - \beta)t} \sum_{u \in N_t} \varphi(X^i_t),$$

(also appearing in [EK04, HHK12]) that arises naturally from the definition of the process. Since $\sum_{u \in N_t} \varphi(X^i_t)$ roughly tells us the size of the system at time $t$, and $M_t$ converges almost surely, the behaviour of the exponential term in (1.6) governs the possible survival or extinction of the process.

We then turn to the proofs of Theorems 1.4 to 1.6. The proof of Theorem 1.4 proceeds by a combination of probabilistic arguments, and the analysis of a system of coupled ordinary differential equations. Naively, we expand the survival probability (as a function of $x$, for each fixed $t$) with respect to the orthonormal basis of $L^2(D)$ given by the eigenfunctions of $-L$. Then because the survival probability satisfies a certain partial differential equation (the FKPP equation for branching Brownian motion, [McK75]) we get a family of coupled ODEs from the coefficients. In fact, we do not explicitly use that the survival probability satisfies this PDE (as we can derive the ODEs for
the coefficients directly and avoid potentially complicated technical assumptions), but this is the motivation behind the proof. Unfortunately however, the system of ODEs is not immediately easy to analyse, and this is where the probabilistic line of reasoning comes into play. Changing measure using the martingale $M_t$ (to get a spine characterisation of the system as discussed in the introduction) allows us to deduce that the survival probability actually just decays like $a(t)\varphi(x)$ as $t \to \infty$, where $a(t)$ is the first coefficient in the expansion. Thus, our problem is reduced to the study of a single ODE. From here elementary analysis, combined with some extra information obtained from the probabilistic arguments, yields the result. Proposition 1.5, Theorem 1.6 and Corollary 1.7 then follow fairly straightforwardly.

The remainder of the paper is devoted to the proof of Theorem 1.1. To do this, we take an i.i.d. sequence of critical processes, and concatenate the height functions of their associated continuous genealogical trees. A key idea is to define a suitable analogue of the Lukasiewicz path for Galton–Watson trees: that is, something that will approximate this concatenated height process well, and will converge after rescaling to a reflected Brownian motion. At first it seems too much to hope that such a precise combinatorial structure survives in this spatial context. However, it turns out that we can exploit the martingale $M_t$ by “exploring it in a different order.” Just as $M_t$ roughly measures the size of our population when we let time evolve in the usual way, when we explore the genealogical tree in a “depth first” order, the martingale corresponding to $M_t$ becomes, perhaps surprisingly, a kind of spatial analogue of the height function. After strengthening Corollary 1.7, we can prove that the quadratic variation of this martingale is essentially linear, and thus obtain an invariance principle.

Of course, we have to prove that this martingale is indeed a good approximation to the height process. This is one of the main difficulties, as the reversibility tools that are key to proving the analagous statement for the Lukasiewicz path in the Galton–Watson case are lost. Instead, we must use precise estimates, and a delicate ergodicity argument related to our spine change of measure. This is one of the reasons that our machinery from the proof of Theorem 1.4 is so essential. Tightness arguments then allow us to conclude.

Acknowledgements I would like to thank Nathanaël Berestycki, for suggesting this problem and for many helpful discussions and suggestions.

2 Preliminaries

2.1 Spectral Theory and Diffusions

Let us first assume that $D \subset \mathbb{R}^d$ is a bounded domain satisfying a uniform exterior cone condition. This means that: (1) $D$ is an open connected set of $\mathbb{R}^d$ with $|D| < \infty$ and boundary $\partial D$; and (2) there exist $r, \kappa > 0$ such that $\forall y \in \partial D$, we can find $\eta \in \mathbb{R}^d$ with $||\eta|| = 1$ and

$$\{ z \in B(y, r) : \eta \cdot (z - y) > 0 \text{ and } |z - \eta \cdot (z - y)| < \kappa|z - y| \} \subset D^c.$$  

Such a condition is satisfied, for example, if $\partial D$ is Lipschitz, see eg. [Dav89, p.27].

Let

$$L = \sum_{i,j=1}^{d} \partial_{x_i}(a^{ij} \partial_{x_j})$$

be a self-adjoint differential operator on $D$, which is uniformly elliptic, meaning that there exists a constant $\theta > 0$ such that for all $\xi \in \mathbb{R}^d$ and a.e. $x \in D$

$$\sum_{i,j=1}^{n} a^{ij}(x)\xi_i\xi_j \geq \theta|\xi|^2,$$

see [Eva98 §6.1]. We also assume that $a$ is symmetric, i.e. $a^{ij} = a^{ji}$, and that $a^{ij} \in C^1(\bar{D})$ for all $1 \leq i, j \leq d$. 


We say that $\tilde{\varphi} \in H^1_0(D)$ is an eigenvector of $-L$ with Dirichlet boundary conditions, and associated eigenvalue $\lambda$, if
\[
\int_D \sum_{i,j=1}^d a^{ij}(x) \partial_{x_i} \tilde{\varphi}(x) \partial_{x_j} v(x) \, dx = \int_D \tilde{\lambda} \tilde{\varphi}(x) v(x) \, dx \tag{2.1}
\]
for every $v \in H^1_0(D)$, as in \[Eva98\, [6.3]. That is, $\tilde{\varphi}$ is a weak solution of $-Lu = \tilde{\lambda}u$ with zero boundary conditions. Given the assumptions made on $L$ and $D$, the following properties then hold (see \[Dav89\, Theorem 1.6.8] and \[GT83\, Theorem 9.30]):

- The eigenvalues of $-L$ are all real and can be written $0 < \lambda_1 < \lambda_2 \leq \lambda_3 \cdots$ (repeated according to their finite multiplicity) with $\lambda_k \to \infty$ as $k \to \infty$.
- The associated eigenfunctions $\{\varphi_i\}_{i \geq 0}$ (normalised correctly) form an orthonormal basis of $L^2(D)$. Moreover, the first eigenfunction $\varphi_1 := \varphi$ is strictly positive in $D$, and $\varphi_i \in C(D)$ for all $i \geq 1$.

Now we consider the diffusion $(X(t))_{t \geq 0}$ associated to $L$ on $D$. This is the Markov process on $D \cup (\partial)$, where we identify the boundary $\partial D$ of $D$ with the single isolated point $(\partial)$, such that:

- $X(t)$ evolves as a diffusion with generator $L$ for all $t < \tau^D := \inf\{s \geq 0 : X(s) \in \partial D\}$; and
- $X(t) = (\partial)$ for all $t \geq \tau^D$.

Thus $(X(t))_{t \geq 0}$ is the diffusion with generator $L$, killed or absorbed upon hitting $\partial D$. We write $P^x$ for its law when started from $x \in D$. Then by \[Dav89\, Theorems 2.1.4 and 2.3.6], the function
\[
p_t(x,y) := \sum_{n \geq 1} \exp(-\lambda_n t) \varphi_n(x) \varphi_n(y); \quad t \in (0, \infty), \; x, y \in D \tag{2.2}
\]
is well defined as a uniform limit on $[\alpha, \infty) \times D \times D$ for any $\alpha > 0$, and is the transition density of the process $X$, restricted to $(x,y) \in D \times D$. We also have the estimate, \[Dav89\, Corollary 3.2.8]
\[
0 \leq p_t(x,y) \leq ct^{-d/2} \tag{2.3}
\]
for some constant $c > 0$. In particular, for any $t > 0$ and any $f \in L^1(D)$ we have
\[
\int_D f(y)p_t(x,y) \, dy = P^x(f(X(t))1_{\{\tau^D > t\}}). \tag{2.4}
\]

The properties (2.2)-(2.4) of the killed diffusion and associated transition kernel $p_t^D$ above, are consequences of the fact that the symmetric Markov semigroup associated with the killed diffusion is ultracontractive (see \[Dav89\, Section 2]) when $D$ satisfies a uniform exterior cone condition. In fact, if we assume some more regularity on the domain, then it will satisfy a certain stronger form of contractivity known as intrinsic ultracontractivity, first defined in \[DS84\,]. Intrinsic ultracontractivity is satisfied by the semigroup of the killed diffusion, for example, if the domain $D$ is bounded and Lipschitz \[Ban99\, Theorem 1]. The key property of intrinsic ultracontractivity that we will use is the following.

**Lemma 2.1.** Suppose that $D$ is a bounded Lipschitz domain (or more generally, a domain such that the semigroup of $X$ is intrinsically ultracontractive). If
\[
K_t^D(x,y) := \frac{e^{\lambda t} p_t^D(x,y) \varphi(y)}{\varphi(x)}
\]

2 We define $H^1_0(D)$ to be the closure of $C^\infty_c(D)$ (the space of infinitely differentiable functions with compact support strictly inside $D$) with respect to the norm $\|u\|_{H^1_0(D)} := \|u\|_{L^2} + \sum_{i=1}^d \|D^i u\|_{L^2}$, where $D^i u$ is the $i$th partial derivative of $u$ in the weak sense.
is the transition density for \((X(t))_{t \geq 0}\) conditioned to remain in \(D\) for all time [Pin85] then for any \(\varepsilon > 0\) there exists a constant \(C_\varepsilon\) depending only on the domain such that

\[
\left| \frac{K^D_t(x,y)}{\varphi(y)^2} - 1 \right| \leq C_\varepsilon e^{-\gamma t}
\]

for all \(t > \varepsilon\) and \(x, y \in D\), where \(\gamma := \lambda_2 - \lambda_1 > 0\) is the spectral gap for \(-L\) on \(D\).

**Proof.** See for example [DS84] or [Bañ99, Equation (1.8)]. □

We also have the following estimate:

**Lemma 2.2.** Suppose we are in the set up of Lemma 2.1. Then for some constants \(c_1, c_2\), we have

\[
p^D_r(w,z) \leq c_1 r^{-d/2} e^{-\frac{|w-z|^2}{c_2 r}},
\]

for all \(z, w \in D\).

**Proof.** See [Bañ99 Eq.(1.2)] or [Dav89 P.89]. □

We also have the following result, which gives us extra regularity on the eigenfunctions of \(L\), if we assume some extra regularity on the domain.

**Lemma 2.3.** Suppose that the boundary of \(D\) is \(C^{2,\alpha}\) for some \(\alpha > 0\) and that \(L\) is a generator satisfying the conditions assumed throughout this section. Let \(\{\varphi_i\}_{i \geq 1}\) be an \(L^2\) orthonormal basis of eigenfunctions for \(-L\). Then

\[
\varphi_i \in C^1(\overline{D})
\]

for all \(i \geq 1\).

**Proof.** [GT83 Theorem 6.31] □

### 2.2 Branching diffusions

As stated in the introduction, we can view a branching diffusion in \(D \subset \mathbb{R}^d\) as a point process

\[
X_t = \{X_u(t) : u \in N_t\}
\]

taking values in \(D\). This is often all that we will need to speak about, but since we are eventually interested in the genealogy of such processes, it will be helpful at various points to view them as elements of a larger state space: the space of marked trees. The set up described in this section closely follows [HH09, HR15].

We begin by recalling the Ulam–Harris labelling system. Let

\[
\Omega := \{\emptyset\} \cup \bigcup_{n \in \mathbb{N}} (\mathbb{N})^n
\]

be the set of finite labels on \(\mathbb{N} = \{1, 2, 3, \ldots\}\). A subset \(T \subset \Omega\) is called a tree if:

- \(\emptyset \in T\);
- \(u, v \in \Omega\) and \(uv \in T\) implies that \(u \in T\); and
- for all \(u \in T\) there exists \(A_u \in \mathbb{N} \cup \{0\}\) such that \(uj \in T \iff 1 \leq j \leq A_u\).
We will refer to elements $u \in T$ as particles or individuals in $T$. We think of the element $\emptyset$ as representing an initial ancestor, and individuals $u \in T$ as describing its descendants. For example, if $u \in T$ is given by the label $(2,1)$ then $u$ would be the first child of the second child of $\emptyset$. For $u, v \in \Omega$ we write $uv$ for the concatenation of the words $u$ and $v$, so for example if $u = (1,2,3), v = (2,1)$, then $uv = (1,2,3,2,1)$. We also set $u\emptyset = \emptyset u = u$ for all $u \in \Omega$. We say that $v$ is an ancestor of $u$ (written $v \prec u$) if there exists $w \in \Omega$ such that $uv = u$, and write $|u|$ for the length (or generation) of $u$, where $|u| = n$ if $u \in \mathbb{N}^n$. Then the above tree condition simply means that: $T$ has an initial ancestor or root $\emptyset$; $T$ contains all of the ancestors of all of its individuals; and finally, each individual $u \in T$ has a finite (possibly 0) number $A_u$ of children, labelled in a consecutive way. We write $\mathbb{T}$ for the set of trees.

We will want to consider marked trees, where the marks will correspond to the behaviour of particles in our branching diffusion. If we have a tree $T \in \mathbb{T}$, we will mark each $u \in T$ with a lifetime $l_u \in [0, \infty)$, and a motion in $D$,

$$X_u : [s_u - l_u, s_u) \to D,$$

where $s_u = \sum_{v \leq u} l_v$ is the death time of the particle $u$.

We write

$$\mathcal{T} := \{(T, l, X) = (T, (l_u)_{u \in T}, (X_u)_{u \in T}) : T \in \mathbb{T}, \text{ and } l_u \in [0, \infty) , X_u : [s_u - l_u, s_u) \to D \text{ for all } u \in T\}$$

for the set of all marked trees on $D$. With an abuse of notation, if we have a marked tree $(T, l, X)$ and $u \in T$ we will also sometimes extend the definition of $X_u$ to the whole of the ancestry of $u$. That is we will set $X_u(t)$ to be equal to $X_v(t)$ if $t \notin [s_u - l_u, s_u)$, and $v$ is the unique ancestor of $u$ alive at time $t$. If $u$ has no children, we write $X_u(t) = (\emptyset)$ for all $t \geq s_u$, where $(\emptyset)$ is an additional cemetery point that we introduce for use later on. Finally, we write

$$N_t := \{u \in T : t \in [s_u - l_u, s_u)\}$$

for the set of particles alive in $T$ at time $t$, and let $|N_t|$ be the number of such particles. As in the introduction, we let

$$\mathcal{X}_t := \{X_u(t) : u \in N_t\}$$

be the point process on $D$ corresponding to the marked tree $(T, l, X)$.

2.2.1 Probability measures on marked trees

Let $(\mathcal{F}_t)_{t \geq 0}$ be the filtration on the space of marked trees defined by

$$\mathcal{F}_t := \sigma(\{(u, X_u, l_u) : s_u \leq t\} \cup \{(u, X_u(s)) : s \in [s_u - l_u, t], t \in [s_u - l_u, s_u)\})$$

and set

$$\mathcal{F}_\infty = \sigma(\cup_{t \geq 0} \mathcal{F}_t).$$

Then $(\mathcal{F}_t)_{t \geq 0}$ is the natural filtration associated with the point process $(\mathcal{X}_t)_{t \geq 0}$. We let $\mathbb{P}_x$ be the probability measure on $(\mathcal{T}, \mathcal{F}_\infty)$ such that:

- $X_\emptyset(t)$ evolves under the law $\mathbb{P}^x$ described in Section 2.1 for $0 \leq t \leq l_\emptyset$, where $l_\emptyset = \tau_\emptyset^D \land \tau_\emptyset^D$, $\tau_\emptyset^D$ is the first time that $X_\emptyset$ hits $\partial D$ and $\tau_\emptyset^D \sim \text{Exp}(\beta)$ is an exponential random variable independent of $X_\emptyset$.

- $A_\emptyset = 0$ on the event that $l_\emptyset = \tau_\emptyset^D$. On the complementary event, $A_\emptyset$ is distributed as an independent copy of the offspring distribution $A$.

- At any branching event where a positive number of children are born, all children repeat stochastically, and independently, the behaviour of their parent, starting from the point of fission.

That is, $\mathbb{P}_x$ is the law of the system described in the introduction, with offspring distribution $A$ and constant branching rate $\beta > 0$. 

8
2.2.2 The many-to-few formulae

One particularly useful property of the branching diffusions considered in this section are the so-called many-to-few formulae, which allow one to calculate certain expectations for the system with relative ease. We state the two simplest cases here; for the more general formula, see for example [HR15 Lemma 1].

Lemma 2.4 (Many-to-one). Suppose that \( f \) is a measurable function on the Borel sets of \( D \). Then

\[
\mathbb{P}^x \left( \sum_{u \in N_t} f(X_u(t)) \right) = e^{(m-1)\beta t} \mathbb{P}^x \left( f(X(t)) \mathbf{1}_{\{\tau^D > t\}} \right)
\]

where we recall that \( m = \mathbb{E}(A) \).

Lemma 2.5 (Many-to-two). Suppose that \( f \) and \( g \) are measurable functions on the Borel sets of \( D \). Then

\[
\mathbb{P}^x \left( \sum_{u \in N_t} f(X_u(t)) \sum_{v \in N_t} g(X_v(t)) \right) = e^{(m-1)\beta t} \mathbb{P}^x \left( f(X(t))g(X(t))\mathbf{1}_{\{\tau^D > t\}} + \beta (\mathbb{E}(A^2) - \mathbb{E}(A)) \int_0^t e^{(2r-s)(m-1)\beta} \mathbb{P}^x \left( \mathbf{1}_{\{\tau^D > r\}} q(X(s), t - s) \right) \right)
\]

where \( q(y, r) = \mathbb{P}^y \left( f(X(r)) \mathbf{1}_{\{\tau^D > r\}} \mathbf{1}_{\{X(r) \neq X(t)\}} \right) \) for \( y \in D \) and \( r \geq 0 \).

For the proof of the above lemmas, see [HR15 Lemma 1], which is stated in a more general setting. For an explanation of how this general statement gives the lemmas above, see [HR15 §4.1, §4.2].

2.2.3 The continuous genealogical plane tree

If we have a marked tree \((T, l, X) \in \mathcal{T}\) corresponding to a branching diffusion, we will also want to associate with it a continuous genealogical tree \( T \). This tree (which we emphasise is different to \( T \)) is the main object of Theorem 1.1 it is the plane tree with branch lengths given by the lifetimes of particles in the system.

We first need to give a few definitions. A metric space \((T, d)\) is said to be a real tree if, for all \( v, w \in T \) the following two conditions hold [LG05]:

1. There exists a unique isometric map \( \phi_{v, w} : [0, d(v, w)] \to T \) with \( \phi_{v, w}(0) = v \) and \( \phi_{v, w}(d(v, w)) = w \).
2. Any continuous injective map \([0, 1] \to T\) that joins \( v \) and \( w \) has the same image as \( \phi_{v, w} \).

One way to define a real tree is the following: take a continuous function \( C : [0, \infty) \to [0, \infty) \) with \( C(0) = 0 \) and define a “distance” function on \([0, \infty) \times [0, \infty)\) by

\[
d_C(s, t) = C(s) + C(t) - 2 \min_{r \in [s, t]} C(r)
\]

whenever \( s \leq t \). It is easy to verify that this defines a pseudometric on \([0, \infty)\). Thus, quotienting by the equivalence relation \( \sim \) that identifies points with \( d_C(\cdot, \cdot) = 0 \) we obtain a metric space \((T_C, d_C) := ([0, \infty) / \sim, d_C)\). One can prove, see for example [LG05], that this metric space is a real tree. The function \( C \) is called the contour function of the tree.

In our set up, if we have a marked tree \((T, l, X) \in \mathcal{T}\) we let \((T(T, l, X), d(T, l, X)) = (T_{C(T,l,X)}, d_{C(T,l,X)})\) be the real tree with contour function \((C(T, X, l)(t))_{t \geq 0} = (C(t))_{t \geq 0}\) described as follows. Let \( \{0 =: u_0, u_1, u_2, \ldots u_{|T|} \} \) be the set of labels of \( T \) in depth first order. This is the ordering on \( T \) such that \( u \) is less than \( v \) if at the first coordinate where the labels of \( u \) and \( v \) differ, the coordinate of \( u \) is less than that of \( v \). For any two individuals \( w, w' \in T \) we let \( w \wedge w' \) denote their most recent common ancestor: that is, the (unique) \( u \) with \(|u| \) largest, such that \( u \preceq w \) and \( u \preceq w' \). We can then define, for \( j \geq 0 \) such that \( u_j \neq u_{j+1} \):

- \( r_j := s_{u_j} - (s_{u_{j+1}} - l_{u_{j+1}}) \) to be the length of time between the birth time of \( u_j \wedge u_{j+1} \) and the death time of \( u_j \), and
\begin{itemize}
\item $r_j' := l_{u_{j+1}} - (s_{u_j \wedge u_{j+1}} - l_{u_j \wedge u_{j+1}})$ to be the length of time between the birth time of $u_j \wedge u_{j+1}$ and the birth time of $u_{j+1}$.
\end{itemize}

We set $r_j = r_j' = 0$ if $u_j < u_{j+1}$. Let $R_j = \sum_{k<j} l_k + r_k + r_k'$, and for $t \in [0, R_{j+1} - R_j)$ set

$$C(t + R_j) - C(R_j) = \begin{cases} t & \text{if } t \in [0, l_j) \\
2l_j - t & \text{if } t \in [l_j, l_j + r_j) \\
-2r_j + t & \text{if } t \in [l_j + r_j, l_j + r_j + r_j') \end{cases}.$$ 

That is: $C(t)$ is positive and linear with unit speed on $[R_j, R_j + l_j]$ and $[R_j + l_j + r_j, R_j + l_j + r_j + r_j')$, and is negative and linear with unit speed on $[R_j + l_j, R_j + l_j + r_j + r_j')$. Finally, for $t \in [R_j, R_j + l_j + r_j + r_j')$ we let $C(t)$ interpolate linearly between $C(R_j)$ and 0. The definition of the function $C$ is probably clearest from a picture: we draw a tree with branch lengths corresponding to lifetimes of individuals in the system, and traverse it (with backtracking) at speed one. $C(t)$ measures how “high” we are in the tree at time $t$.

Figure 2.1: An example of the continuous tree $T = T_C$ generated by a branching diffusion. Here $u_0 = 0, u_1 = (1), u_2 = (1, 1), u_3 = (2)$ and $u_4 = (3)$. Every branch of $T$ corresponds to a particle $u \in T$, and this branch has length $l_u$.

### 2.3 Martingales

Suppose that $D \subset \mathbb{R}^d$ is bounded satisfying a uniform exterior cone condition, and $P^x$, $(X(t))_{t \geq 0}$, $\tau^D$ and $\rho_t^D$ are as in Section 2.1. Then by (2.2), and using the fact that the eigenfunctions $\{\varphi_i\}$ form an orthonormal basis of $L^2(D)$ we see that for any $f \in L^2(D)$ and $t > 0$

$$P^x(f(X(t))1_{\{\tau^D > t\}}) = \sum_{i=1}^{\infty} \exp(-\lambda_i t) \varphi_i(x) \varphi_i(f)$$  \hspace{1cm} (2.5)

where we write $(f, g) := \int_D f(x)g(x)\,dx$ throughout the paper. In particular,

$$P^x[\varphi_i(X(t))1_{\{\tau^D > t\}}] = e^{-\lambda_i t} \varphi_i(x)$$  \hspace{1cm} (2.6)

for all $x \in D$. One consequence of this is that the process

$$\exp(\lambda t)\varphi(X(t))1_{\{\tau^D > t\}}$$  \hspace{1cm} (2.7)

is a (positive) martingale under $P^x$. Furthermore, this single-particle martingale implies the existence of a martingale for the entire branching diffusion under $P^x$. Indeed, a straightforward application of the Markov property for the branching diffusion and Lemma 2.4 yields the following.
Lemma 2.6. The process
\[ M_t = e^{(\lambda - \beta(m-1))t} \sum_{u \in N_t} \varphi(X_u(t)) \]
is a positive martingale under \( \mathbb{P}^x \), for each \( x \in D \). It therefore converges \( \mathbb{P}^x \)-almost surely to an almost surely finite limit \( M_\infty \).

This martingale is the natural analogue of the well-known martingale \( Z_n/n^m \) for Galton–Watson processes (with offspring mean \( m \)). Variants of the martingale for general branching processes have been studied extensively in the literature: see for example [BK77, Lyo97] for the branching random walk case and [CR88, Kyp04, EK04, HH07, HHK12] for branching Brownian motion, among many others.

2.4 Spine theory

It turns out that a helpful approach in many parts of this paper will be to study the behaviour of the system under a change of measure. Precisely, the change of measure defined by the \( \mathbb{P}^x \) martingale \( M_t \) from the previous section. To give a useful description of this, we need to view our process on a yet larger state space: the space of marked trees with spines. This is a classical technique first introduced by [CR88], and since used extensively by many authors [HHK12, HH09, Rob10, HR15]. For a thorough exposition of the subject, we refer the reader to [HH09], and the overview in this section will closely follow that given in [HH09, HR15].

Suppose we have a marked tree \((T, l, X) \in \mathcal{T} \). A spine \( \psi \) on \((T, l, X) \) is a subset of \( T \cup \{\psi\} \) (where we recall that \{\psi\} is an isolated cemetery point) such that:

- \( \emptyset \in \psi \) and \( |\psi \cap (N_t \cup \{\psi\})| = 1 \) for all \( t \);
- \( v \in \psi \) and \( u \prec v \) implies that \( u \in \psi \); and
- if \( v \in \psi \) and \( A_v > 0 \), then there exists a unique \( 1 \leq j \leq A_v \) with \( v_j \in \psi \). If \( A_v = 0 \), then \( \psi \cap N_t = \emptyset \) for all \( t \geq s_v \).

We write
\[ \mathcal{T} := \{(T, l, X, \psi) : (T, l, X) \in \mathcal{T}, \text{ and } \psi \text{ is a spine on } (T, l, X)\} \]
for the space of marked trees with spines. Given \((T, l, X, \psi) \in \mathcal{T} \) we let \( \psi_t := v \) be the unique element \( v \in \psi \cap (N_t \cup \{\psi\}) \) and write \( \xi_t := X_v(t) \) for its position. With a slight abuse of notation, we set \( X_v(t) = (\psi) \) if \( \xi_t = (\psi) \) and also in this case say that \( u \prec \psi_t \) if \( u \in N_t \) and \( u = \psi_s \) for some \( s < t \).

2.4.1 Filtrations

There are several different filtrations we can place on \( \mathcal{T} \). We give brief descriptions of each of these below: see [Rob10] for more rigorous definitions.

- \( (\mathcal{F}_t)_{t \geq 0} \) is the natural filtration of the branching process as before, and does not contain any information about spines. We write \( \mathcal{F}_\infty := \sigma(\cup_{t \geq 0} \mathcal{F}_t) \).

- \( (\tilde{\mathcal{F}}_t)_{t \geq 0} \) is the natural filtration of the branching process plus the spine. We write \( \tilde{\mathcal{F}}_\infty := \sigma(\cup_{t \geq 0} \tilde{\mathcal{F}}_t) \).

- \( (\mathcal{G}_t)_{t \geq 0} := (\sigma(\xi_s : s < t))_t \) is the filtration generated by the spatial motion of the spine. We write \( \mathcal{G}_\infty := \sigma(\cup_{t \geq 0} \mathcal{G}_t) \).

- \( (\tilde{\mathcal{G}}_t)_{t \geq 0} := \sigma(\mathcal{G}_t \cup \{v \in \psi_s : 0 \leq s \leq t\} \cup \{A_u : u < \psi_t\}) \) is the filtration that knows everything about the spine until time \( t \): which individuals are in the spine, their motions, fission times, and family sizes at fission times along the spine. We write \( \tilde{\mathcal{G}}_\infty := \sigma(\cup_{t \geq 0} \tilde{\mathcal{G}}_t) \).
2.4.2 Probability measures

We first want to define the probability measure \( \tilde{\mathbb{P}}^x \) on \((\tilde{T}, \tilde{\mathcal{F}}_\infty)\) under which, informally speaking, the law of the tree \((T, l, X)\) is the same as under \(\mathbb{P}^x\), and then the spine is chosen by picking one of the children uniformly at random at every branching event. More rigorously, if \( Y \) is an \( \tilde{\mathcal{F}}_t \)-measurable random variable, then \( Y \) can be written \([HR15]\) as

\[
\sum_{v \in N_t \cup \{\dagger\}} Y(v) \mathbf{1}_{\{\psi_t = v\}}
\]

where \( Y(v) \) is measurable with respect to \( \mathcal{F}_t \). Given this representation, we define the measure \( \tilde{\mathbb{P}}^x \) on \((\tilde{T}, \tilde{\mathcal{F}}_\infty)\) by setting for each \( \tilde{\mathcal{F}}_t \)-measurable \( Y \):

\[
\tilde{\mathbb{P}}^x(Y) = \mathbb{P}^x(\sum_{v \in N_t} (Y(v) \prod_{u < v} \frac{1}{A_u}) + Y(\{\dagger\}) \sum_{w \in D_t} \prod_{u < w} \frac{1}{A_u}),
\]

where \( D_t := \{w \in \bigcup_{s \leq t} N_s : A_w = 0\} \). Note that \( \tilde{\mathbb{P}}^x |_{\mathcal{F}_\infty} = \mathbb{P}^x \).

2.4.3 Change of measure

Recall from Section 2.3 that \( e^\lambda \frac{\varphi(X(t))}{\varphi(x)} \mathbf{1}_{\{\tau_D^x > t\}} \) defines a mean-one martingale under \( \mathbb{P}^x \). This implies, see \([HH09]\), that

\[
\zeta_t := \mathbf{1}_{\{A_{\psi_s} > 0 \ \forall s < t\}} \mathbf{1}_{\{\tau_D^x > t\}} e^{(\lambda - \beta(m-1))t} \frac{\varphi(\xi_t)}{\varphi(x)} \prod_{v < \psi_t} A_v,
\]

where \( \tau_D^x \) is the first time that \( \xi \) leaves the domain \( D \), is a mean-one \( \tilde{\mathcal{F}}_t \)-martingale under \( \tilde{\mathbb{P}}^x \). In fact, it is easy to check that \( \frac{dQ^x}{d\mathbb{P}^x} |_{\mathcal{F}_t} = \tilde{\mathbb{P}}^x(\zeta_t | \mathcal{F}_t) \). Thus, if we define a new probability measure \( \tilde{\mathbb{Q}}^x \) on \((\tilde{T}, \tilde{\mathcal{F}}_\infty)\) via the martingale change of measure

\[
\frac{d\tilde{\mathbb{Q}}^x}{d\tilde{\mathbb{P}}^x} |_{\mathcal{F}_t} := \zeta_t,
\]

then we have, defining \( Q^x := \tilde{\mathbb{Q}}^x |_{\mathcal{F}_\infty} \), that

\[
\frac{dQ^x}{d\mathbb{P}^x} |_{\mathcal{F}_t} := M_t \frac{\varphi(x)}{\varphi(x)}.
\]

We have the following description for how the branching diffusion with a distinguished spine behaves under \( \tilde{\mathbb{Q}}^x \) (see for example \([CR88]\) or \([HH09]\)):

- we begin with one particle at position \( x \), which is the spine particle;
- the spine particle evolves as if under the changed measure

\[
\frac{dQ^x}{d\mathbb{P}^x} |_{\mathcal{F}_t} := \mathbf{1}_{\{\tau_D^x > t\}} e^\lambda \frac{\varphi(\xi_t)}{\varphi(x)};
\]

- the spine particle branches at rate \( m\beta \) and is replaced by a number of children having the size-biased distribution \( \hat{A} \), where

\[
\mathbb{P}(\hat{A} = k) = \frac{k}{m} \mathbb{P}(A = k);
\]

- given that there are \( k \) children born at such a branching event, one is chosen uniformly at random to be the spine and stochastically repeats the behaviour of its parent. Non-spine particles initiate independent branching diffusions with law \( \mathbb{P}^y \), from the point \( y \) of fission.
We now make a number of remarks about this. Firstly, note that the spine particle under \( \tilde{Q}_x \) never leaves the domain \( D \), and that it always has a positive number of children. This means that the branching diffusion under \( \tilde{Q} \) never becomes extinct, and that the spine particle is never in the cemetery state. Also note that under the change of measure \((2.12)\), the spine particle evolves as a diffusion with transition kernel \( K_n^D(x,y) \): the kernel we defined in Lemma 2.1. Its motion is that of the diffusion \( X(t) \) under \( P^x \), but conditioned to remain in \( D \) for all time, see [Pin85]. In particular, Lemma 2.1 tells us that if the domain \( D \) is Lipschitz, its position converges quickly to an equilibrium distribution with density \( \varphi^2 \). Finally, we record that (by an easy calculation):

\[
\tilde{Q}_x^p(v = \psi_t|\mathcal{F}_t) = \frac{\varphi(X_v(t))}{\sum_{u \in N_t} \varphi(X_u(t))}
\]

for \( v \in N_t \).

3 The Phase Transition

In this section we will provide a proof of Theorem 1.3.

Proof of Theorem 1.3 First suppose that \( \beta > \lambda/(m - 1) \). Then we know by Lemma 2.6 that \( M_t \to M_\infty \) almost surely. Thus, by definition of \( M_t \), we can conclude the proof in this case as soon as we can show that \( \mathbb{P}^x(M_\infty > 0) > 0 \). However, for this it is enough to show that \( (M_t)_t \) is uniformly integrable, and in fact, an elementary calculation using Lemma 2.5 gives that if \( \beta > \lambda \), then \( (M_t)_t \) is uniformly bounded in \( L^2(\mathbb{P}^x) \). We leave this calculation to the reader.

Next suppose that \( \beta < \lambda/(m - 1) \). We write

\[
\mathbb{P}^x(|N_t| > 0) \leq \mathbb{P}^x(|N_t|) = e^{(m-1)\beta} \mathbb{P}^x(\tau^D > t),
\]

where the second equality comes from Lemma 2.4. Observe that by Lemma 2.1 we also have the existence of a constant \( C \), depending only on \( D \), such that

\[
p_t^D(x,y) \leq C\varphi(x)\varphi(y)e^{-\lambda t}
\]

for all \( t \geq 1 \). Since \( \mathbb{P}^x(\tau^D > t) = \int_D p_t^D(x,y)dy \), this implies that when \( \beta < \lambda \), the right hand side of (3.1) converges to 0 as \( t \to \infty \) (uniformly in \( x \in D \)). Hence we have almost sure extinction in this case.

Finally, we deal with the critical case \( \beta = \lambda/(m - 1) \). We make use of the following lemma, which can be found in [Wat65, Lemma 2.1].

Lemma 3.1 ([Wat65]). For all \( x \in D \)

\[
\mathbb{P}^x(|N_t| \to 0 \text{ or } |N_t| \to \infty \text{ as } t \to \infty) = 1.
\]

The proof we give is the same as that in [Wat65] (we include it only to show that it still works with our current assumptions on \( L \) and \( D \)).

Proof. Since \( |N_t| \) is integer-valued, and \( \{ |N_t| = 0 \} \Rightarrow \{ |N_r| > 0 \forall r \geq t \} \), it is sufficient to prove that \( \mathbb{P}^x(|N_t| = k \text{ i.o.}) = 0 \) for every \( k \in \mathbb{N} \). Fix \( k \) and define a sequence of hitting and leaving times \( (L_n, H_n)_{n \geq 1} \), by letting \( L_1 \) be the first time \( t \) that \( |N_t| \neq 0 \) and \( H_1 \) be the first time (necessarily after \( L_1 \)) that \( |N_t| = k \). Then inductively, let \( L_n \) be the first time after \( H_{n-1} \) that \( |N_t| \neq k \), and \( H_n \) the first time after \( L_n \) that \( |N_t| = k \). We have to show that \( \mathbb{P}^x(H_n < \infty) \to 0 \) as \( n \to \infty \). Set

\[
v = \inf_{y \in D} \mathbb{P}^y(\tau^D \leq 1) > 0
\]

which is strictly positive by Lemma 2.1, and let \( p > 0 \) be the probability that an \( Exp(\beta) \) random variable is bigger than 1. Then we have that \( \mathbb{P}^x(H_1 < \infty) \leq \mathbb{P}^x(|N_{L_1}| > 0) \leq (1 - pv) \) (since the probability that the initial particle exits the domain without branching is greater than or equal to \( pv \)). Then inductively, using the Markov property at each time \( H_j \), we see that \( \mathbb{P}^x(H_n < \infty) \leq \mathbb{P}^x(|N_{L_n}| > 0) \leq (1 - pv)(1 - (pv)^k)^{n-1} \to 0 \). This completes the proof. \( \Box \)
With this in hand, to prove that we have almost sure extinction in the critical case, it is enough to show that 
\[ \mathbb{P}^x(|N_t| \to \infty \text{ as } t \to \infty) = 0. \]
To do this, we use the fact that \( M_t = \sum_{u \in N_t} \varphi(X_u(t)) \) converges almost surely to \( M_\infty < \infty \). Then the idea is that if \(|N_t|\) is very big then \( M_t \) should be big as well, and this will give a contradiction.

Consider the event \( A_{t,K,R} := \{|N_t| \geq R \} \cap \{|M_t | \leq K \} \), for \( R, K \geq 1 \). On this event, by definition of \( M_t \), it must be the case that \( N_t \) is non-empty, and that \( \varphi(X_u(t)) \leq K/R \) for some \( u \in N_t \). In other words, we have \( A_{t,K,R} \subset \{\sum_{u \in N_t} \mathbb{1}(\varphi(X_u(t)) \leq K/R) \geq 1\} \). We compute that
\[
\mathbb{P}^x\left( \sum_{u \in N_t} \mathbb{1}(\varphi(X_u(t)) \leq K/R) \geq 1 \right) \leq \mathbb{P}^x\left( \sum_{u \in N_t} \mathbb{1}(\varphi(X_u(t)) \leq K/R) \right) \leq e^{(m-1)\beta t} \int_{y \in D \varphi(y) \leq K/R} p_t^D(x,y) \, dy
\]
where the second inequality follows from Lemma 2.4. Finally, using (3.2), we see that this is less than \( cK/R \) whenever \( t \geq 1 \), for some constant \( c \) depending only on \( D \).

Now we are ready to conclude. We have
\[ \mathbb{P}^x\left( |N_t| \to \infty \right) = \mathbb{P}^x\left( \{|N_t| \to \infty \} \cap \{M_\infty < \infty \} \right) \]
and we have just shown that \( \mathbb{P}^x\left( \cap_{t \geq T} A_{t,K,R} \right) \leq \mathbb{P}^x\left( A_{T,K,R} \right) \leq cK/R \) for all \( T \geq 1 \). Taking limits on the right hand side of (3.3) we see that \( \mathbb{P}^x\left( |N_t| \to \infty \right) = 0. \)

To complete the proof of Theorem 1.3 we must show that the decay of the survival probability in the critical case is uniform in \( D \). To do this we will show that if we set
\[ u(x,t) := \begin{cases} \mathbb{P}^x(|N_t| > 0) & x \in D \\ 0 & x \in \partial D \end{cases} \] (3.4)
then \( u(x,t) \) is a continuous function in \( \bar{D} \) for all \( t > 0 \). Then since the \( u(x,t) \) are decreasing in \( t \) and converge to the continuous function 0 as \( t \to \infty \) for each \( x \in D \), by Dini’s theorem [Rud76, Theorem 7.13]), the decay must indeed be uniform.

To prove this, first we fix \( t > 0 \) and \( x \in D \), and show that \( u(x,t) \) is continuous at \( x \). Indeed, given any \( \varepsilon > 0 \), we can pick \( T > 0 \) such that if \( Z \sim \text{Exp}(\lambda/(m-1)) \) then \( \mathbb{P}(Z < T) \leq \varepsilon/4 \). This means that for any \( y \in D \), by conditioning on whether or not the first branching time is less than or equal to \( T \), we have
\[ |u(x,t) - u(y,t)| \leq \varepsilon/2 + \int_D |p_T^D(x,w) - p_T^D(y,w)| u(w,t-T) \, dw. \]
Then the continuity follows, since for fixed \( T \) the second term on the right hand side will be less than \( \varepsilon/2 \) whenever \( |y-x| \) is small enough. This last claim holds since \( p_T^D(\cdot,w) \) is continuous at \( x \) for each \( w \in D \) (e.g. by (2.2)), and by dominated convergence.

So to complete the proof, we just need to show that \( u(y,t) \to 0 \) as \( y \to \partial D \). However, this follows since
\[
\mathbb{P}^y(|N_t| > 0) \leq 1 - \mathbb{P}^y\text{the process becomes extinct before the first branching time) } \leq 1 - \mathbb{P}(Z > s) \mathbb{P}^y(\tau^D \leq s)
\]
for any \( s > 0 \), where \( Z \) is the same random variable described above. The last line can be made arbitrarily small by first taking \( s \) to 0, and then \( y \to \partial D \). Thus \( u(\cdot, t) \in C(D) \) for each \( t > 0 \).

We finish this section by recording a useful lemma in the critical case.

**Lemma 3.2.** For any \( T > 0 \) there exists \( C_T > 0 \) such that \( \mathbb{P}^x(|N_t|^2) \leq TC_T \varphi(x) \) for all \( t \geq T \) and \( x \in D \).

**Proof.** This essentially follows from the many-to-two lemma, along with a couple of estimates that have been developed in this section. In the proof, \( c_T, c_T', c_T'' \) all represent constants depending only on \( T \).

First, by (3.2) we know that there exists \( c_T \) such that for all \( r \geq T/2 \) and \( y \in D \),
\[
\mathbb{P}^y(\tau^D > r) \leq c_T \varphi(y) e^{-\lambda r}. \] (3.5)
We can also write, by Lemma 2.5
\[ \mathbb{P}^x(|N_t|^2) = e^{\lambda t} \mathbb{P}^x(\tau^D > t) + \frac{\lambda}{m-1} (\mathbb{E}(A^2) - \mathbb{E}(A)) \int_0^t e^{\lambda(2t-s)} \mathbb{P}^x(\mathbb{1}_{\{\tau^D > s\}} \mathbb{P}^{X}(\mathbb{1}_{\{\tau^D > t-s\}})^2) \, ds \]
where by (3.5) the first term on the right hand side of this expression is bounded by \(c_T \varphi(x)\). Furthermore, using (3.5) again, the integrand in the second term is bounded by \(c_T \varphi(x)\) for some \(c_T\), whenever \(s \geq T/2\) and \(t - s \geq T/2\). To conclude, notice that if \(s < T/2\) then \(t - s\) must be \(\geq T/2\). Therefore, the integrand on this region is bounded by \(c_T e^{\lambda s} \mathbb{P}^x(\varphi(X_s) \mathbb{1}_{\{\tau^D > s\}})\) (for some \(c_T\)), which is less than or equal to \(c_T' \varphi(x)\) by (2.6). Finally, if \(t - s < T/2\), then again we must have \(s \geq T/2\), and the integrand on this region (also using that using that \(e^{\lambda(2t-s)} \leq e^{\lambda T} e^{\lambda s}\) in this case) is less than \(e^{\lambda T} \varphi(x)\). Putting all of this together gives the lemma.

\[ \square \]

4 Survival at Criticality: Proof of Theorem 1.4

Throughout this section, we will work in the critical case \(\beta = \lambda/(m - 1)\).

4.1 Asymptotics for the Survival Probability

We will first prove that the survival probability \(\mathbb{P}^x(|N_t| > 0)\) decays asymptotically like \(\varphi(x)a(t)\), where
\[ a(t) := \int_D \mathbb{P}^x(|N_t| > 0) \varphi(z) \, dz. \]
In fact, we prove a more general Proposition (see below) from which we will obtain this as a special case.

**Proposition 4.1.** Suppose that the conditions of Theorem 1.3 are satisfied. For a measurable function \(f\) on \(D\) with \(0 \leq f < 1\), set
\[ a_f(t) := \int_D \varphi(x) \mathbb{P}^x(1 - \hat{f}(N_t)) \, dx; \quad \hat{f}(N_t) := \begin{cases} \prod_{u \in N_t} f(X_u(t)) & \text{if } N_t \neq \emptyset \\ 1 & \text{if } N_t = \emptyset. \end{cases} \]
Then
\[ \left| \frac{\mathbb{P}^x(1 - \hat{f}(N_t))}{\varphi(x) a_f(t)} - 1 \right| \to 0 \quad (4.1) \]
as \(t \to \infty\), uniformly in \(x \in D\) and \(f \) with \(0 \leq f < 1\).

**Remark 4.2.** When \(f = 0\) we see that \(\mathbb{P}^x(1 - \hat{f}(N_t)) = \mathbb{P}^x(|N_t| > 0)\) and write \(a_0(t) := a(t)\).

**Remark 4.3.** Note that \(a_f(t) \neq 0\) for all \(t \geq 0\). Indeed for any \(t \geq 0\), \(1 - \hat{f}(N_t)\) is a positive (\(F_t\)-measurable) random variable, and for every \(x \in D\) there is a set of strictly positive \(\mathbb{P}^x\) measure on which \(1 - \hat{f}(N_t) > 0\) (i.e. on the event that \(|N_t| > 0\). Thus, \(\mathbb{P}^x(1 - \hat{f}(N_t)) > 0\) for every \(x \in D\). Since this is a measurable function of \(x\), and \(\varphi(x)\) is also measurable and strictly positive in \(D\), we obtain that \(a_f(t) > 0\).

**Proof.** The idea for the proof of this is to write, recalling the changes of measure from Section 2.4.3
\[ \frac{\mathbb{P}^x(1 - \hat{f}(N_t))}{\varphi(x)} = \mathbb{Q}^x(\frac{1 - \hat{f}(N_t)}{\sum_{u \in N_t} \varphi(X_u(t))}) = \hat{\mathbb{Q}}^x(\frac{1 - \hat{f}(N_t)}{\sum_{u \in N_t} \varphi(X_u(t))}) \quad (4.2) \]
see (2.11), and then show that the right hand side essentially does not depend on \(x\) for large \(t\). The intuition behind this is that under \(\hat{\mathbb{Q}}^x\), the position of the spine particle will converge very quickly to equilibrium. Moreover, contributions to \((1 - \hat{f}(N_t))/\sum_{u \in N_t} \varphi(X_u(t))\) from subprocesses branching off the spine before its position has become well mixed are unlikely to occur, as these have the law of branching diffusions under \(\mathbb{P}\), which we know are unlikely to survive for a long time.
To turn this into a rigorous proof, we first pick $1 \leq t_0 < t$ (we may assume $t > 1$) and decompose

$$N_t = N_t^1 \cup N_t^2 := \{ u \in N_t : \psi_{t_0} \leq u \} \cup \{ u \in N_t : \psi_{t_0} \neq u \}.$$  

Then we see using (4.2) that

$$\frac{P_x(1 - \hat{f}(N_t))}{\varphi(x)} \geq \tilde{Q}^x \left( \frac{1 - \hat{f}(N_t^1)}{\sum_{u \in N_t^1} \varphi(X_u(t) - 0)} \right) = \tilde{Q}^x \left( \varphi(X_u(t) - 0) \right)$$

where we recall the definition of the $\sigma$-algebra $\mathcal{G}_{t_0}$ from Section 2.4.1 that knows everything about the spine’s motion $(\xi_u)_{u \leq t_0}$ and about branching points along the spine up to time $t_0$. Using the description of the behaviour of the system under $\tilde{Q}^x$, we see that this is equal to

$$\tilde{Q}^x \left( \sum_{u \in N_{t-t_0}} \frac{1 - \hat{f}(N_{t-t_0})}{\varphi(X_u(t-t_0))} \right) \prod_{u \sim \psi_{t_0}} (P_{N_t}(|N_{t-u}| = 0))^{A_u-1}$$

(4.3)

where recalling the notation from Section 2.2.3, this product is over all branching points along the spine before time $t_0$ (at times $s_u$, and positions $X_v(s_u) := (\xi(s_u))$, with $A_v$ children being born). Now we make the simple bound

$$1 - \sum_{u \sim \psi_{t_0}} A_u \sup_{w \in D} P_w(|N_{t-u}| > 0) \leq \prod_{u \sim \psi_{t_0}} (P_{N_t}(|N_{t-u}| = 0))^{A_u-1}$$

which tells us, by independence of $\xi_{t_0}$ and $\sum_{u \sim \psi_{t_0}} A_u$, that (4.3) is greater than or equal to

$$\left[ \tilde{Q}^x \left( \sum_{u \in N_{t-t_0}} \frac{1 - \hat{f}(N_{t-t_0})}{\varphi(X_u(t-t_0))} \right) \right] \times \left[ 1 - \sup_{w \in D} \left( P_w(|N_{t-u}| > 0) \right) \tilde{Q}^x \left( \sum_{u \sim \psi_{t_0}} A_u \right) \right].$$

We make a few observations. Firstly, $\tilde{Q}^x \left( \sum_{u \sim \psi_{t_0}} A_u \right) \leq ct_0$ for some constant $c$ depending only on the branching diffusion. Secondly,

$$\tilde{Q}^x \left( \sum_{u \in N_{t-t_0}} \frac{1 - \hat{f}(N_{t-t_0})}{\varphi(X_u(t-t_0))} \right) = \int_D K^D_{t_0}(x, y) \frac{\varphi(y)}{\varphi(y)} dy$$

by (4.2) again, and the fact that the spine particle has transition density $K^D$ (defined in Lemma 2.1). Using Lemma 2.1, which tells us that $K^D_{t_0}(x, y)$ converges to $\varphi(y)^2$ exponentially fast, uniformly in $x$ and $y$, we see that the right hand side of the above is greater than or equal to $(1 - e^{-\gamma t_0})a_f(t-t_0)$, for another constant $0 < c' < \infty$, where $\gamma > 0$ is the spectral gap for $L$ on $D$. Overall, we obtain that

$$\frac{P_x(1 - \hat{f}(N_t))}{\varphi(x)} a_f(t-t_0) \geq 1 - C(t_0 \sup_{w \in D} P_{N_t}(|N_t| > 0) + e^{-\gamma t_0})$$

where we emphasise that the constant $C$ depends only on the branching diffusion, and not on $x$ or $f$.

We now move on to the upper bound. This is simpler and we do not need to use the spine change of measure. Indeed we can write

$$P_x(1 - \hat{f}(N_t)) = P_x(P_x(1 - \hat{f}(N_t) | \mathcal{F}_{t_0})) = P_x(1 - \prod_{u \in N_{t_0}} (1 - P_{X_u(t_0)}(1 - \hat{f}(N_{t-t_0}))))$$

$$\leq P_x \left( \sum_{u \in N_{t_0}} P_{X_u(t_0)}(1 - \hat{f}(N_{t-t_0})) \right),$$
Thus, to complete the proof we need only show that $a_f(t - t_0) \leq 1 + c e^{-\gamma t_0}$, and so combining the upper and lower bounds we have

$$\frac{\mathbb{P}^x(1 - \hat{f}(N_t))}{\varphi(x) a_f(t - t_0)} - 1 \leq C \left( e^{-\gamma t_0} + t_0 \sup_{w \in D} \mathbb{P}^w(|N_{t - t_0}| > 0) \right)$$

(4.4)

where again $C$ is a constant not depending on $x$ or $f$.

From here we can conclude. Because we know by Theorem 1.3 that $\sup_{w \in D} \mathbb{P}^w(|N_s| > 0) \to 0$ as $s \to 0$, we can pick a function $1 \leq t_0(t) < t$ (defined for all $t > 1$) such that the right hand side of (4.4) tends to $0$ as $t \to 0$. Thus, to complete the proof we need only show that $a_f(t) \sim a_f(t - t_0(t))$ as $t \to \infty$. However, this follows since

$$\left| \frac{a_f(t)}{a_f(t - t_0(t))} - 1 \right| = \left| \int_D \left( \frac{\mathbb{P}^x(1 - \hat{f}(N_t))}{\varphi(x) a_f(t - t_0(t))} - 1 \right) \varphi(x)^2 \chi_1 \right|$$

which converges to $0$, uniformly in $f$, by (4.4) and the choice of $t_0(t)$.

\[ \square \]

### 4.2 Asymptotics for $a(t)$

To complete the proof of Theorem 1.4, we need to show that $a_0(t) = a(t) \sim bt^{-1}$ as $t \to \infty$, where

$$b := 2(m - 1) / (\lambda, 1 - (1/2) \mathbb{E}(A^2) - \mathbb{E}(A)).$$

(4.6)

The idea is the following: if we write $\mathbb{P}^x(|N_t| > 0) = u(x, t)$ as before, then $u$ will satisfy a certain partial differential equation in $D$ (this is the FKPP equation if $L = \frac{1}{2}\Delta$). This will give us an ordinary differential equation that is satisfied by $a(t)$, and consequently we will be able to deduce the desired asymptotic. In fact, we will never explicitly use the fact that $u$ is a solution of this PDE (and instead derive the equation for $a$ directly) but this is the motivation behind our approach. Again we state the relevant lemma in a more general setting, as this will be helpful later on. We let $u_f(x, t) = \mathbb{P}^x(1 - \hat{f}(N_t))$ for any measurable $0 \leq f < 1$ on $D$, and note that $u_0$ corresponds to $u$ defined above.

**Lemma 4.4.** Suppose that $0 \leq f < 1$ and $f \in C(\bar{D})$. Then $a_f(t) := \int_D \varphi(x) u_f(x, t)$ is differentiable for all $t > 0$ and

$$\frac{da_f}{dt}(t) = \frac{\lambda}{m - 1} \int_D \left( G(1 - u_f(x, t)) + m u_f(x, t) - 1 \right) \varphi(x) dx$$

(4.7)

where $G(s) = \mathbb{E}(s^A)$ is the generating function for $A$.

**Proof.** The idea is to rewrite $u_f(x, t) = \mathbb{P}^x(1 - \hat{f}(N_t))$, by conditioning on the first branching time for the system. This gives us that $a_f(t)$ is equal to

$$\int_D \varphi(x) e^{-\frac{\lambda}{m - 1} \mathbb{P}^x(1 - f(X(t)))} dx$$

$$+ \int_D \varphi(x) \int_0^t \frac{\lambda}{m - 1} e^{-\frac{\lambda s}{m - 1} \mathbb{P}^x(1 - G(1 - u_f(X(s), t - s)))} ds dx.$$

17
The key fact that allows us to simplify this expression is that for all $y \in D$ and $t > 0$ we have $\int_D p_t^D(x,y)\varphi(x)\,dx = e^{-\lambda t} \varphi(y)$ by (2.5). Hence, by Fubini, we have

\[
\begin{align*}
a_f(t) &= e^{-\lambda(t/(m-1))} \int_D \varphi(y)(1 - f(y)) \\
&\quad + e^{-\lambda(t/(m-1))} \int_0^t \int_D \frac{\lambda}{m-1} e^{\lambda s/(m-1)} \varphi(y)(1 - G(1 - u_f(y,s))) \,dy \,ds
\end{align*}
\]

where we have also made the change of variables $s \leftrightarrow t - s$ in the integral. Now we claim that for every $y \in D$, $s \mapsto u_f(y,s)$ is a continuous function on $[0, \infty)$. Assuming this is true (we will prove it momentarily) we see by dominated convergence (and the definition of $G$) that the time integrand in (4.8) is continuous in $s$. Thus, by the fundamental theorem of calculus, $a_f(t)$ is differentiable and

\[
\frac{da_f(t)}{dt} = -\lambda \left( \frac{1}{m-1} + 1 \right) a_f(t) + \frac{\lambda}{m-1} \int_D \varphi(y)(1 - G(1 - u_f(y,t))) \,dy.
\]

Writing $a_f(t) = \int_D \varphi(x)u_f(x,t)\,dx$ to make the right hand side of the above into a single integral over $D$, the result follows.

So, we only need to prove the claim concerning continuity. Fix $t \in [0, \infty)$ and $y \in D$. Then for any $s \in [0, \infty)$

\[
|u_f(y,s) - u_f(y,t)| = |\mathbb{P}^y(\hat{f}(N_s) - \hat{f}(N_t))| = |\mathbb{P}^y(\mathbb{I}_{\{\tau \wedge s > 0\}} \left( \bigwedge_{u \in N_{t \wedge s}} f(X_u(t \wedge s)) - \bigwedge_{u \in N_{t \wedge s}} \mathbb{P}^{X_u(t \wedge s)}(\hat{f}(N_{t - s})) \right))|
\]

where for the second equality we have conditioned on $\mathcal{F}_{t \wedge s}$. By the triangle inequality (and using the fact that $f < 1$), this is less than

\[
\mathbb{P}^y \left( \sum_{u \in N_{t \wedge s}} |f(X_u(t \wedge s)) - \mathbb{P}^{X_u(t \wedge s)}(\hat{f}(N_{t - s}))| \right), \tag{4.9}
\]

which we can also write as

\[
e^{\lambda(t \wedge s)} \mathbb{E}^{y}(\mathbb{I}_{\{t \wedge s \geq \tau \wedge s\}} f(X(t \wedge s)) - \mathbb{P}^{X_u(t \wedge s)}(\hat{f}(N_{t - s})))] = e^{\lambda(t \wedge s)} \int_D p^{D}_{t \wedge s}(y,z)|f(z) - \mathbb{P}^z(\hat{f}(N_{t - s})| \,dz
\]

by the many-to-one lemma. Therefore by dominated convergence, since by Lemma 2.1 we can bound $p^D_r(y,z)$ above by something integrable in $z$, uniformly in $r > t/2$ for example, it is enough to show that

\[
|f(z) - \mathbb{P}^z(\hat{f}(N_r))| \to 0
\]

pointwise in $z \in D$ as $r \to 0$.

To do this, we bound

\[
\mathbb{P}^z(f(z) - \hat{f}(N_r)) \leq (1 - e^{-\lambda/(m-1)r}) + f(z)\mathbb{P}^z(\tau^D \leq r) + \int_D p^D_r(z,w)|f(z) - f(w)| \,dz, \tag{4.10}
\]

which follows since the probability of the first branching time being less than $r$ is equal to $(1 - e^{(m-1)^{-1}\lambda r})$, and on this event $|f(z) - \hat{f}(N_r)| \leq 1$. Clearly the first two terms of (4.10) go to 0 as $r \to 0$. Moreover so does the third, since $p^D_r(z,w) \to 0$ as $r \to 0$ on $|z - w| > \eta$ for any $\eta > 0$, and $\hat{f}$ is uniformly continuous in $D$. The claim concerning $p^D_r(z,w)$ follows because $D$ is Lipschitz and therefore $p^D_r$ is intrinsically ultracontractive, which means that there exist constants $c_1, c_2$ such that

\[
p^D_r(w,z) \leq c_1 r^{-d/2} e^{-\frac{|w-z|^2}{c_2 r}},
\]

for all $w, z \in D$ (see for example [Bañ99 eq. (1.2)].)
Now let us see how this allows us to conclude the proof of Theorem 1.4.

Proof of Theorem 1.4. First we observe that since \( A \) has finite variance, \( s^{-2}(G(1-s) + ms - 1) \to \frac{1}{2}(\mathbb{E}[A^2] - \mathbb{E}[A]) \) as \( s \to 0 \) (by dominated convergence). This means, by Lemma 4.4, that we can write, for \( a(t) = a_0(t) \) and denoting \( \dot{a}(t) = \frac{da}{dt}(t) \):

\[
\frac{\dot{a}(t)}{a(t)^2} = -b^{-1}(1, \varphi^3)^{-1} \int_D \varphi(x) \frac{u^2(x, t)}{a(t)^2} (1 + E(x, t)) \, dx
\]

where \( E(x, t) \to 0 \) uniformly in \( x \) as \( t \to \infty \) (here we have also used that the survival probability decays uniformly to 0 in \( D \)). Now by Proposition 4.1, we can conclude that

\[
\frac{d}{dt} \left( -\frac{1}{a(t)} \right) = -\frac{\dot{a}(t)}{a(t)^2} = b^{-1} + \dot{E}(t)
\]

where \( \dot{E}(t) \to 0 \) as \( t \to \infty \). Using the fundamental theorem of calculus, [Rud76, Theorem 6.21] (note we only need integrability of the derivative), we deduce the result. \( \square \)

5 The Conditioned System

Theorem 1.4 allows us to study the law of our branching diffusions conditioned to survive for a long time in much greater depth. One aspect of the limiting behaviour is captured by what happens to the law of the process run up to some fixed time \( T \), if it is then conditioned to survive until a much larger time \( t \). It turns out that this limiting description is given precisely by the evolution of the process under \( Q_x \), as described in Section 2.4.3.

Proof of Proposition 1.5. Recall, we would like to prove that for any \( T \geq 0, x \in D \) and \( B \in \mathcal{F}_T \), we have that

\[
\lim_{t \to \infty} \mathbb{P}^x(B||N_t| > 0) = \mathbb{Q}_x(B).
\]

Conditioning on \( \mathcal{F}_T \), we see that

\[
\mathbb{P}^x(B||N_t| > 0) = \frac{\mathbb{P}^x(1_B \mathbb{P}^x(\{N_t| > 0| \mathcal{F}_T \})}{\mathbb{P}^x(\{N_t| > 0)} := \frac{\mathbb{P}^x(1_B Y)}{\mathbb{P}^x(\{N_t| > 0)}
\]

where we have defined

\[
Y := \mathbb{P}^x(\{N_t| > 0| \mathcal{F}_T) = \sum_{u \in N_T} \mathbb{P}^{X_u(T)}(\{N_{t-T}| > 0} \left( \prod_{v \neq u \in N_T} \mathbb{P}^{X_v(t)}(\{N_{t-T}| = 0} \right).
\]

Then, from our asymptotic for the survival probability, Theorem 1.4, and the fact that \( \frac{t}{t-T} \to 1 \) as \( t \to \infty \), it follows that

\[
\mathbb{P}^x(\{N_t| > 0) \to \sum_{u \in N_T} \varphi(X_u(T)) \frac{\mathbb{P}^{X_u(T)}(\{N_{t-T}| > 0}}{\varphi(x)} \mathbb{P}^{X_u(t)}(\{N_{t-T}| = 0} \right)
\]

almost surely, as \( t \to \infty \). Moreover, we have that \( Y \leq \sum_{u \in N_T} \mathbb{P}^{X_u(T)}(\{N_{t-T}| > 0} \leq C \frac{M_T}{t-T} \) for all large enough \( t \) and some \( C > 0 \) depending only on the diffusion. This means we can dominate \( \mathbb{P}^x(\{N_t| > 0) \) by an integrable random variable, namely a constant multiple of \( M_T \). The dominated convergence theorem then provides the result. \( \square \)

Given the asymptotic for the survival probability, it is also not too much work to prove Theorem 1.6, which gives some limiting information on the positions of particles at time \( t \), given survival. Recall we would like to show
that for any $f$ with $(f^2, \varphi) < \infty$ (this condition is required because we will apply the many-to-two Lemma in the proof) we have
\[
(t^{-1} \sum_{u \in N_t} f(X_u(t)) \big| |N_t| > 0) \to Z
\]
in distribution as $t \to \infty$, where $Z \sim \text{Exp}(b(\varphi, f)^{-1})$ and $b = \frac{2(m-1)}{\lambda(m-1)(E(A^2) - E(A))}$ as before.

We first need an auxiliary Lemma.

**Lemma 5.1.** Recall the definition of $a_f(t)$ from Lemma 4.4 Then
\[
\frac{1}{t} \left( \frac{1}{a_f(t)} - \frac{1}{a_f(0)} \right) \to b^{-1},
\]
uniformly over all $f \in C(\bar{D}, [0, 1)) := \{ f \in C(\bar{D}) : 0 \leq f < 1 \}$.

**Proof.** Note that $a_f(t) > 0$ for all $t$, by Remark 4.3. Exactly as in the proof of Theorem 1.4, we obtain that
\[
\frac{d}{dt} \left( \frac{1}{a_f(t)} \right) = b^{-1} + \hat{E}_f(t)
\]
where $\hat{E}_f(t) \to 0$ as $t \to \infty$, uniformly in $f \in C(\bar{D}, [0, 1))$. The uniformity comes from the fact that $u^f(x,t) \leq \mathbb{P}^x(|N_t| > 0)$ tends to 0 uniformly over $D \times C(\bar{D}, [0, 1))$, and that the convergence in Proposition 4.1 is uniform over $C(\bar{D}, [0, 1))$. The result then follows by the fundamental theorem of calculus. \(\square\)

**Proof of Theorem 1.6.** We first complete the proof in the special case $f = \varphi$. We will show that for any $\alpha > 0$
\[
\mathbb{P}^x \left( e^{-\frac{\alpha}{t} \sum_{u \in N_t} \varphi(X_u(t))} \big| |N_t| > 0 \right) \to \frac{b}{b + \alpha} \tag{5.1}
\]
uniformly in $x$ as $t \to \infty$ (which is enough, by Lévy’s continuity theorem for the Laplace transform).

Fix $\alpha > 0$. To prove (5.1), we first observe that if we define $f_t(x) = e^{-\frac{x}{t} \varphi(x)}$, then
\[
\mathbb{P}^x \left( e^{-\frac{\alpha}{t} \sum_{u \in N_t} \varphi(X_u(t))} \big| |N_t| > 0 \right) = 1 - \frac{u_{f_t}(x,t)}{\mathbb{P}^x(|N_t| > 0)}.
\]

By Proposition 4.1 and Theorem 1.4, we also know that
\[
\frac{u_{f_t}(x,t)}{\varphi(x)a_{f_t}(t)} \to 1 \quad \text{and} \quad \frac{b\varphi(x)}{t\mathbb{P}^x(|N_t| > 0)} \to 1
\]
as $t \to \infty$, uniformly in $x \in D$. Note that it doesn’t matter that $f_t$ depends on $t$, since the convergence in Proposition 4.1 is uniform over $C(\bar{D}, [0, 1))$. Hence, we are required to show that
\[
\frac{1}{ta_{f_t}(t)} \to \frac{1}{b} + \frac{1}{\alpha}
\]
as $t \to \infty$. However, this follows directly from Lemma 5.1 and the fact that $t(1 - e^{-\frac{\alpha}{t} \varphi(z)}) \to \alpha \varphi(z)$ as $t \to \infty$ for any $z \in D$.

To deal with general $f$ such that $(f^2, \varphi) < \infty$, we write $\tilde{f} = f - (\varphi, f) \varphi$. We will show that for any $\varepsilon > 0$
\[
\mathbb{P}^x( |t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t))| > \varepsilon \big| |N_t| > 0 ) \to 0 \tag{5.2}
\]
as $t \to \infty$, uniformly in $x$. This clearly implies the result by writing $f = \tilde{f} + (\varphi, f) \varphi$ and applying the special case of Theorem 1.6 (with $\varphi$) that we have just proved.
To prove (5.2), it is enough by conditional Markov’s inequality to show that
\[ P^x \left( \left( t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t)) \right)^2 \mid |N_t| > 0 \right) \to 0 \quad \text{as} \quad t \to \infty, \quad \text{uniformly in} \quad x. \]
Moreover, the many-to-two lemma, Lemma 2.5, tells us that
\[ P^x \left( \left( t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t)) \right)^2 \mid |N_t| > 0 \right) = \frac{\varphi(x)}{t^{\varphi_x(|N_t| > 0)}} \quad (5.3) \]
where we know by Theorem 1.4 that the expression outside of the brackets on the right hand side is uniformly bounded in \( t \) (\( t \geq 1 \), say) and \( x \). We also know by (3.2) and Lemma 2.1 that
\[ \frac{e^{\lambda r} P^x \left( \tilde{f}(X(r)) \mathbf{1}_{\{r > r\}} \right)}{\varphi(x)} \leq C \varphi(f^2) \quad \text{and} \quad \frac{e^{\lambda r} P^x \left( \tilde{f}(X(r)) \mathbf{1}_{\{r > r\}} \right)}{\varphi(x)} \leq C e^{-\gamma r} (\varphi, f^2)^{1/2} \quad (5.4) \]
for all \( r > 1 \) and some constant \( C \) depending only on the branching diffusion. Note that the second equality uses the fact that \( (\varphi, \tilde{f}) = 0 \) (and Cauchy–Schwarz). Finally, we observe that for \( s > 1 \), by Cauchy–Schwarz and (3.2) again,
\[ \varphi(x)^{-1} P^x \left( \mathbf{1}_{\{r > s\}} \left[ P^x(s) \left( \tilde{f}(X(t-s)) \right)^2 \right] \right) \leq \varphi(x)^{-1} P^x \left( \mathbf{1}_{\{r > s\}} P^x(s) \left( \tilde{f}(X(t-s)) \right)^2 \right) \]
\[ = \varphi(x)^{-1} P^x \left( \mathbf{1}_{\{r > s\}} \left[ \tilde{f}(X(t)) \right]^2 \right) \]
\[ \leq C e^{-\lambda t} \varphi(f^2) \]
where \( C \) is another, possibly different, constant. This tells us that
\[ \lim_{t \to \infty} \frac{\int_{t-1}^t \lambda e^{\lambda (2t-s)} P^x \left( \mathbf{1}_{\{r > s\}} \left[ P^x(s) \left( \tilde{f}(X(t-s)) \right)^2 \right] \right) ds}{\varphi(x)t} \to 0 \]
as \( t \to \infty \), uniformly in \( x \). Using (5.4) on the remaining parts of the second line of (5.3), together with the fact that
\[ P^x \left( \mathbf{1}_{\{r > s\}} \varphi(X(s))^2 \right) \leq \sup_{w \in D} |\varphi(w)| P^x \left( \mathbf{1}_{\{r > s\}} \varphi(X(s)) \right) \leq \varphi(x) \sup_{w \in D} |\varphi(w)| e^{-\lambda s} \quad (5.5) \]
for all \( s > 0 \), it is easy to conclude that \( P^x \left( \left( t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t)) \right)^2 \mid |N_t| > 0 \right) \) does indeed tend to 0, uniformly in \( x \).

We conclude by explaining how one can obtain Corollary 1.7 from here, which describes the asymptotic distribution of a particle picked at random from the population, given survival.

**Proof of Corollary 1.7.** To prove the Corollary we first show that for any \( \varphi \) with \( (\varphi, \tilde{f}) < \infty \)
\[ P_x \left( \left( \sum_{u \in N_t} \frac{f(X_u(t))}{\varphi(X_u(t))} - (\varphi, \tilde{f}) > \varepsilon \mid |N_t| > 0 \right) \to 0 \quad (5.6) \]
as \( t \to \infty \). Defining \( \tilde{f} \) as in the proof of Theorem 1.6, the left hand side of (5.6) is equal to
\[ \begin{align*}
P^x \left( \left( t^{-1} \sum_{u \in N_t} \frac{\tilde{f}(X_u(t))}{\varphi(X_u(t))} > \varepsilon \mid |N_t| > 0 \right) \leq P^x \left( t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t)) > \delta \mid |N_t| > 0 \right) \right.
\end{align*} \]
\[ \begin{align*}
+ P^x \left( t^{-1} \sum_{u \in N_t} \varphi(X_u(t)) < \frac{\delta}{\varepsilon} \mid |N_t| > 0 \right) \right)
\]
for any $\delta > 0$. From (5.2) and Theorem 1.6 if we take a limit as $t \to \infty$ on the right hand side, we are left with simply the probability that an exponential random variable is less than $\delta / \varepsilon$. Taking $\delta \to 0$ proves (5.6). The corollary then follows by applying the above with both $f$ and the constant function 1, and writing $\sum f(X_u(t))/|N_t| = \sum f(X_u(t))/\sum \varphi(X_u(t)) \times \sum \varphi(X_u(t))/|N_t|$. \hfill $\Box$

Note that by (5.2), for any fixed $\delta$ the first term on the right hand side of (5.7) converges to 0 uniformly in $x$. Moreover, by Markov’s inequality we can write

$$\mathbb{P}^x \left( \sum_{u \in N_t} \varphi(X_u(t)) < \frac{\delta}{\varepsilon} |N_t| > 0 \right) \leq e^{1/\varepsilon} \mathbb{P}^x \left( e^{-\frac{1}{\varepsilon} \sum_{u \in N_t} \varphi(X_u(t))} | |N_t| > 0 \right)$$

which by (5.1) converges to $e^{1/\varepsilon} b(b + \delta^{-1})^{-1}$ as $t \to \infty$, uniformly in $x$. This implies the following:

**Corollary 5.2.** For $f$ as in Corollary 1.7 and any $\varepsilon > 0$, the convergence to 0 in (5.6) is uniform in the starting position $x$.

We will use this to prove a stronger version of Corollary 1.7. Corollary 1.7 tells us that the average value of $f(X_v(t))$ among all vertices $v \in N_t$, given survival, converges to $(f, \varphi)/(1, \varphi)$. The next lemma will tell us that in fact we need only look at the average over a large enough subset of $N_t$ (we will see precisely what this means in a moment). First we need some notation. Recall from Section 2.2.2 that we can view our branching diffusion as a marked tree $(T, t, X) \in \mathcal{T}$. This means that for every $t$ such that $N_t \neq \emptyset$, we can write $N_t = \{w_1(t), w_2(t), \ldots, w_{|N_t|}(t)\}$, where the indices correspond to the depth first ordering of the particles (see Section 2.2.3). For any $M \in \mathbb{N}$ we can then define the set

$$N_{t,M} = \begin{cases} \{w_i(t) : M \leq i \leq |N_t|\} & \text{if } M \leq |N_t| \\ \emptyset & \text{if } M > |N_t|. \end{cases}$$

**Lemma 5.3.** Let $f$ be a bounded measurable function on $D$. Then for any $\varepsilon, \rho > 0$ and $x \in D$

$$\mathbb{P}^x \left( B_{t,\varepsilon}^\rho(\rho) \mid |N_t| > 0 \right) := \mathbb{P}^x \left( \sum_{u \in N_t} \frac{f(X_u(t))}{M} - \frac{(f, \varphi)}{(1, \varphi)} > \varepsilon \right) \mid |N_t| > 0$$

converges to 0 as $t \to \infty$.

**Proof.** We will prove the lemma by looking at the system conditioned to survive until time $t$, and dividing $N_t$ into families, depending on whether or not particles have the same ancestor at some earlier time. This earlier time will be chosen such that with high probability, the average value of $f(X_v(t))$ over any one of these families is close to $(f, \varphi)/(1, \varphi)$. This will show that there are many subsets of $N_t$ over which the average value of $f(X_v(t))$ is close to what we want. To extend this to all large enough subsets of $N_t$ (as in the statement of the lemma), we will show that the size of each of these families is very small compared to $t$.

To do this, fix $\varepsilon > 0$ and write

$$p(t, \varepsilon) := \sup_{x \in D} \mathbb{P}^x \left( \sum_{u \in N_t} \frac{f(X_u(t))}{|N_t|} - \frac{(f, \varphi)}{(1, \varphi)} > \varepsilon/2 \right) = 0$$

which by Corollary 5.2 converges to 0 as $t \to \infty$. This means that we can choose a function $g(t, \varepsilon) \leq t$ such that $g(t, \varepsilon) \to \infty$ as $t \to \infty$, but

$$g(t, \varepsilon) \mathbb{P} \left( \frac{t}{g(t, \varepsilon)}, \varepsilon \right) \to 0$$

as $t \to \infty$. \hfill $\Box$

\footnote{Indeed, since $\tilde{p}(t) := \sup \{ p(u, \varepsilon) : u \geq t \}$ converges monotonically to 0 as $t \to \infty$, you can choose $g(t, \varepsilon) \leq \sqrt{t}$ but still converging to $\infty$, such that $g(t, \varepsilon) \mathbb{P}(\sqrt{t}) \to 0$ as $t \to \infty$. Then since $p(t/g(t, \varepsilon), \varepsilon) \leq \tilde{p}(t/g(t, \varepsilon)) \leq \tilde{p}(\sqrt{t})$, the function $g$ will satisfy (5.8).}
As mentioned above, we will break up the set \( N_t \) into families. Two vertices will be in the same family if they have a common ancestor at time \( t - t/g(t, \varepsilon) \). For \( 1 \leq i \leq |N_{t-t/g(t, \varepsilon)}| \) we define

\[
m_i := \frac{\sum_{v \in N_t : w_i(t - t/g(t, \varepsilon)) \prec v} f(X_v(t))}{\sigma_i}
\]

for \( \sigma_i := |\{ v \in N_t : w_i(t - t/g(t, \varepsilon)) \prec v \}| \) to be the average value of \( f \) among the \( i \)th of these families. If \( w_i(t - t/g(t, \varepsilon)) \) has no descendants at time \( t \) we set \( m_i = (f, \varphi)/(1, \varphi) \). The key to the proof of this lemma will be to show that

\[
\mathbb{P}^x(A_t \mid |N_t| > 0) := \mathbb{P}^x\left(\bigcup_{i=1}^{\lfloor N_{t-t/g(t, \varepsilon)} \rfloor} \left\{ m_i - \frac{\{f, \varphi\}}{\{1, \varphi\}} > \varepsilon/2 \right\} \mid |N_t| > 0 \right) \to 0
\]

as \( t \to \infty \). To do this, it turns out to be more convenient to look at the unconditioned probability, and then use Theorem 1.4. We write

\[
\mathbb{P}^x(A_t \cap \{|N_t| > 0\}) = \mathbb{P}^x\left( A_t \cap \{|N_t| > 0\} \mid \mathcal{F}_{t-t/g(t, \varepsilon)} \right)
\]

\[
\leq C\mathbb{P}^x(|N_{t-t/g(t, \varepsilon)}|) \mathbb{P}(\frac{t}{g(t, \varepsilon)}, \varepsilon) g(t, \varepsilon)
\]

for some \( C \) depending only on the branching diffusion, where the inequality follows from Theorem 1.4, a union bound, and the definition of \( p(s, \varepsilon) \). Moreover, by Lemma 2.4 and (3.2), we have that \( \mathbb{P}^x(|N_{t-t/g(t, \varepsilon)}|) \leq C' \varphi(x) \) for some further \( C' \), and so the expression on the second line of (5.10) is less than or equal to a constant times \( t^{-1} \varphi(x) p(t/g(t, \varepsilon), \varepsilon) g(t, \varepsilon) \). Finally, dividing left hand side of (5.10) by \( \mathbb{P}^x(|N_t| > 0) \sim t^{-1} \varphi(x) \) and then applying (5.8), we obtain (5.9).

Now we prove that

\[
\mathbb{P}_x(A_t' \mid |N_t| > 0) := \mathbb{P}_x\left( \bigcup_{i=1}^{\lfloor N_{t-t/g(t, \varepsilon)} \rfloor} \left\{ \sigma_i > \frac{t}{(g(t, \varepsilon))^{1/3}} \right\} \mid |N_t| > 0 \right) \to 0
\]

as \( t \to \infty \). For this we apply a similar argument to above. We write

\[
\mathbb{P}^x(A_t' \cap \{|N_t| > 0\}) = \mathbb{P}^x\left( A_t' \cap \{|N_t| > 0\} \mid \mathcal{F}_{t-t/g(t, \varepsilon)} \right)
\]

\[
\leq C\mathbb{P}^x(|N_{t-t/g(t, \varepsilon)}|) \sup_{y \in D} \mathbb{P}^y\left( |N_{t/g(t, \varepsilon)}| > \frac{t}{g(t, \varepsilon)^{1/3}} \right) \leq \frac{C\varphi(x)}{g(t, \varepsilon)^{1/3} t}
\]
where the last inequality follows because for any $y \in D$

$$
\mathbb{P}^y(\{|N_{t}\} > \frac{t}{g(t,\varepsilon)}) \leq \frac{t}{g(t,\varepsilon)^{2/3}} \leq C \frac{g(t,\varepsilon)^{2/3}}{t^2}
$$

by Markov's inequality and the uniform bound on $\mathbb{P}^y(\{|N_{t}\}^2)$ from Lemma 3.2. Now dividing through by $\mathbb{P}^y(|N| > 0)$ on the right hand side of (5.12) gives (5.11).

Finally, we are in a position to prove that $\mathbb{P}_x(B_t^{f,\varepsilon}(\rho) \mid |N| > 0) \to 0$ as $t \to \infty$. By the above work, and a union bound it is enough to show that

$$B_t^{f,\varepsilon}(\rho) \subset A_t \cup A_t'$$

for all $t$ large enough. Suppose we are on the event $\{A_t \cup A_t'\}^c$, and $\rho t \leq |N_t|$ (it is enough to consider this and show that we must be on $\{B_t^{f,\varepsilon}(\rho)\}^c$, since if we are on the event $\{A_t \cup A_t'\}^c$ and $\rho t > |N_t|$ then by definition we are not on $B_t^{f,\varepsilon}(\rho)$). For every $M \leq |N_t|$, set $M \leq k(M) := \sigma_0 + \sigma_1 + \cdots + \sigma_{i+1}$, where $\sigma_0 := 0$, and $i$ is the unique integer such that $\sigma_0 + \cdots + \sigma_i < M \leq \sigma_0 + \cdots + \sigma_{i+1}$. Then because we are on the event $\{A_t\}^c$ we have that

$$\left(\sum_{u \in N_{t,k(M)}} f(X_u(t)) / k(M) \right) - \left(\frac{f(u)}{1}\right) \leq \varepsilon / 2$$

for all $\rho t \leq M \leq |N_t|$ simultaneously (in fact, for all $1 \leq M \leq |N_t|$). Furthermore, since we are on the event $\{A_t'\}^c$, for $M \geq \rho t$ we have

$$\frac{\sum_{u \in N_{t,M}} f(X_u(t))}{k(M)} - \frac{\sum_{u \in N_{t,k(M)}} f(X_u(t))}{k(M)} \leq \left| k(M) - M \right| \left( \frac{\sum_{u \in N_{t,k(M)}} f(X_u(t))}{k(M)} + \sup_{w \in D} |f(w)| \right)$$

$$\leq \frac{1}{\rho g(t,\varepsilon)^{1/3}} \left( \frac{2}{1}\right) \left( \frac{f}{1}\right) + \sup_{w \in D} |f(w)|.$$

Thus, for all $t$ large enough, because $g(t,\varepsilon) \to \infty$, this must be less than $\varepsilon / 2$ for all $M \geq \rho t$, and consequently we must be on the event $\{B_t^{f,\varepsilon}(\rho)\}^c$. 

\section{Convergence to the Brownian CRT}

From this point onwards, we will assume that $\varphi \in C^1(D)$ as in the statement of Theorem 1.1. In particular this means that all the first order partial derivatives of $\varphi$ are bounded on $D$.

Recall from Section 2.2.3 that we associate to our branching diffusion $(T, l, X)$ a continuous plane tree $T = T(T, l, X)$. This is the continuum tree with branch lengths given by lifetimes of particles in the system. Let us first describe the continuous time depth first exploration of $T$. Recall that if $|T|$ is the total number of individuals in $T$, then we can order them $\{u_0, u_1, \ldots, u_{|T|}\}$ with respect to the depth first order. The depth first exploration is then the continuous time process $v_t$ on $[0, \sum_{0 \leq j \leq |T|} l_{u_j}) := [0, L(T))$, taking values in $T$, defined by

$$v_t := u_i \quad \text{if} \quad \sum_{j < i} l_{u_j} \leq t < \sum_{j \leq i} l_{u_j} \quad (6.1)$$

where we recall the definition of the lifetime $l_v$ of an individual $v$ from Section 2.2.3. So this process visits the individuals in depth first order, and spends time $l_u$ visiting individual $u$. We let $\kappa_t = \sum_{j < i} l_{u_j}$ if $v_t = u_i$, so $\kappa_t = \inf\{s \leq t : v_s = v_i\}$ is the time at which the depth first exploration “started visiting” the individual $v_t$. Although this depth first exploration takes values in $T$, we should think of it as a continuous exploration of $T$, where between visiting vertices, the branches of $T$ are traversed at speed one. See Figure 6.1 for an illustration: note that we shall always refer to the exploration as an exploration of $T$ rather than an exploration of $T$, in order to keep the correct intuition.

24
Figure 6.1: A sketch of the depth first exploration of a continuous tree. Full arrowed lines represent motion at speed one in the vertical direction. Dotted arrowed lines represent instantaneous jumps.

Now we describe the height function associated to the depth first exploration. This is defined by

\[ H_t := (s_{v_t} - l_{v_t}) + (t - \kappa_t), \]  

(6.2)

so describes the evolution of height if we think of our depth first exploration process as described in the previous paragraph (again see Figure 6.1). Remember that in Theorem 1.1 we want to describe the properties of \((T, d)\) as a metric space (see Section 2.2.3 for the definition of \(d\)). The height function will be important to achieving this for the following reason:

**Remark 6.1.** Consider the interval \([0, L(T))\) on which the depth first exploration is defined, and for \(r, w \in [0, L(T))\) set

\[ d^*(r, w) = H_r + H_w - 2 \inf_{s \in [r \wedge w, r \vee w]} H_s. \]

(6.3)

Then this defines a pseudo-metric on \([0, L(T))\): we can have \(d^*(r, w) = 0\) if \(r\) and \(w\) correspond to two times at which a “branch point” on \(T\) is visited. We can therefore define a metric space by quotienting \([0, L(T))\) using the equivalence relation \(r \sim w \iff d^*(r, w) = 0\). It is easy to check that \((0, L(T)), d^*)\) is \(\mathbb{P}^x\)-almost surely isometrically equivalent to \((T, d)\).

In fact, for the proof of Theorem 1.1 in the end we will prove that if we look at \((0, L(T)), \frac{1}{n}d^*)\) under the law \(\mathbb{P}^x(\cdot \mid |N_n| > 0)\), then these random metric spaces will converge, in distribution as \(n \to \infty\), to the Brownian CRT \((T_e, d_e)\).

There is one more important definition to make. It should be clear that any time \(t\) in our exploration is naturally associated with a position in \(D\). That is, if \(v_t\) has associated motion \(X_{v_t} : [s_{v_t} - l_{v_t}, s_{v_t}) \to D\), then we can write

\[ V_t := X_{v_t}((s_{v_t} - l_{v_t}) + (t - \kappa_t)). \]

(6.4)

This process \(V_t\) then describes the evolution of position in \(D\), if we do a continuous depth first exploration of \(T\), and we follow the motion of an individual \(u\) over its lifetime during the time that we are visiting it.

Now we will start setting up for the proof of Theorem 1.1. A classical technique in proofs of this sort (see for example [LGD02, LG05]) is to, instead of considering one tree conditioned to be large, to consider an i.i.d. sequence of trees without any conditioning. Describing the scaling limit of this process then allows one to also describe the scaling limit of a single “large” tree.

So, we write \(\mathbb{P}^x\) for the law of a sequence of i.i.d trees \(((T^1, l^1, X^1), (T^2, l^2, X^2), \cdots)\), where each \((T^i, l^i, X^i)\) is distributed as \((T, l, X)\) under \(\mathbb{P}^x\). We concatenate the continuous depth first explorations of the trees in the natural way, and write \(H_t\) for the associated concatenation of the height processes. To show the convergence in Theorem 1.1 it will be important to show that \(H_t\) when rescaled appropriately, looks like a reflected Brownian motion. To do this, we introduce a further process, \(\tilde{S}_t\), which will turn out to be a martingale. This can be thought of as an analogue of the Lukasiewicz path used in [LGDO2].
In the following, we write
\[ \Lambda_t := \inf_{j \geq 1} \sum_{i=1}^j L(T^j) > t \]  
for the index of the tree being visited at time \( t \) in the continuous depth first exploration. We also write
\[ \bar{v}_t := (v_t, \Lambda_t) \in T^{\Lambda_t} \times \{\Lambda_t\} \subset \Omega \times \mathbb{N}, \]
so \( v_t \) is the individual in \( T^{\Lambda_t} \) being visited at time \( t \) in the exploration of the sequence of trees. We still write \( V_t \)
for the associated position as defined in \((6.4)\). Finally, we say that for any \( u, v \in \Omega \), \( w <_s u \), if there exist \( j < k \in \mathbb{N} \) and \( v \in \Omega \) such that \( u = v_j \) and \( w = v_k \). If we have a tree \( T \) and an individual \( u \in T \) we write
\[ Y(u, T) = \{w \in T : w <_s v \text{ for some } v \leq u\} \]
for the set of younger siblings in \( T \) along the ancestry of \( u \).

**Definition 6.2.** Suppose we have a sequence of trees \(((T^i, l^i, X^i))_{i \geq 0}\). We define for \( t \geq 0 \)
\[ \bar{S}_t = \bar{\varphi}(V_t) - \sum_{t \leq \Lambda_t} \varphi(X^i_0(0)) + \sum_{w \in \bar{Y}(v_t, T^{\Lambda_t})} \varphi(X^i_w(s^i_w - l^i_w)). \] \((6.6)\)

Informally, the process \( \bar{S}_t \) can be defined as follows. We do our continuous depth first exploration of the sequence of trees and, at first, \( \bar{S}_t \) is just equal to \( -\varphi(x) \), plus a term that follows \( \varphi(V_t) \): that is, \( \varphi \) applied to the position of the individual we are visiting at time \( t \). The \( -\varphi(x) \) term is just there so that \( \bar{S}_0 = 0 \). As the exploration evolves, we always keep the term that follows \( \varphi(V_t) \), but also, whenever we pass a branch point at time \( t \), we add on \( \varphi(V_t) \) times the number of particles born minus 1. Whenever we reach a leaf (but not the end of a tree) we jump down to the next particle to be visited in the depth first exploration, and then \( \varphi \) of the position of that particle, which was before included in the third term of \((6.6)\), becomes the \( \varphi(V_t) \) term: the first term of \((6.6)\). When we reach the end of a tree, we subtract \( \varphi(x) \), but then start the depth first exploration of the next tree (so in particular \( \varphi(V_t) \) at such a time becomes \( \varphi(x) \)).

In fact, \( \bar{S}_t \) is very closely related to the martingale \( M_t \) defined in Section 2.3. Essentially they are the same process, but “explored” in different orders, and we will see that this is enough to preserve the martingale property. The overall idea is that we would like to approximate the height process by \( \bar{S}_t \), and then apply an invariance principle for this martingale.

This is an analogous idea to that used to prove convergence of Galton–Watson processes to the CRT in [LGD02], where \( \bar{S}_t \) here plays the role of the Lukasievicz path. We first record a property of this process, which will be essential to showing a relationship with the height function:

**Lemma 6.3.** Let \( \bar{S}'_t = \bar{S}_t - \varphi(V_t) \) and \( \bar{I}'_t = \inf_{0 \leq s \leq t} \bar{S}'_s \). Then
\[ S_t := \bar{S}'_t - \bar{I}'_t = \sum_{w \in \bar{Y}(v_t, T^{\Lambda_t})} \varphi(X^i_w(s^i_w - l^i_w)). \]

**Proof.** Since \( \varphi \) is positive, it is clear that \( \bar{I}'_t = -\sum_{t \leq \Lambda_t} \psi(X^i_0(0)) \). This implies the result. \( \square \)

**Definition 6.4.** If we have a single tree \((T, X, l)\), with continuous depth first exploration process \( v_t \), we define the corresponding processes
\[ S_t := \varphi(V_t) + \sum_{w \in \bar{Y}(v_t, T)} \varphi(X_w(s_w - l_w)) \quad \text{and} \quad S_t := \sum_{w \in \bar{Y}(v_t, T)} \varphi(X_w(s_w - l_w)). \]

**Remark 6.5.** Note that \( (S_t)_t \) is a process that starts and ends at 0 (at time \( L(T) \)) and is positive in between, i.e. an excursion.

For \( y > 0 \) the law of \( (S_t)_t \) under \( \mathbb{P}^x \), conditioned on \( \{\sup_t S_t \geq y\} \), is the same as the law of \( (\bar{S}_t)_t \) under \( \mathbb{P}^x \), restricted to the first excursion in which it exceeds \( y \).

---

\( ^4 \)that is, a point on the tree corresponding to one individual dying, and being replaced with a strictly positive number of offspring
6.1 Martingale Convergence

We let

\[ \mathcal{F}_t := \sigma \{ \bigcup_{i < \Lambda_t} (T_i, l_i, X_i^i) \}, \{(l_w^0, X_w^0, A^0_w) : \bar{v}_r = (w, \Lambda_t), r < \kappa_t \}, \{X_{v_t}^i(r) : r \in [s_{v_t}^i - l_{v_t}^i, s_{v_t}^i - l_{v_t}^i + (t - \kappa_t)] \} \]

be the filtration that knows everything about the continuous depth first search of the sequence of trees \((T_i, l_i, X_i^i)\) up to time \(t\). That is, \(\mathcal{F}_t\) is the filtration encoding: (i) which individuals in which trees are visited and when before time \(t\); (ii) their spatial motions (although for \(v_t\) only up until the point in its lifetime that has been explored by time \(t\)); and (iii) their progenies, or offspring, (apart from that of \(v_t\)).

**Lemma 6.6.** Under \(\mathbb{P}^x\) the process \((\bar{S}_t)_{t \geq 0}\) is a martingale with respect to \((\mathcal{F}_t)_{t \geq 0}\).

**Proof.** There are three steps to this proof. In **Step 1** we show that for the single tree process \(S_t\) under \(\mathbb{P}^x\) (see Definition 6.4), \(\mathbb{P}^x(S_t)\) is differentiable with derivative 0 at \(t = 0\). In **Step 2** we show how this implies that

\[ \mathbb{P}^x(S_t) = \varphi(x). \]  

(6.7)

for all \(x \in D\) and \(t \geq 0\). Finally, in **Step 3** we deduce the martingale property.

**Step 1.** For \(\delta > 0\), we write

\[ \mathbb{P}^x(S_\delta) = \mathbb{P}^x(S_\delta 1_{\{v_0 = \emptyset\}}) + \mathbb{P}^x(S_\delta 1_{\{v_0 = 1\}}) + \mathbb{P}^x(S_\delta 1_{\{v_0 \in \bigcup_{n \geq 2} \{1\}^n\}}) + \mathbb{P}^x(S_\delta 1_{\{v_0 \neq 0, v_0 \in \bigcup_{n \geq 1} \{1\}^n\}}) \]  

(6.8)

where \(\bigcup_{n \geq k} \{1\}^k\) is the set of labels of the form \((1 \cdots 1)\) with length \(\geq k\). The idea is that the final two terms are order \(o(\delta)\) because the probability of the events in the indicators are very small. It is also the case that the sum of the first two terms is \(\mathbb{P}^x(S_0) + o(\delta)\), which follows from a straightforward calculation (all of this should not be too suprising if you convince yourself that \(M_\delta\) and \(S_\delta\) are very close when \(\delta\) is small).

We begin with this calculation. We have

\[ \mathbb{P}^x(S_\delta 1_{\{v_0 = \emptyset\}}) = e^{-\frac{\lambda}{m-1}\delta} \mathbb{P}^x(\varphi(X(\delta)) 1_{\{D > \delta\}}) = e^{-\frac{\lambda(1 + \frac{1}{m-1})}{m-1} \delta} \varphi(x) \]

by conditioning on the first branching time being greater than \(\delta\), and applying (2.6). Furthermore, by noticing that for \(\{v_0 = 1\}\) to occur it must be that \(\{l_0 < \delta\}\) and \(\{A_0 \geq 1\}\), we can first condition on \(\{l_0, A_0\}\) and then take expectation over them on this event, to write

\[ \mathbb{P}^x(S_\delta 1_{\{v_0 = 1\}}) = \int_0^\delta \frac{\lambda}{m-1} e^{-\frac{\lambda}{m-1}(m-1)u} (1 - e^{-\lambda u} + e^{-\delta}) du. \]

The second line here follows by \((2.6)\) again. Putting these two expressions together, we have

\[ \mathbb{P}^x(S_\delta) = \mathbb{P}^x(S_\delta 1_{\{v_0 = \emptyset\}}) + \mathbb{P}^x(S_\delta 1_{\{v_0 = 1\}}) = \varphi(x) \left( e^{-\delta} + \frac{1 - e^{-\lambda(1+(m-1)^{-1})\delta}}{1 + (m-1)^{-1}} \right) = \varphi(x) (1 + o(\delta)) \]

where the \(o(\delta)\) term bounded by \(c\delta^2\) for some \(c\) independent of \(\delta\) and \(x\), using standard error bounds for the exponential series.

Now we have to deal with the last two terms in (6.8). For the penultimate term, we observe that for the event \(\{v_0 \in \bigcup_{n \geq 2} \{1\}^n\}\) to occur, it must be the case that \(\{l_0 < \delta\}, \{A_0 \geq 1\}, \{l_1 \leq \delta - l_0\}\) and \(\{A_1 \geq 1\}\). On this event, we can also clearly bound \(S_\delta\) by \(\|\varphi\|_\infty (1 + \sum_{u < v_0} (A_u - 1))\), where \(\|\varphi\|_\infty := \sup_{u \in D} \varphi(u)\). Therefore, by conditioning on \(l_0, l_1, A_0\) and \(A_1\), we can write

\[ \mathbb{P}^x(S_\delta 1_{\{v_0 \in \bigcup_{n \geq 2} \{1\}^n\}}) \leq \|\varphi\|_\infty \left( \int_0^\delta \frac{\lambda}{m-1} e^{-\frac{\lambda}{m-1} u} \int_0^{\delta-u} \frac{\lambda}{m-1} e^{-\frac{\lambda}{m-1} (2(m-1) + \lambda(\delta - r - u)) dr du} \right). \]

(6.9)
The expression $\lambda(\delta - r - u)$ in the integral is the expectation, given $l_0 \in du$, $l(1) \in dr$, of $\sum_{u=v_3,|u|>1}(A_u - 1)$. An easy calculation shows that the right hand side of (6.9) is bounded by $c\delta^2$ for some $c$ not depending on $x$ or $\delta$.

Finally we deal with the term $\mathbb{P}^x(S_{\delta}\mathbb{1}\{v_3\neq\emptyset,v_3\not\in\cup_{n\geq 1}\{1\}^n\})$. On the event $\{v_3 \neq \emptyset, v_3 \not\in \cup_{n\geq 1}\{1\}^n\}$ we clearly must have $\{l_0 < \delta\}$. Moreover, if $A_0 = 0$, then $S_{\delta} = 0$, so we need only consider the event where $\{l_0 < \delta\}$ and $\{A_0 \geq 1\}$. Then we also have $\{l(1) \leq \delta - l_0\}$, else we would have $v_4 = (1)$. We claim that on the event $\{l_0 < \delta\} \cap \{A_0 \geq 1\} \cap \{l(1) \leq \delta - l_0\}$

$$\mathbb{P}^x(S_{\delta}\mathbb{1}\{A_0, A(1), l_0, l(1)\}) \leq \|\varphi\|_{\infty}(A_0 - 1 + A(1))\tilde{E}(N_0),$$

where $\tilde{E}$ is the law of a continuous time branching process (with no spatial motion, branching at rate $\frac{\lambda}{m-1}$ and with offspring distribution $A$) and $N_0$ is the total progeny of this branching process up to time $\delta$. Indeed, on the event $\{l_0 < \delta, A_0 \geq 1, l(1) \leq \delta - l_0, A_0 = j, A(1) = k\}$ for any $j \geq 1, k \geq 0$, a moment’s thought tells us that we can couple our branching diffusion with $j - 1 + k$ independent continuous time branching processes under $\tilde{E}$ (in the obvious way, attaching one to each individual in $(2, \ldots, j), (11, \ldots, (1k))$ if this set is non empty) such that $S_{\delta}$ will be less than $\|\varphi\|_{\infty}$ times the total progeny of all of these processes before time $\delta$. Since $\mathbb{E}(N_0) \leq e^{2\lambda\delta}$, by taking expectation over $A_0, A(1)$ and then $l_0, l(1)$ we obtain that

$$\mathbb{P}^x(S_{\delta}\mathbb{1}\{v_3\neq\emptyset,v_3\not\in\cup_{n\geq 1}\{1\}^n\}) \leq \mathbb{E}^x\left(\mathbb{P}^x(S_{\delta}\mathbb{1}\{l_0<\delta,A_0\geq 1,l(1)\leq\delta-l_0\} \mid \{A_0, A(1)\})\right) \leq \|\varphi\|_{\infty}2me^{2\lambda\delta}\left(\int_0^\delta \lambda e^{-\frac{\lambda}{m-1}u} \left(\mathbb{P}^x(u < \tau^D \leq \delta) + \int_0^{\delta-u} \frac{\lambda}{m-1} e^{-\frac{\lambda}{m-1}r} m dr\right) du\right) \leq c\delta(\mathbb{P}^x(\tau^D \leq \delta) + \delta)$$

for some $c$ not depending on $x$ or $\delta$ (for any $\delta$ with $\delta \leq 1$, say). The terms $\mathbb{P}^x(u < \tau^D \leq \delta)$ and $\int_0^{\delta-u} \frac{\lambda}{m-1} e^{-\frac{\lambda}{m-1}r} m dr$ inside the integral on the second line above correspond to the possibilities $\{A(1) = 0\}$ and $\{A(1) > 0\}$ respectively.

Putting all of this together we deduce that for some $c > 0$

$$||\mathbb{P}^x(S_{\delta}) - \varphi(x)|| \leq c e^{\lambda\delta} \delta(\mathbb{P}^x(\tau^D \leq \delta) + \delta) \quad (6.10)$$

for all $0 < \delta \leq 1$ and $x \in D$. This clearly implies that $\mathbb{P}^x(S_\delta)$ has (right) derivative $0$ at time $t = 0$.

Step 2. Now we use this to show that $\mathbb{P}^x(S_t)$ is a constant function of $t$ for each fixed $x$, which implies (6.7).

To do this, we prove that it is differentiable everywhere (in $t$) with derivative $0$. Fix $t > 0$ and write

$$\mathbb{P}^x(S_t) - \mathbb{P}^x(S_0) = \mathbb{P}^x(\mathbb{P}^x(S_{tv}|\mathcal{H}_{t\wedge s}) - S_{t\wedge s}) \quad (6.11)$$

where $\mathcal{H}_u = \sigma(\{l_w, X_w, A_w\} : v_r = w, r < \kappa_u \cup \{X_{v u}(r) : r \in [s_{v u} - l_{vu}, s_{vu} - l_{vu} + (u - \kappa_u)]\}$ is the same as $\mathcal{F}_u$, but just for the exploration of a single tree. We have $S_{t\wedge s} = \varphi(V_{t\wedge s}) + \sum_{w \in Y(v_{s\wedge t}, T)} \varphi(X_w(s_{w} - l_w))$, and by the branching Markov property we can also write

$$\mathbb{P}^x(S_{tv}|\mathcal{H}_{t\wedge s}) = \mathbb{P}^{V_{t\wedge s}}(S_{[t-s]} + \sum_{w \in Y(v_{s\wedge t}, T)} \mathbb{P}^{X_w(s_{w} - l_w)}(S_{\sigma(w)})$$

where $\otimes$ represents the independent product measure and for each $w \in Y(v_{s\wedge t}, T)$, $\mu_w$ is a random, $\mathcal{H}_{t\wedge s}$-measurable law on $\sigma(w)$, such that $\sigma(w)$ is $\mu_w$-almost surely less than $|t-s|$. For example, if $w_1$ is the next individual after $v_{s\wedge t}$ in the depth first ordering of $T$, then $\sigma(w_1)$ is the conditional law of $(|t-s| - L^{v_{s\wedge t}}) \vee 0$ given $\mathcal{H}_{t\wedge s}$, where $L^{v_{s\wedge t}}$ is the total length of the continuous depth first exploration of the subtree rooted at $v_{s\wedge t}$. Similarly, if $w_2$ is the next individual after $w_1$ in the depth first ordering of $T$ then $\sigma(w_1)$ is the conditional law of $(|t-s| - L^{v_{s\wedge t}} - L^{w_1}) \vee 0$ where $L^{w_1}$ is the total length of the continuous depth first exploration of the subtree rooted at $w_1$, and so on.

\footnote{That is, if we view this branching process as a tree $T$ as in Section 2.2 then $N_0 = |\cup_{0 \leq \delta} N_0|$}

\footnote{Writing $f(t) = \mathbb{E}(N_t)$ we see by conditioning on the first branching time that $f(t) \leq 1 + \int_0^t \frac{m}{m-1} f(s), ds$ and so obtain the result using Gronwall’s inequality}

28
Therefore, by (6.10) we see that
\[
\limsup_{s \to t} \frac{|\mathbb{P}^x(S_t) - \mathbb{P}^x(S_s)|}{|t - s|} \leq c_t \limsup_{s \to t} \mathbb{P}^x(\mathbb{P}^{V_{c,t}}(\tau_D \leq |t - s|) + \sum_{w \in Y(v_{s,t}, t)} \mathbb{P}^{X_w}(s_w - l_w)(\tau_D \leq |t - s|))
\]
where \(c_t < \infty\) is a constant depending on \(t\) but not \(|t - s|\). Now, the expression inside the expectation on the right hand side converges to 0 almost surely as \(s \to t\). Moreover it is bounded above, for example, by \(1 + |\bigcup_{r \leq t-1} N_r|\), which we know by comparing to a continuous time branching process (as in Step 1) has expectation less than \((1 + e^{2\lambda(t+1)}\)). Hence, applying dominated convergence we see that \(\mathbb{P}^x(S_t)\) is differentiable at \(t\) with derivative 0.

**Step 3.** To prove the martingale property we write for \(0 \leq s \leq t\) (similarly to in Step 2, but now we have a whole infinite sequence of trees)
\[
\mathbb{P}^x(\bar{S}_t - \bar{S}_s \mid \mathcal{F}_s) = \mathbb{P}^{V_s}(S_{t-s}) - \varphi(V_s) + \sum_{w \in Y(v_s, \Lambda_s)} \left( (\mathbb{P}^{X^{\Lambda_w}_s(s^{\Lambda_w}_w - l^{\Lambda_w}_w)} \otimes \mu_w)(S_{\sigma(w)}) - \varphi(X^{\Lambda_w}_s(s^{\Lambda_w}_w - l^{\Lambda_w}_w)) \right)
+ \sum_{i=1}^{\infty} \left( (\mathbb{P}^x \otimes \mu_i)(S_{\sigma_i}) - \varphi(x) \right)
\]
where the \(\mu_w\) are random laws on \(\sigma(w)\) for each \(w \in Y(v_s, \Lambda_s)\), and the \(\mu_i\) are random laws on \(\sigma_i\) for each \(i \geq 1\). By Step 2, we see that the right hand side of the above is equal to 0. \(\square\)

**Lemma 6.7.** For any \(x \in D\), \((\bar{S}_t)_{t \geq 0}\) is a locally square-integrable martingale under \(\mathbb{P}^x\). Its predictable quadratic variation is given by
\[
\langle \bar{S} \rangle_t = \int_0^t \frac{\lambda}{m - 1} \mathbb{E}(A^2 - 2A + 1) \varphi(V_s)^2 + \sum_{i,j} a^{ij}(V_s) \frac{\partial \varphi}{\partial x_i} \frac{\partial \varphi}{\partial x_j}(V_s) \, ds.
\]

**Proof.** For the square integrability we make some very crude bounds. First note that, under \(\mathbb{P}^x\), the second term in the definition of \(\bar{S}_t\) (6.6) is equal to \(-\varphi(x) \Lambda_t\), and that the sum of the first and third terms of \(\bar{S}_t\) is positive, and equal to \(S_{\Lambda_t}^{\Lambda_t} \cdot L(T^{\Lambda_t}) + \cdots + L(T^{\Lambda_t-1})\), where \(S_{\Lambda_t}^{\Lambda_t}\) is the process \(S\) for the single tree \(T^{\Lambda_t}\). Thus
\[
\mathbb{P}^x(S_t^2) \leq \|\varphi\|_\infty^2 \mathbb{P}^x(\Lambda_t^2) + \mathbb{P}^x((S_{\Lambda_t}^{\Lambda_t} \cdot L(T^{\Lambda_t}) + \cdots + L(T^{\Lambda_t-1}))^2).
\]
Now fix \(T \geq 0\). For any \(t \leq T\) we can stochastically dominate \(\Lambda_t\) by a geometric random variable with success probability \(\mathbb{P}^x(\tau_D > T)\), and hence \(\mathbb{P}^x(\Lambda_t^2)\) is uniformly bounded for \(t \leq T\). For the second term, we write
\[
\mathbb{P}^x((S_{\Lambda_t}^{\Lambda_t} \cdot L(T^{\Lambda_t}) + \cdots + L(T^{\Lambda_t-1}))^2) = \mathbb{P}^x((S_{\Lambda_t}^{\Lambda_t} \cdot L(T^{\Lambda_t}) + \cdots + L(T^{\Lambda_t-1}))^2 \mid L(T) + \cdots + L(T^{\Lambda_t-1})) \leq \frac{\sup_{r \in [0,T]} \mathbb{P}^x(S_r^2)}{\mathbb{P}^x(\tau_D > T)}.
\]
The final inequality follows since given \(L(T) + \cdots + L(T^{\Lambda_t-1})\), the tree \(T^{\Lambda_t}\) has the same law as under \(\mathbb{P}^x\), but conditioned on the event \(L(T) > t - L(T) + \cdots + L(T^{\Lambda_t-1})\), and the probability of this event is greater than or equal to \(\mathbb{P}^x(\tau_D > T)\). Thus to complete the proof of local square integrability we just need to show that \(\mathbb{P}^x(S_t^2)\) is uniformly bounded on \([0, T]\). However, this follows by bounding \(\mathbb{P}^x(S_t^2)\) above by \(\|\varphi\|_\infty^2 \mathbb{E}(N_t^2)\) for \(t \in [0, T]\), where \(\mathbb{E}\) and \(N\) are as in the proof of Lemma 6.6. Again it can be shown that \(\mathbb{E}(N_t^2)\) is finite (similarly to in the proof of Lemma 6.6, by conditioning on the first branching time, using the fact that \(A\) is assumed to have finite variance, and by Gronwall’s inequality).

To calculate the predictable quadratic variation, let us look at the definition of \(\bar{S}_t\) a little more closely. It is clear that all the terms in (6.6) apart from the \(\varphi(V_t)\) term, are constant except at times where \(v_t\) (the vertex we are visiting at time \(t\) in the depth-first exploration) changes. It is also clear that \(\varphi(V_t)\) is continuous away from these times. Note that the number of such times is a.s. finite on any finite time interval (for example, since the expectation of the number of such times is finite). We write \(V_{t-} = \lim_{s \uparrow t} V_s\), which exists since \((V_t)_t\) is cadlag. Similarly, since the process \(t \mapsto v_t = (v_t, \Lambda_t)\) is cadlag with respect to the discrete topology on \(\Omega \times \mathbb{N}\), we can define \(v_{t-} = \lim_{s \uparrow t} v_s\). By the above discussion, any discontinuities of \(S_t\) must occur at times \(t\) such that \(v_{t-} \neq v_t\).

We consider the different possibilities for such times:
(I) $V_{t-} \in \partial D$. This means that $\varphi(V_{t-}) = 0$, and so $\varphi(V_t) - \varphi(V_{t-}) = \varphi(V_t)$. In this case, there are again two possibilities: (a) $\Lambda_{t-} = \Lambda_t$ and $\bar{v}_t = (w, \Lambda_{t-})$ for some $w \in Y(v_{t-}, T^{\Lambda -})$ or (b) $\bar{v}_t = (0, \Lambda_t)$ with $\Lambda_t \neq \Lambda_{t-}$. In case (a) we see that the third term on the right hand side of (6.6) also decreases by $\varphi(V_t)$ at time $t$, and so $\bar{S}_{t -} - \bar{S}_{t-} = 0$. Similarly, in case (b) the second term on the right hand side of (6.6) decreases by $\varphi(x) = \varphi(V_t)$ at time $t$, so again $\bar{S}_{t -} - \bar{S}_{t-}$ is equal to 0. Overall, such a situation does not actually correspond to a discontinuity of $\bar{S}_t$.

(II) $V_{t-} \notin \partial D$ but $V_{t-} \neq V_t$. Then $A_{v_{t-}} = 0$ and again $\bar{v}_t$ satisfies either (a) or (b) defined above. By the same reasoning, we have that $\bar{S}_{t -} - \bar{S}_{t-} = -\varphi(V_{t-}) = (A_{v_{t-}} - 1)\varphi(V_{t-})$.

(III) $V_{t-} = V_t$ but $\bar{v}_t \neq \bar{v}_{t-}$. This corresponds to the continuous depth first exploration reaching a branching point at time $t$ that is not the end of a branch. In this case we have $A_{v_{t-}} \geq 1$ and $\bar{S}_{t -} - \bar{S}_{t-} = (A_{v_{t-}} - 1)\varphi(V_t)$.

These are all the possibilities, so we can conclude that we have the decomposition $\bar{S}_t = \bar{S}_t^{(c)} + \bar{S}_t^{(j)}$ where

$$\bar{S}_t^{(c)} := \varphi(V_t) - \varphi(x) - \sum_{s \leq t} (\varphi(V_s) - \varphi(V_{s-})) \quad \text{and} \quad \bar{S}_t^{(j)} := \sum_{(w, i) = \bar{v}_s} \varphi(X_j^i(s^i_w))(A^i_w - 1).$$

(6.12)

Indeed, it is clear that $\bar{S}_t - \bar{S}_t^{(c)}$ is equal to the sum of the jumps of $\varphi(V_s)$ that occur before time $t$, plus the sum of the jumps of $\bar{S}_s - \varphi(V_s)$ that occur before time $t$. The above considerations tell us that at times where (I) is satisfied, these two types of jump cancel each other out, and at times where (II) or (III) is satisfied, the jumps are precisely the jumps of $\bar{S}_t^{(j)}$.

Now, it is clear that $(\bar{S}_t^{(c)})_t$ is continuous and $(\bar{S}_t^{(j)})_t$ is a jump process. This means that if we write $v(\omega, dt, dx)$ for the predictable compensator of $\bar{S}_t^{(j)}$ [JS87, §II, Theorem 1.8] then we have [JS87, §II, Proposition 2.29]

$$\langle \bar{S} \rangle_t = [\bar{S}^{(c)}]_t + \int_0^t \int_\mathbb{R} x^2 v(\omega, ds, dx).$$

(6.13)

where $[\cdot]$ denotes the ordinary quadratic variation. Since $\bar{S}^{(c)}$ is continuous, and its increments are that of $\varphi$ applied to an $L$-diffusion in $D$, we have that

$$[\bar{S}^{(c)}]_t = \int_0^t \sum_{i,j} a_{ij}(V_s) \partial_\varphi \partial_\varphi (V_s) ds.$$

Moreover, one can check (straight from the definition in [JS87]) that

$$v(\omega, ds, dx) = \sum_{k \geq 0} \mathbb{P}(A = k) \delta_{(k-1)} \varphi(V_{s-}(\omega))(dx) \times \frac{\lambda}{m-1} ds.$$

(6.14)

From this and (6.13), the lemma then follows. □

**Proposition 6.8.** Let

$$\sigma^2 = \frac{2b^{-1}}{(1, \varphi)} = \frac{\lambda(1, \varphi^3)(\mathbb{E}(A^2) - \mathbb{E}(A))}{(m - 1)(1, \varphi)}.$$

Then

$$\left( \frac{\bar{S}_{nt}}{\sqrt{n}} \right)_{t \geq 0} \rightarrow (\sigma B_t)_{t \geq 0}$$

in distribution as $n \rightarrow \infty$, with respect to the Skorohod topology, where $(B_t)_t$ is a standard one dimensional Brownian motion.
Proof. This follows from the functional central limit theorem for martingales [JS87, §VIII, Theorem 3.22] once we can show that for all $t \geq 0$

$$\langle \tilde{S}^n_t \rangle_t \rightarrow \sigma t$$  \hspace{1cm} (6.15)

in probability as $n \rightarrow \infty$. Here $(\tilde{S}^n_t)_{t \geq 0} = (\tilde{S}_{nt}/\sqrt{n})_{t \geq 0}$. In fact, since $(\tilde{S}^n_t)_t$ may have jumps, we also need to verify an extra condition in order to apply the functional central limit theorem. However this condition, see [JS87, §VIII, Eq.(3.23)], is simply that we have, for every $t > 0$,

$$\frac{\lambda}{m-1} \int_0^t \varphi(V_{s-}(w))^2 \mathbb{E}((A - 1)^2 \mathbb{1}_{\{A-1)^2\varphi(V_{s-}(w))^2 > \epsilon_n\}}) \, ds \rightarrow 0$$

almost surely as $n \rightarrow \infty$ (where the expectation $\mathbb{E}$ is only over $A$). Since $A$ has finite variance, this is clearly true.

To show the convergence of the predictable quadratic variation (6.15), we observe that we can write

$$\frac{(\tilde{S}^n_t)^2}{t} = \frac{1}{nt} \int_0^{nt} \frac{\lambda}{m-1} \mathbb{E}(A^2 - 2A + 1) \varphi(V_s)^2 + \sum_{i,j} a_{ij} \partial_x^i \varphi \partial_x^j (V_s) \, ds.$$ 

Then, since $\varphi$ and all of its first order derivatives are bounded (recall we are assuming this in this section), the result follows immediately from Proposition 6.9 below. We recover $\sigma^2$ by applying (2.1) with $v(x) = \varphi(x)^2$ (equivalently, integrating by parts).

**Proposition 6.9.** Suppose that $f$ is a bounded, measurable function. Then

$$Q_t := \frac{1}{t} \int_0^t f(V_s) \, ds \rightarrow \frac{(f, \varphi)}{(1, \varphi)}$$

in $\mathbb{P}^\alpha$-probability as $t \rightarrow \infty$.

Before we prove Proposition 6.9, let us record some of the consequences of Proposition 6.8

**Proposition 6.10.** Under $\mathbb{P}^\alpha$, we have the joint convergence

$$\left( \frac{S_{nt}}{\sqrt{n}}, \frac{L_{nt}}{\sqrt{n}} \right)_{t \geq 0} \rightarrow \left( \sigma |\beta|, \frac{\sigma}{\varphi(x)} L^0_t(\beta) \right)_{t \geq 0}$$  \hspace{1cm} (6.16)

as $n \rightarrow \infty$, in distribution with respect to the Skorohod topology. Here, $\beta$ is a standard Brownian motion started at 0 and $L^0_t(\beta)$ is the local time of $\beta$ at 0. Furthermore, for any $y > 0$, under $\mathbb{P}^\alpha(\cdot \mid \{\sup_t S_t \geq \sqrt{n}y\})$ we have

$$\left( \frac{S_{nt}}{\sqrt{n}} \right)_{t \geq 0} \overset{\text{d}}{\underset{n \rightarrow \infty}{\rightarrow}} \left( \sigma e^{\frac{y}{\sigma}} \right)_{t \geq 0}$$  \hspace{1cm} (6.17)

where $(e^{\frac{y}{\sigma}})_{t \geq 0}$ is a Brownian excursion conditioned to reach (at least) height $y/\sigma$.

**Proof.** From the definition of $S'_{nt} = S_t - \varphi(V_t)$, we have that $|S'_{nt}/\sqrt{n} - \tilde{S}_{nt}/\sqrt{n}| \leq ||\varphi||_\infty/\sqrt{n}$ for all $t \geq 0$ and so Proposition 6.8 implies that

$$\left( \frac{S'_{nt}}{\sqrt{n}} \right)_{t \geq 0} \rightarrow (\sigma B_t)_{t \geq 0}$$

as $n \rightarrow \infty$ as well. Writing $B_t = \inf_{0 \leq s \leq t} B_s$ this implies the joint convergence

$$\left( \frac{S_{nt}}{\sqrt{n}}, \frac{L'_{nt}}{\sqrt{n}} \right)_{t \geq 0} = \left( \frac{S'_{nt}/\sqrt{n}}{\sqrt{n}}, \frac{L'_{nt}/\sqrt{n}}{\sqrt{n}} \right)_{t \geq 0} \overset{\text{d}}{\underset{n \rightarrow \infty}{\rightarrow}} (\sigma(B_t - B), -\sigma B_t)_{t \geq 0}$$  \hspace{1cm} (6.18)

where the right hand side by Lévy’s theorem, see for example [RY91, §VI, Theorem VI.2.3], is equal in distribution to

$$\left( \sigma |\beta|, \sigma L^0_t(\beta) \right)_{t \geq 0}.$$
However, we know that \( \Lambda_t = -I_t^*/\varphi(x) \), and so (6.16) follows. For the second claim of the Proposition, we follow [LGD02, Proposition 2.5.2]. It is well known that you can construct the process \( e^{y/\sigma} \) from a standard Brownian motion \( \beta \) by taking

\[
e^{y/\sigma} = |\beta(t) + e|D|
\]

where \( T = \inf\{ t \geq 0 : |\beta_t| \geq y/\sigma \} \), \( G = \sup\{ t \leq T : \beta_t = 0 \} \) and \( D = \inf\{ t \geq T : \beta_t = 0 \} \). By the Skorohod representation theorem and (6.16), we also know that there exists a process

\[
\left( Z_t^{(n)}, \Lambda_t^{(n)} \right)_{t \geq 0} \overset{d}{=} \left( \frac{S_{nt}^n - I_{nt}^n}{\sqrt{n}}, \frac{\Lambda_{nt}^n}{\sqrt{n}} \right)_{t \geq 0}
\]

such that

\[
\left( Z_t^{(n)}, \Lambda_t^{(n)} \right)_{t \geq 0} \xrightarrow{n \to \infty} \left( \sigma|\beta_t|, \frac{\sigma}{\varphi(x)} L_0^T(\beta) \right)_{t \geq 0}
\]

uniformly on every compact set almost surely. This is because Skorohod convergence is equivalent to local uniform convergence when the limit is continuous. Define \( T^{(n)} = \inf\{ t \geq 0 : Z_t^{(n)} \geq y \} \) for this sequence of processes, and \( G^{(n)}, D^{(n)} \) in the same way as \( G \) and \( D \) above. By Remark 6.5 and (6.16), if we can prove that \( G^{(n)} \to G \) and \( D^{(n)} \to D \) almost surely, we will be done. To do this, first note that since \( \beta \) must exceed \( x/\sigma \) immediately after time \( T \), we have that \( T^{(n)} \to T \) almost surely. This implies straight away that for all \( t < D \) we have \( t \leq D^{(n)} \) for all \( n \) large enough almost surely. Now we must show that for all \( t > D \) we have \( t \geq D^{(n)} \) for all \( n \) large enough almost surely. These facts together (along with the corresponding results for \( G \)) are enough to prove the convergence. To see the final claim, we use the convergence of the local time. For any \( t > D \), we have using basic properties of Brownian local time that \( L_t^0 > L_D^0 = L_T^0 \). The convergence of the local time therefore tells us that \( \Lambda_t^{(n)} \to \frac{\sigma}{\varphi(x)} L_T^0 \) for all \( n \) large enough almost surely. Since

\[
\Lambda_{T^{(n)}} \to \frac{\sigma}{\varphi(x)} L_T^0
\]

almost surely, this implies that we have also have \( \Lambda_t^{(n)} > \Lambda_{T^{(n)}}^{(n)} \) for all \( n \) large enough almost surely. Using the fact that \( \Lambda^{(n)} \) stays constant on \( [T^{(n)}, D^{(n)}] \), we see that \( t \geq D^{(n)} \).

### 6.1.1 Proof of Proposition 6.9

The rest of this subsection will be devoted to the proof of Proposition 6.9 but we will first need some preliminary estimates. Recall the definition of \( L(T) \) from just before (6.1).

**Lemma 6.11.** For any \( t_0 > 0 \), there exists a constant \( c_{t_0} \in (0, \infty) \) such that for all \( x \in D \) and all \( t \geq t_0 \)

\[
\mathbb{P}^x(L(T) > t) \geq c_{t_0} \frac{\varphi(x)}{\sqrt{t}}.
\]

Now, let \( U_t \) be a sample from the uniform distribution on \( [0, t] \), and write \( \mu_t^U \) for its law. In the following we write

\[
H_t^* = H_{U_t} \quad ; \quad \Lambda_t^* = \Lambda_{U_t} \quad \text{and} \quad V_t^* = V_{U_t} \quad (6.19)
\]

to simplify notation.

**Lemma 6.12.** For all \( x \in D \)

\[
\mathbb{P}^x \otimes \mu_t^U \left( H_t^* \geq C \sqrt{t} \right) \to 0
\]

as \( C \to \infty \), uniformly in \( \{ t \geq t_0 \} \) for any \( t_0 > 0 \), where \( \otimes \) represents the independent product measure.

We will now show how we may deduce Proposition 6.9 from these lemmas, and then go on to prove them. In the following we set \( N_s^i = \{ u \in T^i : s \in [s_u - l_u, s_u] \} \) and \( \bar{N}_{s,t}^i = \{ u \in T^i : s \in [s_u - l_u, s_u] \cap \{ (u, T^i) = \bar{v}_r \text{ for some } r \leq t \} \} \). So \( N_s^i \) is simply the set \( N_s \) for the \( i \)th tree, and \( \bar{N}_{s,t}^i \subset N_s^i \) is the set of individuals in \( N_s^i \) that have been visited in the depth first exploration of the sequence of trees before time \( t \).
Proof of Proposition 6.9} We first show that \( \bar{\mathbb{P}}^x(Q_t) \to \frac{f}{(1, \varphi)} \) as \( t \to \infty \). Define

\[
m_t := \left( |\tilde{N}^*_t| \right)^{-1} \sum_{u \in \tilde{N}^*_t} f(X^*_u) (H^*_t)
\]

It is the average value of \( f \) among the vertices at height \( H^*_t \) of the \( \Lambda^*_t \)th tree, that have been visited before time \( t \) in the depth first exploration. Observe that

\[
\bar{\mathbb{P}}^x(Q_t) = \bar{\mathbb{P}}^x \otimes \mu_t^U \left( f(V^*_t) \right) = \bar{\mathbb{P}}^x \otimes \mu_t^U \left( \bar{\mathbb{P}}^x \otimes \mu_t^U \left( f(V^*_t) \mid H^*_t, \Lambda^*_t, (T^{\Lambda^*_t}, X_{t}^{\Lambda^*_t}, t^{\Lambda^*_t}) \right) \right) = \bar{\mathbb{P}}^x \otimes \mu_t^U \left( m_t \right).
\]

We will show that for any fixed \( \varepsilon > 0 \),

\[
\bar{\mathbb{P}}^x \otimes \mu_t^U \left( A_t \right) := \bar{\mathbb{P}}^x \otimes \mu_t^U \left( |m_t - \frac{(\varphi, f)}{(\varphi, 1)}| > \varepsilon \right) \to 0 \tag{6.20}
\]

as \( t \to \infty \). Then since \( m_t \) is bounded (because \( f \) is) the convergence in expectation will follow.

Now also fix \( \delta > 0 \). By Lemma 6.12 we can choose \( C \) such that \( \bar{\mathbb{P}}^x \otimes \mu_t^U \left( H^*_t \geq C \sqrt{t} \right) < \frac{\delta}{3} \) for all \( t \geq 1 \). Similarly we can define \( R(t) > 0 \) such that \( \bar{\mathbb{P}}^x \otimes \mu_t^U \left( H^*_t \leq R(t) \right) = \frac{\delta}{3} \). We claim that \( R(t) \to \infty \) as \( t \to \infty \) by the law of large numbers. Indeed, for any fixed \( K > 0 \), we have

\[
\bar{\mathbb{P}}^x \otimes \mu_t^U \left( H^*_t \leq K \right) = \bar{\mathbb{P}}^x \left( \frac{1}{t} \sum_{i=1}^{\Lambda_t} \int_0^K |\tilde{N}^*_i| \, dr \right)
\]

by conditioning on the entire sequence of trees, and using that the amount of time before \( t \) that the depth first exploration spends below height \( K \) in the \( i \)th tree is equal to \( \int_0^K |T^{\Lambda^*_i}| \, dr \) for every \( i \leq \Lambda_t \) (this is easy to check). Moreover,

\[
\frac{1}{t} \sum_{i=1}^{\Lambda_t} \int_0^K |\tilde{N}^*_i| \, dr \leq \frac{1}{t} \sum_{i=1}^{\Lambda_t} \int_0^K |N^*_i| \, dr \leq \frac{\sum_{i=1}^{\Lambda_t} \int_0^K |N^*_i| \, dr}{\sum_{i=1}^{\Lambda_t} L(T^i)} \tag{6.21}
\]

since \( t \geq \sum_{i \leq \Lambda_t} L(T^i) \) by definition of \( \Lambda_t \). However, for each \( i \) the random variables \( \int_0^K |N^*_i| \, dr \) are i.i.d with finite variance (by Fubini), and the random variables \( L(T^i) \) are i.i.d with infinite mean (by Lemma 6.11). Since \( \Lambda_t \to \infty \) a.s. as \( t \to \infty \) (each tree is finite almost surely), the strong law of large numbers allows us to conclude that the right (and therefore left) hand side of (6.21) converges to 0 almost surely under \( \bar{\mathbb{P}}^x \). By dominated convergence (the random variable on the left hand side of (6.21) is a.s. less than or equal to 1) we therefore have \( \bar{\mathbb{P}}^x \otimes \mu_t^U \left( H^*_t \leq K \right) \to 0 \), and hence \( R(t) \to \infty \).

For the \( i \)th tree in our exploration and \( s, t > 0 \), let

\[
m^i_{s,t} := |\tilde{N}^i_{s,t}|^{-1} \sum_{u \in \tilde{N}^i_{s,t}} f(X^i_u(s))
\]

be the average value of \( f \) among the individuals of tree \( i \) at height \( s \) that are visited before time \( t \) (so that \( m_t = m^\Lambda^*_t \)). Also write \( A^i_{s,t} \) for the event that \( |m^i_{s,t} - (\varphi, f)/(\varphi, 1)| > \varepsilon \). Then the above considerations tell us that we can write

\[
\bar{\mathbb{P}}^x \otimes \mu_t^U \left( m_t \right) \leq \frac{2 \delta}{3} + \frac{1}{t} \bar{\mathbb{P}}^x \left( \int_{R(t)} \sum_{i=1}^{\infty} 1_{\{i \leq \Lambda_t\}} 1_{A^i_{s,t}} |\tilde{N}^i_{s,t}| \, ds \right), \tag{6.22}
\]

for all \( t \geq 1 \), where we have obtained the expression for the second term on the right hand side by first conditioning on \( \{H^*_t, \Lambda^*_t\} \), and then taking expectation over them.

Now by Fubini (since everything is bounded by 1) we can rewrite

\[
\frac{1}{t} \bar{\mathbb{P}}^x \left( \int_{R(t)} \sum_{i=1}^{\infty} 1_{\{i \leq \Lambda_t\}} 1_{A^i_{s,t}} |\tilde{N}^i_{s,t}| \, ds \right) = \frac{1}{t} \int_{R(t)} \sum_{i=1}^{\infty} \bar{\mathbb{P}}^x (1_{\{i \leq \Lambda_t\}} 1_{A^i_{s,t}} |\tilde{N}^i_{s,t}|) \, ds. \tag{6.23}
\]
The idea is, for each $i \geq 1$, to condition on $\mathcal{K}_{i-1} = \sigma(\cup_{j \leq i-1}(T^j, X^j, t^j))$. Note that $\mathcal{K}_{i-1}$ is independent of the $i$th tree. Moreover, the event $\{i \leq \Lambda_t\}$ and the start time $\tau_i = L(T^1) + \cdots + L(T^{i-1})$ of the $i$th tree, are measurable with respect to $\mathcal{K}_{i-1}$. Thus we can write, for each $i \geq 1$,

$$
\mathbb{P}^x(\mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} \mid N_{s,t}^i, \mathcal{K}_{i-1}) = \mathbb{P}^x(\mathbb{1}_{A_{s,t}^i} \mid N_{s,t}^i, \mathcal{K}_{i-1}) \mathbb{1}_{\{i \leq \Lambda_t\}} = \mathbb{P}^x(\mathbb{1}_{A_{s,t}^i} \mid N_{s,t}^i, \mathcal{K}_{i-1}) \mathbb{1}_{\{i \leq \Lambda_t\}}
$$

(6.24)

where the expectation in the final term is now with respect to a single branching diffusion under $\mathbb{P}^x$, $N_s(r) := \{u \in N_s : u = v_h \text{ for some } h \leq r\}$ is the number of particles at level $s$ in this tree that are visited before time $r$ in a depth first exploration of it, and $A_s(r) = \{|N_s(r)|^{-1} \sum_{u \in N_s(r)} f(X_u(s)) - (\varphi, f)/\varphi \mid \geq \epsilon\}$ is the event that the average of $f$ among the positions of these particles at time $s$ is more than $\epsilon$ away from $(\varphi, f)/\varphi, 1$.

We note the simple fact here that by Theorem 1.4 and Lemma 5.2 there exists a $K$ such that

$$
\mathbb{E}_x[|N_t|^2 \mid |N_t| > 0]^{1/2} \leq Kt, \quad \mathbb{P}_x(|N_t| > 0) \leq K/t
$$

(6.25)

for all $t \geq 1$. We can also, by Lemma 6.11 choose this $K$ such that

$$
\mathbb{P}^x(\Lambda_t) \leq \frac{K \sqrt{t}}{\varphi(x)}
$$

(6.26)

for all $t \geq 1$. Indeed, $\Lambda_t$ can be stochastically dominated by a geometric random variable, whose success rate is $\mathbb{P}^x(L(T) > t) \geq c_1 \varphi(x)/\sqrt{t}$.

Now, decomposing on whether or not $|N_s(t - \tau_i)|$ is bigger than $\delta s/6CK^2$ (recall the definition of $C$ from earlier in the proof) we have

$$
\mathbb{P}^x(\mathbb{1}_{A_{s,t}^i} \mid N_{s,t}^i, \mathcal{K}_{i-1}) \leq \frac{\delta s}{6CK^2} + \mathbb{P}^x(\mathbb{1}_{B_{s,t}^{f,\epsilon}(\delta/6CK^2)} \mid N_{s,t}^i, \mathcal{K}_{i-1}) \leq \frac{\delta s}{6CK^2} + \mathbb{P}^x(\mathbb{1}_{|N_s^i|^2 > 0}^{1/2} \mathbb{P}^x(B_{s,t}^{f,\epsilon}(\delta/6CK^2) \mid N_{s,t}^i, \mathcal{K}_{i-1})^{1/2})
$$

(6.24), (6.25) and the above we can therefore deduce that (as long as $s \geq 1$ and $t$ is big enough that $R(t) \geq 1$)

$$
\mathbb{P}^x(\mathbb{1}_{i \leq \Lambda_t} \mathbb{1}_{A_{s,t}^i} \mid N_{s,t}^i) \leq \left(\frac{\delta}{6CK} + K^2 \mathbb{P}^x(B_{s,t}^{f,\epsilon}(\delta/6CK^2) \mid N_{s,t}^i, \mathcal{K}_{i-1})^{1/2}\right) \mathbb{P}^x(\mathbb{1}_{i \leq \Lambda_t}).
$$

(6.27)

This means that the right hand side of (6.23) is less than

$$
\left(\frac{\delta}{6CK} + K^2 \mathbb{P}^x(B_{s,t}^{f,\epsilon}(\delta/6CK^2) \mid N_{s,t}^i, \mathcal{K}_{i-1})^{1/2}\right) \frac{1}{t} \int_{R(t)}^\infty \sum_{i=1}^\infty \mathbb{P}^x(\mathbb{1}_{i \leq \Lambda_t}).
$$

However, by Fubini and (6.26), this is less than

$$
\varphi(x)^{-1}\left(\frac{\delta}{6} + CK^3 \sup_{s \geq R(t)} \left\{\mathbb{P}^x(B_{s,t}^{f,\epsilon}(\delta/6CK^2) \mid N_{s,t}^i, \mathcal{K}_{i-1})^{1/2}\right\}\right).
$$

Using Lemma 5.3 and the fact that $R(t) \to \infty$, we see than this is less than $\frac{\delta}{3}$ for all $t$ large enough. Substituting in to (6.22) proves (6.20).

To complete the proof of the proposition, we must also show that

$$
\mathbb{P}^x(Q_t^2) \to \frac{(f, \varphi)^2}{(1, \varphi)^2}
$$
as \( t \to \infty \). However, letting \( U^1_t \) and \( U^2_t \) be two independent uniform random variables on \([0, t]\), we see that
\[
\mathbb{P}_x(Q^t_t) = \mathbb{P}_x \otimes \mu^U_t \otimes \mu^U_t (f(V^1_t) f(V^2_t))
\]
Let \( m^t_i \) and \( A^t_i \) correspond to \( m_t \) and \( A_t \) for \( U^i_t, i = 1, 2 \). Then the same reasoning as above tells us that
\[
\mathbb{P}_x \otimes \mu^U_t \otimes \mu^U_t (f(V^1_t) f(V^2_t)) = \mathbb{P}_x \otimes \mu^U_t \otimes \mu^U_t (m^1_t m^2_t)
\]
and we also have, by a union bound, that
\[
\mathbb{P}_x \otimes \mu^U_t \otimes \mu^U_t (A^1_t \cup A^2_t) \to 0
\]
as \( t \to \infty \). The result follows in exactly the same way.

Proof of Lemma 6.17: In fact, we will prove a slightly stronger statement, as it will also be of use later on. We will show that
\[
\mathbb{P}_x(L(T) \leq c s^2 \mid |N_s| > 0) \to 0
\]
(6.28)
as \( c \to 0 \), uniformly in \( s \geq 0 \) and \( x \in D \). Then by choosing \( c \) such that \( \mathbb{P}_x(L(T) \leq c s^2 \mid |N_s| > 0) \geq 1/2 \) for all \( s \geq 0 \) we can write
\[
\mathbb{P}_x(L(T) \geq t) \geq \mathbb{P}_x(L(T) \geq t \mid |N_s| > 0) \mathbb{P}_x(|N_s| > 0) \geq \frac{1}{2} \mathbb{P}_x(|N_s| > 0)
\]
for all \( t \). Applying Theorem 1.4, Lemma 6.11 follows.

To prove (6.28) we will show that for any \( \delta > 0 \) there exists a \( c > 0 \) and \( S > 0 \) such that \( \mathbb{P}_x(L(T) \leq c s^2 \mid |N_s| > 0) < \delta \) for all \( x \in D \) and \( s \geq S \). Clearly we can then redefine \( c \) such that this holds for all \( s \geq 0 \) (indeed for all \( s \leq S \), we have \( \mathbb{P}_x(L(T) \leq c s^2 \mid |N_s| > 0) \leq \mathbb{P}_x(|N_s| > 0)^{-1} \mathbb{P}_x(L(T) \leq c s^2) \) which tends to 0 as \( c \to 0 \).

So, we fix \( \delta > 0 \) and also pick some \( D' \subseteq D \) (meaning that \( D' \) is compact and contained in \( D \)). We write \( N^D_s := \{ u \in N_s : X_u(s) \in D' \} \) and choose \( \eta > 0 \) and \( S_1 > 0 \) such that
\[
\mathbb{P}_x(|N^D_s| < \eta s \mid |N_s| > 0) < \delta/2
\]
(6.29)
for all \( x \in D \) and \( s \geq S_1 \). We can do this because
\[
\mathbb{P}_x(|N^D_s| < \eta s \mid |N_s| > 0) = \mathbb{P}_x\left(e^{-\frac{1}{s} \sum_{u \in N_s} 1_{(X_u(s) \in D')} > e^{-1} \mid |N_s| > 0}\right)
\]
\[
\leq e \mathbb{P}_x\left(e^{-\frac{1}{\bar{\eta}} \sum_{u \in C_{N_s}} 1_{(X_u(s) \in D')} \mid |N_s| > 0}\right)
\]
where the final inequality follows by conditional Markov’s inequality. We know by Theorem 1.6 that this converges to \( e(b \int_{D'} \varphi)(b \int_{D'} \varphi + \bar{\eta}^{-1})^{-1} \) as \( s \to \infty \), uniformly in \( x \) (uniformity is a consequence of the proof). Hence, we can choose \( \eta > 0 \) small enough and \( S_1 > 0 \) large enough so that (6.29) holds for all \( x \in D \) and \( s \geq S_1 \).

Now we pick \( a > 0 \) (very large) such that \( \inf_{y \in D'} \mathbb{P}_y(\mid |N_r| \mid \geq 2/a\eta) \) for all \( r \geq 0 \) (which is possible by the Many-to-One Lemma 2.4 Lemma 2.1 and the fact that \( D' \in D \)). For \( c > 0 \) and \( u \in \Omega \), we set \( L^u_{c,s} = \int_0^{a \eta s} |\{ w \in N_{s+r} : u \leq w \}| dr \), so that no matter what the value of \( c, L(T) \geq \sum_{u \in N_s} L^u_{c,s} \). This means that we can write
\[
\mathbb{P}_x(L(T) \leq c s^2 \mid |N_s| > 0) \leq \frac{\delta}{2} + \frac{\mathbb{P}_x\left(\big\{ \sum_{u \in N_s} L^u_{c,s} \leq c s^2 \big\} \cap \{ |N^D_s| \geq \eta s \} \right)}{\mathbb{P}_x(|N_s| > 0)}
\]
(6.30)
Moreover, conditionally on \( F_s \), the random variables \( \{ L^u_{c,s} : u \in N^D_s \} \) are independent, and all have mean \( \geq 2cs/\eta \) by definition of \( a \) and \( L^u_{c,s} \). They also have variance \( \leq Ke^{3s^2} a^3 \) for some \( K < \infty \) not depending on \( c \) or \( s \). This follows from an easy integration once you observe that for any \( r \geq r' \geq 1 \) and \( y \in D \)
\[
\mathbb{P}_y(|N_r||N_{r'}|) \leq \sup_{z \in D} \sup_{u > 0} \mathbb{P}_z(\mid |N_u| \mid) \times \mathbb{P}_y(\mid |N_r| \mid) \leq K_1 r'
\]
for some $K_1$ not depending on $y, r, r'$. Here the first inequality follows by conditioning on $F_{r'}$, and the second inequality follows from Lemma 3.2 and the fact that $\sup_{z \in D} \sup_{u > 0} \mathbb{P}^z (|N_u|) < \infty$ (an easy consequence of (3.2)).

Putting all of this together we see that

$$\mathbb{P}^x \left( \sum_{u \in N_s \cap D'} L_{i, s}^u \leq cs^2 \mid F_s \right)$$

is the probability that a sum of $N_{\infty}^{|C - D|'}$ independent positive, random variables (with means and variances as above) is less than or equal to $cs^2$. Let us suppose that $N_{\infty}^{|C - D|'} = n$ and $n \geq \eta s$. Then since the conditional mean of $\sum_{u \in N_s \cap D'} L_{i, s}^u$ is bigger than or equal to $2csn/\eta$, we see that the probability in (6.31) is less than or equal to the conditional probability that $\sum_{u \in N_s \cap D'} L_{i, s}^u$ minus its conditional mean is greater than or equal to $2csn/\eta - cs^2$.

By Markov’s inequality, using $\leq$ independence and the upper bound on the variances, we see that this is less than or equal to

$$\frac{Knc^3s^3}{(2csn/\eta - cs^2)^2} \leq \frac{cK\eta a^3}{2 \cdot \frac{n}{m}} \leq cK\eta a^3$$

where for both inequalities we have used the fact that $n \geq ms$. Substituting this in, we see that the second term on the right hand side of (6.30) is less than or equal to $\frac{c}{2}$ for all $c$ small enough. This concludes the proof. \qed

Proof of Lemma 6.12. Pick $t_0 > 0$. As explained in the proof of Proposition 6.9, Lemma 6.11 immediately implies that $\mathbb{P}^x (\Lambda_t) \leq K\sqrt{t}$ for some $K = K(t_0)$ and all $x \in D$, $t \geq t_0$. By Theorem 1.4 we can also choose this $K$ such that $\mathbb{P}^x (|N_t| > 0) \leq K\varphi(x)/r$ for all $x \in D$ and $r \geq \sqrt{t_0}$. Then we write

$$\mathbb{P}^x \otimes \mu^U_t (H_t^x \geq C\sqrt{t}) \leq \mathbb{P}^x \left( \sum_{i=1}^{\infty} \mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{\{|N_{\infty}^{|C - D|'}| > 0\}} \right).$$

We will show that we can apply Fubini to the right hand side of the above, and that the expression we get converges to 0 as $C \to \infty$. To do this, we write for each $i \geq 1$, since the event $\{i \leq \Lambda_t\}$ does not depend on the $i$th tree:

$$\mathbb{P}^x (\{i \leq \Lambda_t\} \cap \{|N_{\infty}^i| > 0\}) \leq \mathbb{P}^x (i \leq \Lambda_t) \mathbb{P}^x (|N_{\infty}^{|C - D|'}| > 0).$$

Then as long as $t \geq t_0$ and $C > 1$, our definition of $K$ means that the right hand side of the above is less than or equal to $K\mathbb{P}^x (i \leq \Lambda_t) \frac{\varphi(x)}{C\sqrt{t}}$. Therefore by Fubini, since $\sum_{i=1}^{\infty} \mathbb{P}^x (i \leq \Lambda_t) = \mathbb{P}^x (\Lambda_t) \leq K\sqrt{t}/\varphi(x)$, we have

$$\mathbb{P}^x \otimes \mu^U_t (H_t^x \geq C\sqrt{t}) \leq \frac{K^2}{C}$$

for all $x \in D$, $t \geq t_0$ and $C \geq 1$. This concludes the proof. \qed

### 6.2 Connection with the height function

In order to make use of the above invariance principle, we will now make the connection between the height function $H_t$, and the process

$$S_t := \sum_{w \in Y(v_t, T)} \varphi(X_w(s_w - l_w))$$

from Definition 6.4 for the depth first exploration of a single tree $T$ conditioned to be large. For a vertex $u \in T$ we also use the notation

$$S(u) := \sum_{w \in Y(u, T)} \varphi(X_w(s_w - l_w))$$

so that $S_t = S(v_t)$. We will show that for large $t$, and for an overwhelming proportion of the individuals $u \in N_t$ (given survival), $S(u)$ is close to a constant times $t$. Our approach will use an ergodicity property for the spine particle in the system under $\mathbb{Q}^x$, and is inspired from [HR14].
In the following, given \( \eta > 0 \), we will say that a pair \((u, t)\) with \( u \in N_t\) is \( \eta\)-bad if

\[
\left| \frac{S(u)}{t} - b^{-1} \right| > \eta
\]  

(recalling the definition of \( b \) from \((4.6)\)). We also say, for given \( T \geq 0 \) and \( t \geq T \), that \((u, t)\) is \( \eta_T\)-bad if some \((v, s)\) with \( T \leq s \leq t \), \( v \in N_s \) and \( v \leq u \) is \( \eta\)-bad. That is, \((u, t)\) is \( \eta_T\)-good if all of its ancestors after time \( T \) are \( \eta\)-good. We have the following estimate for the proportion of \( u \in N_t\) such that \((u, t)\) is \( \eta_T\)-bad:

**Proposition 6.13.** Fix \( \varepsilon, \eta > 0 \) and write \( N_t^{\eta_T} := \{ u \in N_t : (u, t) is \eta_T - bad \} \). Then we have

\[
\sup_{t \geq T} \mathbb{P}^x \left( \frac{|N_t^{\eta_T}|}{|N_t|} > \varepsilon \left| |N_t| > 0 \right) \right) \to 0
\]

as \( T \to \infty \), for any \( x \in D \).

**Proof.** We will first show that for any \( \varepsilon > 0 \), setting \( E^\varepsilon_{T,t} := \left\{ \frac{\sum_{u \in N_t} \varphi(X_u(t)) \mathbb{1}_{\{(u,t) is \eta_T - bad\}}}{\sum_{u \in N_t} \varphi(X_u(t))} > \varepsilon \right\} \)

\[
\sup_{t \geq T} \mathbb{P}^x \left( E^\varepsilon_{T,t} \left| |N_t| > 0 \right) \right) \to 0
\]

as \( T \to \infty \). To do this, we will use the description of the system under the measure \( \tilde{Q}^x \) given in Section 2.4.3. Recalling that \( \tilde{Q}^x = \tilde{Q}^x_{\mathcal{F}_\infty} \) has \((d\tilde{Q}^x/d\mathbb{P}^x)\) \( \mathcal{F}_t = \varphi(x)^{-1} \sum_{u \in N_t} \varphi(X_u(t)) \) we see that

\[
\mathbb{P}^x \left( E^\varepsilon_{T,t} \left| |N_t| > 0 \right) \right) = \tilde{Q}^x \left[ \varphi(x)/\mathbb{P}^x(|N_t| > 0) \sum_{u \in N_t} \varphi(X_u(t)) \mathbb{1}_{E^\varepsilon_{T,t}} \right] = \tilde{Q}^x \left[ Y_t \mathbb{1}_{E^\varepsilon_{T,t}} \right],
\]

where \( Y_t := \frac{\varphi(x)/\mathbb{P}^x(|N_t| > 0)}{\sum_{u \in N_t} \varphi(X_u(t))} \). To see that this converges to 0 it is enough to prove that:

- \( \sup_{t \geq T} \tilde{Q}_x(E^\varepsilon_{T,t}) \to 0 \) as \( T \to \infty \) and

- For every \( \delta > 0 \), there exists \( T' \) and \( K \) positive, such that \( \tilde{Q}^x \left( Y_t \mathbb{1}_{|Y_t| > K} \right) \leq \delta \) for all \( t \geq T' \).

The first point comes from the fact that if \( \psi_t \) is the spine particle under \( \tilde{Q}^x \) at time \( t \), and \( t \geq T \), then \( (\psi_t, t) \) is very unlikely to be \( \eta_T\)-bad for large \( T \). More precisely, by (2.13), we have that

\[
\frac{\sum_{u \in N_t} \varphi(X_u(t)) \mathbb{1}_{\{(u,t) is \eta_T - bad\}}}{\sum_{u \in N_t} \varphi(X_u(t))} = \tilde{Q}^x((\psi_t, t) is \eta_T - bad \mid \mathcal{F}_t)
\]

and so by Markov’s inequality

\[
\tilde{Q}^x(E^\varepsilon_{T,t}) = \tilde{Q}^x(E^\varepsilon_{T,t}) \leq \varepsilon^{-1} \tilde{Q}^x \left( \frac{\sum_{u \in N_t} \varphi(X_u(t)) \mathbb{1}_{\{(u,t) is \eta_T - bad\}}}{\sum_{u \in N_t} \varphi(X_u(t))} \right) = \varepsilon^{-1} \tilde{Q}^x((\psi_t, t) is \eta_T - bad)
\]

where

\[
\sup_{t \geq T} \tilde{Q}^x((\psi_t, t) is \eta_T - bad) \to 0
\]

as \( T \to \infty \). This proves the first point. The reason that (6.35) is true is because under \( \tilde{Q}^x \) we know (from Section 2.4.3) that the spine particle evolves as a diffusion conditioned to remain in \( D \), and that at constant rate \( \frac{m}{m-1} \lambda \) a random number of “younger siblings” (see just before Definition 6.2) is born along its trajectory (each with initial position equal to the position of the spine at the time that they are born). The random number of younger siblings born at each time is always independent from everything else and has finite variance and mean \((2m)^{-1}(\mathbb{E}(A^2) - \mathbb{E}(A))\), which comes from the description in Section 2.4.3 again. We also have that for any \( r \geq 0 \), \( S(\psi_r) = \sum_{y \in Y} \varphi(y) \), where \( Y \) is the set of initial positions (counted with multiplicity) of younger siblings born.
along the trajectory of the spine before time \(r\). Since by Lemma \[2.1\] and \[Pin85\], the trajectory of the spine is a Markov process with invariant density \(\varphi^2\), (6.35) follows by a straightforward ergodicity argument.

The second point essentially says that \((Y_t)_{t \geq 0}\) is \(\mathbb{Q}^x\)-uniformly integrable (in fact it is, but we only need the weaker statement). To prove it, one can use the change of measure between \(\varphi\) and \(\mathbb{P}^x\) again to write

\[
\mathbb{Q}^x((Y_t \mathbbm{1}_{\{Y_t > K\}}) = \frac{\mathbb{P}^x\left(\{Y_t > K\} \cap \{\|N_t| > 0\}\right)}{\mathbb{P}^x(\|N_t| > 0)} = \mathbb{P}^x\left(Y_t > K \mid \|N_t| > 0\right).
\]

Since \(\varphi(x)/(t\mathbb{P}^x(\|N_t| > 0))\) is uniformly bounded above for \(t \geq 1\) (say), by Theorem \[1.4\] we just need to show that for any \(\delta > 0\) there exists \(K\) and \(T\) such that

\[
\sup_{t \geq T} \mathbb{P}^x\left(\frac{\sum_{u \in N_t} \varphi(X_u(t))}{t} < 1/K \mid \|N_t| > 0\right) \leq \delta. \tag{6.36}
\]

However, this is a direct consequence of the convergence given by Theorem \[1.5\] since we know that for fixed \(K\) the probability above converges, as \(t \to \infty\), to the probability that an exponential random variable is less than \(1/K\). This completes the proof of (6.34).

We must now deduce from (6.34) that

\[
\sup_{t \geq T} \mathbb{P}^x\left(\frac{|N^y_t|}{|N_t|} > \varepsilon \mid \|N_t| > 0\right) \to 0 \text{ as } T \to \infty.
\]

The idea behind this is that \(\sum_{u \in N_t} \varphi(X_u(t))\mathbb{1}_{\{(u,t) \in \eta_T\text{-bad}\}} / \sum_{u \in N_t} \varphi(X_u(t))\) is a reasonable approximation to \(|N^y_t|/|N_t|\) at large times, on the event of survival. Indeed, by Corollary \[1.7\] we know that given \(\delta > 0\), there exists \(r > 0\) such that

\[
\sup_{t \geq T} \mathbb{P}^x\left(\frac{|D^y_t|}{|N_t|} > \frac{\varepsilon}{2} \mid \|N_t| > 0\right) \leq \delta/2, \tag{6.37}
\]

for all \(T\) large enough, where \(D^y = \{y \in D : \varphi(y) < 1/r\}\) and \(N^y_t := \{u \in N_t : X_u(t) \in D^y\}\). Moreover, we can write \(|N^y_t| \leq |N^y_t| + |N^y_t \cap N^D_{t}\}|, where we have

\[
\frac{|N^y_t| \cap (N^y_t \cap N^D_{t})}{|N_t|} \leq \sum_{u \in N_t} \varphi(X_u(t))\mathbb{1}_{\{(u,t) \in \eta_T\text{-bad}\}} \|\varphi\|_{\infty r},
\]

so that also, by (6.34),

\[
\sup_{t \geq T} \mathbb{P}^x\left(\frac{|N^y_t| \cap (N^y_t \cap N^D_{t})}{|N_t|} > \varepsilon/2\right) < \delta/2
\]

for all \(T\) large enough. Putting this together with (6.37), we can conclude. \(\square\)

**Remark 6.14.** Now suppose we are given \(c > 0\) and \(x \in D\). Then for any \(t\) such that \(ct \wedge t \geq T\)

\[
\mathbb{P}^x\left(\left\{\frac{|N^y_t|}{|N^y_t|} > \varepsilon\right\} \cap \{|N^y_t| > 0\} \mid |N_t| > 0\right) = \frac{\mathbb{P}^x\left(\left\{\frac{|N^y_t|}{|N^y_t|} > \varepsilon\right\} \cap \{|N^y_t| > 0\} \cap \{\|N_t| > 0\}\right)}{\mathbb{P}^x(\|N_t| > 0)}
\]

\[
\leq \mathbb{P}^x\left(\left\{\frac{|N^y_t|}{|N^y_t|} > \varepsilon\right\} \cap \{\|N_t| > 0\} \mid |N_t| > 0\right) \frac{\mathbb{P}^x(\|N_t| > 0)}{\mathbb{P}^x(\|N_t| > 0)}
\]

\[
\leq \left(\sup_{s \geq T} \mathbb{P}^x\left(\left\{\frac{|N^y_s|}{|N^y_s|} > \varepsilon\right\} \mid |N_s| > 0\right) \right) \frac{\mathbb{P}^x(\|N_t| > 0)}{\mathbb{P}^x(\|N_t| > 0)}
\]

\[
\leq \frac{F(T)}{c} \left(\sup_{s \geq T} \mathbb{P}^x\left(\left\{\frac{|N^y_s|}{|N^y_s|} > \varepsilon\right\} \mid |N_s| > 0\right) \right)
\]

where

\[
F(T) = \sup_{s \geq T} sb^{-1}\varphi(x)^{-1}\mathbb{P}^x(\|N_s| > 0) \to 1
\]

as \(T \to \infty\), by Theorem \[1.4\]. Note that we are allowing \(c < 1\) here.
The next lemma provides the key connection between \((S_t)\) and the height function \((H_t)\) for a critical branching diffusion under \(\mathbb{P}^x\), that is conditioned to survive for a long time. We write \(\mathcal{P}^x_T\) for the law of a tree \((T, X, l)\) under \(\mathbb{P}^x\) plus a random variable \(t^1\), which conditionally on \((T, X, l)\) is chosen uniformly from \([0, L(T)) := [0, L)\).

**Lemma 6.15.** For any \(\eta > 0\) we have

\[
\lim_{T \to \infty} \lim_{t \to \infty} \mathcal{P}^x_T((v_1, H_t) \text{ is } \eta_T - \text{bad } | |N_t| > 0) \to 0.
\]

In the case that \(H_{t_1} < T\) we also say that \((v_1, H_t)\) is \(\eta_T\)-bad (with an abuse of notation; recall we only defined the notion of \(\eta_T\)-badness for \((u, t)\) with \(t \geq T\)). However, since the probability of this event goes to 0 as \(t \to \infty\) for any fixed \(T\), this will not play a role.

**Proof.** We write for any \(a, c, C > 0\)

\[
\mathcal{P}^x_T((v_1, H_t) \text{ is } \eta_T - \text{bad } | |N_t| > 0) \leq \mathcal{P}^x_T(L \leq at^2 | |N_t| > 0) + \mathcal{P}^x_T \left( \{H_{t_1} \notin [ct, Ct]\} \cap \{L > at^2\} \cap \{H_t > 0\} \right)
\]

and begin by claiming that the first two terms on the right hand side above tend to 0 as \(a \to 0\) and then \(c \to 0, C \to \infty, \text{ uniformly in } t \geq 1\) (say). Note that they do not depend on \(T\) at all. This claim follows by (6.28), the fact that

\[
\mathbb{P}^x(|N_{Ct}| > 0 | |N_t| > 0) \to 0
\]

as \(C \to \infty\) uniformly in \(t \geq 1\) by Theorem 1.4 and the estimate

\[
\mathcal{P}^x_T(\{H_t^1 \leq ct\} \cap \{L > at^2\} | |N_t| > 0) \leq (\mathbb{P}^x(|N_t| > 0))^{-1} \mathbb{P}^x(1_{\{L \geq at^2\}} \mathcal{P}^x(H_t^1 \leq ct | \mathcal{F}_\infty))
\]

for some finite \(K\) and all \(t \geq 1\). The above inequalities follow by first expanding the conditional probability; then conditioning on the \(\sigma\)-algebra \(\mathcal{F}_\infty\) generated by \((T, l, X)\); and finally using the lower bound on \(L\) from the indicator function, the lower bound on \(\mathbb{P}^x(|N_t| > 0)\) that one obtains from Theorem 1.4 and the fact that \(\sup_{x \in D} \sup_{u > 0} \mathbb{P}^x(|N_u|) < \infty\) (which we saw in the proof of Lemma 6.11).

Thus, we are left to prove that for any fixed \(a, c, C\)

\[
\lim_{T \to \infty} \lim_{t \to \infty} \mathcal{P}^x_T((v_1, H_t) \text{ is } \eta_T - \text{bad } \cap \{L > at^2\} \cap \{H_{t_1} \in [ct, Ct]\} \cap \{H_t > 0\} = 0.
\]  

(6.38)

From now on we assume that \(ct \geq T\) (which is fine, since we are letting \(t \to \infty\) first). By conditioning on \(\mathcal{F}_\infty\) again, we see that

\[
\mathcal{P}^x_T((v_1, H_t) \text{ is } \eta_T - \text{bad } \cap \{L > at^2\} \cap \{H_{t_1} \in [ct, Ct]\} | |N_t| > 0)
\]

\[
= \mathbb{P}^x(1_{\{L > at^2\}}) \int_{ct}^{Ct} \frac{|N_u^\eta_T|}{|N_u|} \frac{|N_u|}{L} du | |N_t| > 0
\]

\[
\leq \delta + (at^2)^{-1} \mathbb{P}^x \left( \int_{ct}^{Ct} 1_{\frac{|N_u^\eta_T|}{|N_u|} > \delta} |N_u| du | |N_t| > 0 \right)
\]

for any \(\delta > 0\). Therefore, we just need to prove that the second term above (for fixed \(\delta\)) converges to 0 as \(t \to \infty\) and then \(T \to \infty\). However, we can write this term (by Fubini) as

\[
(at^2)^{-1} \int_{ct}^{Ct} \mathbb{P}^x(|N_u| 1_{\frac{|N_u^\eta_T|}{|N_u|} > \delta} | |N_t| > 0) du \leq \sqrt{c} \sup_{s \geq T} \mathbb{P}^x \left( \frac{|N_u^\eta_T|}{|N_u|} > \delta \right) |N_s| > 0 \right)^{1/2}
\]

\[
\times (at^2)^{-1} \int_{ct}^{Ct} \mathbb{P}^x(|N_u|^2 | |N_t| > 0)^{1/2} du
\]
where the inequality follows by conditional Cauchy–Schwarz, Remark \(6.14\) and the assumption that \(ct \geq T\). Finally, we observe that for any \(u \in [ct, Ct]\) and \(ct \geq T \geq 1\), \(\mathbb{P}^x([N_u]^2|N_t| > 0)^{1/2} \leq Mu\) for some constant \(M = M(c, C')\), by \((6.25)\) and Theorem \(1.4\). Hence, by integrating and applying Proposition \(6.13\) we obtain \((6.38)\).

Now we let \(\mathcal{P}_k^x\) be the law of a tree \((T, X, l)\) under \(\mathbb{P}^x\), together with \(k\) random variables \((t^1, \cdots, t^k)\) chosen conditionally uniformly at random from \([0, L(T)] = [0, L]\). We also define the \(k \times k\) matrices

\[
(D^S_t^i)_{ij} := t^{-1}(S(v^i_t) + S(v^j_t) - 2S(v^{ij}_t)) \quad \text{and} \quad (D^H_t^i)_{ij} := t^{-1}(H^i_t + H^j_t - 2h^{ij}_t)
\]

where \(v^{ij}_t = v^i_t \land v^j_t\) is the most recent common ancestor of \(v^i_t\) and \(v^j_t\), and \(h^{ij}_t = s_{v^{ij}_t}\) is its “death time” (see Section \(2.2\) for definitions of these objects.) The next proposition says that conditioned on survival up to a large time \(t\), these matrices are essentially the same up to a constant.

**Proposition 6.16.** Let \(k \geq 1\) and \(D^S_t^i\) and \(D^H_t^i\) be as defined above. Then for any \(\varepsilon > 0\)

\[
\mathcal{P}_k^x \left( \|b^{-1}D^H_t - D^S_t\| > \varepsilon \parallel N_t > 0 \right) \to 0
\]

as \(t \to \infty\), where the distance is the Euclidean distance between \(k \times k\) matrices.

**Proof.** We prove this in the case \(k = 2\): the general result following by a union bound. Note that since \((D^H_t)^{ii} = (D^S_t)^{ii} = 0\) for \(i = 1, 2\), and by symmetry, we need only control the distance \(|b^{-1}(D^H_t)_{12} - (D^S_t)_{12}|\).

Given \(\delta > 0\), we first choose \(R \geq 1\) such that \(\lim_{t \to \infty} \mathbb{P}^x([|N_{Rt}| > 0] \mid |N_t| > 0) \leq \delta/4\), which is possible by Theorem \(1.4\). We then set \(\eta := \varepsilon/(4R)\) and pick \(T\) large enough that \(\lim_{t \to \infty} \mathbb{P}^x(\{v^i_1, H^i_1\} \text{ is } \eta_T - \text{bad} \cup \{v^i_2, H^i_2\} \text{ is } \eta_T - \text{bad} \}) \leq \delta/4\), which is possible by a union bound, and Lemma \(6.15\). Finally, we pick \(K\) large enough that \(\lim_{t \to \infty} \mathbb{P}^x(\int_0^T |N_s| \geq K \mid |N_t| > 0) \leq \delta/4\). This is possible by Theorem \(1.5\) and Markov’s inequality.

Putting this together, we see that for all \(t\) large enough,

\[
\mathcal{P}_2^x(A) := \mathcal{P}_2^x(\{N_{Rt} > 0\} \cup \{\int_0^T |N_s| \geq K\} \cup \{(v^i_1, H^i_1)\} \text{ is } \eta_T - \text{bad} \cup \{(v^i_2, H^i_2)\} \text{ is } \eta_T - \text{bad}\} \leq \delta
\]

Now suppose we are on the complementary event \(A^c\). Straight from the definitions, we know that on this event

\[
H^i_1 \leq Rt, \quad H^i_2 \leq Rt, \quad \frac{|S(v^i_t)|}{H^i_t} - b^{-1} \leq \eta, \quad \text{and } \quad \frac{|S(v^i_t)|}{H^i_t} - b^{-1} \leq \eta
\]

where \(\eta = \varepsilon/(4R)\). Then there are two possibilities: we are either on the event \(B := \{h^{ij} \geq T\}\) or the event \(C := \{h^{ij} \leq T\}\). On \(\{B \cap A^c\}\), by definition of \(A\) and “\(\eta_T\)-badness”, we also have that

\[
h^{12} \leq Rt \quad \text{and } \quad \frac{|S(v^{12}_t)|}{h^{12}_t} - b^{-1} \leq \eta
\]

Putting this together with \((6.41)\) we see that on \(\{B \cap A^c\}\) we have the deterministic bound

\[
|0(D^S_t)_{12} - b^{-1}(D^H_t)_{12}| \leq \frac{1}{t} \times 4Rt \times \eta \leq \varepsilon.
\]

Furthermore, on the event \(\{C \cap A^c\}\), we have \(h^{ij} \leq T\) and \(S(v^{ij}_t) \leq ||\varphi||_\infty K\) (by a very crude bound). Therefore, we have

\[
|S(v^{12}_t)| - b^{-1}h^{12}_t \leq K + Tb^{-1}
\]

and so

\[
|0(D^S_t)_{12} - b^{-1}(D^H_t)_{12}| \leq \frac{\varepsilon}{2} + \frac{K + Tb^{-1}}{t}.
\]

Since the second term on the right hand side above is less than \(\varepsilon/2\) for all \(t\) large enough, the proof is complete. \(\square\)
6.3 Convergence to the CRT

6.3.1 Preliminaries on converge of metric measure spaces

Before we can prove Theorem 6.1, we must introduce various notions of convergence for metric spaces, and more generally, for metric measure spaces. Although our aim is to prove convergence of conditioned genealogical trees in the sense of Gromov–Hausdorff distance between metric spaces, it turns out to be helpful to go through the framework of metric measure spaces. We first recall the definition of the Gromov–Hausdorff metric on $\mathbb{X}$ in the sense of Gromov–Hausdorff distance between metric spaces, it turns out to be helpful to go through the generally, for metric measure spaces. Although our aim is to prove convergence of conditioned genealogical trees

Definition 6.17. The Gromov–Hausdorff distance between $(X, r_X)$ and $(Y, r_Y)$ in $\mathbb{X}$ is given by

$$d_{GH}((X, r_X), (Y, r_Y)) = \inf_{g_{X,Y}} d_{H}^{(Z,r_Z)}(g_X(X), g_Y(Y)),$$

where the infimum is taken over all isometric embeddings $g_X, g_Y$ from $X$ and $Y$ to a common metric space $(Z, r_Z)$, and $d_{H}^{(Z,r_Z)}$ is the usual Hausdorff distance on $(Z, r_Z)$.

For us, a metric measure space $(X, r, \mu)$ will be a compact metric space $(X, r)$ equipped with a finite Borel measure $\mu$. These will be considered modulo the equivalence $\sim$, where $(X, r, \mu) \sim (X', r', \mu')$ if there exists a measure preserving isometry between $X$ and $X'$. We denote the set of (equivalence classes) of these spaces by $\mathbb{X}$. We will be interested in the Gromov–Prohorov metric and the Gromov–Hausdorff–Prohorov metric on $\mathbb{X}$. We will begin by defining the so-called Gromov–weak topology.

Definition 6.18. [GPW09, Definition 2.3] We will call a function $\Phi : \mathbb{X} \to \mathbb{R}$ a polynomial if there exists an $k \in \mathbb{N}$ and a bounded continuous function $\phi : [0, \infty)^{(\frac{k}{2})} \to \mathbb{R}$ such that

$$\Phi((X, r, \mu)) = \int \mu^\otimes k (d(x_1, \ldots, x_n)) \phi((r(x_i, x_j))_{1 \leq i < j \leq k}),$$

where $\mu^\otimes k$ is the product measure of $\mu$. We write $\Pi$ for the set of all polynomials.

Definition 6.19. [GPW09, Definition 2.8] A sequence $\mathcal{X}_n \in \mathbb{X}$ is said to converge to $\mathcal{X} \in \mathbb{X}$ with respect to the Gromov–weak topology if and only if $\Phi(\mathcal{X}_n)$ converges to $\Phi(\mathcal{X})$ in $\mathbb{R}$, for all polynomials $\Phi \in \Pi$.

It is known, see [GPW09, Theorem 5], that this topology is metrised by the Gromov–Prohorov metric, that we now define.

Definition 6.20. The Gromov–Prohorov distance between $\mathcal{X} = (X, r_X, \mu_X)$ and $\mathcal{Y} = (Y, r_Y, \mu_Y)$ in $\mathbb{X}$ is given by

$$d_{GP}(\mathcal{X}, \mathcal{Y}) = \inf_{g_{X,Y}} d_{P}^{(Z,r_Z)}((g_X)_*(\mu_X), (g_Y)_*(\mu_Y)),$$

where the infimum is as in Definition 6.17 and $d_{P}^{(Z,r_Z)}$ is the Prohorov distance between probability measures on $(Z, r_Z)$.

Finally, we define the Gromov–Hausdorff–Prohorov metric [ADH13, Mic09] on $\mathbb{X}$.

Definition 6.21. Let $\mathcal{X}, \mathcal{Y}$ be as in Definition 6.20. The Gromov–Hausdorff–Prohorov distance between $\mathcal{X}$ and $\mathcal{Y}$ is defined by

$$d_{GHP}(\mathcal{X}, \mathcal{Y}) = \inf_{g_{X,Y}} \left( d_{P}^{(Z,r_Z)}((g_X)_*(\mu_X), (g_Y)_*(\mu_Y)) + d_{H}^{(Z,r_Z)}(g_X(X), g_Y(Y)) \right).$$

Remark 6.22. It is clear from the above definitions that convergence in the Gromov–Hausdorff–Prohorov metric implies convergence in both the Gromov–Hausdorff metric and the Gromov–Prohorov metric.
We will need a couple of facts for our proof:

**Lemma 6.23.** [GPW09 Corollary 3.1] A sequence \( \{P_n\}_{n \in \mathbb{N}} \) of probability measures on \( X \) converges weakly to a probability measure \( P \) with respect to the Gromov–weak topology, if and only if

(i) The family \( \{P_n\}_{n \in \mathbb{N}} \) is relatively compact in the space of probability measures on \( X \).

(ii) For all polynomials \( \Phi \in \Pi, P_n[\Phi] \to P[\Phi] \) in \( \mathbb{R} \) as \( n \to \infty \).

and

**Lemma 6.24.** [ADHI13 Theorem 2.4], [BBI01 Theorem 7.4.15] A set \( K \subset X \) is relatively compact with respect to the Gromov–Hausdorff–Prohorov metric if and only if

(i) There is a constant \( D \) such that \( \text{diam}(X) < D \) for all \( X \in K \).

(ii) For all polynomials \( \Phi \in \Pi, P_n[\Phi] \to P[\Phi] \) in \( \mathbb{R} \) as \( n \to \infty \).

(iii) \( \sup_{X \in K} \mu_X(X) < +\infty \)

### 6.3.2 Proof of the main theorem

Recall the definitions of \( d^* \) and \( \sim^* \) from Remark 6.1. In this remark we observed that \( \left( \frac{[0,L(T)]}{\sim^*}, d^* \right) \) is \( \mathbb{P} \)-almost surely isometrically equivalent to \( (T(T,X,l),d(T,X,l)) \) as defined in Section 2.2.3 Furthermore, if \( (T_\sigma, d_\sigma) \) is the real tree with contour function given by \( \sigma \), a Brownian excursion conditioned to reach height 1, then it is clear from Brownian scaling that \( (T_\sigma, d_\sigma) \) is isometrically equivalent to \( (T_\sigma, d_\sigma) \), where \( \sigma \) is the time a Brownian excursion conditioned to reach height \( (b\sigma)^{-1} \) and \( \sigma \) is the constant from Proposition 6.8.

Therefore, to prove Theorem 1.1 it is enough to prove that if \( (T^{(n)}, d^{(n)}) \) has the law of \( \left( \frac{[0,L(T)]}{\sim^*}, d^* \right) \) under \( \mathbb{P} \{ \cdot \mid |N_n| > 0 \} \) then

\[
(T^{(n)}, d^{(n)}) \xrightarrow{d} (T_\sigma, d_\sigma)
\]

as \( n \to \infty \), with respect to the Gromov–Hausdorff topology.

We equip \( (T^{(n)}, d^{(n)}) \) with the measure \( \mu^{(n)}(\cdot) \), which is defined to be the push forward of uniform measure on \([0,L(T)]\) under the equivalence relation \( \sim^* \). We also equip \( (T_\sigma, d_\sigma) \) with the measure \( \mu_\sigma \), where if \( \tau \) is the length of \( \sigma \) and \( (T_\sigma, d_\sigma) = \left( \frac{[0,\tau]}{\sim^*}, d_\sigma \right) \) as described in Section 2.2.3, then \( \mu_\sigma \) is the push forward of uniform measure on \([0,\tau]\) under \( \sim \). For the proof of Theorem 1.1 we then need the following two ingredients:

**Lemma 6.25.** The family

\[
(T^{(n)}, d^{(n)}, \mu^{(n)})(n \geq 0)
\]

is tight with respect to the Gromov–Hausdorff–Prohorov metric.

**Lemma 6.26.** \( (T^{(n)}, d^{(n)}, \mu^{(n)}) \) converges to \( (T_\sigma, d_\sigma, \mu_\sigma) \) as \( n \to \infty \), with respect to the Gromov–Prohorov metric.

Let us first see how this gives Theorem 1.1.

**Proof of Theorem 1.1** Since Gromov–Hausdorff–Prohorov convergence implies Gromov–Prohorov convergence, (2) characterises subsequential limits with respect to the Gromov–Hausdorff–Prohorov metric. Thus we have the convergence in distribution

\[
(T^{(n)}, d^{(n)}, \mu^{(n)}) \xrightarrow{d} (T_\sigma, d_\sigma, \mu_\sigma)
\]

with respect to the Gromov–Hausdorff–Prohorov metric. This then implies (6.42) by Remark 6.22 (GHP convergence implies GH convergence).
All that remains therefore is to verify Lemmas \ref{lem:6.25} and \ref{lem:6.26}. In the following, we write \(\mathbb{P}^{(n)}\) for the law of \((T^{(n)}, d^{(n)}, \mu^{(n)})\) and \(\mathbb{P}_{\delta}\) for the law of \((\hat{T}_\delta, \hat{d}_\delta, \hat{\mu}_\delta)\).

**Proof of Lemma \ref{lem:6.25}** We need to show that for any \(\varepsilon > 0\) there exists a relatively compact \(\mathbb{K} \subset \mathbb{X}\) (wrt the Gromov–Hausdorff–Prohorov metric) such that

\[
\inf_n \mathbb{P}^{(n)}((T^{(n)}, d^{(n)}, \mu^{(n)}) \in \mathbb{K}) \geq 1 - \varepsilon.
\]

Fix \(\varepsilon > 0\) and let

\[\mathbb{K}_{R,M} := \{(X, r, \mu) \in \mathbb{X} : \text{diam}(X) \leq 2R, \mu(X) = 1, \forall k \geq 1 \text{ can cover } X \text{ with less than } 2^{4k} M \text{ balls of radius } 2^{-k}\},\]

which is a relatively compact subset of \(\mathbb{X}\) by Lemma \ref{lem:6.24}. We will prove that we can pick \(R\) and \(M\) large enough such that \(\mathbb{P}^{(n)}((T^{(n)}, d^{(n)}, \mu^{(n)}) \in \mathbb{K}_{R,M}) \geq 1 - \varepsilon\) for all \(n \in \mathbb{N}\). To do this, first observe that \(\mu^{(n)}(T^{(n)}) = 1\) for all \(n\). We also claim that for \(M = M(\varepsilon)\) and \(R = R(\varepsilon)\) large enough, we have

\[
\mathbb{P}^{(n)}(\text{diam}(T^{(n)}) > 2R) \leq \varepsilon/2; \quad \text{and}
\]

\[
\mathbb{P}^{(n)}(\{\text{cannot cover } (T^{(n)}, d^{(n)}) \text{ with less than } \delta^{-4} M \text{ balls of radius } \delta\}) \cap \{\text{diam}(T^{(n)}) \leq 2R\} \leq \delta\varepsilon/2 \quad (6.44)
\]

for all \(\delta > 0\) and \(n \in \mathbb{N}\). Summing over \(\delta = 2^{-k}\) then concludes the proof of tightness.

To prove the above claim, we first reformulate the probabilities in \(6.43\) and \(6.44\) as conditional probabilities under \(\mathbb{P}^x(\cdot || N_n > 0)\). Recall that the law of \((T^{(n)}, d^{(n)})\) under \(\mathbb{P}^{(n)}\) is that of \((\frac{[0, L(T)]}{\delta}, \frac{d^*}{\delta})\) under \(\mathbb{P}^x(\cdot || N_n > 0)\). It is therefore immediate from the definition of \(d^*\) that

\[
\mathbb{P}^{(n)}(\text{diam}(T^{(n)}) > 2R) \leq \mathbb{P}^x(\text{diam}(T^{(n)} > 2R) || N_n > 0)
\]

and so by Theorem \ref{thm:1.14} we can take \(R = R(\varepsilon)\) large enough that this is \(\leq \varepsilon/2\) for all \(n\). Then, fixing this \(R\) and given \(\delta > 0\), we divide \([0, 2Rn]\) into intervals of length \(n\delta/2 := \delta^n\) (so there are \(4R/\delta\) of them). We claim that the probability on the left hand side of \(6.44\) is less than or equal to

\[
\mathbb{P}^x \left( \bigcup_{j=0}^{(4R/\delta)-1} \{ |S_j| > \frac{M}{4R\delta^3} \} \bigg| |N_n| > 0 \right) \quad (6.45)
\]

where \(S_j := \{ u \in N_{\delta^n j} : \exists w \in N_{\delta^{j+1}} \delta^n u \text{ with } u \leq w \}\). To see why this reformulation is true, observe that each \(r \in [0, L(T)]\) corresponds to a pair \((v_r, H_r)\) where \(H_r \geq 0\) and \(v_r \in N_{H_r}\) (remember \(v_r\) is the particle being visited at time \(r\) in the depth first exploration of the branching diffusion genealogical tree, and \(H_r\) is the height of this exploration at time \(r\)). We also have that \(r_1 \sim_s r_2\) if and only if the time \(H_{r_1} = H_{r_2}\) corresponds to a branching time for the branching diffusion (in its normal time parameterisation) and \(v_{r_1}, v_{r_2}\) are two of the particles born at this time.

Now, suppose we are on the event

\[
\{ \text{diam}(\frac{[0, L(T)]}{\delta^n}, \frac{d^*}{n}) \leq 2R \} \cap \cap_{j=0}^{(4R/\delta)-1} \{ |S_j| \leq M(4R\delta^3)^{-1} \} \quad (6.46)
\]

We just need to show that that implies that we can cover \((\frac{[0, L(T)]}{\delta^n}, \frac{d^*}{n})\) with less than \(\delta^{-4} M\) balls of radius \(\delta\). Let \(S\) be the set of equivalence classes of \((\frac{[0, L(T)]}{\delta^n}, \frac{d^*}{n})\) that contain a point \(s\) such that \((v_s, H_s) = (u, j\delta^{n,j})\) for some \(u \in S_j\) and some \(j \in \{0, \ldots, (4R/\delta) - 1\}\). Then from the definition of the event \(6.46\), \(|S| \leq \delta^{-4} M\). We will show that any element of \((\frac{[0, L(T)]}{\delta^n}, \frac{d^*}{n})\) has \(\frac{d^*}{n}\) distance less than \(\delta\) from some point in \(S\). Indeed, any such element corresponds to some (not necessarily unique but it doesn’t matter) point \(r \in [0, L(T)]\), and since \(\text{diam}((\frac{[0, L(T)]}{\delta^n}, \frac{d^*}{n})) \leq 2R\) we have \(H_r \leq 2R\). Therefore \(H_r \in [j\delta^{n,j}, (j + 1)\delta^{n,j})\) for some \(j \in \{0, \ldots, (4R/\delta) - 1\}\), and so \(u \leq v_r\) for some \(u \in S_j\) (by definition of \(S_j\)). This means that there is some \(s \geq 0\) such that \((v_s, H_s) = (u, \delta^{n,j})\) and also that the equivalence class of \(s\) under \(\sim_s\) is contained in \(S\). It is clear that we have \(d^*(s, r) \leq \delta/2 < \delta\), and so we are done.
So, we estimate (6.45). For any \( j, n, \delta \) we write
\[
\mathbb{P}^x(|S_j|) = \mathbb{P}^x\left( \mathbb{P}^x\left(|S_j| \mid \mathcal{F}_{jbn} \right) \right)
\]
\[
= \mathbb{P}^x\left( \sum_{u \in N_{jbn,\delta}} \mathbb{P}^x(N_u(jbn,\delta) \mid |N_{bn,\delta}| > 0) \right) \leq C(n\delta)^{-1}
\]
where \( C \) is a constant that does not depend on \( j, n \) or \( \delta \). The final inequality follows from the fact that \( \sup_x \mathbb{P}^x(|N_r|) < \infty \), and Theorem 1.4. Then by Theorem 1.4 again (now we are conditioning on the event \( |N_n| > 0 \))
\[
\mathbb{P}^x(|S_j| \mid |N_n| > 0) \leq \hat{C} \delta^{-1}
\]
for another \( \hat{C} \) which may depend on \( x \), but does not depend on \( j, n, \delta \). Therefore, by conditional Markov’s inequality we see that
\[
\mathbb{P}^x(|S_j| > \frac{M}{4R\delta^3} \mid |N_n| > 0) \leq \frac{4R\hat{C} \delta^2}{M}
\]
and so by a union bound
\[
\mathbb{P}^x\left( \bigcup_{j=0}^{(4R/\delta)^{-1}} \left\{ \{u \in N_{jbn,\delta} : \exists w \in N_{(j+1)bn,\delta} \text{ with } u \leq w\} \mid |N_n| > 0 \right\} \right) \leq \frac{16R^2\hat{C} \delta}{M}.
\]
Choosing \( M = M(R, \varepsilon) = M(\varepsilon) \) so that \( 16R^2\hat{C} \delta M^{-1} < \varepsilon/2 \) then gives (6.44). \( \square \)

**Proof of Lemma 6.26.** For this we would like to show that \((T^{(n)}, d^{(n)}, \mu^{(n)})\) converges to \((\mathcal{T}_{\hat{e}}, d_{\hat{e}}, \mu_{\hat{e}})\) in the Gromov–Prohorov metric, or equivalently, with respect to the Gromov–weak topology. We will consider the latter formulation, in order to use the characterisation given by Lemma 6.23. Since convergence in the Gromov–Hausdorff–Prohorov sense implies convergence in the Gromov–Prohorov/Gromov–weak sense, Lemma 6.25 also shows that part (i) of the characterisation (relative compactness of the laws) is satisfied. Therefore, we need only show that for any polynomial \( \Phi \in \Pi \), we have \( \mathbb{P}^{(n)}[\Phi] \to \mathbb{P}^{\hat{e}}[\Phi] \) as \( n \to \infty \). To do this, we use Proposition 6.10 and Proposition 6.16.

Fix a polynomial
\[
\Phi((X,r,\mu)) = \int \mu^{\otimes k}(d(x_1, \cdots, x_n))\phi((r(x_i, x_j))_{1 \leq i < j \leq k}),
\]
for \( \phi : [0, \infty)^{(k)} \to \mathbb{R} \) continuous and bounded. Examining the definitions, we see that
\[
\mathbb{P}^{(n)}[\Phi] = \mathcal{P}^x_k(\phi(D_n^H) \mid |N_n| > 0) \tag{6.47}
\]
where \( \mathcal{P}^x_k \) and the matrix \( D_n^H \) are defined just before Proposition 6.16 (interpreting \( \phi(D_n^H) = \phi(((D_n^H)_{ij})_{1 \leq i < j \leq k}) \)). Using this Proposition, and also recalling the definition of the matrix \( D_n^S \), we see that since \( \phi \) is continuous and bounded,
\[
\mathcal{P}^x_k(\phi(D_n^H) \mid |N_n| > 0) - \mathcal{P}^x_k(\phi(bD_n^S) \mid |N_n| > 0) \to 0 \tag{6.48}
\]
as \( n \to \infty \). Now we also claim that
\[
\mathcal{P}^x_k(\phi(bD_n^S) \mid |N_n| > 0) - \mathcal{P}^x_k(\phi(bD_n^S) \mid A_n) \to 0 \tag{6.49}
\]
as \( n \to \infty \), where \( A_n := \{ \sup_t S_t \geq b^{-1}n \} \). To prove this (because \( \phi \) is bounded) it is enough to show that
\[
\mathbb{P}^x(A_n \mid |N_n| > 0) \to 1 \text{ and } \mathbb{P}^x(|N_n| > 0 \mid A_n) \to 1 \tag{6.50}
\]
as \( n \to \infty \). The first convergence follows by observing that
\[
\mathbb{P}^x(A_n \mid |N_n| > 0) \geq \mathbb{P}^x(A_n \mid |N_n(1+\delta)| > 0) \frac{\mathbb{P}^x(|N_n(1+\delta)| > 0)}{\mathbb{P}^x(|N_n| > 0)}
\]
\[44\]
for any $\delta > 0$, where the first term in the product on the right hand side converges to 1 as $n \to \infty$ (for any $\delta$) by Proposition 6.13 and the second converges to $(1 + \delta)^{-1}$ as $n \to \infty$, by Theorem 1.4. To finish showing (6.50), it is therefore enough to show that $P_x[(A_n)] \sim P_x(|N_n| > 0)$ as $n \to \infty$. For this, we note that Proposition 6.10 allows us to compute an exact asymptotic for $P_x(A_n)$, just as in for example [LG05, Section 1.4, p263]. It is easy to check that this is asymptotic is the same as that satisfied by $P_x(|N_n| > 0)$ (which we know by Theorem 1.4).

To conclude, by (6.47), (6.48) and (6.49), it is enough to show that

$$P_x(k \mid \sup_t S_t \geq b^{-1} n) \to P_x[\Phi]$$

as $n \to \infty$. However, this is a direct consequence of the convergence given by (6.17).

□

References

[ADH13] R. Abraham, J.F. Delmas, and P. Hoscheit. A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces. Electronic Journal of Probability, 18, 2013.

[AH83] S. Asmussen and H. Hering. Branching processes. Birkhäuser, Boston, MA, 1983.

[Ald91] D. Aldous. The continuum random tree I. Annals of Probability, 19:1–28, 1991.

[Ald93] D. Aldous. The continuum random tree III. Annals of Probability, 21:248–289, 1993.

[Bañ99] R. Bañuelos. Intrinsic ultracontractivity and eigenfunction estimates for Schrödinger operators. Journal of Functional Analysis, 100:181–205, 1999.

[BBI01] D. Burago, Y. Burago, and S. Ivanov. A course in metric geometry. AMS, Boston MA, 2001.

[BBS11] J. Berestycki, N. Berestycki, and J. Schweinsberg. Survival of near-critical branching Brownian motion with absorption. Journal of Statistical Physics, 143:833–854, 2011.

[BBS13] J. Berestycki, N. Berestycki, and J. Schweinsberg. The genealogy of branching brownian motion with absorption. Annals of Probability, 41(2):527–618, 2013.

[BBS14] J. Berestycki, N. Berestycki, and J. Schweinsberg. Critical branching Brownian motion with absorption: survival probability. Probability Theory and Related Fields, 160:489–520, 2014.

[BBS15] J. Berestycki, N. Berestycki, and J. Schweinsberg. Critical branching brownian motion with absorption: particle configurations. Ann. Inst. H. Poincar, 51 (4):1215–1250, 2015.

[BDMM] E. Brunet, B. Derrida, A.H. Mueller, and S. Munier. Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. Phys. Rev. E (3), 76.

[BDMM06] E. Brunet, B. Derrida, A.H. Mueller, and S. Munier. Noisy travelling waves: effect of selection on genealogies. Europhys. Lett., 76:1–7, 2006.

[Ber09] N. Berestycki. Recent Progress in Coalescent Theory. Ensaios Matemáticos [Mathematical Surveys]. Sociedade Brasileira de Matemática, 2009.

[BK77] J.D. Biggins and A.E. Kyprianou. Martingale convergence in the branching random walk. J. Appl. Probability, 14(1):25–37, 1977.

[BS98] E. Bolthausen and A.-S Sznitman. On rueelle’s probability cascades and an abstract cavity method. Comm. Math. Phys., 197:247–276, 1998.
[CR88] B. Chauvin and A. Rouault. KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees. Probability Theory and Related Fields, 80:299–314, 1988.

[Dav89] E.B. Davies. Heat kernels and Spectral Theory. Cambridge University Press, 1989.

[DS84] E.B. Davies and B. Simon. Ultracontractivity and heat kernels for Schrödinger operators and Dirichlet Laplacians. Journal of Functional Analysis, 59:335–395, 1984.

[EK04] J. Engländer and A.E. Kyprianou. Local extinction vs. local exponential growth for spatial branching processes. Annals of Probability, 32:78–99, 2004.

[Eva98] L.C. Evans. Partial Differential Equations. Graduate Studies in Mathematics, Volume 19. AMS, 1998.

[GPW09] A. Greven, P. Pfaffelhuber, and A. Winter. Convergence in distribution of random metric measure spaces (λ-coalescent measure trees). Probability Theory and Related Fields, 145(1-2):285–322, 2009.

[GT83] D. Gilbarg and N.S. Trudinger. Elliptic partial differential equations of second order. Springer-Verlag, Berlin, 2nd edition, 1983.

[Her74] H. Hering. Limit theorem for critical branching diffusion processes with absorbing barriers. Math. Biosci., 19, 1974.

[Her78a] H. Hering. Multigroup branching diffusions. Adv. Probability, 5, 1978.

[Her78b] H. Hering. Uniform primitivity of semigroups generated by perturbed elliptic differential operators. Math. Proc. Cambr. Phil. Soc, 83, 1978.

[HH07] J.W. Harris and S.C. Harris. Survival probabilities for branching Brownian motion with absorption. Electronic Communications in Probability, 12:81–92, 2007.

[HH09] R. Hardy and S.C. Harris. A spine approach to branching diffusions with applications to Lp-convergence of martingales, pages 280–330. Séminaire de Probabilité XLII, volume 1979, Lecture Notes in Math. Springer, Berlin, 2009.

[HHK12] S.C. Harris, M. Hesse, and A.E. Kyprianou. Branching Brownian motion in a strip: Survival near criticality. Annals of Probability, 40(1):53–72, 2012.

[HR14] S.C. Harris and M. Roberts. A strong law of large numbers for branching processes: almost sure spine events. Electronic Communications in Probability, 19,28, 2014.

[HR15] S.C. Harris and M. Roberts. The many-to-few lemma and multiple spines. Annales de l’Institut Henri Poincaré (B) Probability and Statistics, 2015.

[JS87] J.Jacod and A.N. Shiryaev. Limit Theorems for Stochastic Processes. Springer-Verlag, 1987.

[Kes78] H. Kesten. Branching Brownian motion with absorption. Stochastic Processes and Applications, 7:9–47, 1978.

[Kol38] A.N. Kolmogorov. Zur Lösung einer biologischen Aufgabe. Izv. Naučn.-Issled. Inst. Mat. i Meh. Tomsk. Gosudarstv. Univ., 2:1–6, 1938.

[Kyp04] A. E. Kyprianou. Travelling wave solutions to the K-P-P equation: alternatives to Simon Harris probabilistic analysis. Ann. Inst. H. Poincaré Probab. Statist., 40(1):53–72, 2004.

[LG05] J.F. Le Gall. Random trees and applications. Probability Surveys, 2, 2005.
[LGD02] J.F. Le Gall and T. Duquesne. Random trees, Lévy processes and spatial branching processes. *Astérisque*, 281, 2002.

[Lyo97] R. Lyons. A simple path to Biggins’ martingale convergence for branching random walk, pages 217–221. Number 84 in IMA Volumes in Mathematics and its Applications. Springer, New York, 1997.

[McK75] H.P. McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovski-Piscounov. *Communications in Pure and Applied Mathematics*, 28:323–331, 1975.

[Mie08] G. Miermont. Invariance principles for spatial multitype Galton-Watson trees. *Annales de l’institut Henri Poincaré (B) Probabilités et Statistiques*, 44.6:1128–1161, 2008.

[Mie09] G. Miermont. Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér.*, 42(5):725–781, 2009.

[Pin85] R.G. Pinsky. On the convergence of diffusion processes conditioned to remain in a bounded region for large time to limiting positive recurrent diffusion processes. *Annals of Probability*, 13:363–378, 1985.

[Rob10] M.I. Roberts. Spine changes of measure and branching diffusions. *PhD thesis, University of Bath*, http://people.bath.ac.uk/mir20/thesis.pdf, 2010.

[Rud76] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill Book Co., New York, 3rd edition, 1976.

[Ry91] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer, 1991.

[Sev58] B.A. Sevst’yanov. Branching stochastic processes for particles diffusing in a bounded domain with absorbing boundaries. *Theory of Probability and its Applications*, 1,3:111–126, 1958.

[Wat65] S. Watanabe. On the branching process for Brownian particles with an absorbing boundary. *Jour. Math. Kyoto Univ.*, 4:385–398, 1965.

DPMMS, UNIVERSITY OF CAMBRIDGE, CAMBRIDGE, CB3 0WA, ENGLAND.
E-mail address: egpowell12@gmail.com