

An invariance principle for branching diffusions in bounded domains

Ellen Powell¹

Received: 11 October 2017 / Revised: 16 April 2018 / Published online: 27 April 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract We study branching diffusions in a bounded domain D of \mathbb{R}^d in which particles are killed upon hitting the boundary ∂D . It is known that any such process undergoes a phase transition when the branching rate β 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.

Mathematics Subject Classification 60J60 · 60J80

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

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

✉ Ellen Powell
egpowell12@gmail.com

¹ Eidgenössische Technische Hochschule Zurich, Zurich, Switzerland

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

$$\mathbb{E}(A) = m > 1 \quad \text{and} \quad \mathbb{E}(A^2) < \infty, \tag{1.1}$$

and that the generator

$$L = \frac{1}{2} \sum_{i,j} \partial_{x_j} (a^{ij} \partial_{x_i}) \tag{1.2}$$

of the diffusion is uniformly elliptic (see Sect. 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 β there is a positive probability of survival for all time, but for small β , including at criticality, there is almost sure extinction. The critical value of β is equal to $\frac{\lambda}{m-1}$, where λ 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 $t \geq 0$ we condition the critical branching diffusion to survive until time *at least* t , and look at the associated genealogical tree \mathbf{T} , equipped with its natural distance d . Rescaling distances by a factor t gives us a sequence of laws on random compact metric spaces:

$$(\mathbf{T}_{t,x}, d_{t,x}) \stackrel{(\text{law})}{:=} \left(\mathbf{T}, \frac{1}{t}d \right) \text{ under } \mathbb{P}^x(\cdot \mid |N_t| > 0). \tag{1.3}$$

We will prove that this sequence converges in distribution to a conditioned Brownian continuum random tree as $t \rightarrow \infty$, with respect to the Gromov-Hausdorff topology. Indeed, if we let \mathbf{e} be a Brownian excursion conditioned to reach height at least 1, and write $(\mathbf{T}_\mathbf{e}, d_\mathbf{e})$ for the real tree whose contour function is given by \mathbf{e} , then we obtain the following result.

Theorem 1.1 *Suppose that $D \subset \mathbb{R}^d$ is a C^1 domain¹ and that L as in (1.2) is uniformly elliptic with coefficients $a^{ij} = a^{ji} \in C^1(\bar{D})$ for $1 \leq i, j \leq d$. Further suppose that A satisfies (1.1), and that $\varphi \in C^1(\bar{D})$ where φ is the first eigenfunction of $-L$ on D (see Sect. 2.1). Then, at the critical branching rate $\beta = \lambda/(m - 1)$, and for any starting point $x \in D$,*

$$(\mathbf{T}_{t,x}, d_{t,x}) \xrightarrow[t \rightarrow \infty]{} (\mathbf{T}_\mathbf{e}, d_\mathbf{e})$$

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

¹ We say that $D \subset \mathbb{R}^d$ is a $\{\text{Lipschitz} / C^k / C^{k,\alpha}\}$ 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 $\{\text{Lipschitz} / C^k / C^{k,\alpha}\}$ function $\gamma : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ such that relabelling and reorienting axes as necessary, $D \cap B(x_0, r) = \{x \in B(x_0, r) : x^d > \gamma(x^1, \dots, x^{d-1})\}$.

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 [45] and Watanabe [46], in the case when L is a constant multiple of Δ . However, it has also been reworked and generalised since then. In [4, 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 [31] the criterion is shown to be satisfied, for example, if the manifold has C^3 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 [20], which studies local extinction versus local growth on compactly contained subsets of a (possibly infinite) domain D , and in [26], where D is taken to be a bounded interval of the real line.

Theorem 1.3 ([45], [46]) *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 λ 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 $\mathbb{P}^x(|N_t| > 0) \rightarrow 0$ as $t \rightarrow \infty$, 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, 1.6 and Corollary 1.7 have also been shown in [4, Chapter 6] in the same framework discussed above (see also [30, 31], and [29] 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).

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*

$$\mathbb{P}^x (|N_t| > 0) \sim \frac{1}{t} \times \frac{2(m-1)\varphi(x)}{\lambda (\mathbb{E}[A^2] - \mathbb{E}[A]) \int_D \varphi(y)^3 dy} \tag{1.4}$$

as $t \rightarrow \infty$. Here φ 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 [41]). 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 \mathbb{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*

$$\lim_{t \rightarrow \infty} \mathbb{P}^x (B \mid |N_t| > 0) = \mathbb{Q}^x (B). \tag{1.5}$$

To our knowledge, Proposition 1.5 does not appear in the existing literature, although the idea behind it is well-known: see [25, Theorem 5].

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*

$$\left(t^{-1} \sum_{u \in N_t} f(X_u(t)) \mid |N_t| > 0 \right) \rightarrow Z$$

in distribution as $t \rightarrow \infty$, where Z is an exponential random variable with mean

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

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 φ , normalised to be a probability distribution.

Corollary 1.7 For $t \geq 0$, let μ_t be the random point measure with law

$$\mathcal{L} \left(\frac{1}{|N_t|} \sum_{u \in N_t} \delta_{X_u(t)} \mid |N_t| > 0 \right).$$

That is, μ_t has the law of the empirical measure of particles alive at time t , given survival. Then for each f as in Theorem 1.6, we have that

$$\mu_t(f) \rightarrow \mu(f)$$

in distribution, and hence in probability, as $t \rightarrow \infty$, 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 [34] proved an asymptotic for the probability of survival up to time n :

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

where Z_n is the population size at time n , and the constant depends on the variance of the offspring distribution. Moreover, Aldous [2,3] and Duquesne and Le Gall [19] 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, [2]. In fact, this result can be extended to multitype Galton–Watson processes with a finite number of types, as in [39], 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, [6–9,33] 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 set up there is a critical value of the branching rate β , equal to $\beta_c = \mu^2/2$, such that for $\beta \leq \beta_c$ extinction occurs with probability one. In [8], an asymptotic for the survival probability is calculated, which is very different from that of Theorem 1.4. Moreover, results of [7] in the *near critical regime* suggest

that the critical genealogy in this case should be closely related to the Bolthausen–Sznitman coalescent [12]. This is also related to non-rigorous physics predictions of Brunet, Derrida, Mueller and Munier, [13, 14]: see [10] 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 [4, 45, 46], is to exploit the existence of the martingale

$$(M_t)_{t \geq 0} := \left(e^{(\lambda - \beta(m-1))t} \sum_{u \in N_t} \varphi(X_t^i) \right)_{t \geq 0} \quad (1.6)$$

(also appearing in [20, 26]) that arises naturally from the definition of the process. Since $\sum_{u \in N_t} \varphi(X_t^i)$ roughly tells us the size of the system at time t , and M_t converges almost surely as $t \rightarrow \infty$, 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, [38]) 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 $(M_t)_{t \geq 0}$ (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 $t \rightarrow \infty$, 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 Lukazewicz 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 $(M_t)_{t \geq 0}$ by “exploring it in a different order”. Just as $(M_t)_{t \geq 0}$ 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 and define a new process analogously to $(M_t)_{t \geq 0}$, this process remains a martingale and, perhaps surprisingly, becomes a kind of spatial analogue of the height function. After strengthening Corollary 1.7, we can prove that the quadratic variation of this new martingale is essentially linear, and thus obtain an invariance principle.

Of course, we have to prove that this new 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 analogous 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.

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 ∂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

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

Such a condition is satisfied, for example, if ∂D is Lipschitz, see eg. [17, 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 [21, §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 $\tilde{\varphi} \in H_0^1(D)^2$ is an eigenfunction of $-L$ with Dirichlet boundary conditions, and associated eigenvalue $\tilde{\lambda}$, 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 $v \in H_0^1(D)$, as in [21, §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 [17, Theorem 1.6.8] and [22, Theorem 9.30].):

- The eigenvalues of $-L$ are all real and can be written $0 < \lambda := \lambda_1 < \lambda_2 \leq \lambda_3 \dots$ (repeated according to their finite multiplicity) with $\lambda_k \rightarrow \infty$ as $k \rightarrow \infty$.
- 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 $\varphi_i \in C(\bar{D})$ 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 ∂D of D with the single isolated point (∂) , 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 ∂D . We write \mathbf{P}^x for its law when started from $x \in D$. Then by [17, Theorems 2.1.4 and 2.3.6], the function

$$p_t^D(x, y) := \sum_{n \geq 1} \exp(-\lambda_n t) \varphi_n(x) \varphi_n(y); \quad t \in (0, \infty), \quad x, y \in D \tag{2.2}$$

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, [17, Corollary 3.2.8]

$$0 \leq p_t^D(x, y) \leq ct^{-d/2} \tag{2.3}$$

for some constant $c > 0$.

² We define $H_0^1(D)$ to be the closure of $C_c^\infty(D)$ (the space of infinitely differentiable functions with compact support strictly inside D) with respect to the norm $\|u\|_{H^1(D)} := \|u\|_{L^2} + \sum_{i=1}^d \|D^{x_i}(u)\|_{L^2}$, where $D^{x_i}u$ is the i th partial derivative of u in the weak sense.

In particular, for any $t > 0$ and any $f \in L^1(D)$ we have

$$\int_D f(y) p_t^D(x, y) dy = \mathbf{P}^x(f(X(t))\mathbb{1}_{\{\tau^D > t\}}). \tag{2.4}$$

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 [17, §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 [18]. Intrinsic ultracontractivity is satisfied by the semigroup of the killed diffusion, for example, if the domain D is bounded and Lipschitz [5, 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*

$$K_t^D(x, y) := \frac{e^{\lambda t} p_t^D(x, y)\varphi(y)}{\varphi(x)}$$

is the transition density for $(X(t))_{t \geq 0}$ conditioned to remain in D for all time [41] then for any $\varepsilon > 0$ there exists a constant C_ε depending only on the domain such that

$$\left| \frac{K_t^D(x, y)}{\varphi(y)^2} - 1 \right| \leq C_\varepsilon e^{-\gamma t}$$

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 [18] or [5, 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*

$$p_r^D(w, z) \leq c_1 r^{-d/2} e^{-\frac{|w-z|^2}{c_2 r}},$$

for all $z, w \in D$.

Proof See [5, Eq. (1.2)] or [17, 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*

$$\varphi_i \in C^1(\bar{D})$$

for all $i \geq 1$.

Proof [22, 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

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

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 [24,28].

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 $\mathbb{N} = \{1, 2, 3 \dots\}$. 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 $v < u$) if there exists $w \in \Omega$ such that $vw = u$, 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 ,

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

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

We write

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

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 $X_u(t) = (\dagger)$ for all $t \geq s_u$, where (\dagger) 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

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

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

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

for $t \geq 0$ 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 $(\mathbb{X}_t)_{t \geq 0}$. 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 Sect. 2.1 for $0 \leq t \leq l_\emptyset$, where $l_\emptyset = \tau_\emptyset^D \wedge \nu_\emptyset$, τ_\emptyset^D is the first time that X_\emptyset hits ∂D and $\nu_\emptyset \sim \text{Exp}(\beta)$ 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 $\beta > 0$.

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 [28, Lemma 1].

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

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

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*

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

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

For the proof of the above lemmas, see [28, Lemma 1], which is stated in a more general setting. For an explanation of how this general statement gives the lemmas above, see [28, § 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 \mathbf{T} . This tree (which we emphasise is *different* from 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 (\mathbf{T}, d) is said to be a real tree if, for all $v, w \in \mathbf{T}$ the following two conditions hold [36]:

- (1) There exists a unique isometric map $\phi_{v,w} : [0, d(v, w)] \rightarrow \mathbf{T}$ with $\phi_{v,w}(0) = v$ and $\phi_{v,w}(d(v, w)) = w$.
- (2) Any continuous injective map $[0, 1] \rightarrow \mathbf{T}$ 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 $C : [0, \infty) \rightarrow [0, \infty)$ 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

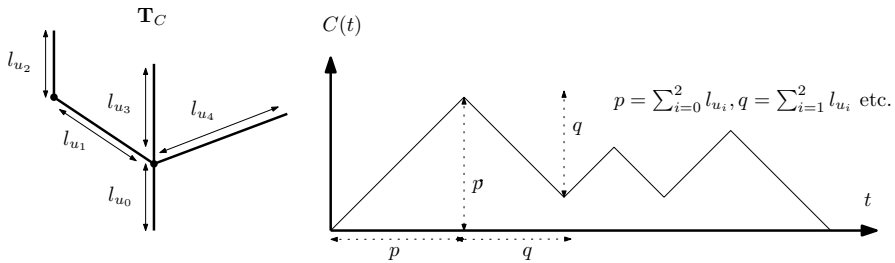


Fig. 1 An example of the continuous tree $\mathbf{T} = \mathbf{T}_C$ generated by a branching diffusion. Here $u_0 = \emptyset, u_1 = (1), u_2 = (1, 1), u_3 = (2)$ and $u_4 = (3)$. Every branch of \mathbf{T} corresponds to a particle $u \in T$, and this branch has length l_u

obtain a metric space $(\mathbf{T}_C, d_C) := ([0, \infty) / \sim, d_C)$. One can prove, see for example [36], 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 $(\mathbf{T}(T, l, X), d(T, l, X)) = (\mathbf{T}_{C(T,l,X)}, d_{C(T,l,X)})$ be the real tree with contour function $(C(T, X, l)(t))_{t \geq 0} =: (C(t))_{t \geq 0}$ described as follows. Let $\{\emptyset =: u_0, u_1, u_2, \dots, u_{|T|}\}$ be the set of labels of T in *depth first* order. This is the ordering on T such that u is less than v iff 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 $u < w$ and $u < w'$. We can then define, for $j \geq 0$ such that $u_j \not\prec u_{j+1}$:

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

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 $[R_j + l_j, R_j + l_j + r_j)$. Finally, for $t \in [R_{|T|}, R_{|T|} + l_{|T|} + s_{|T|}]$ we let $C(t)$ interpolate linearly between $C(R_{|T|})$ 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 (Fig. 1).

2.3 Martingales

Suppose that $D \subset \mathbb{R}^d$ is bounded satisfying a uniform exterior cone condition, and \mathbb{P}^x , $(X(t))_{t \geq 0}$, τ^D and p_t^D are as in Sect. 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 $t > 0$

$$\mathbb{P}^x \left(f(X(t)) \mathbb{1}_{\{\tau^D > t\}} \right) = \sum_1^\infty \exp(-\lambda_i t) \varphi_i(x) (\varphi_i, f) \tag{2.5}$$

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

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

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

$$\left(\exp(\lambda t) \varphi(X(t)) \mathbb{1}_{\{\tau^D > t\}} \right)_{t \geq 0} \tag{2.7}$$

is a (positive) martingale under \mathbb{P}^x . Furthermore, this *single-particle* martingale implies the existence of a martingale for the entire branching diffusion under \mathbb{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*

$$\left(M_t = e^{(\lambda - \beta(m-1))t} \sum_{u \in N_t} \varphi(X_u(t)) \right)_{t \geq 0}$$

is a positive martingale under \mathbb{P}^x , for each $x \in D$. It therefore converges \mathbb{P}^x -almost surely as $t \rightarrow \infty$ to an almost surely finite limit M_∞ .

This martingale is the natural analogue of the well-known martingale $(Z_n/m^n)_{n \in \mathbb{N}}$ 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 [11,37] for the branching random walk case and [16,20,25,26,35] 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 $(M_t)_{t \geq 0}$ 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 in [16], and

since used extensively by many authors [24,26,28,43]. For a thorough exposition of the subject, we refer the reader to [24], and the overview in this section will closely follow that given in [24,28].

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

- $\emptyset \in \psi$ and $|\psi \cap (N_t \cup (\dagger))| = 1$ for all t ;
- $v \in \psi$ and $u < v$ 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

$$\tilde{\mathcal{T}} := \{(T, l, X, \psi) : (T, l, X) \in \mathcal{T}, \text{ and } \psi \text{ is a spine on } (T, l, X)\}$$

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

2.4.1 Filtrations

There are several different filtrations we can place on $\tilde{\mathcal{T}}$. We give brief descriptions of each of these below : see [43] 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 $(\tilde{\mathcal{T}}, \tilde{\mathcal{F}}_\infty)$ 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 [28] as

$$\sum_{v \in N_t \cup (\dagger)} Y(v) \mathbb{1}_{\{\psi_t = v\}} \tag{2.8}$$

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

$$\tilde{\mathbb{P}}^x(Y) = \mathbb{P}^x \left(\sum_{v \in N_t} \left(Y(v) \prod_{u < v} \frac{1}{A_u} \right) + Y((\dagger)) \sum_{w \in D_t} \prod_{u < w} \frac{1}{A_u} \right), \tag{2.9}$$

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

2.4.3 Change of measure

Recall from Sect. 2.3 that $e^{\lambda t} \frac{\varphi(X(t))}{\varphi(x)} \mathbb{1}_{\{\tau^D > t\}}$ defines a mean-one martingale under \mathbf{P}^x . This implies, see [24], that

$$\zeta_t := \mathbb{1}_{\{A_{\psi_s} > 0 \forall s < t\}} \mathbb{1}_{\{\tau_\xi^D > t\}} e^{(\lambda - \beta(m-1))t} \frac{\varphi(\xi_t)}{\varphi(x)} \prod_{v < \psi_t} A_v,$$

where τ_ξ^D is the first time that ξ 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 $\frac{M_t}{\varphi(x)} = \tilde{\mathbb{P}}^x(\zeta_t | \mathcal{F}_t)$. Thus, if we define a new probability measure $\tilde{\mathbb{Q}}^x$ on $(\tilde{\mathcal{T}}, \tilde{\mathcal{F}}_\infty)$ via the martingale change of measure

$$\left. \frac{d\tilde{\mathbb{Q}}^x}{d\tilde{\mathbb{P}}^x} \right|_{\tilde{\mathcal{F}}_t} := \zeta_t, \tag{2.10}$$

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

$$\left. \frac{d\mathbb{Q}^x}{d\mathbb{P}^x} \right|_{\mathcal{F}_t} := \frac{M_t}{\varphi(x)}. \tag{2.11}$$

We have the following description for how the branching diffusion with a distinguished spine behaves under \mathbb{Q}^x (see for example [16] or [24]):

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

$$\left. \frac{d\mathbb{Q}^x}{d\mathbf{P}^x} \right|_{\mathcal{G}_t} := \mathbb{1}_{\{\tau_\xi^D > t\}} e^{\lambda t} \frac{\varphi(\xi_t)}{\varphi(x)}; \tag{2.12}$$

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

$$\mathbb{P}(\hat{A} = k) = \frac{k}{m} \mathbb{P}(A = k); \text{ and}$$

- 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 $\tilde{\mathbb{Q}}_x$ never leaves the domain D , and that it always has a positive number of children. This means that the branching diffusion under $\tilde{\mathbb{Q}}$ 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 $K_t^D(x, y)$: the kernel we defined in Lemma 2.1. Its motion is that of the diffusion $X(t)$ under \mathbf{P}^x , but *conditioned to remain in D for all time*, see [41]. In particular, Lemma 2.1 tells us that if the domain D is Lipschitz, its position converges quickly to an equilibrium distribution with density φ^2 . Finally, we record that (by an easy calculation):

$$\tilde{\mathbb{Q}}^x(v = \psi_t | \mathcal{F}_t) = \frac{\varphi(X_v(t))}{\sum_{u \in N_t} \varphi(X_u(t))} \tag{2.13}$$

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 $M_t \rightarrow M_\infty$ 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 $\beta > \lambda/(m - 1)$, 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

$$\mathbb{P}^x(|N_t| > 0) \leq \mathbb{P}^x(|N_t|) = e^{(m-1)\beta} \mathbf{P}^x(\tau^D > t), \tag{3.1}$$

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

$$p_t^D(x, y) \leq C\varphi(x)\varphi(y) e^{-\lambda t} \tag{3.2}$$

for all $t \geq 1$. Since $\mathbf{P}^x(\tau^D > t) = \int_D p_t^D(x, y) dy$, this implies that when $\beta < \lambda/(m - 1)$, the right hand side of (3.1) converges to 0 as $t \rightarrow \infty$ (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 [46, Lemma 2.1]. □

Lemma 3.1 ([46]) *For all $x \in D$*

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

The proof we give is the same as that in [46] (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 $|N_t| \neq 1$, 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 $\mathbb{P}^x(H_n < \infty) \rightarrow 0$ as $n \rightarrow \infty$. Set

$$v = \inf_{y \in D} \mathbf{P}^y(\tau^D \leq 1) > 0$$

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 $\mathbb{P}^x(H_n < \infty) \leq \mathbb{P}^x(|N_{L_n}| > 0) \leq (1 - pv)(1 - (pv)^k)^{n-1} \rightarrow 0$. This completes the proof. \square

With this in hand, to prove that we have almost sure extinction in the critical case, it is enough to show that $\mathbb{P}^x(|N_t| \rightarrow \infty \text{ as } t \rightarrow \infty) = 0$. 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

$$\begin{aligned} \mathbb{P}^x \left(\sum_{u \in N_t} \mathbb{1}_{\{\varphi(X_u(t)) \leq K/R\}} \geq 1 \right) &\leq \mathbb{P}^x \left(\sum_{u \in N_t} \mathbb{1}_{\{\varphi(X_u(t)) \leq K/R\}} \right) \\ &\leq e^{(m-1)\beta t} \int_{y \in D: \varphi(y) \leq K/R} p_t^D(x, y) dy \end{aligned}$$

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

$$\begin{aligned} \mathbb{P}^x(|N_t| \rightarrow \infty) &= \mathbb{P}^x(\{|N_t| \rightarrow \infty\} \cap \{M_\infty < \infty\}) \\ &= \mathbb{P}^x(\cup_K \cap_R \cup_T \cap_{t \geq T} A_{t,K,R}), \end{aligned} \tag{3.3}$$

and we have just shown that $\mathbb{P}^x(\cap_{t \geq T} A_{t,K,R}) \leq \mathbb{P}^x(A_{T,K,R}) \leq cK/R$ for all $T \geq 1$. Taking limits on the right hand side of (3.3) we see that $\mathbb{P}^x(|N_t| \rightarrow \infty) = 0$.

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

$$u(x, t) := \begin{cases} \mathbf{P}^x(|N_t| > 0) & x \in D \\ 0 & x \in \partial D \end{cases} \tag{3.4}$$

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 $t \rightarrow \infty$ for each $x \in \bar{D}$, by Dini's theorem [44, 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 $u(y, t) \rightarrow 0$ as $y \rightarrow \partial D$. However, this follows since

$$\begin{aligned} \mathbb{P}^y(|N_t| > 0) &\leq 1 - \mathbb{P}^y(\text{the process becomes extinct before the first branching time}) \\ &\leq 1 - \mathbb{P}(Z > s)\mathbf{P}^y(\tau^D \leq s) \end{aligned}$$

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 $y \rightarrow \partial D$. Thus $u(\cdot, t) \in C(\bar{D})$ 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 and 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 $y \in D$,

$$\mathbf{P}^y(\tau^D > r) \leq c_T\varphi(y)e^{-\lambda r}. \tag{3.5}$$

We can also write, by Lemma 2.5

$$\mathbb{P}^x \left(|N_t|^2 \right) = e^{\lambda t} \mathbf{P}^x(\tau^D > t) + \frac{\lambda}{m - 1} (\mathbb{E}(A^2) - \mathbb{E}(A)) \int_0^t e^{\lambda(2t-s)} \mathbf{P}^x(\mathbb{1}_{\{\tau^D > s\}} \mathbf{P}^{X(s)}(\mathbb{1}_{\{\tau^D > t-s\}})^2) ds$$

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 $c''_T e^{\lambda s} \mathbf{P}^x(\varphi(X_s) \mathbb{1}_{\{\tau^D > s\}})$ (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*

$$a_f(t) := \int_D \varphi(x) \mathbb{P}^x \left(1 - \hat{f}(N_t) \right) dx; \quad \hat{f}(N_t) := \begin{cases} \prod_{u \in N_t} f(X_u(t)) & \text{if } N_t \neq \emptyset \\ 1 & \text{if } N_t = \emptyset. \end{cases}$$

Then

$$\left| \frac{\mathbb{P}^x \left(1 - \hat{f}(N_t) \right)}{\varphi(x) a_f(t)} - 1 \right| \rightarrow 0 \tag{4.1}$$

as $t \rightarrow \infty$, uniformly in $x \in D$ and f with $0 \leq f < 1$.

Remark 4.2 When $f = 0$ we see that $\mathbb{P}^x \left(1 - \hat{f}(N_t) \right) = \mathbb{P}^x(|N_t| > 0)$ 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 (\mathcal{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 Sect. 2.4.3,

$$\frac{\mathbb{P}^x(1 - \hat{f}(N_t))}{\varphi(x)} = \mathbb{Q}^x\left(\frac{1 - \hat{f}(N_t)}{\sum_{u \in N_t} \varphi(X_u(t))}\right) = \tilde{\mathbb{Q}}^x\left(\frac{1 - \hat{f}(N_t)}{\sum_{u \in N_t} \varphi(X_u(t))}\right) \tag{4.2}$$

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 $\tilde{\mathbb{Q}}^x$, 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

$$N_t = N_t^1 \cup N_t^2 := \{u \in N_t : \psi_{t_0} \prec u\} \cup \{u \in N_t : \psi_{t_0} \not\prec u\}.$$

Then we see using (4.2) that

$$\begin{aligned} \frac{\mathbb{P}^x(1 - \hat{f}(N_t))}{\varphi(x)} &\geq \tilde{\mathbb{Q}}^x\left(\frac{1 - \hat{f}(N_t^1)}{\sum_{u \in N_t^1} \varphi(X_u(t))} \mathbb{1}_{\{N_t^2 = \emptyset\}}\right) \\ &= \tilde{\mathbb{Q}}^x\left(\tilde{\mathbb{Q}}^x\left(\frac{1 - \hat{f}(N_t^1)}{\sum_{u \in N_t^1} \varphi(X_u(t))} \mathbb{1}_{\{N_t^2 = \emptyset\}} \mid \tilde{\mathcal{G}}_{t_0}\right)\right) \end{aligned}$$

where we recall the definition of the σ -algebra $\tilde{\mathcal{G}}_{t_0}$ from Sect. 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 $\tilde{\mathbb{Q}}^x$, we see that this is equal to

$$\tilde{\mathbb{Q}}^x\left(\tilde{\mathbb{Q}}^{\xi_{t_0}}\left(\frac{1 - \hat{f}(N_{t-t_0})}{\sum_{u \in N_{t-t_0}} \varphi(X_u(t-t_0))}\right) \prod_{v \prec \psi_{t_0}} (\mathbb{P}^{X_v(s_v)}(|N_{t-s_v}| = 0))^{A_v - 1}\right) \tag{4.3}$$

where recalling the notation from Sect. 2.2.3, this product is over all branching points along the spine before time t_0 (at times s_v , and positions $X_v(s_v) (= \xi(s_v))$), with A_v

children being born). Now we make the simple bound

$$1 - \sum_{v < \psi_{t_0}} A_v \sup_{w \in D} \mathbb{P}^w (|N_{t-t_0}| > 0) \leq \prod_{v < \psi_{t_0}} (\mathbb{P}^{X_v(s_v)} (|N_{t-s_v}| = 0))^{A_v - 1}$$

which tells us, by independence of ξ_{t_0} and $\sum_{v < \psi_{t_0}} A_v$, that (4.3) is greater than or equal to

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

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

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

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 $K_{t_0}^D(x, y)$ 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 $(1 - c' e^{-\gamma t_0})a_f(t - t_0)$, for another constant $0 < c' < \infty$, where $\gamma > 0$ is the spectral gap for L on D . Overall, we obtain that

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

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

$$\begin{aligned} \mathbb{P}^x (1 - \hat{f}(N_t)) &= \mathbb{P}^x \left(\mathbb{P}^x (1 - \hat{f}(N_t) | \mathcal{F}_{t_0}) \right) = \mathbb{P}^x \left(1 - \prod_{u \in N_{t_0}} (1 - \mathbb{P}^{X_u(t_0)} (1 - \hat{f}(N_{t-t_0}))) \right) \\ &\leq \mathbb{P}^x \left(\sum_{u \in N_{t_0}} \mathbb{P}^{X_u(t_0)} (1 - \hat{f}(N_{t-t_0})) \right), \end{aligned}$$

where the last inequality follows because $0 \leq f < 1$. Then applying the Many-to-one Lemma 2.4 and dividing through by $\varphi(x)$ we obtain

$$\frac{\mathbb{P}^x (1 - \hat{f}(N_t))}{\varphi(x)} \leq \int_D \frac{e^{\lambda t_0} p_{t_0}^D(x, y)}{\varphi(x)\varphi(y)} \varphi(y) \mathbb{P}^y (1 - \hat{f}(N_{t-t_0})) dy.$$

Finally, by Lemma 2.1 again, we see this is bounded above by $a_f(t - t_0)(1 + c' e^{-\gamma t_0})$, and so combining the upper and lower bounds we have

$$\left| \frac{\mathbb{P}^x \left(1 - \hat{f}(N_t) \right)}{\varphi(x)a_f(t - t_0)} - 1 \right| \leq C \left(e^{-\gamma t_0} + t_0 \sup_{w \in D} \mathbb{P}^w (|N_{t-t_0}| > 0) \right) \tag{4.4}$$

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 $\sup_{w \in D} \mathbb{P}^w (|N_s| > 0) \rightarrow 0$ as $s \rightarrow \infty$, 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 $t \rightarrow \infty$. Thus, to complete the proof we need only show that $a_f(t) \sim a_f(t - t_0(t))$ as $t \rightarrow \infty$. However, this follows since

$$\left| \frac{a_f(t)}{a_f(t - t_0(t))} - 1 \right| = \left| \int_D \left(\frac{\mathbb{P}^x \left(1 - \hat{f}(N_t) \right)}{\varphi(x)a_f(t - t_0(t))} - 1 \right) \varphi(x)^2 dx \right| \tag{4.5}$$

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

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 $t \rightarrow \infty$, where

$$b := \frac{2(m - 1)}{\lambda(1, \varphi^3)(\mathbb{E}(A^2) - \mathbb{E}(A))}. \tag{4.6}$$

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 $u_f(x, t) = \mathbb{P}^x \left(1 - \hat{f}(N_t) \right)$ 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*

$$\frac{da_f}{dt}(t) = -\frac{\lambda}{m - 1} \int_D (G(1 - u_f(x, t)) + mu_f(x, t) - 1)\varphi(x) dx \tag{4.7}$$

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

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

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

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) (and (2.2), which shows that p_t^D is symmetric in x and y). Hence, by Fubini, we have

$$a_f(t) = e^{-\lambda t(\frac{1}{m-1}+1)} \int_D \varphi(y)(1 - f(y)) dy + e^{-\lambda t(\frac{1}{m-1}+1)} \int_0^t \int_D \frac{\lambda}{m-1} e^{\lambda s(\frac{1}{m-1}+1)} \varphi(y)(1 - G(1 - u_f(y, s))) dy ds \tag{4.8}$$

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 $s \in [0, \infty)$

$$\begin{aligned} |u_f(y, s) - u_f(y, t)| &= \left| \mathbb{P}^y \left(\hat{f}(N_t) - \hat{f}(N_s) \right) \right| \\ &= \left| \mathbb{P}^y \left(\mathbb{1}_{\{|N_{t \wedge s}| > 0\}} \left(\prod_{u \in N_{t \wedge s}} f(X_u(t \wedge s)) \right. \right. \right. \\ &\quad \left. \left. \left. - \prod_{u \in N_{t \wedge s}} \mathbb{P}^{X_u(t \wedge s)} \left(\hat{f}(N_{|t-s|}) \right) \right) \right) \right| \end{aligned}$$

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

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

which we can also write as

$$\begin{aligned} & e^{\lambda(t \wedge s)} \mathbf{E}^y \left(\mathbb{1}_{\{\tau^D > t \wedge s\}} \left| f(X(t \wedge s)) - \mathbb{P}^{X_u(t \wedge s)} \left(\hat{f}(N_{|t-s|}) \right) \right| \right) \\ &= e^{\lambda(t \wedge s)} \int_D p_{t \wedge s}^D(y, z) |f(z) - \mathbb{P}^z \left(\hat{f}(N_{|t-s|}) \right)| dz \end{aligned}$$

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

$$\left| f(z) - \mathbb{P}^z \left(\hat{f}(N_r) \right) \right| \rightarrow 0$$

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

To do this, we bound

$$\begin{aligned} \mathbb{P}^z \left(f(z) - \hat{f}(N_r) \right) &\leq (1 - e^{-\lambda/(m-1)r}) + f(z) \mathbf{P}^z(\tau^D \leq r) \\ &\quad + \int_D p_r^D(z, w) |f(z) - f(w)| dw, \end{aligned} \tag{4.10}$$

which follows since the probability of the first branching time being less than r is equal to $(1 - e^{-(m-1)\lambda r})$, and on this event $|f(z) - \hat{f}(N_r)| \leq 1$. Clearly the first two terms of (4.10) go to 0 as $r \rightarrow 0$. Moreover so does the third. Indeed, by continuity of f , for any $\varepsilon > 0$ we can choose $\eta = \eta(\varepsilon)$ such that $|w - z| < \eta \Rightarrow |f(w) - f(z)| < \varepsilon$, and hence the third term can be made less than $\varepsilon + \mathbf{P}^z(|X(r) - z| > \eta)$. By continuity of the process X , we can then choose r small enough that this is less than 2ε . \square

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, $s^{-2}(G(1 - s) + ms - 1) \rightarrow \frac{1}{2}(\mathbb{E}[A^2] - \mathbb{E}[A])$ as $s \rightarrow 0$ (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 $E(x, t) \rightarrow 0$ uniformly in x as $t \rightarrow \infty$ (here we have also used that the survival probability decays uniformly to 0 in D). Now by Proposition 4.1, we can conclude that

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

where $\hat{E}(t) \rightarrow 0$ as $t \rightarrow \infty$. Using the fundamental theorem of calculus, [44, 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 \mathbb{Q}^x , as described in Sect. 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

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

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

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

where we have defined, ordering $N_T = \{u_1, \dots, u_{|N_T|}\}$ (for example with the depth first order described in Sect. 2.2.3),

$$Y := \mathbb{P}^x (|N_t| > 0 | \mathcal{F}_T) = \sum_{i=1}^{|N_T|} \mathbb{P}^{X_{u_i}(T)} (|N_{t-T}| > 0) \left(\prod_{j>i} \mathbb{P}^{X_{u_j}(t)} (|N_{t-T}| = 0) \right).$$

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

$$\frac{\mathbb{1}_B Y}{\mathbb{P}^x (|N_t| > 0)} \xrightarrow{t \rightarrow \infty} \frac{\sum_{u \in N_T} \varphi(X_u(T))}{\varphi(x)} \mathbb{1}_B = \frac{M_T}{M_0} \mathbb{1}_B$$

almost surely, as $t \rightarrow \infty$. 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 $\mathbb{1}_B Y / \mathbb{P}^x (|N_t| > 0)$ 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

$$\left(t^{-1} \sum_{u \in N_t} f(X_u(t)) \mid |N_t| > 0 \right) \rightarrow Z$$

in distribution as $t \rightarrow \infty$, where $Z \sim \text{Exp}(b(\varphi, f)^{-1})$ and $b = \frac{2(m-1)}{\lambda(1, \varphi^3)(\mathbf{E}(A^2) - \mathbf{E}(A))}$ as before.

We first need an auxilliary Lemma.

Lemma 5.1 *Recall the definition of $a_f(t)$ from Lemma 4.4. Then*

$$\frac{1}{t} \left(\frac{1}{a_f(t)} - \frac{1}{a_f(0)} \right) \rightarrow b^{-1},$$

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 $\hat{E}_f(t) \rightarrow 0$ as $t \rightarrow \infty$, 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 $\alpha > 0$

$$\mathbb{P}^x \left(e^{-\frac{\alpha}{t} \sum_{u \in N_t} \varphi(X_u(t))} \mid |N_t| > 0 \right) \rightarrow \frac{b}{b + \alpha} \tag{5.1}$$

uniformly in x as $t \rightarrow \infty$ (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 $f_t(x) = e^{-\frac{\alpha}{t}\varphi(x)}$, then

$$\mathbb{P}^x \left(e^{-\frac{\alpha}{t} \sum_{u \in N_t} \varphi(X_u(t))} \mid |N_t| > 0 \right) = 1 - \frac{u_{f_t}(x, t)}{\mathbb{P}^x(|N_t| > 0)}.$$

By Proposition 4.1 and Theorem 1.4, we also know that

$$\frac{u_{f_t}(x, t)}{\varphi(x)a_{f_t}(t)} \rightarrow 1 \text{ and } \frac{b\varphi(x)}{t\mathbb{P}^x(|N_t| > 0)} \rightarrow 1$$

as $t \rightarrow \infty$, 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

$$\frac{1}{ta_{f_t}(t)} \rightarrow \frac{1}{b} + \frac{1}{\alpha}$$

as $t \rightarrow \infty$. However, this follows directly from Lemma 5.1 and the fact that $t(1 - e^{-\frac{\alpha}{t}\varphi(z)}) \rightarrow \alpha\varphi(z)$ as $t \rightarrow \infty$ 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 $\varepsilon > 0$

$$\mathbb{P}^x \left(|t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t))| > \varepsilon \mid |N_t| > 0 \right) \rightarrow 0 \tag{5.2}$$

as $t \rightarrow \infty$, 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 φ) that we have just proved.

To prove (5.2), it is enough by conditional Markov's inequality to show that $\mathbb{P}^x \left((t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t)))^2 \mid |N_t| > 0 \right) \rightarrow 0$ as $t \rightarrow \infty$, uniformly in x . Moreover, the many-to-two lemma, Lemma 2.5, tells us that

$$\begin{aligned} \mathbb{P}^x \left((t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t)))^2 \mid |N_t| > 0 \right) &= \frac{\varphi(x)}{t\mathbb{P}^x(|N_t| > 0)} \\ &\times \left(\frac{e^{\lambda t} \mathbf{P}^x \left(\tilde{f}(X(t))^2 \mathbb{1}_{\{\tau^D > t\}} \right) + \frac{\mathbb{E}(A^2) - \mathbb{E}(A)}{m-1} \int_0^t \lambda e^{\lambda(2t-s)} \mathbf{P}^x \left(\mathbb{1}_{\{\tau^D > s\}} \left[\mathbf{P}^{X(s)} \left(\mathbb{1}_{\{\tau^D > t-s\}} \tilde{f}(X(t-s))^2 \right) \right]^2 \right) ds}{t\varphi(x)} \right) \end{aligned} \tag{5.3}$$

where we know by Theorem 1.4 that the expression outside of the brackets on the right hand side is uniformly bounded in t (≥ 1 , say) and x . We also know, by (3.2), that for all $r > 1$ and some constant C depending only on the branching diffusion, we have

$$\frac{e^{\lambda r} \mathbf{P}^x \left(\tilde{f}(X(r))^2 \mathbb{1}_{\{\tau^D > r\}} \right)}{\varphi(x)} \leq C(\varphi, \tilde{f}^2), \tag{5.4}$$

and by Lemma 2.1, plus Cauchy–Schwarz and the fact that that $(\varphi, \tilde{f}) = 0$, that

$$\begin{aligned} \frac{e^{\lambda r} \mathbf{P}^x \left(\tilde{f}(X(r)) \mathbb{1}_{\{\tau^D > r\}} \right)}{\varphi(x)} &= \int_D \left(\frac{K_r^D(x, y)}{\varphi(y)^2} - 1 \right) \tilde{f}(y)\varphi(y) dy + \int_D \tilde{f}(y)\varphi(y) dy \\ &\leq C e^{-\gamma r} (\varphi, \tilde{f}^2)^{1/2}. \end{aligned} \tag{5.5}$$

Finally, we observe that for $s > 1$, by Cauchy–Schwarz and (3.2) again,

$$\varphi(x)^{-1} \mathbf{P}^x \left(\mathbb{1}_{\{\tau^D > s\}} \left[\mathbf{P}^{X(s)} \left(\mathbb{1}_{\{\tau^D > t-s\}} \tilde{f}(X(t-s)) \right) \right]^2 \right)$$

$$\begin{aligned} &\leq \varphi(x)^{-1} \mathbf{P}^x \left(\mathbb{1}_{\{\tau^D > s\}} \mathbf{P}^{X(s)} \left(\mathbb{1}_{\{\tau^D > t-s\}} \left[\tilde{f}(X(t-s)) \right]^2 \right) \right) \\ &= \varphi(x)^{-1} \mathbf{P}^x \left(\mathbb{1}_{\{\tau^D > t\}} \left[\tilde{f}(X(t)) \right]^2 \right) \\ &\leq C e^{-\lambda t} (\varphi, \tilde{f}^2) \end{aligned}$$

where C is another, possibly different, constant. This tells us that

$$\frac{\int_{t-1}^t \lambda e^{\lambda(2t-s)} \mathbf{P}^x \left(\mathbb{1}_{\{\tau^D > s\}} \left[\mathbf{P}^{X(s)} \left(\mathbb{1}_{\{\tau^D > t-s\}} \tilde{f}(X(t-s)) \right) \right]^2 \right) ds}{\varphi(x)t} \rightarrow 0$$

as $t \rightarrow \infty$, uniformly in x . Using (5.4) and (5.5) on the remaining parts of the second line of (5.3), together with the fact that

$$\mathbf{P}^x \left(\mathbb{1}_{\{\tau^D > s\}} \varphi(X(s))^2 \right) \leq \sup_{w \in D} |\varphi(w)| \mathbf{P}^x \left(\mathbb{1}_{\{\tau^D > s\}} \varphi(X(s)) \right) \leq \varphi(x) \sup_{w \in D} |\varphi(w)| e^{-\lambda s} \tag{5.6}$$

for all $s > 0$, it is easy to conclude that $\mathbb{P}^x \left((t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t)))^2 \mid |N_t| > 0 \right)$ does indeed tend to 0, uniformly in x . \square

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 f with $(\varphi, f^2) < \infty$

$$\mathbb{P}_x \left(\left| \frac{\sum_{u \in N_t} f(X_u(t))}{\sum_{u \in N_t} \varphi(X_u(t))} - (\varphi, f) \right| > \varepsilon \mid |N_t| > 0 \right) \rightarrow 0 \tag{5.7}$$

as $t \rightarrow \infty$. Defining \tilde{f} as in the proof of Theorem 1.6, the left hand side of (5.7) is equal to

$$\begin{aligned} &\mathbb{P}^x \left(\left| \frac{t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t))}{t^{-1} \sum_{u \in N_t} \varphi(X_u(t))} \right| > \varepsilon \mid |N_t| > 0 \right) \\ &\leq \mathbb{P}^x \left(t^{-1} \sum_{u \in N_t} \tilde{f}(X_u(t)) > \delta \mid |N_t| > 0 \right) \\ &\quad + \mathbb{P}^x \left(t^{-1} \sum_{u \in N_t} \varphi(X_u(t)) < \frac{\delta}{\varepsilon} \mid |N_t| > 0 \right) \end{aligned} \tag{5.8}$$

for any $\delta > 0$. From (5.2) and Theorem 1.6, if we take a limit as $t \rightarrow \infty$ on the right hand side, we are left with simply the probability that an exponential random variable

is less than δ/ε . Taking $\delta \rightarrow 0$ proves (5.7). The corollary then follows by applying the above with both f and the constant function 1, and writing $\sum f(X_u(t))/|N_t| = \sum f(X_u(t))/\sum \varphi(X_u(t)) \times \sum \varphi(X_u(t))/|N_t|$. \square

Note that by (5.2), for any fixed δ the first term on the right hand side of (5.8) converges to 0 uniformly in x . Moreover, by Markov’s inequality we can write

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

which by (5.1) converges to $e^{1/\varepsilon} b(b + \delta^{-1})^{-1}$ as $t \rightarrow \infty$, 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 (5.7) 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 Sect. 2.2 that we can view our branching diffusion as a marked tree $(T, l, X) \in \mathcal{T}$. This means that for every t such that $N_t \neq \emptyset$, we can write $N_t = \{w_1(t), w_2(t), \dots, w_{|N_t|}(t)\}$, where the indices correspond to the depth first ordering of the particles (see Sect. 2.2.3). For any $1 \leq M \leq |N_t|$ we can then define the set

$$N_{t,M} = \{w_i(t) : 1 \leq i \leq M\},$$

and have $|N_{t,M}| = M$.

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

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

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

Proof (see Fig. 2 for a sketch) 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 v in any one of these families is close to $(f, \varphi) / (1, \varphi)$. This already shows 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

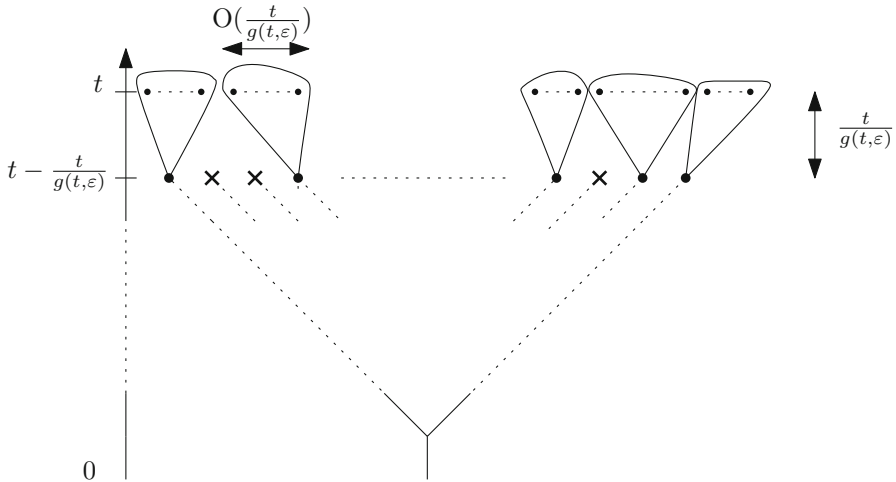


Fig. 2 Sketch of the argument. There are $O(g(t, \varepsilon))$ particles at time $t - t/g(t, \varepsilon)$ with descendants at time t (marked with dots). Each of these families is likely to be *good* and has size $O(t/g(t, \varepsilon))$

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

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

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

$$g(t, \varepsilon) p \left(\frac{t}{g(t, \varepsilon)}, \varepsilon \right) \rightarrow 0 \tag{5.9}$$

as $t \rightarrow \infty$.³

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

$$m_i := \frac{\sum_{\{v \in N_t : w_i(t-t/g(t, \varepsilon)) < v\}} f(X_v(t))}{\sigma_i} \quad \text{for} \quad \sigma_i := |\{v \in N_t : w_i(t-t/g(t, \varepsilon)) < v\}|$$

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

³ Indeed, since $\hat{p}(t) := \sup\{p(u, \varepsilon) : u \geq t\}$ converges monotonically to 0 as $t \rightarrow \infty$, we can choose $g(t, \varepsilon) \leq \sqrt{t}$ but still converging to ∞ , such that $g(t, \varepsilon)\hat{p}(\sqrt{t}) \rightarrow 0$ as $t \rightarrow \infty$. Then since $p(t/g(t, \varepsilon), \varepsilon) \leq \hat{p}(t/g(t, \varepsilon)) \leq \hat{p}(\sqrt{t})$, the function g will satisfy (5.9).

$$\mathbb{P}^x (A_t \mid |N_t| > 0) := \mathbb{P}^x \left(\bigcup_{i=1}^{|N_{t-t/g(t,\varepsilon)}|} \left\{ \left| m_i - \frac{(f, \varphi)}{(1, \varphi)} \right| > \varepsilon/2 \right\} \mid |N_t| > 0 \right) \rightarrow 0 \tag{5.10}$$

as $t \rightarrow \infty$.

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

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

which by Theorem 1.4, a union bound, and the definition of $p(s, \varepsilon)$, yields the inequality

$$\mathbb{P}^x (A_t \cap \{|N_t| > 0\}) \leq C \mathbb{P}^x (|N_{t-t/g(t,\varepsilon)}|) p \left(\frac{t}{g(t, \varepsilon)}, \varepsilon \right) \frac{g(t, \varepsilon)}{t} \tag{5.11}$$

for some C depending only on the branching diffusion. 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 right hand side of (5.11) 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 (5.11) by $\mathbb{P}^x (|N_t| > 0) \sim t^{-1} \varphi(x)$ and then applying (5.9), we obtain (5.10).

Now we prove that

$$\mathbb{P}_x (A'_t \mid |N_t| > 0) := \mathbb{P}_x \left(\bigcup_{i=1}^{|N_{t-t/g(t,\varepsilon)}|} \left\{ \sigma_i > \frac{t}{(g(t, \varepsilon))^{1/3}} \right\} \mid |N_t| > 0 \right) \rightarrow 0 \tag{5.12}$$

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

$$\begin{aligned} \mathbb{P}^x (A'_t \cap \{|N_t| > 0\}) &= \mathbb{P}^x (\mathbb{P}^x (A'_t \cap \{|N_t| > 0\} \mid \mathcal{F}_{t-t/g(t,\varepsilon)})) \\ &\leq C \mathbb{P}^x (|N_{t-t/g(t,\varepsilon)}|) \sup_{y \in D} \mathbb{P}^y \left(|N_{t/g(t,\varepsilon)}| > \frac{t}{g(t, \varepsilon)^{1/3}} \right) \\ &\leq \frac{C \varphi(x)}{g(t, \varepsilon)^{1/3} t} \end{aligned} \tag{5.13}$$

where the last inequality follows because for any $y \in D$, by Markov's inequality and the uniform bound on $\mathbb{P}^y (|N_{t/g(t,\varepsilon)}|^2)$ from Lemma 3.2,

$$\mathbb{P}^y \left(|N_{t/g(t,\varepsilon)}| > \frac{t}{g(t, \varepsilon)^{1/3}} \right) \leq \frac{\mathbb{P}^y (|N_{t/g(t,\varepsilon)}|^2) g(t, \varepsilon)^{2/3}}{t^2} \leq C \frac{\frac{t}{g(t,\varepsilon)} g(t, \varepsilon)^{2/3}}{t^2}.$$

Now dividing through by $\mathbb{P}^x (|N_t| > 0)$ on the right hand side of (5.13) gives (5.12).

Finally, we are in a position to prove that $\mathbb{P}_x(B_t^{f,\varepsilon}(\rho) \mid |N_t| > 0) \rightarrow 0$ as $t \rightarrow \infty$. 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. In fact, we will show that

$$(A_t)^c \cap (A_t')^c \cap \{\rho t \leq |N_t|\} \subset (B_t^{f,\varepsilon}(\rho))^c,$$

for all t large enough, which suffices, since by definition $(A_t)^c \cap (A_t')^c \cap \{\rho t > |N_t|\} \subset (B_t^{f,\varepsilon}(\rho))^c$.

So, suppose we are on the event $(A_t)^c \cap (A_t')^c \cap \{\rho t \leq |N_t|\}$ and for every $M \leq |N_t|$, set $M \leq k(M) := \sigma_0 + \sigma_1 + \dots + \sigma_{i+1}$, where $\sigma_0 := 0$, and i is the unique integer such that $\sigma_0 + \dots + \sigma_i < M \leq \sigma_0 + \dots + \sigma_{i+1}$. Then because we are on the event $(A_t)^c$ we have that

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

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

$$\begin{aligned} & \left| \frac{\sum_{u \in N_{t,M}} f(X_u(t))}{M} - \frac{\sum_{u \in N_{t,k(M)}} f(X_u(t))}{k(M)} \right| \\ & \leq \left| \frac{k(M) - M}{M} \right| \left(\left| \frac{\sum_{u \in N_{t,k(M)}} f(X_u(t))}{k(M)} \right| + \sup_{w \in D} |f(w)| \right) \\ & \leq \frac{1}{\rho g(t, \varepsilon)^{1/3}} \left(2 \frac{(f, \varphi)}{(1, \varphi)} + \sup_{w \in D} |f(w)| \right). \end{aligned}$$

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

6 Convergence to the Brownian CRT

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

Recall from Sect. 2.2.3 that we associate to our branching diffusion (T, l, X) a continuous plane tree $\mathbf{T} = \mathbf{T}(T, l, X)$. 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 \mathbf{T} ; this is very closely related to the continuous exploration of \mathbf{T} used to define the contour function in Sect. 2.2.3, but in this case we *do not* “backtrack” along branches of the tree (see Fig. 3).

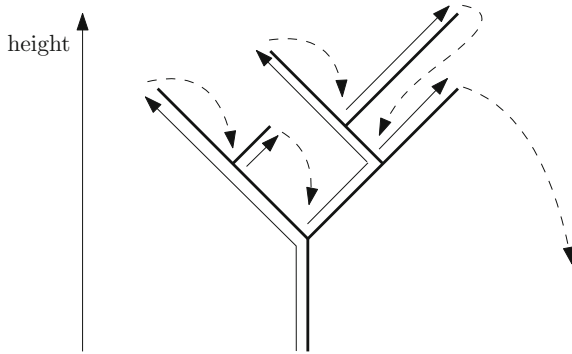


Fig. 3 A sketch of the depth first exploration of a continuous tree. Full arrowed lines represent motion at speed one in the vertical direction. Dotted arrowed lines represent instantaneous jumps

Recall that if $|T|$ is the total number of individuals in T , then we can order them $\{u_\emptyset := u_0, u_1, \dots, u_{|T|}\}$ 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

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

where we recall the definition of the lifetime l_v of an individual v from Sect. 2.2.3. Informally, this process visits the individuals of T 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 $\kappa_t = \inf\{s \leq t : v_s = v_t\}$ is the time at which the depth first exploration “starts visiting” the individual v_t . Although this depth first exploration takes values in T , we should think of it as a continuous exploration of \mathbf{T} , where between visiting vertices, the branches of \mathbf{T} are traversed at speed one. See Figure 3 for an illustration: note that we shall always refer to the exploration as an exploration of \mathbf{T} rather than an exploration of T , in order to keep the correct intuition.

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

$$H_t := (s_{v_t} - l_{v_t}) + (t - \kappa_t), \tag{6.2}$$

so describes the evolution of *height* as we traverse, if we think of our depth first exploration process as described in the previous paragraph (again see Fig. 3). We can think of this as a “non-backtracking” version of the contour function from Sect. 2.2.3. The height function will be useful for the proof of Theorem 1.1 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

$$d^*(r, w) = H_r + H_w - 2 \inf_{s \in [r \wedge w, r \vee w]} H_s. \tag{6.3}$$

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 \mathbf{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 $(\frac{[0, L(T))}{\sim^*}, d^*)$ is \mathbb{P}^x -almost surely isometrically equivalent to (\mathbf{T}, d) .

In fact, for the proof of Theorem 1.1, in the end we will prove that if we look at $(\frac{[0, L(T))}{\sim^*}, \frac{1}{t}d^*)$ under the law $\mathbb{P}^x(\cdot \mid |N_t| > 0)$, then *these* random metric spaces will converge, in distribution as $t \rightarrow \infty$, to the Brownian CRT (\mathbf{T}_e, d_e) . Remark 6.1 tells us that this implies the theorem. The reason we choose to study the height function rather than the contour function described in Sect. 2.2.3 is because, as we will soon see, there is a natural way of defining a martingale that is very closely related to the height function (and as soon as we introduce backtracking into the exploration of the tree, the martingale structure is lost).

Finally, it should be clear that any time t in our depth first exploration of the tree is naturally associated with a position in D . That is, if v_t has associated motion $X_{v_t} : [s_{v_t} - l_{v_t}, s_{v_t}) \rightarrow D$, then we can write

$$V_t := X_{v_t}((s_{v_t} - l_{v_t}) + (t - \kappa_t)). \tag{6.4}$$

This process V_t then describes the evolution of position in D , if we do a continuous depth first exploration of \mathbf{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 [19,36]) is, 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 $\tilde{\mathbb{P}}^x$ for the law of a sequence of i.i.d trees $((T^1, l^1, X^1), (T^2, l^2, X^2), \dots)$, where each (T^i, l^i, X^i) has the same distribution as (T, l, X) under \mathbb{P}^x . We concatenate the continuous depth first explorations of the trees in the natural way, and write \tilde{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 \tilde{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 [19].

In the following, we write

$$\Lambda_t := \inf \left\{ j \geq 1 : \sum_{i=1}^j L(T^i) > t \right\} \tag{6.5}$$

for the index of the tree being visited at time t in the continuous depth first exploration. We also write

$$\tilde{v}_t := (v_t, \Lambda_t) \in T^{\Lambda_t} \times \{\Lambda_t\} \subset \Omega \times \mathbb{N},$$

where v_t is the individual in T^{Λ_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 is a *younger sibling* of u , denoted $w <_s u$, if there exist $j < k \in \mathbb{N}$ and $v \in \Omega$ such that $u = vj$ and $w = vk$. If we have a tree T and an individual $u \in T$ we write

$$Y(u, T) = \{w \in T : w <_s v \text{ for some } v < u\}$$

for the set of younger siblings of ancestors of u .

Definition 6.2 Suppose we have a sequence of trees $((T^i, l^i, X^i))_{i \geq 0}$. We define for $t \geq 0$

$$\bar{S}_t = \varphi(V_t) - \sum_{i \leq \Lambda_t} \varphi(X_{\emptyset}^i(0)) + \sum_{w \in Y(v_t, T^{\Lambda_t})} \varphi(X_w^{\Lambda_t}(s_w^{\Lambda_t} - l_w^{\Lambda_t})). \tag{6.6}$$

Informally, the process $(\bar{S}_t)_{t \geq 0}$ can be defined as follows. We do our continuous depth first exploration of the sequence of trees and, until the first “branching” or “death” time, \bar{S}_t is just equal to $-\varphi(x)$, plus a term that follows $(\varphi(V_t))_{t \geq 0}$: that is, φ applied to the position of the individual we are visiting at time t . The initial $-\varphi(x)$ term is just included for convenience, 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 point⁴ 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 φ 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, $(\bar{S}_t)_{t \geq 0}$ is very closely related to the martingale $(M_t)_{t \geq 0}$ defined in Sect. 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 $(\bar{H}_t)_{t \geq 0}$ by $(\bar{S}_t)_{t \geq 0}$, and then apply an invariance principle for the latter, which we can do because it is a martingale. This is an analogous idea to that used to prove convergence of Galton–Watson processes to the CRT in [19], where $(\bar{S}_t)_{t \geq 0}$ here plays the role of the *Lukasiewicz path*. We first record a property of this process, which will be essential to showing a relationship with the height function. It states that a slight modification of $(\bar{S}_t)_{t \geq 0}$ is equal to a certain explicit process, that will later turn out to be (roughly) proportional to $(H_t)_{t \geq 0}$.

Lemma 6.3 For $t \geq 0$, let $\bar{S}'_t = \bar{S}_t - \varphi(V_t)$ and $\bar{I}'_t = \inf_{0 \leq s \leq t} \bar{S}'_s$. Then

$$\bar{S}_t := \bar{S}'_t - \bar{I}'_t = \sum_{w \in Y(v_t, T^{\Lambda_t})} \varphi(X_w^{\Lambda_t}(s_w^{\Lambda_t} - l_w^{\Lambda_t}))$$

⁴ That is, a point on the tree corresponding to one individual dying, and being replaced with a strictly positive number of offspring.

Proof Since φ is positive, it is clear that $\bar{I}_t^i = -\sum_{i \leq \Lambda_t} \varphi(X_{\bar{y}}^i(0))$. This implies the result. \square

Definition 6.4 If we have a single tree (T, X, l) , with continuous depth first exploration process $(v_t)_{t \geq 0}$, we define the corresponding processes $(S_t)_{t \geq 0}$ and $(\mathbf{S}_t)_{t \geq 0}$ by

$$S_t := \begin{cases} \varphi(V_t) + \sum_{w \in Y(v_t, T)} \varphi(X_w(s_w - l_w)) - \varphi(x); & t \leq L(T) \\ -\varphi(x); & t > L(T) \end{cases}$$

and

$$\mathbf{S}_t := \begin{cases} \sum_{w \in Y(v_t, T)} \varphi(X_w(s_w - l_w)); & t \leq L(T) \\ 0; & t > L(T). \end{cases}$$

Remark 6.5 Note that $(\mathbf{S}_t)_{t \geq 0}$ is a process that starts and ends at 0 (at time $L(T)$) and is positive in between, i.e. an excursion. Moreover, for $y > 0$ the law of $(\mathbf{S}_t)_t$ under \mathbb{P}^x , conditioned on $\{\sup_t \mathbf{S}_t \geq y\}$, is the same as the law of $(\bar{\mathbf{S}}_t)_{t \geq 0}$ under $\bar{\mathbb{P}}^x$, restricted to the first excursion in which it exceeds y .

6.1 Martingale convergence

We let $(\bar{\mathcal{F}}_t)_{t \geq 0}$ be defined for each $t \geq 0$ by $\bar{\mathcal{F}}_t :=$

$$\sigma\{\{\cup_{i < \Lambda_t} (T^i, l^i, X^i)\}, \{(l_w^{\Lambda_t}, X_w^{\Lambda_t}, A_w^{\Lambda_t}) : \bar{v}_r = (w, \Lambda_t), r < \kappa_t\}, \\ \{X_{v_t}^{\Lambda_t}(r) : r \in [s_{v_t}^{\Lambda_t} - l_{v_t}^{\Lambda_t}, s_{v_t}^{\Lambda_t} - l_{v_t}^{\Lambda_t} + (t - \kappa_t)]\}\}.$$

That is $(\bar{\mathcal{F}}_t)_{t \geq 0}$ is 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 . It encodes: (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 $\bar{\mathbb{P}}^x$ the process $(\bar{S}_t)_{t \geq 0}$ is a martingale with respect to $(\bar{\mathcal{F}}_t)_{t \geq 0}$.

Before we prove the lemma, we show that $(\bar{S}_t)_{t \geq 0}$ is locally square integrable, which will be of technical importance in what follows.

Lemma 6.7 For every $x \in D$ and $R \geq 0$, $\sup_{t \leq R} \bar{\mathbb{P}}^x[(\bar{S}_t)^2] < \infty$.

Proof of Lemma 6.7 Fix $R \geq 0$. Since φ is bounded it suffices to show that Λ_t and $|Y(v_t, \Lambda_t)|$ have uniformly bounded second moments under $\bar{\mathbb{P}}^x$ for $t \leq R$.

- For Λ , we simply note that we can stochastically dominate Λ_t for any $t \leq R$ by a geometric random variable with success probability $\mathbf{P}^x(\tau_D > R)$.
- For $|Y|$, it is a little more complicated. First observe that, for any $t \geq 0$, if we condition on the time $L(T^1) + \dots + L(T^{\Lambda_t-1})$ that we start exploring the Λ_t^{th} tree in the sequence, the conditional law of the Λ_t^{th} tree is just that of a single tree under $\bar{\mathbb{P}}^x$, conditioned on having $|N_{t-L(T^1)+\dots+L(T^{\Lambda_t-1})}| > 0$. Since the probability of

this event is bounded below by $\mathbb{P}^x(\tau^D > R)$, and $|Y(v_t, \Lambda_t)|$ is bounded above by $\max\{|Y(u, \Lambda_t)| : u \in T^{\Lambda_t}, s_u^{\Lambda_t} - l_u^{\Lambda_t} \leq R\}$, it is enough to show that

$$\mathbb{P}^x[(\max\{|Y(u, T)| : u \in T, s_u - l_u \leq R\})^2] < \infty \tag{6.7}$$

(note this is now an expectation for a single branching process). However this follows from the assumption that A has finite variance. Indeed, consider a pure (i.e. no spatial motion) branching process in continuous time, that branches at rate $\lambda/(m - 1)$ and has offspring distribution $A' = A + 1$ (so that there is always a positive number of offspring born). Then if \mathcal{N}_R is the total size of the population of this process at time R , it is clear that \mathcal{N}_R stochastically dominates $\max\{|Y(u, T)| : u \in T, s_u - l_u \leq R\}$. Moreover, it is easy to check that \mathcal{N}_R has finite variance (for example, by conditioning on the first branching time).

□

Proof of Lemma 6.6 Before commencing with the technicalities of the proof, we would like to emphasise that the reason $(\bar{S}_t)_{t \geq 0}$ is a martingale is exactly the same as the reason that $(M_t)_{t \geq 0}$ from (1.6) is a martingale. The branching Markov structure of the process is preserved when we explore in the depth first order, and this means that $(\bar{S}_t)_{t \geq 0}$ being a martingale is equivalent to it having constant expectation. The fact that it has constant expectation is intuitively clear because, after a small time δ , $\mathbb{P}^x(\varphi(V_\delta)) \approx e^{-\lambda\delta} \varphi(x) \approx \varphi(x)(1 - \lambda\delta)$ and branching contributes $\lambda\delta\varphi(x)$ to the expectation (the fact that we are exploring in the depth first order does not affect this).

Let us now be more precise. The first step in the proof is to show that if $(S_t)_{t \geq 0}$ is the process $(\bar{S}_t)_{t \geq 0}$ for a single tree (as described in Definition 6.4) then

$$\mathbb{P}^x[S_t] = 0 \quad \forall t \geq 0 \tag{6.8}$$

We will then show how this implies the martingale property.

To see (6.8) we define a sequence of times R_1, R_2, \dots as follows: let $R_1 = s_\emptyset$, and given $\{R_1, \dots, R_k\}$ set $R_{k+1} = R_k + l_{v_{R_k}}$. In words, $(R_k)_{k \geq 0}$ is the sequence of times at which the particle we are visiting during the depth first exploration of the tree either branches or exits the domain. We will prove by induction that for any $n \geq 1$

$$\mathbb{P}^x[S_{t \wedge R_n}] = 0. \tag{6.9}$$

Then since $\mathbb{P}^x[S_t] = \mathbb{P}^x[S_{t \wedge R_n}] + \mathbb{P}^x[S_t \mathbb{1}_{\{t > R_n\}}] - \mathbb{P}^x[S_{t \wedge R_n} \mathbb{1}_{\{t > R_n\}}]$ and the second two terms tend to 0 as $n \rightarrow \infty$ (by Cauchy–Schwarz, (6.7), and the fact that $\mathbb{P}(t > R_n) \rightarrow 0$) this implies (6.8).

For the induction, first note that in the case $n = 1$ we can immediately write $\mathbb{P}^x[S_{t \wedge R_1}] = \mathbb{P}^x[M_{t \wedge R_1}] - \varphi(x) = 0$. Then for the inductive step we consider $\mathbb{P}^x[S_{t \wedge R_{n+1}} - S_{t \wedge R_n}]$, and condition on everything that has happened in the depth first exploration up to time $t \wedge R_n$. In the case that $t \leq R_n$ or $t \geq L(T)$ the conditional expectation of $S_{t \wedge R_{n+1}} - S_{t \wedge R_n}$ is clearly equal to 0. Moreover, on the complementary event we can use exactly the same argument as for the $n = 1$ case to see that the conditional expectation is equal to 0. This proves (6.9) for all $n \geq 1$.

Finally, we must use this to prove the martingale property. The idea is that for $t > s$, $\bar{\mathbb{P}}^x(\bar{S}_t - \bar{S}_s | \bar{\mathcal{F}}_s)$ can be written as a sum of expectations of the form (6.8): one for each of the subtrees rooted at the elements of $Y(v_s, \Lambda_s)$ plus one for each $(T^j)_{j > \Lambda_s}$, where the time t in (6.8) is replaced in each expectation by a random time that is (importantly) independent of the tree/subtree being considered. More precisely, we can write for $0 \leq s \leq t$

$$\begin{aligned} \bar{\mathbb{P}}^x(\bar{S}_t - \bar{S}_s | \bar{\mathcal{F}}_s) = & \mathbb{P}^{V_s}(S_{(t-s)}) - \varphi(V_s) + \sum_{w \in Y(v_s, \Lambda_s)} \left(\left(\mathbb{P}^{X_w^{\Lambda_t}(s_w^{\Lambda_t} - l_w^{\Lambda_t})} \otimes \mu_w \right) (S_{\sigma(w)}) \right. \\ & \left. - \varphi(X_w^{\Lambda_t}(s_w^{\Lambda_t} - l_w^{\Lambda_t})) \right) + \sum_{i=1}^{\infty} \left((\mathbb{P}^x \otimes \mu_i)(S_{\sigma_i}) - \varphi(x) \right) \end{aligned}$$

where \otimes represents the independent product measure, the μ_w are random laws on times $\sigma(w)$ for each $w \in Y(v_s, \Lambda_s)$, and the μ_i are random laws on times σ_i for each $i \geq 1$.⁵ By (6.8) we see that the right hand side of the above is equal to 0. \square

Lemma 6.8 *For any $x \in D$, $(\bar{S}_t)_{t \geq 0}$ is a locally square-integrable martingale under $\bar{\mathbb{P}}^x$. Its predictable quadratic variation is given by*

$$\langle \bar{S} \rangle_t = \int_0^t \frac{\lambda}{m-1} \mathbb{E}(A^2 - 2A + 1) \varphi(V_s)^2 + 2 \sum_{i,j} a^{ij}(V_s) \frac{\partial \varphi}{\partial x_i} \frac{\partial \varphi}{\partial x_j}(V_s) ds.$$

Proof We already know from Lemmas 6.6 and 6.7 that $(\bar{S}_t)_{t \geq 0}$ is a locally square-integrable martingale. To calculate its predictable quadratic variation, let us look at the definition a little more closely, and try to break up $(\bar{S}_t)_{t \geq 0}$ into continuous and jump parts.

It is clear that all the terms in (6.6), apart from the $\varphi(V_t)$ term, are constant except at times where \bar{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 $t \mapsto \bar{v}_t = (v_t, \Lambda_t)$ is cadlag with respect to the discrete topology on $\Omega \times \mathbb{N}$, we can define $\bar{v}_{t-} = \lim_{s \uparrow t} \bar{v}_s$. By the above discussion, any discontinuities of \bar{S}_t must occur at times t such that $\bar{v}_{t-} \neq \bar{v}_t$.

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 two further possibilities: (a) $\Lambda_{t-} = \Lambda_t$ and $\bar{v}_t = (w, \Lambda_{t-})$ for some $w \in Y(v_{t-}, T^{\Lambda_{t-}})$ or (b) $\bar{v}_t = (\emptyset, \Lambda_t)$ 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 .

⁵ For example, if w is the oldest younger sibling of v_t (so $v_t = u_j$ and $w = u(j+1)$ for some $j \geq 1$) then $\sigma(w)$ would be $(t-s)$ minus the minimum of $(t-s)$ and the total length of the subtree rooted at v_s .

- (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 $\bar{S}_t - \bar{S}_{t-} = (A_{v_{t-}} - 1)\varphi(V_t)$.

With this in mind, we define

$$\bar{S}_t^{(c)} := \varphi(V_t) - \varphi(x) - \sum_{s \leq t} (\varphi(V_s) - \varphi(V_{s-})) \quad \text{and} \quad \bar{S}_t^{(j)} := \sum_{\substack{(w,i)=\bar{v}_s \\ \text{for some } s < \kappa_t}} \varphi(X_w^i(s_w^i))(A_w^i - 1). \tag{6.10}$$

Observe that

$$\bar{S}_t - \bar{S}_t^{(c)} := \sum_{s \leq t} (\varphi(V_s) - \varphi(V_{s-})) - \sum_{i \leq \Lambda_t} \varphi(X_{\emptyset}^i(0)) + \sum_{w \in Y(v_t, T^{\Lambda_t})} \varphi(X_w^{\Lambda_t}(s_w^{\Lambda_t} - l_w^{\Lambda_t}))$$

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 . This implies that $(\bar{S}_t^{(c)})_{t \geq 0}$ is a continuous process. Moreover, together with our previous considerations, it tells us that $\bar{S}_t - \bar{S}_t^{(c)}$ is equal to $\bar{S}_t^{(j)}$ for each t . Indeed, at times where (I) is satisfied, the two types of jump that make up $\bar{S}_t - \bar{S}_t^{(c)}$ cancel each other out, and at times where (II) or (III) is satisfied, the jumps are precisely the jumps of $\bar{S}_t^{(j)}$.

In conclusion, we can write $\bar{S}_t = \bar{S}_t^{(c)} + \bar{S}_t^{(j)}$ for $t \geq 0$, thus decomposing \bar{S} into a continuous process and a jump process. This means that if we write $\nu(\omega, dt, dx)$ for the predictable compensator of $\bar{S}_t^{(j)}$ (see [32, § II, Theorem 1.8] for the definition) then we have [32, § II, Proposition 2.29]

$$\langle \bar{S} \rangle_t = [\bar{S}^{(c)}]_t + \int_0^t \int_{\mathbb{R}} x^2 \nu(\omega, ds, dx). \tag{6.11}$$

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

$$[\bar{S}^{(c)}]_t = 2 \int_0^t \sum_{i,j} a^{ij}(V_s) \frac{\partial \varphi}{\partial x_i} \frac{\partial \varphi}{\partial x_j}(V_s) ds.$$

Moreover, one can check (straight from the definition in [32]) that

$$\nu(\omega, ds, dx) = \sum_{k \geq 0} \mathbb{P}(A = k) \delta_{(k-1)\varphi(V_{s-}(\omega))}(dx) \times \frac{\lambda}{m-1} ds. \tag{6.12}$$

From this and (6.11), the lemma then follows. □

Proposition 6.9 *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.

Remark 6.10 In fact, n need not be an integer here (ie. we can let $n \rightarrow \infty$ in \mathbb{R}). We use the notation n so as to clearly distinguish it from the continuous time parameters used to describe the evolution of the process. The same remark applies to Propositions 6.12 and 6.13 below.

Proof This follows from the functional central limit theorem for martingales [32, §VIII, Theorem 3.22] once we can show that for all $t \geq 0$

$$\langle \bar{S}^n \rangle_t \rightarrow \sigma^2 t \tag{6.13}$$

in probability as $n \rightarrow \infty$. Here $(\bar{S}^n_t)_{t \geq 0} = (\bar{S}_{nt}/\sqrt{n})_{t \geq 0}$.⁶

To show the convergence (6.13), we observe that we can write

$$\frac{\langle \bar{S}^n \rangle_t}{t} = \frac{1}{nt} \int_0^{nt} \frac{\lambda}{m - 1} \mathbb{E}(A^2 - 2A + 1) \varphi(V_s)^2 + 2 \sum_{i,j} a^{ij}(V_s) \frac{\partial \varphi}{\partial x_i} \frac{\partial \varphi}{\partial x_j}(V_s) ds.$$

Then, since φ and all of its first order derivatives are bounded (recall we are assuming this in this section), the result follows immediately from Proposition 6.11 below. Indeed the proposition gives that

$$\frac{\langle \bar{S}^n \rangle_t}{t} \rightarrow \frac{\lambda \mathbb{E}[A^2 - 2A + 1]}{(1, \varphi)(m - 1)} \int_D \varphi(x)^3 dx + \frac{2}{(1, \varphi)} \int_D \varphi(x) \sum_{i,j} a^{ij}(x) \frac{\partial \varphi}{\partial x_i} \frac{\partial \varphi}{\partial x_j}(x) dx,$$

⁶ In fact, since $(\bar{S}^n_t)_t$ may have jumps, we also need to verify an extra condition in order to apply the functional central limit theorem. However this condition, see [32, § VIII, Eq.(3.23)], is simply that we have, for every $t > 0$,

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

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

as $t \rightarrow \infty$, where the right hand side, by applying (2.1) with $v(x) = \varphi(x)^2$, is equal to

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

□

Proposition 6.11 *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}^x -probability as $t \rightarrow \infty$.

Before we prove Proposition 6.11, let us record some of the consequences of Proposition 6.9.

Proposition 6.12 *Under \mathbb{P}^x , we have the joint convergence*

$$\left(\frac{\bar{S}_{nt}}{\sqrt{n}}, \frac{\Lambda_{nt}}{\sqrt{n}} \right)_{t \geq 0} \rightarrow \left(\sigma |\beta_t|, \frac{\sigma}{\varphi(x)} L_t^0(\beta) \right)_{t \geq 0} \tag{6.14}$$

as $n \rightarrow \infty$, in distribution with respect to the Skorohod topology. Here, β is a standard Brownian motion started at 0 and $L_t^0(\beta)$ is the local time of β at 0.

Proposition 6.13 *For any $y > 0$, under $\mathbb{P}^x(\cdot \mid \{\sup_t S_t \geq \sqrt{n}y\})$ (note that we are now considering the law of a single tree), we have*

$$\left(\frac{S_{nt}}{\sqrt{n}} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{(d)} \left(\sigma \mathbf{e}_t^{\geq y/\sigma} \right)_{t \geq 0} \tag{6.15}$$

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

Proof of Proposition 6.12 From the definition of $\bar{S}'_t = \bar{S}_t - \varphi(V_t)$, we have that $|\bar{S}'_{nt}/\sqrt{n} - \bar{S}_{nt}/\sqrt{n}| \leq \|\varphi\|_\infty/\sqrt{n}$ for all $t \geq 0$ and so Proposition 6.9 implies that

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

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

$$\left(\frac{\bar{S}_{nt}}{\sqrt{n}}, -\frac{\bar{I}'_{nt}}{\sqrt{n}} \right)_{t \geq 0} = \left(\frac{\bar{S}'_{nt} - \bar{I}'_{nt}}{\sqrt{n}}, -\frac{\bar{I}'_{nt}}{\sqrt{n}} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{(d)} (\sigma(B_t - \underline{B}_t), -\sigma \underline{B}_t)_{t \geq 0} \tag{6.16}$$

where the right hand side by Lévy’s theorem, see for example [42, §VI, Theorem VI.2.3], is equal in distribution to

$$\left(\sigma|\beta_t|, \sigma L_t^0(\beta)\right)_{t \geq 0}.$$

However, we know that $\Lambda_t = -\bar{I}'_t/\varphi(x)$, and so (6.14) follows. □

Proof of Proposition 6.13 For this result, we follow [19, Proposition 2.5.2]. It is well known that you can construct the process $e^{\geq y/\sigma}$ from a standard Brownian motion β by taking

$$e_t^{\geq y/\sigma} = |\beta_{(G+t) \wedge D}|$$

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 (6.14) we also know that there exists a process

$$\left(Z_t^{(n)}, \Lambda_t^{(n)}\right)_{t \geq 0} \stackrel{(d)}{=} \left(\frac{\bar{S}'_{nt} - \bar{I}'_{nt}}{\sqrt{n}}, \frac{\Lambda_{nt}}{\sqrt{n}}\right)_{t \geq 0}$$

such that

$$\left(Z_t^{(n)}, \Lambda_t^{(n)}\right)_{t \geq 0} \xrightarrow{n \rightarrow \infty} \left(\sigma|\beta_t|, \frac{\sigma}{\varphi(x)} L_t^0(\beta)\right)_{t \geq 0}$$

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 (6.14), if we can prove that $G^{(n)} \rightarrow G$ and $D^{(n)} \rightarrow D$ almost surely, we will be done. To do this, first note that since β must exceed y/σ immediately after time T , we have that $T^{(n)} \rightarrow T$ 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 $\Lambda_t^{(n)} > \frac{\sigma}{\varphi(x)} L_T^0$ for all n large enough almost surely. Since

$$\Lambda_{T^{(n)}}^{(n)} \rightarrow \frac{\sigma}{\varphi(x)} L_T^0$$

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 6.11

The rest of this subsection will be devoted to the proof of Proposition 6.11. We begin by stating some preliminary lemmas that we will need for the proof. The first lemma gives a lower bound of order $t^{-1/2}$ for the probability that a single tree T under \mathbb{P}^x has total size (as measured by the total length $L(T)$ of the continuous depth first exploration of T) bigger than t . This is a consequence (see the proof) of the fact that, if T is conditioned on reaching height at least s with s large, T will have size of order s^2 : the same scaling relation that holds for a Brownian excursion. The second and third say that if we pick a time uniformly in $[0, t]$ and look at the height of the particle visited at this time in the concatenated depth first exploration of a sequence of trees under \mathbb{P}^x , then with high probability the height will not be of order bigger than \sqrt{t} , but it will go to ∞ . Again this is heuristically what one should expect if Theorem 1.1 is true, since the same holds for the value of a reflected Brownian motion at a uniformly chosen time in $[0, t]$.

Lemma 6.14 *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 μ_t^U for its law. In the following we write

$$H_t^* = H_{U_t} \quad ; \quad \Lambda_t^* = \Lambda_{U_t} \quad \text{and} \quad V_t^* = V_{U_t} \tag{6.17}$$

to simplify notation.

Lemma 6.15 *For all $x \in D$*

$$\tilde{\mathbb{P}}^x \otimes \mu_t^U \left(H_t^* \geq C\sqrt{t} \right) \rightarrow 0$$

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

Lemma 6.16 *For $\delta > 0$, define $R^\delta(t) > 0$ to be such that $\tilde{\mathbb{P}}^x \otimes \mu_t^U \left(H_t^* \leq R^\delta(t) \right) = \frac{\delta}{3}$. Then $R^\delta(t) \rightarrow \infty$ as $t \rightarrow \infty$ for any $x \in D$.*

We will now show how we may deduce Proposition 6.11 from these lemmas, and then go on to prove them. In the following we set $N_s^i = \{u \in T^i : s \in [s_u^i - l_u^i, s_u^i)\}$ and $\tilde{N}_{s,t}^i = \{u \in T^i : \{s \in [s_u^i - l_u^i, s_u^i)\} \cap \{(u, T^i) = \bar{v}_r \text{ for some } r \leq t\}\}$. In words, N_s^i is simply the set N_s for the i th tree, and $\tilde{N}_{s,t}^i \subset N_s^i$ 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 6.11 The main idea of the proof is to view Q_t as the ‘‘average value’’ of f taken over all the positions $(V_s)_{s \in [0,t]}$ discovered before time t in the depth first exploration of the i.i.d. sequence of trees. Recall from Corollary 5.2 that if we take the average value of f over all positions of particles at height t in a single tree (conditioned to reach height at least t) then this converges to $(f, \varphi)/(1, \varphi)$ in probability as $t \rightarrow \infty$. Roughly speaking, this convergence carries over to the average Q_t because the depth first exploration spends most of its time in large trees at relatively large heights.

To prove this properly, we define

$$m_t := \left(|\tilde{N}_{H_t^*, t}^{\Lambda_t^*}| \right)^{-1} \sum_{u \in \tilde{N}_{H_t^*, t}^{\Lambda_t^*}} f(X_u^{\Lambda_t^*}(H_t^*))$$

to be the average value of f among the vertices at height H_t^* of the Λ_t^{*th} tree, that have been visited *before time t* in the depth first exploration. Observe that

$$Q_t = \bar{\mathbb{P}}^x \otimes \mu_t^U \left(m_t \mid \sigma((T^i, X^i, l^i)_{i \geq 1}) \right).$$

This means that if m_t converges in $\bar{\mathbb{P}} \otimes \mu_t^U$ -probability to $(f, \varphi)/(1, \varphi)$, then it will also converge in $L^2(\bar{\mathbb{P}} \otimes \mu_t^U)$ (because f is bounded), and so by conditional Jensen’s inequality, Q_t will converge in L^2 and hence in probability to $(f, \varphi)/(1, \varphi)$ as well. Thus if we can show that for any fixed $\varepsilon > 0$,

$$\bar{\mathbb{P}}^x \otimes \mu_t^U (A_t) := \bar{\mathbb{P}}^x \otimes \mu_t^U \left(\left| m_t - \frac{(\varphi, f)}{(\varphi, 1)} \right| > \varepsilon \right) \rightarrow 0 \tag{6.18}$$

as $t \rightarrow \infty$, then Proposition 6.11 will follow.

To do this, also fix a $\delta > 0$: we will show that $\bar{\mathbb{P}}^x \otimes \mu_t^U (A_t) < \delta$ for all large enough t , and hence prove (6.18). First, by Lemma 6.15 we can choose C such that $\bar{\mathbb{P}}^x \otimes \mu_t^U (H_t^* \geq C\sqrt{t}) < \frac{\delta}{3}$ for all $t \geq 1$. We can also define $R(t) := R^\delta(t)$ to be such that $\bar{\mathbb{P}}^x \otimes \mu_t^U (H_t^* \leq R(t)) = \frac{\delta}{3}$, and by Lemma 6.16, we have that $R(t) \rightarrow \infty$ as $t \rightarrow \infty$. Now, for the i th tree in our exploration and $s, t > 0$, let

$$m_{s,t}^i = |\tilde{N}_{s,t}^i|^{-1} \sum_{u \in \tilde{N}_{s,t}^i} f(X_u^i(s))$$

be the average value of f among the individuals of tree i at height s that are visited before time t in the depth first exploration (so that $m_t = m_{H_t^*, t}^{\Lambda_t^*}$). Also write $A_{s,t}^i$ for the event that $|m_{s,t}^i - (\varphi, f) / (\varphi, 1)| > \varepsilon$. Then the above considerations tell us that we can write

$$\bar{\mathbb{P}}^x \otimes \mu_t^U (A_t) \leq \frac{2\delta}{3} + \frac{1}{t} \bar{\mathbb{P}}^x \left(\int_{R(t)}^{C\sqrt{t}} \sum_{i=1}^{\infty} \mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} |\tilde{N}_{s,t}^i| ds \right), \tag{6.19}$$

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

$$\frac{1}{t} \bar{\mathbb{P}}^x \left(\int_{R(t)}^{C\sqrt{t}} \sum_{i=1}^{\infty} \mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} |\tilde{N}_{s,t}^i| ds \right) = \frac{1}{t} \int_{R(t)}^{C\sqrt{t}} \sum_{i=1}^{\infty} \bar{\mathbb{P}}^x \left(\mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} |\tilde{N}_{s,t}^i| \right) ds. \tag{6.20}$$

The idea is, for each $i \geq 1$, to condition on $\mathcal{K}_{i-1} = \sigma(\cup_{j \leq i-1} (T^j, X^j, l^j))$. Note that \mathcal{K}_{i-1} is independent of the i th tree. Moreover, the event $\{i \leq \Lambda_t\}$ and the start time $\tau_i = L(T^1) + \dots + L(T^{i-1})$ of the i th tree, are measurable with respect to \mathcal{K}_{i-1} . Thus we can write, for each $i \geq 1$,

$$\begin{aligned} \bar{\mathbb{P}}^x \left(\mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} |\tilde{N}_{s,t}^i| \mid \mathcal{K}_{i-1} \right) &= \mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{P}^x \left(\mathbb{1}_{A_s(t-\tau_i)} |N_s(t - \tau_i)| \right) \\ &= \mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{P}^x \left(\mathbb{1}_{A_s(t-\tau_i)} |N_s(t - \tau_i)| \mid |N_s| > 0 \right) \mathbb{P}^x (|N_s| > 0) \end{aligned} \tag{6.21}$$

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 $A_s(r) = \{ \left| |N_s(r)|^{-1} \sum_{u \in N_s(r)} f(X_u(s)) - (\varphi, f)/(\varphi, 1) \right| > \varepsilon \}$ is the event that the average of f among the positions of these particles is more than ε away from $(\varphi, f)/(\varphi, 1)$.

We record here that by Theorem 1.4 and Lemma 3.2, there exists a K such that

$$\mathbb{P}^x \left[|N_t|^2 \mid |N_t| > 0 \right]^{1/2} \leq Kt, \quad \mathbb{P}^x (|N_t| > 0) \leq K/t \tag{6.22}$$

for all $t \geq 1$. We can also, by Lemma 6.14, choose this K such that

$$\bar{\mathbb{P}}^x (\Lambda_t) \leq \frac{K\sqrt{t}}{\varphi(x)} \tag{6.23}$$

for all $t \geq 1$. Indeed, Λ_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

$$\begin{aligned} &\mathbb{P}^x \left(\mathbb{1}_{A_s(t-\tau_i)} |N_s(t - \tau_i)| \mid |N_s| > 0 \right) \\ &\leq \frac{\delta s}{6CK^2} + \mathbb{P}^x \left(\mathbb{1}_{B_s^{f,\varepsilon}(\delta/6CK^2)} |N_s(t - \tau_i)| \mid |N_s| > 0 \right) \\ &\leq \frac{\delta s}{6CK^2} + \mathbb{P}^x \left(|N_s|^2 \mid |N_s| > 0 \right)^{1/2} \mathbb{P}^x \left(B_s^{f,\varepsilon}(\delta/6CK^2) \mid |N_s| > 0 \right)^{1/2} \end{aligned}$$

where $B_s^{f,\varepsilon}(\cdot)$ is the event from Lemma 5.3. Using (6.21), (6.22) and the above we can therefore deduce that (as long as $s \geq 1$ and t is big enough that $R(t) \geq 1$)

$$\mathbb{P}^x \left(\mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{A_{s,t}^i} | \tilde{N}_{s,t}^i \right) \leq \left(\frac{\delta}{6CK} + K^2 \mathbb{P}^x \left(B_s^{f,\varepsilon}(\delta/6CK^2) \mid |N_s| > 0 \right)^{1/2} \right) \bar{\mathbb{P}}^x (i \leq \Lambda_t). \tag{6.24}$$

This means that the right hand side of (6.20) is less than

$$\left(\frac{\delta}{6CK} + K^2 \sup_{s \geq R(t)} \left\{ \mathbb{P}^x \left(B_s(\delta/6CK^2) \mid |N_s| > 0 \right)^{1/2} \right\} \right) \frac{1}{t} \int_{R(t)}^{C\sqrt{t}} \sum_{i=1}^{\infty} \bar{\mathbb{P}}^x (\mathbb{1}_{\{i \leq \Lambda_t\}}).$$

However, by Fubini and (6.23), this is less than

$$\varphi(x)^{-1} \left(\frac{\delta}{6} + CK^3 \sup_{s \geq R(t)} \left\{ \mathbb{P}^x \left(B_s(\delta/6CK^2) \mid |N_s| > 0 \right)^{1/2} \right\} \right).$$

Using Lemma 5.3, and the fact that $R(t) \rightarrow \infty$, we see than this is less than $\frac{\delta}{3}$ for all t large enough. Substituting in to (6.19) proves that $\bar{\mathbb{P}}^x \otimes \mu_t^U(A_t) < \delta$ for such t , and hence completes the proof. \square

Proof of Lemma 6.14 In fact, we will prove a slightly stronger statement, as it will also be of use later on. We will show that

$$\mathbb{P}^x(L(T) \leq cs^2 \mid |N_s| > 0) \rightarrow 0 \tag{6.25}$$

as $c \rightarrow 0$, uniformly in $s \geq 0$ and $x \in D$. Then by choosing c such that $\mathbb{P}^x(L(T) \geq cs^2 \mid |N_s| > 0) \geq 1/2$ for all $s \geq 0$ we can write

$$\mathbb{P}^x(L(T) \geq t) \geq \mathbb{P}^x \left(L(T) \geq t \mid |N_{\sqrt{t/c}}| > 0 \right) \mathbb{P}^x \left(|N_{\sqrt{t/c}}| > 0 \right) \geq \frac{1}{2} \mathbb{P}^x \left(|N_{\sqrt{t/c}}| > 0 \right)$$

for all t . Applying Theorem 1.4, Lemma 6.14 follows.

To prove (6.25) we will show that for any $\delta > 0$ there exists a $c > 0$ and $S > 0$ such that $\mathbb{P}^x(L(T) \leq cs^2 \mid |N_s| > 0) < \delta$ 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 $\mathbb{P}^x(L(T) \leq cs^2 \mid |N_s| > 0) \leq \mathbb{P}^x(|N_S| > 0)^{-1} \mathbb{P}^x(L(T) \leq cS^2)$ which tends to 0 as $c \rightarrow 0$).

The idea of the proof is that, conditionally on $|N_s| > 0$, N_s will have size $O(s)$ and in fact, $O(s)$ of the particles alive at time s will be well away from the boundary of D (cf. Theorem 1.6). Then, at least in expectation, each of subtrees rooted at the individuals of N_s will contribute $O(s)$ to $L(T)$. Putting this together with a law of large numbers type argument tells us that $L(T) \geq cs^2$ with high probability for small enough c .

So, we fix $\delta > 0$ and also pick some $D' \Subset D$ (meaning that \bar{D}' is compact and contained in D). We write $N_s^{C D'} := \{u \in N_s : X_u(s) \in D'\}$ and first claim that we can choose $\eta > 0$ and $S_1 > 0$ such that

$$\mathbb{P}^x \left(|N_s^{C D'}| < \eta s \mid |N_s| > 0 \right) < \delta/2 \tag{6.26}$$

for all $x \in D$ and $s \geq S_1$. To see that this is possible, write

$$\mathbb{P}^x \left(|N_s^{C D'}| < \eta s \mid |N_s| > 0 \right) = \mathbb{P}^x \left(e^{-\frac{1}{\eta s} \sum_{u \in N_s} \mathbb{1}_{\{X_u(s) \in D'\}}} > e^{-1} \mid |N_s| > 0 \right)$$

and note that by conditional Markov’s inequality, the right hand side is less than or equal to

$$e \mathbb{P}^x \left(e^{-\frac{1}{\eta s} \sum_{u \in N_s} \mathbb{1}_{\{X_u(s) \in D'\}}} \mid |N_s| > 0 \right). \tag{6.27}$$

Since we know from Theorem 1.6 that (6.27) converges to $e(b \int_{D'} \varphi)(b \int_{D'} \varphi + \eta^{-1})^{-1}$ as $s \rightarrow \infty$, uniformly in x (uniformity is a consequence of the proof), we can choose $\eta > 0$ small enough and $S_1 > 0$ large enough so that (6.26) holds for all $x \in D$ and $s \geq S_1$.

Next we pick $a > 0$ (very large) such that $\inf_{y \in D'} \mathbb{P}^y (|N_r|) \geq 2/a\eta$ for all $r \geq 0$ (which is possible by the many-to-one Lemmas 2.4 and 2.1, and the fact that $D' \Subset D$). For $c > 0$ and $u \in \Omega$, we set $L_{c,s}^u = \int_0^{acs} |\{w \in N_{s+r} : u < w\}| dr$, so that no matter what the value of c , $L(T) \geq \sum_{u \in N_s^{C D'}} L_{c,s}^u$. This means that we can write

$$\mathbb{P}^x \left(L(T) \leq cs^2 \mid |N_s| > 0 \right) \leq \frac{\delta}{2} + \frac{\mathbb{P}^x \left(\left\{ \sum_{u \in N_s^{C D'}} L_{c,s}^u \leq cs^2 \right\} \cap \left\{ |N_s^{C D'}| \geq \eta s \right\} \right)}{\mathbb{P}^x (|N_s| > 0)} \tag{6.28}$$

To proceed, we will apply a (conditional) law of large numbers type argument (in fact in the end we will just use conditional Markov’s inequality) to the random variables $\{L_{c,s}^u; u \in N_s^{C D'}\}$. For this, we make the following observations:

- conditionally on \mathcal{F}_s , the random variables $\{L_{c,s}^u : u \in N_s^{C D'}\}$ are **independent**;
- by definition of a they all have **conditional mean** $\geq 2cs/\eta$;
- for any $r \geq r' \geq 1$ and $y \in D$, by conditioning on $\mathcal{F}_{r'}$ we have

$$\mathbb{P}^y (|N_r| \mid |N_{r'}|) \leq \sup_{z \in D} \sup_{u > 0} \mathbb{P}^z (|N_u|) \times \mathbb{P}^y (|N_{r'}|^2),$$

which by 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)) is less than or equal to $K_1 r'$ for some K_1 not depending on y, r, r' . From here, an easy integration gives that each of the $\{L_{c,s}^u : u \in N_s^{C D'}\}$, has **conditional variance** given \mathcal{F}_s less than $Kc^3 s^3 a^3$ for some $K < \infty$ not depending on c or s .

Putting all of this together we see that

$$\mathbb{P}^x \left(\sum_{u \in N_s^{C D'}} L_{c,s}^u \leq cs^2 \mid \mathcal{F}_s \right) \tag{6.29}$$

is the probability that a sum of $N_s^{C D'}$ independent positive, random variables (with means and variances as above) is less than or equal to cs^2 . Let us suppose that $N_s^{C D'} =$

n and $n \geq \eta s$. Then since the conditional mean of $\sum_{u \in N_s^{cD'}} L_{c,s}^u$ is bigger than or equal to $2csn/\eta$, we see that the probability in (6.29) is less than or equal to the conditional probability that “the absolute value of $(\sum_{u \in N_s^{cD'}} L_{c,s}^u$ minus its conditional mean) is greater than or equal to $2csn/\eta - cs^2$ ”. By conditional Markov’s inequality, using conditional independence and the upper bound on the conditional variances, we see that this is less than or equal to

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

where for both inequalities we have used the fact that $n \geq \eta s$. Substituting this in, we see that the second term on the right hand side of (6.28) is less than or equal to $\frac{\delta}{2}$ for all c small enough. This concludes the proof. \square

Proof of Lemma 6.15 Pick $t_0 > 0$. As explained in the proof of Proposition 6.11, see (6.23), Lemma 6.14 immediately implies that $\bar{\mathbb{P}}^x(\Lambda_t) \leq \frac{K\sqrt{t}}{\varphi(x)}$ 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 $\mathbb{P}^x(|N_r| > 0) \leq K\varphi(x)/r$ for all $x \in D$ and $r \geq \sqrt{t_0}$. To proceed, we observe that $\bar{\mathbb{P}}^x \otimes \mu_t^U(H_t^* \geq C\sqrt{t})$ is less than the probability that *some* tree visited before time t in the depth first exploration has height exceeding $C\sqrt{t}$. Thus, by a union bound we can write

$$\bar{\mathbb{P}}^x \otimes \mu_t^U(H_t^* \geq C\sqrt{t}) \leq \bar{\mathbb{P}}^x \left(\sum_{i=1}^{\infty} \mathbb{1}_{\{i \leq \Lambda_t\}} \mathbb{1}_{\{|N_{C\sqrt{t}}^i| > 0\}} \right).$$

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 $C \rightarrow \infty$. 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,

$$\bar{\mathbb{P}}^x \left(\{i \leq \Lambda_t\} \cap \{|N_{C\sqrt{t}}^i| > 0\} \right) \leq \bar{\mathbb{P}}^x(i \leq \Lambda_t) \mathbb{P}^x(|N_{C\sqrt{t}}| > 0).$$

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 $K\bar{\mathbb{P}}^x(i \leq \Lambda_t) \frac{\varphi(x)}{C\sqrt{t}}$. Therefore by Fubini, since $\sum_{i=1}^{\infty} \bar{\mathbb{P}}^x(i \leq \Lambda_t) = \bar{\mathbb{P}}^x(\Lambda_t) \leq K\sqrt{t}/\varphi(x)$, we have

$$\bar{\mathbb{P}}^x \otimes \mu_t^U(H_t^* \geq C\sqrt{t}) \leq \frac{K^2}{C}$$

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

Proof of Lemma 6.16 This follows from the law of large numbers. Indeed, observe that for any fixed $K > 0$, the amount of time in $[0, t]$ that the depth first exploration spends below height K in the i th tree is equal to $\int_0^K |\tilde{N}_{r,t}^i| dr$ for every $i \leq \Lambda_t$. Thus by conditioning on the entire sequence of trees, we have

$$\bar{\mathbb{P}}^x \otimes \mu_t^U (H_t^* \leq K) = \bar{\mathbb{P}}^x \left(\frac{1}{t} \sum_{i=1}^{\Lambda_t} \int_0^K |\tilde{N}_{r,t}^i| dr \right).$$

Moreover, since $t \geq \sum_{i \leq \Lambda_t} L(T^i)$, by definition of Λ_t , we can bound

$$\frac{1}{t} \sum_{i=1}^{\Lambda_t} \int_0^K |\tilde{N}_{r,t}^i| dr \leq \frac{1}{t} \sum_{i=1}^{\Lambda_t} \int_0^K |N_r^i| dr \leq \frac{\sum_{i=1}^{\Lambda_t} \int_0^K |N_r^i| dr}{\sum_{i=1}^{\Lambda_t} L(T^i)}. \tag{6.30}$$

However, for each i the random variables $\int_0^K |N_r^i| dr$ 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 6.14). Since $\Lambda_t \rightarrow \infty$ a.s. as $t \rightarrow \infty$ (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 (6.30) converges to 0 almost surely under $\bar{\mathbb{P}}^x$. By dominated convergence (the random variable on the left hand side of (6.30) is a.s. less than or equal to 1) we therefore have $\bar{\mathbb{P}}^x \otimes \mu_t^U (H_t^* \leq K) \rightarrow 0$, and hence $R(t) \rightarrow \infty$ \square

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 $(H_t)_{t \geq 0}$, and the process $(S_t)_{t \geq 0}$ from Definition 6.4, corresponding to the depth first exploration of a single tree T conditioned to be large. For a vertex $u \in T$ we use the notation

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

so that $\mathbf{S}_t = \mathbf{S}(v_t)$. We will show that for large t , and for an overwhelming proportion of the individuals $u \in N_t$ (given survival), $\mathbf{S}(u)$ is close to a constant times t . Our approach will use an ergodicity property for the spine particle in the system under $\tilde{\mathbb{Q}}^x$, and is inspired from [27].

In the following, given $\eta > 0$, we will say that a pair (u, t) with $u \in N_t$ is η -bad if

$$\left| \frac{\mathbf{S}(u)}{t} - b^{-1} \right| > \eta \tag{6.31}$$

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

Proposition 6.17 Fix $\varepsilon, \eta > 0$ and write $N_t^{\eta R} := \{u \in N_t : (u, t) \text{ is } \eta_R\text{-bad}\}$. Then we have

$$\sup_{t \geq R} \mathbb{P}^x \left(\frac{|N_t^{\eta R}|}{|N_t|} > \varepsilon \mid |N_t| > 0 \right) \rightarrow 0 \tag{6.32}$$

as $R \rightarrow \infty$, for any $x \in D$.

Proof We will first show that for any $\varepsilon > 0$, setting $E_{R,t}^\varepsilon := \left\{ \frac{\sum_{u \in N_t} \varphi(X_u(t)) \mathbb{1}_{\{(u,t) \text{ is } \eta_R\text{-bad}\}}}{\sum_{u \in N_t} \varphi(X_u(t))} > \varepsilon \right\}$, we have

$$\sup_{t \geq R} \mathbb{P}^x \left(E_{R,t}^\varepsilon \mid |N_t| > 0 \right) \rightarrow 0 \tag{6.33}$$

as $R \rightarrow \infty$. The reason for this approach is that it is technically more convenient to work with $E_{R,t}^\varepsilon$ than $\{|N_t^{\eta_R}|/|N_t| > \varepsilon\}$ and, as we will see in the end, the events are essentially the same.

To show (6.33), we will use the description of the system under the measure $\tilde{\mathbb{Q}}^x$ given in Sect. 2.4.3. Recalling that $\mathbb{Q}^x = \tilde{\mathbb{Q}}^x|_{\mathcal{F}_\infty}$ has $(d\mathbb{Q}^x/d\mathbb{P}^x)|_{\mathcal{F}_t} = \varphi(x)^{-1} \sum_{u \in N_t} \varphi(X_u(t))$ we see that

$$\mathbb{P}^x(E_{R,t}^\varepsilon \mid |N_t| > 0) = \mathbb{Q}^x \left[\frac{\varphi(x)/\mathbb{P}^x(|N_t| > 0)}{\sum_{u \in N_t} \varphi(X_u(t))} \mathbb{1}_{E_{R,t}^\varepsilon} \right] = \mathbb{Q}^x \left[Y_t \mathbb{1}_{E_{R,t}^\varepsilon} \right],$$

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:

- (i) $\sup_{t \geq R} \mathbb{Q}^x(E_{R,t}^\varepsilon) \rightarrow 0$ as $R \rightarrow \infty$; and
- (ii) For every $\delta > 0$, there exists R' and K positive, such that $\mathbb{Q}^x(Y_t \mathbb{1}_{|Y_t| > K}) \leq \delta$ for all $t \geq R'$.

Point (i) comes, roughly speaking, from the fact that if ψ_t is the *spine particle* under $\tilde{\mathbb{Q}}^x$ at time t , and $t \geq R$, then (ψ_t, t) is very unlikely to be η_R -bad for large R . More precisely, by (2.13), we have that

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

and so by Markov’s inequality

$$\begin{aligned} \mathbb{Q}^x(E_{R,t}^\varepsilon) &= \tilde{\mathbb{Q}}^x(E_{R,t}^\varepsilon) \leq \varepsilon^{-1} \tilde{\mathbb{Q}}^x \left(\frac{\sum_{u \in N_t} \varphi(X_u(t)) \mathbb{1}_{\{(u,t) \text{ is } \eta_R\text{-bad}\}}}{\sum_{u \in N_t} \varphi(X_u(t))} \right) \\ &= \varepsilon^{-1} \tilde{\mathbb{Q}}^x((\psi_t, t) \text{ is } \eta_R\text{-bad}). \end{aligned}$$

Moreover, we claim (see explanation below) that

$$\sup_{t \geq R} \tilde{\mathbb{Q}}^x((\psi_t, t) \text{ is } \eta_R\text{-bad}) \rightarrow 0 \tag{6.34}$$

as $R \rightarrow \infty$. To see this, recall that under $\tilde{\mathbb{Q}}^x$ we know (from Sect. 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 mean $(2m)^{-1} (\mathbb{E}(A^2) - \mathbb{E}(A))$

(coming from the description in Sect. 2.4.3 again). On the other hand, we have that for any $r \geq 0$, $\mathbf{S}(\psi_r) = \sum_{y \in Y} \varphi(y)$, 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 [41], the trajectory of the spine is a Markov process with invariant density φ^2 , (6.34) follows by a straightforward ergodicity argument. This proves (i).

Point (ii) 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 \mathbb{Q}^x and \mathbb{P}^x again to write

$$\mathbb{Q}^x(Y_t \mathbf{1}_{\{|Y_t| > K\}}) = \frac{\mathbb{P}^x(\{|Y_t| > K\} \cap \{|N_t| > 0\})}{\mathbb{P}^x(|N_t| > 0)} = \mathbb{P}^x(|Y_t| > K \mid |N_t| > 0).$$

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 R' such that

$$\sup_{t \geq R'} \mathbb{P}^x \left(\frac{\sum_{u \in N_t} \varphi(X_u(t))}{t} < 1/K \mid |N_t| > 0 \right) \leq \delta. \tag{6.35}$$

However, this is a direct consequence of the convergence given by Theorem 1.5, since we know that for fixed K the probability in (6.35) converges, as $t \rightarrow \infty$, to the probability that an exponential random variable is less than $1/K$. This completes the proof of (6.33).

We must now deduce from (6.33) that

$$\sup_{t \geq R} \mathbb{P}^x \left(\frac{|N_t^{\eta R}|}{|N_t|} > \varepsilon \mid |N_t| > 0 \right) \rightarrow 0 \text{ as } R \rightarrow \infty.$$

The idea behind this is that $\sum_{u \in N_t} \varphi(X_u(t)) \mathbf{1}_{\{(u,t) \text{ is } \eta R\text{-bad}\}} / \sum_{u \in N_t} \varphi(X_u(t))$ is a reasonable approximation to $|N_t^{\eta R}|/|N_t|$ at large times. Roughly speaking they should at least be proportional, since φ is bounded above, and on any subdomain (where most of the particles will be at large times, given survival) it is also bounded below.

More precisely, by Corollary 1.7 we know that given $\delta > 0$, if $D_r = \{y \in D : \varphi(y) < 1/r\}$ and $N_t^{D_r} := \{u \in N_t : X_u(t) \in D_r\}$, there exists $r > 0$ such that

$$\sup_{t \geq R} \mathbb{P}^x \left(\frac{|N_t^{D_r}|}{|N_t|} > \frac{\varepsilon}{2} \mid |N_t| > 0 \right) \leq \delta/2, \tag{6.36}$$

for all R large enough. Moreover, we can write $|N_t^{\eta R}| \leq |N_t^{D_r}| + |N_t^{\eta R} \setminus (N_t^{\eta r} \cap N_t^{D_r})|$, where we have

$$\frac{|N_t^{\eta R} \setminus (N_t^{\eta r} \cap N_t^{D_r})|}{|N_t|} \leq \frac{\sum_{u \in N_t} \varphi(X_u(t)) \mathbf{1}_{\{(u,t) \text{ is } \eta R\text{-bad}\}}}{\sum_{u \in N_t} \varphi(X_u(t))} \|\varphi\|_\infty r.$$

Finally, by (6.33) we can bound

$$\sup_{t \geq R} \mathbb{P}^x \left(\frac{|N_t^{\eta R} \setminus (N_t^{\eta t} \cap N_t^{D_r})|}{|N_t|} > \varepsilon/2 \right) < \delta/2$$

for all R large enough. Putting this together with (6.36) allows us to conclude. \square

Remark 6.18 Now suppose we are given $c > 0$ and $x \in D$. Then for any t such that $ct \wedge t \geq R$

$$\begin{aligned} & \mathbb{P}^x \left(\left\{ \frac{|N_{ct}^{\eta R}|}{|N_{ct}|} > \varepsilon \right\} \cap \{|N_{ct}| > 0\} \mid |N_t| > 0 \right) \\ &= \frac{\mathbb{P}^x \left(\left\{ \frac{|N_{ct}^{\eta R}|}{|N_{ct}|} > \varepsilon \right\} \cap \{|N_{ct}| > 0\} \cap \{|N_t| > 0\} \right)}{\mathbb{P}^x(|N_t| > 0)} \\ &\leq \mathbb{P}^x \left(\left\{ \frac{|N_{ct}^{\eta R}|}{|N_{ct}|} > \varepsilon \right\} \cap \{|N_t| > 0\} \mid |N_{ct}| > 0 \right) \frac{\mathbb{P}^x(|N_{ct}| > 0)}{\mathbb{P}^x(|N_t| > 0)} \\ &\leq \left(\sup_{s \geq R} \mathbb{P}^x \left(\frac{|N_s^{\eta R}|}{|N_s|} > \varepsilon \mid |N_s| > 0 \right) \right) \frac{\mathbb{P}^x(|N_{ct}| > 0)}{\mathbb{P}^x(|N_t| > 0)} \\ &\leq \frac{F(R)}{c} \left(\sup_{s \geq R} \mathbb{P}^x \left(\frac{|N_s^{\eta R}|}{|N_s|} > \varepsilon \mid |N_s| > 0 \right) \right) \end{aligned}$$

where

$$F(R) = \frac{\sup_{s \geq R} s b^{-1} \varphi(x)^{-1} \mathbb{P}^x(|N_s| > 0)}{\inf_{s \geq R} s b^{-1} \varphi(x)^{-1} \mathbb{P}^x(|N_s| > 0)} \rightarrow 1$$

as $R \rightarrow \infty$, by Theorem 1.4. Note that we are allowing $c < 1$ here.

The next lemma provides the key connection between $(S_t)_{t \geq 0}$ and the height function $(H_t)_{t \geq 0}$ for a critical branching diffusion under \mathbb{P}^x , that is conditioned to survive for a long time. We write \mathcal{P}_1^x 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 6.19 *For any $\eta > 0$ we have*

$$\lim_{R \rightarrow \infty} \lim_{t \rightarrow \infty} \mathcal{P}_1^x((v_{t^1}, H_{t^1}) \text{ is } \eta_R\text{-bad} \mid |N_t| > 0) \rightarrow 0.$$

Remark 6.20 In the case that $H_{t^1} < R$ we also say that (v_{t^1}, H_{t^1}) is η_R -bad (with an abuse of notation; recall we only defined the notion of η_R -badness for (u, t) with $t \geq R$). However, since the probability of this event goes to 0 as $t \rightarrow \infty$ for any fixed R , this will not play a role; we only introduce the convention for notational convenience.

Proof By decomposing the probability space into the three disjoint events $\{L \leq at^2\}$, $\{L > at^2\} \cap \{H_{t1} \notin [ct, Ct]\}$ and $\{L > at^2\} \cap \{H_{t1} \in [ct, Ct]\}$ we can write for any $a, c, C > 0$

$$\begin{aligned} &\mathcal{P}_1^x((v_{t1}, H_{t1}) \text{ is } \eta_R - \text{bad} \mid |N_t| > 0) \leq \mathcal{P}_1^x(L \leq at^2 \mid |N_t| > 0) \\ &\quad + \mathcal{P}_1^x(\{H_{t1} \notin [ct, Ct]\} \cap \{L > at^2\} \mid |N_t| > 0) \\ &\quad + \mathcal{P}_1^x(\{(v_{t1}, H_{t1}) \text{ is } \eta_R - \text{bad}\} \cap \{L > at^2\} \cap \{H_{t1} \in [ct, Ct]\} \mid |N_t| > 0). \end{aligned} \tag{6.37}$$

We deal with each term separately. To begin, by putting together (6.25) and the fact that $\mathbb{P}^x(|N_{Ct}| > 0 \mid |N_t| > 0) \rightarrow 0$ we see that the first term on the right hand side of (6.37) converges to 0, uniformly in $t \geq 1$ (say), as $a \rightarrow 0$ and then $c \rightarrow 0, C \rightarrow \infty$. For the second, by definition of the conditional probability we have

$$\mathcal{P}_1^x(\{H_{t1}^1 \leq ct\} \cap \{L > at^2\} \mid |N_t| > 0) \leq (\mathbb{P}^x(|N_t| > 0))^{-1} \mathbb{P}^x(\{L \geq at^2\} \cap \{H_{t1} \leq ct\})$$

which, by conditioning on $\mathcal{F}_\infty = \sigma((T, l, X))$, is equal to

$$\mathbb{P}^x(\mathbb{1}_{\{L \geq at^2\}} \mathcal{P}_1^x(H_{t1} \leq ct \mid \mathcal{F}_\infty)).$$

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 $\sup_{z \in D} \sup_{u > 0} \mathbb{P}^z(|N_u|) < \infty$ (which we saw in the proof of Lemma 6.14) we obtain that the above is less than or equal to

$$(\mathbb{P}^x(|N_t| > 0))^{-1} \mathbb{P}^x\left(\mathbb{1}_{\{L \geq at^2\}} \frac{\int_0^{ct} |N_s| ds}{L}\right) \leq \frac{cK}{a\varphi(x)}$$

for some finite K and all $t \geq 1$. This means that the second term on the right hand side of (6.37) also converges to 0 uniformly in $t \geq 1$, as $a \rightarrow 0$ and then $c \rightarrow 0, C \rightarrow \infty$.

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

$$\lim_{R \rightarrow \infty} \lim_{t \rightarrow \infty} \mathcal{P}_1^x(\{(v_{t1}, H_{t1}) \text{ is } \eta_R - \text{bad}\} \cap \{L > at^2\} \cap \{H_{t1} \in [ct, Ct]\} \mid |N_t| > 0) = 0. \tag{6.38}$$

From now on we assume that $ct \geq R$ (which is without loss of generality, since we are letting $t \rightarrow \infty$ first). By conditioning on \mathcal{F}_∞ again, we see that

$$\begin{aligned} &\mathcal{P}_1^x(\{(v_{t1}, H_{t1}) \text{ is } \eta_R - \text{bad}\} \cap \{L > at^2\} \cap \{H_{t1} \in [ct, Ct]\} \mid |N_t| > 0) \\ &= \mathbb{P}^x\left(\mathbb{1}_{\{L > at^2\}} \int_{ct}^{Ct} \frac{|N_u^{\eta_R}| |N_u|}{|N_u| L} du \mid |N_t| > 0\right) \\ &\leq \delta + (at^2)^{-1} \mathbb{P}^x\left(\int_{ct}^{Ct} \mathbb{1}_{\{\frac{|N_u^{\eta_R}|}{|N_u|} > \delta\}} |N_u| du \mid |N_t| > 0\right) \end{aligned} \tag{6.39}$$

for any $\delta > 0$. Therefore, we just need to prove that the second term in the last expression of (6.39), for fixed δ , converges to 0 as $t \rightarrow \infty$ and then $R \rightarrow \infty$. However, we can write this term (by Fubini) as

$$(at^2)^{-1} \int_{ct}^{Ct} \mathbb{P}^x \left(|N_u| \mathbb{1}_{\left\{ \frac{|N_s^{\eta R}|}{|N_u|} > \delta \right\}} \mid |N_t| > 0 \right) du$$

which by conditional Cauchy–Schwarz, Remark 6.18, and the assumption that $ct \geq R$ is less than

$$\frac{\sqrt{F(R)}}{\sqrt{c}} \sup_{s \geq R} \mathbb{P}^x \left(\frac{|N_s^{\eta R}|}{|N_s|} > \delta \mid |N_s| > 0 \right)^{1/2} \times (at^2)^{-1} \int_{ct}^{Ct} \mathbb{P}^x (|N_u|^2 \mid |N_t| > 0)^{1/2} du.$$

Finally, we observe that by (6.22) and Theorem 1.4, we have $\mathbb{P}^x (|N_u|^2 \mid |N_t| > 0)^{1/2} \leq Mu$ for some constant $M = M(c, C')$ and for any $u \in [ct, Ct]$, $ct \geq R \geq 1$. Hence, by integrating and applying Proposition 6.17, we obtain (6.38). \square

Now we extend this result. 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, \dots, t^k) chosen conditionally independently and uniformly at random from $[0, L(T)) = [0, L)$. We also define the $k \times k$ matrices

$$\begin{aligned} (D_t^S)_{ij} &:= t^{-1} (\mathbf{S}(v_{t^i}) + \mathbf{S}(v_{t^j}) - 2\mathbf{S}(v^{ij})) \text{ and} \\ (D_t^H)_{ij} &:= t^{-1} (H_{t^i} + H_{t^j} - 2h^{ij}) \end{aligned} \tag{6.40}$$

where $v^{ij} = v_{t^i} \wedge v_{t^j}$ is the most recent common ancestor of v_{t^i} and v_{t^j} , and $h^{ij} = s_{v^{ij}}$ is its “death time” (see Sect. 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 6.21 *Let $k \geq 1$ and D_t^S and D_t^H be as defined above. Then for any $\varepsilon > 0$*

$$\mathcal{P}_k^x \left(\|b^{-1} D_t^H - D_t^S\| > \varepsilon \mid |N_t| > 0 \right) \rightarrow 0 \tag{6.41}$$

as $t \rightarrow \infty$, 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 by symmetry, and since $(D_t^H)_{ii} = (D_t^S)_{ii} = 0$ for $i = 1, 2$, we need only control the distance $|b^{-1}(D_t^H)_{12} - (D_t^S)_{12}|$. Given $\delta > 0$, we:

- choose $M \geq 1$ such that $\lim_{t \rightarrow \infty} \mathbb{P}^x (|N_{Mt}| > 0 \mid |N_t| > 0) \leq \delta/4$, which is possible by Theorem 1.4;
- set $\eta := \varepsilon/(4M)$ and pick R large enough that $\lim_{t \rightarrow \infty} \mathcal{P}_2^x (\{(v_{t^1}, H_{t^1}) \text{ is } \eta_R - \text{bad}\} \cup \{(v_{t^2}, H_{t^2}) \text{ is } \eta_R - \text{bad}\} \mid |N_t| > 0) \leq \delta/4$, which is possible by a union bound and Lemma 6.19;

- pick K large enough that $\lim_{t \rightarrow \infty} \mathbb{P}^x \left(\int_0^R |N_s| \geq K \mid |N_t| > 0 \right) \leq \delta/4$, which is possible by Theorem 1.5 and Markov’s inequality.

Putting this together, we see that for all t large enough, if we set

$$A_t := \{|N_{Mt}| > 0\} \cup \left\{ \int_{[0,R]} |N_s| \geq K \right\} \cup \{(v_{t1}, H_{t1}) \text{ is } \eta_R\text{-bad}\} \cup \{(v_{t2}, H_{t2}) \text{ is } \eta_R\text{-bad}\}$$

then $\mathbb{P}_2^x(A_t \mid |N_t| > 0) \leq \delta$. This implies that $\mathbb{P}_2^x(A_t \mid |N_t| > 0) \rightarrow 0$ as $t \rightarrow \infty$.

Now consider the complementary event A_t^c . Straight from the definitions, we know that on this event we have

$$H_{t1} \leq Mt, \quad H_{t2} \leq Mt, \quad \left| \frac{\mathbf{S}(v_{t1})}{H_{t1}} - b^{-1} \right| \leq \eta, \quad \text{and} \quad \left| \frac{\mathbf{S}(v_{t2})}{H_{t2}} - b^{-1} \right| \leq \eta \quad (6.42)$$

where $\eta = \varepsilon/(4M)$. Then there are two possibilities: we are either on the event $B := \{h^{ij} > R\}$ or the event $B^c := \{h^{ij} \leq R\}$. We will show that for large enough t , $B \cap A_t^c = B^c \cap A_t^c = \emptyset$, thus completing the proof.

To do this, observe that on $\{B \cap A_t^c\}$, by definition of A_t and “ η_R -badness”, we have that

$$h^{12} \leq Mt \quad \text{and} \quad \left| \frac{\mathbf{S}(v^{12})}{h^{12}} - b^{-1} \right| \leq \eta.$$

Putting this together with (6.42) gives the deterministic bound

$$\left| (D_t^S)_{12} - b^{-1}(D_t^H)_{12} \right| \leq \frac{1}{t} \times 4Mt \times \eta \leq \varepsilon \quad \text{on} \quad \{B \cap A_t^c\}$$

and so we have $B \cap A_t^c = \emptyset$. Furthermore, on the event $\{B^c \cap A_t^c\}$, we have $h^{ij} \leq R$ and $\mathbf{S}(v^{ij}) \leq \|\varphi\|_\infty K$ (by a very crude bound). This means that

$$\left| \mathbf{S}(v^{12}) - b^{-1}h^{12} \right| \leq K + Rb^{-1}$$

and so

$$\left| (D_t^S)_{12} - b^{-1}(D_t^H)_{12} \right| \leq \frac{\varepsilon}{2} + \frac{K + Rb^{-1}}{t}.$$

Since the second term on the right hand side above is less than $\varepsilon/2$ for all t large enough, we see that $B^c \cap A_t^c = \emptyset$ for all such t . □

6.3 Convergence to the CRT

6.3.1 Preliminaries on convergence of metric measure spaces

Before we can prove Theorem 1.1, we need to 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}_c : the space of (isometry classes of) compact metric spaces.

Definition 6.22 The *Gromov–Hausdorff* distance between (X, r_X) and (Y, r_Y) in \mathbb{X}_c is given by

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

where the infimum is taken over all isometric embeddings g_X and 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, μ) will be a compact metric space (X, r) equipped with a finite Borel measure μ . 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 begin by defining the so-called *Gromov–weak* topology.

Definition 6.23 [23, Definition 2.3] We will call a function $\Phi : \mathbb{X} \rightarrow \mathbb{R}$ a *polynomial* if there exists an $k \in \mathbb{N}$ and a bounded continuous function $\phi : [0, \infty)^{\binom{k}{2}} \rightarrow \mathbb{R}$ such that

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

where $\mu^{\otimes k}$ is the product measure of μ . We write Π for the set of all polynomials.

Definition 6.24 [23, 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 [23, Theorem 5], that this topology is metrised by the *Gromov–Prohorov metric*, that we now define. In the following, if $f : X_1 \rightarrow X_2$ is an embedding and μ is a measure on X_1 , we write $f_*\mu$ for the image measure of μ under f on X_2 : defined by $f_*\mu(A) = \mu(f^{-1}(A))$ for $A \subset X_2$.

Definition 6.25 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

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

where the infimum is as in Definition 6.22 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 [1, 40] on \mathbb{X} .

Definition 6.26 Let \mathcal{X}, \mathcal{Y} be as in Definition 6.25. The Gromov–Hausdorff–Prohorov distance between \mathcal{X} and \mathcal{Y} is defined by

$$d_{GHP}(\mathcal{X}, \mathcal{Y}) = \inf_{g_X, g_Y, Z} \left(d_{Pr}^{(Z, rz)}((g_X)_*(\mu_X), (g_Y)_*(\mu_Y)) + d_H^{(Z, rz)}(g_X(X), g_Y(Y)) \right).$$

Remark 6.27 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 6.28 [23, Corollary 3.1] *A sequence $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ of probability measures on \mathbb{X} converges weakly to a probability measure \mathbb{P} with respect to the Gromov–weak topology, if and only if*

- (i) *The family $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ is relatively compact in the space of probability measures on \mathbb{X} .*
- (ii) *For all polynomials $\Phi \in \Pi$, $\mathbb{P}_n[\Phi] \rightarrow \mathbb{P}[\Phi]$ in \mathbb{R} as $n \rightarrow \infty$.*

Lemma 6.29 [1, Theorem 2.4], [15, Theorem 7.4.15] *A set $\mathbb{K} \subset \mathbb{X}$ is relatively compact with respect to the Gromov–Hausdorff–Prohorov metric if and only if*

- (i) *There is a constant D such that $\text{diam}(\mathcal{X}) < D$ for all $\mathcal{X} \in \mathbb{K}$.*
- (ii) *For every $\delta > 0$ there exists $N = N_\delta$ such that for all $\mathcal{X} \in \mathbb{K}$, X can be covered by N_δ balls of radius δ .*
- (iii) $\sup_{\mathcal{X} \in \mathbb{K}} \mu_X(X) < +\infty$

6.3.2 Proof of the main theorem

We first need to reformulate the main theorem into a statement that we are better equipped to prove. Recall the definitions of d^* and \sim^* from Remark 6.1: in this remark we observed that $(\frac{[0, L(T))}{\sim^*}, d^*)$ is \mathbb{P}^X -almost surely isometrically equivalent to $(\mathbf{T}(T, X, l), d(T, X, l))$ as defined in Sect. 2.2.3. Also write \mathbf{e} for a Brownian excursion conditioned to reach height 1 and $\hat{\mathbf{e}}$ for $b\sigma$ times a Brownian excursion conditioned to reach height $(b\sigma)^{-1}$ (where σ is the constant from Proposition 6.9). Then it is clear from Brownian scaling that $(\mathbf{T}_{\mathbf{e}}, d_{\mathbf{e}})$ is isometrically equivalent to $(\mathbf{T}_{\hat{\mathbf{e}}}, d_{\hat{\mathbf{e}}})$ (where as in Sect. 2.2.3, (\mathbf{T}_C, d_C) is the real tree with contour function C).

Therefore, to prove Theorem 1.1, it is enough to prove that if $(\mathbf{T}^{(t)}, d^{(t)})$ has the law of $(\frac{[0, L(T))}{\sim^*}, \frac{d^*}{t})$ under $\mathbb{P}^X(\cdot \mid |N_t| > 0)$ then

$$(\mathbf{T}^{(t)}, d^{(t)}) \xrightarrow{(d)} (\mathbf{T}_{\hat{\mathbf{e}}}, d_{\hat{\mathbf{e}}}) \tag{6.43}$$

as $n \rightarrow \infty$, with respect to the Gromov–Hausdorff topology.

We equip $(\mathbf{T}^{(t)}, d^{(t)})$ with the measure $\mu^{(t)}$, which is defined to be the push forward of uniform measure on $[0, L(T))$ under the equivalence relation \sim^* . We also equip $(\mathbf{T}_{\hat{\mathbf{e}}}, d_{\hat{\mathbf{e}}})$ with the measure $\mu_{\hat{\mathbf{e}}}$, where if \hat{t} is the length of $\hat{\mathbf{e}}$ and $(\mathbf{T}_{\hat{\mathbf{e}}}, d_{\hat{\mathbf{e}}}) = (\frac{[0, \hat{t})}{\sim}, d_{\hat{\mathbf{e}}})$ as described in Sect. 2.2.3, then $\mu_{\hat{\mathbf{e}}}$ is the push forward of uniform measure on $[0, \hat{t})$ under \sim . For the proof of Theorem 1.1 we then need the following two ingredients:

Lemma 6.30 *The family*

$$(\mathbf{T}^{(t)}, d^{(t)}, \mu^{(t)})_{t \geq 0}$$

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

Lemma 6.31 $(\mathbf{T}^{(t)}, d^{(t)}, \mu^{(t)})_t$ *converges to* $(\mathbf{T}_{\hat{e}}, d_{\hat{e}}, \mu_{\hat{e}})$ *as* $t \rightarrow \infty$, *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, Lemma 6.31 characterises subsequential limits with respect to the Gromov–Hausdorff–Prohorov metric. Thus we have the convergence in distribution

$$(\mathbf{T}^{(t)}, d^{(t)}, \mu^{(t)}) \xrightarrow[t \rightarrow \infty]{} (\mathbf{T}_{\hat{e}}, d_{\hat{e}}, \mu_{\hat{e}})$$

with respect to the Gromov–Hausdorff–Prohorov metric. This then implies (6.43) by Remark 6.27 (GHP convergence implies GH convergence). \square

All that remains therefore is to verify Lemmas 6.30 and 6.31. In the following, we write \mathbb{P}_t^x for the law of $(\mathbf{T}^{(t)}, d^{(t)}, \mu^{(t)})$ and $\mathbb{P}_{\hat{e}}$ for the law of $(\mathbf{T}_{\hat{e}}, d_{\hat{e}}, \mu_{\hat{e}})$.

Proof of Lemma 6.30 We need to show that for any $\varepsilon > 0$ there exists a relatively compact $\mathbb{K} \subset \mathbb{X}$ (with respect to the Gromov–Hausdorff–Prohorov metric) such that

$$\inf_{t \geq 0} \mathbb{P}_t^x((\mathbf{T}^{(t)}, d^{(t)}, \mu^{(t)}) \in \mathbb{K}) \geq 1 - \varepsilon.$$

To do this, fix $\varepsilon > 0$ and let $\mathbb{K}_{R,M}$ be the relatively compact subset of \mathbb{X} (by Lemma 6.29), defined by

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

We will prove that we can pick R and M large enough such that $\mathbb{P}_t^x((\mathbf{T}^{(t)}, d^{(t)}, \mu^{(t)}) \in \mathbb{K}_{R,M}) \geq 1 - \varepsilon$ for all $t \geq 0$. The key to the proof is the following claim:

Claim 6.32 *For* $M = M(\varepsilon)$ *and* $R = R(\varepsilon)$ *large enough, we have*

$$\mathbb{P}_t^x(\text{diam}(T^{(t)}) > 2R) \leq \varepsilon/2; \text{ and} \tag{6.44}$$

$$\mathbb{P}_t^x(\{\text{cannot cover } (\mathbf{T}^{(t)}, d^{(t)}) \text{ with } < \delta^{-4}M \text{ balls of radius } \delta\} \cap \{\text{diam}(\mathbf{T}^{(t)}) \leq 2R\}) \leq \delta\varepsilon/2 \tag{6.45}$$

for all $\delta > 0$ and $t \geq 0$.

With this in hand, since $\mu^{(t)}(\mathbf{T}^{(t)}) = 1$ for all $t \geq 0$, summing over $\delta = 2^{-k}$ concludes the proof of tightness.

To prove Claim 6.32, we will reformulate the probabilities in (6.44) and (6.45) as conditional probabilities under $\mathbb{P}^x(\cdot \mid |N_t| > 0)$. Recall that the law of $(\mathbf{T}^{(t)}, d^{(t)})$ under \mathbb{P}_t^x is that of $(\frac{[0, L(T)]}{\sim^*}, \frac{d^*}{t})$ under $\mathbb{P}^x(\cdot \mid |N_t| > 0)$. It is therefore immediate from the definition of d^* that

$$\mathbb{P}_t^x(\text{diam}(T^{(t)}) > 2R) \leq \mathbb{P}^x(|N_{Rt}| > 0 \mid |N_t| > 0)$$

and so by Theorem 1.4, we can take $R = R(\varepsilon)$ large enough that this is less than $\varepsilon/2$ for all t .

Next, we look at (6.45). Fixing R from the previous paragraph, and given $\delta > 0$, we divide $[0, 2Rt]$ into intervals of length $t\delta/2 =: bt^{\delta}$ (so there are $4R/\delta$ of them.) We will show that on the event

$$\left\{ \text{diam}\left(\frac{[0, L(T)]}{\sim^*}, \frac{d^*}{t}\right) \leq 2R \right\} \cap \bigcap_{j=0}^{(4R/\delta)-1} \{|S_j| \leq M(4R\delta^3)^{-1}\}, \tag{6.46}$$

where $S_j := \{u \in N_{jb^{\delta}} : \exists w \in N_{(j+1)b^{\delta}} \text{ with } u \prec w\}$, we can cover $(\frac{[0, L(T)]}{\sim^*}, \frac{d^*}{t})$ with fewer than $\delta^{-4}M$ balls of radius δ . To do this, let S be the set of equivalence classes of $\frac{[0, L(T)]}{\sim^*}$ that contain a point s such that $(v_s, H_s) = (u, jb^{\delta})$ for some $u \in S_j$ and $j \in \{0, \dots, (4R/\delta) - 1\}$. On the event (6.46) we have $|S| \leq \delta^{-4}M$, and so it suffices to show that any element of $(\frac{[0, L(T)]}{\sim^*})$ has (d^*/t) distance less than δ from some point in S . However, this holds since any point of $\frac{[0, L(T)]}{\sim^*}$ corresponds to an individual $u \in T$, which must be in N_r for some $r \in [jb^{\delta}, (j+1)b^{\delta}]$, and $0 \leq j \leq (4R/\delta) - 1$. Such an individual must be a descendant of an individual $v \in S_{j-1}$ (or of \emptyset if $j = 0$), which in turn corresponds to a point of $\frac{[0, L(T)]}{\sim^*}$ contained in S . Because $u \in N_r$ and $v \in N_{r'}$ with $|r - r'| \leq \delta t$, the (d^*/t) distance between the corresponding points of $\frac{[0, L(T)]}{\sim^*}$ is less than δ , as required.

The above considerations mean that the probability on the left hand side of (6.45) is less than or equal to

$$\mathbb{P}^x\left(\bigcup_{j=0}^{(4R/\delta)-1} \left\{|S_j| > \frac{M}{4R\delta^3}\right\} \mid |N_t| > 0\right) \tag{6.47}$$

and hence, we just need to estimate (6.47). For any given j, t, δ we can bound, using Theorem 1.4 and the fact that $\sup_r \mathbb{P}^x(|N_r|) < \infty$,

$$\mathbb{P}^x(|S_j|) = \mathbb{P}^x(\mathbb{P}^x(|S_j| \mid \mathcal{F}_{jb^{\delta}})) = \mathbb{P}^x\left(\sum_{u \in N_{jb^{\delta}}} \mathbb{P}^{X_u(jb^{\delta})}(|N_{b^{\delta}}| > 0)\right) \leq C(t\delta)^{-1}$$

for a constant C that does not depend on j, t or δ (and may now change from line to line). Then by Theorem 1.4 again, conditioning on the event $|N_t| > 0$, we have $\mathbb{P}^x(|S_j| \mid |N_t| > 0) \leq C\delta^{-1}$, and so by conditional Markov's inequality,

$$\mathbb{P}^x \left(|S_j| > \frac{M}{4R\delta^3} \mid |N_t| > 0 \right) \leq \frac{4RC\delta^2}{M}.$$

Finally, by a union bound we can write

$$\begin{aligned} &\mathbb{P}^x \left(\bigcup_{j=0}^{(4R/\delta)-1} \{ |\{u \in N_{jb^{t,\delta}} : \exists w \in N_{(j+1)b^{t,\delta}} \text{ with } u \leq w\}| > \frac{M}{4R\delta^3} \} \mid |N_t| > 0 \right) \\ &\leq \frac{16R^2C\delta}{M}, \end{aligned}$$

and so choosing $M = M(R, \varepsilon) = M(\varepsilon)$ such that $16R^2CM^{-1} \leq \varepsilon/2$ then gives (6.45) and proves Claim 6.32. \square

Proof of Lemma 6.31 For this we would like to show that $(\mathbf{T}^{(t)}, d^{(t)}, \mu^{(t)})$ 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 6.28. Since convergence in the Gromov–Hausdorff–Prohorov sense implies convergence in the Gromov–Prohorov/Gromov–weak sense, Lemma 6.30 already shows that part (i) of the characterisation in Lemma 6.28 (relative compactness of the laws) is satisfied. Therefore, we need only show that for any polynomial $\Phi \in \Pi$, we have $\mathbb{P}_t^x[\Phi] \rightarrow \mathbb{P}_{\hat{e}}[\Phi]$ as $t \rightarrow \infty$. To do this, we use Proposition 6.13, and Proposition 6.21.

Fix a polynomial

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

where $\phi : [0, \infty)^{\binom{k}{2}} \rightarrow \mathbb{R}$ is continuous and bounded. Examining the definitions, we see that

$$\mathbb{P}_t^x[\Phi] = \mathcal{P}_k^x(\phi(D_t^H) \mid |N_t| > 0) \tag{6.48}$$

where \mathcal{P}_k^x and the matrix D_t^H are defined just before Proposition 6.21.⁷ Using this proposition, and also recalling the definition of the matrix D_t^S , we see that since ϕ is continuous and bounded,

$$\mathcal{P}_k^x(\phi(D_t^H) \mid |N_t| > 0) - \mathcal{P}_k^x(\phi(bD_t^S) \mid |N_t| > 0) \rightarrow 0 \tag{6.49}$$

as $t \rightarrow \infty$. Let us suppose for now that conditioning on $\{|N_t| > 0\}$ is comparable to conditioning on the event $A_t := \{\sup_t \mathbf{S}_t \geq b^{-1}t\}$, in the sense that we have

$$\mathcal{P}_k^x(\phi(bD_t^S) \mid |N_t| > 0) - \mathcal{P}_k^x(\phi(bD_t^S) \mid A_t) \rightarrow 0 \tag{6.50}$$

⁷ Interpreting $\phi(D_t^H) = \phi(((D_t^H)_{ij})_{1 \leq i < j \leq k})$.

as $t \rightarrow \infty$. Then since by (6.15) and Remark 6.10 we have

$$\mathcal{P}_k^x(\phi(bD_t^S) \mid \sup_t \mathbf{S}_t \geq b^{-1}t) \rightarrow \mathbb{P}_{\hat{\mathbf{e}}}[\Phi]$$

as $t \rightarrow \infty$, we can combine (6.48), (6.49) and (6.50) to conclude that $\mathbb{P}_t^x[\Phi] \rightarrow \mathbb{P}_{\hat{\mathbf{e}}}[\Phi]$ as desired.

Therefore, we are left to justify (6.50). In fact, it is enough (because ϕ is bounded by assumption) to show that

$$\mathbb{P}^x(A_t \mid |N_t| > 0) \rightarrow 1 \text{ and } \mathbb{P}^x(|N_t| > 0 \mid A_t) \rightarrow 1 \quad (6.51)$$

as $t \rightarrow \infty$. For the first statement, observe that

$$\mathbb{P}^x(A_t \mid |N_t| > 0) \geq \mathbb{P}^x(A_t \mid |N_{t(1+\delta)}| > 0) \frac{\mathbb{P}^x(|N_{t(1+\delta)}| > 0)}{\mathbb{P}^x(|N_t| > 0)}$$

for any $\delta > 0$, where for any $\delta > 0$ the first term in the product on the right hand side converges to 1 as $t \rightarrow \infty$ by Proposition 6.17, and the second converges to $(1 + \delta)^{-1}$ as $t \rightarrow \infty$ by Theorem 1.4. To finish showing (6.51), it is therefore enough to show that $\mathbb{P}^x(A_t) \sim \mathbb{P}^x(|N_t| > 0)$ as $t \rightarrow \infty$. For this, we observe that we have an asymptotic for $\mathbb{P}^x(|N_t| > 0)$ from Theorem 1.4, and also that Proposition 6.12 allows us to compute an asymptotic for $\mathbb{P}^x(A_t)$, just as in for example [36, Section 1.4, p. 263]. It is easy to check that these asymptotics agree, and hence the proof is complete. \square

Acknowledgements I would particularly like to thank Nathanaël Berestycki, for suggesting this problem and for many helpful discussions. I am also grateful to the anonymous referee for numerous useful comments and suggestions.

References

1. Abraham, R., Delmas, J.-F., Hoscheit, P.: A note on the Gromov–Hausdorff–Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.* **18**(14), 21 (2013)
2. Aldous, D.: The continuum random tree. I. *Ann. Probab.* **19**(1), 1–28 (1991)
3. Aldous, D.: The continuum random tree. III. *Ann. Probab.* **21**(1), 248–289 (1993)
4. Asmussen, S., Hering, H.: *Branching Processes*, Volume 3 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc, Boston (1983)
5. Bañuelos, R.: Intrinsic ultracontractivity and eigenfunction estimates for Schrödinger operators. *J. Funct. Anal.* **100**(1), 181–206 (1991)
6. Berestycki, J., Berestycki, N., Schweinsberg, J.: Survival of near-critical branching Brownian motion. *J. Stat. Phys.* **143**(5), 833–854 (2011)
7. Berestycki, J., Berestycki, N., Schweinsberg, J.: The genealogy of branching Brownian motion with absorption. *Ann. Probab.* **41**(2), 527–618 (2013)
8. Berestycki, J., Berestycki, N., Schweinsberg, J.: Critical branching Brownian motion with absorption: survival probability. *Probab. Theory Relat. Fields* **160**(3–4), 489–520 (2014)
9. Berestycki, J., Berestycki, N., Schweinsberg, J.: Critical branching Brownian motion with absorption: particle configurations. *Ann. Inst. Henri Poincaré Probab. Stat.* **51**(4), 1215–1250 (2015)
10. Berestycki, N.: *Recent Progress in Coalescent Theory*, Volume 16 of *Ensaïos Matemáticos [Mathematical Surveys]*. Sociedade Brasileira de Matemática, Rio de Janeiro (2009)

11. Biggins, J.D.: Martingale convergence in the branching random walk. *J. Appl. Probab.* **14**(1), 25–37 (1977)
12. Bolthausen, E., Sznitman, A.-S.: On Ruelle's probability cascades and an abstract cavity method. *Commun. Math. Phys.* **197**(2), 247–276 (1998)
13. Brunet, É., Derrida, B., Mueller, A.H., Munier, S.: Noisy traveling waves: effect of selection on genealogies. *Europhys. Lett.* **76**(1), 1–7 (2006)
14. Brunet, É., Derrida, B., Mueller, A.H., Munier, S.: Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys. Rev. E* (3) **76**(4), 041104, 20 (2007)
15. Burago, D., Burago, Y., Ivanov, S.: *A Course in Metric Geometry*, Volume 33 of Graduate Studies in Mathematics. American Mathematical Society, Providence (2001)
16. Chauvin, B., Rouault, A.: KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees. *Probab. Theory Relat. Fields* **80**(2), 299–314 (1988)
17. Davies, E.B.: *Heat Kernels and Spectral Theory*, Volume 92 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge (1989)
18. Davies, E.B., Simon, B.: Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. *J. Funct. Anal.* **59**(2), 335–395 (1984)
19. Duquesne, T., Le Gall, J.-F.: Random trees, Lévy processes and spatial branching processes. *Astérisque* **281**, vi+147 (2002)
20. Engländer, J., Kyprianou, A.E.: Local extinction versus local exponential growth for spatial branching processes. *Ann. Probab.* **32**(1A), 78–99 (2004)
21. Evans, L.C.: *Partial Differential Equations*, Volume 19 of Graduate Studies in Mathematics, 2nd edn. American Mathematical Society, Providence (2010)
22. Gilbarg, D., Trudinger, N.S.: *Elliptic Partial Differential Equations of Second Order*, volume 224 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 2nd edn. Springer, Berlin (1983)
23. Greven, A., Pfaffelhuber, P., Winter, A.: Convergence in distribution of random metric measure spaces (Λ -coalescent measure trees). *Probab. Theory Relat. Fields* **145**(1–2), 285–322 (2009)
24. Hardy, R., Harris, S.C.: A spine approach to branching diffusions with applications to L^p -convergence of martingales. In: *Séminaire de Probabilités XLII*, volume 1979 of Lecture Notes in Math., pp 281–330. Springer, Berlin (2009)
25. Harris, J.W., Harris, S.C.: Survival probabilities for branching Brownian motion with absorption. *Electron. Commun. Probab.* **12**, 81–92 (2007)
26. Harris, S.C., Hesse, M., Kyprianou, A.E.: Branching Brownian motion in a strip: survival near criticality. *Ann. Probab.* **44**(1), 235–275 (2016)
27. Harris, S.C., Roberts, M.I.: A strong law of large numbers for branching processes: almost sure spine events. *Electron. Commun. Probab.* **19**(28), 6 (2014)
28. Harris, S.C., Roberts, M.I.: The many-to-few lemma and multiple spines. *Ann. Inst. Henri Poincaré Probab. Stat.* **53**(1), 226–242 (2017)
29. Hering, H.: Limit theorem for critical branching diffusion processes with absorbing barriers. *Math. Biosci.* **19**, 355–370 (1974)
30. Hering, H.: Multigroup branching diffusions. In: *Branching processes Conf.*, Saint Hippolyte, Que., 1976), volume 5 of *Advanced Probab. Related Topics*, pp. 177–217. Dekker, New York (1978)
31. Hering, H.: Uniform primitivity of semigroups generated by perturbed elliptic differential operators. *Math. Proc. Camb. Philos. Soc.* **83**(2), 261–268 (1978)
32. Jacod, J., Shiryaev, A.N.: *Limit Theorems for stochastic processes*. volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Berlin (1987)
33. Kesten, H.: Branching Brownian motion with absorption. *Stoch. Processes Appl.* **7**(1), 9–47 (1978)
34. Kolmogorov, A.N.: Zur Lösung einer biologischen Aufgabe. *Izv. Nauč. Issled. Inst. Mat. i Meh. Tomsk. Gosudarstv. Univ.* **2**, 1–6 (1938)
35. Kyprianou, A.E.: Travelling wave solutions to the K-P-P equation: alternatives to Simon Harris' probabilistic analysis. *Ann. Inst. Henri Poincaré Probab. Stat.* **40**(1), 53–72 (2004)
36. Le Gall, J.-F.: Random trees and applications. *Probab. Surv.* **2**, 245–311 (2005)
37. Lyons, R.: A simple path to Biggins' martingale convergence for branching random walk. In: *Classical and modern branching processes* (Minneapolis, MN, 1994), volume 84 of *IMA Vol. Math. Appl.*, pp. 217–221. Springer, New York (1997)

38. McKean, H.P.: Application of Brownian motion to the equation of Kolmogorov–Petrovskii–Piskunov. *Comm. Pure Appl. Math.* **28**(3), 323–331 (1975)
39. Miermont, G.: Invariance principles for spatial multitype Galton–Watson trees. *Ann. Inst. Henri Poincaré Probab. Stat.* **44**(6), 1128–1161 (2008)
40. Miermont, G.: Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér. (4)* **42**(5), 725–781 (2009)
41. Pinsky, R.G.: On the convergence of diffusion processes conditioned to remain in a bounded region for large time to limiting positive recurrent diffusion processes. *Ann. Probab.* **13**(2), 363–378 (1985)
42. Revuz, D., Yor, M.: *Continuous Martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, 3rd edn. Springer, Berlin (1999)
43. Roberts, M. I.: Spine changes of measure and branching diffusions. PhD thesis, University of Bath. <http://people.bath.ac.uk/mir20/thesis.pdf> (2010)
44. Rudin, W.: *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics, 3rd edn. McGraw-Hill Book Co., New York–Auckland–Düsseldorf (1976)
45. Sevast'yanov, B.A.: Branching stochastic processes for particles diffusing in a bounded domain with absorbing boundaries. *Teor. Veroyatnost. i Primenen.* **3**, 121–136 (1958)
46. Watanabe, S.: On the branching process for Brownian particles with an absorbing boundary. *J. Math. Kyoto Univ.* **4**, 385–398 (1965)