



# Long Paths in First Passage Percolation on the Complete Graph II. Global Branching Dynamics

Maren Eckhoff<sup>1</sup> · Jesse Goodman<sup>2</sup> · Remco van der Hofstad<sup>3</sup> · Francesca R. Nardi<sup>3</sup>

Received: 15 November 2019 / Accepted: 2 June 2020 / Published online: 6 August 2020  
© The Author(s) 2020

## Abstract

We study the random geometry of first passage percolation on the complete graph equipped with independent and identically distributed positive edge weights. We consider the case where the lower extreme values of the edge weights are highly separated. This model exhibits strong disorder and a crossover between local and global scales. Local neighborhoods are related to invasion percolation that display self-organised criticality. Globally, the edges with relevant edge weights form a barely supercritical Erdős–Rényi random graph that can be described by branching processes. This near-critical behaviour gives rise to optimal paths that are considerably longer than logarithmic in the number of vertices, interpolating between random graph and minimal spanning tree path lengths. Crucial to our approach is the quantification of the extreme-value behavior of small edge weights in terms of a sequence of parameters  $(s_n)_{n \geq 1}$  that characterises the different universality classes for first passage percolation on the complete graph. We investigate the case where  $s_n \rightarrow \infty$  with  $s_n = o(n^{1/3})$ , which corresponds to the barely supercritical setting. We identify the scaling limit of the weight of the optimal path between two vertices, and we prove that the number of edges in this path obeys a central limit theorem with mean approximately  $s_n \log(n/s_n^3)$  and variance  $s_n^2 \log(n/s_n^3)$ . Remarkably, our proof also applies to  $n$ -dependent edge weights of the form  $E^{s_n}$ , where  $E$  is an exponential random variable with mean 1, thus settling a conjecture of Bhamidi et al. (Weak disorder asymptotics in the stochastic meanfield model of distance. *Ann Appl Probab* 22(1):29–69, 2012). The proof relies on a decomposition of the smallest-weight tree into an initial part following invasion percolation dynamics, and a main part following branching process dynamics. The initial part has been studied in Eckhoff et al. (Long paths in first passage percolation on the complete graph I. Local PWIT dynamics. *Electron. J. Probab.* 25:1–45, 2020. <https://doi.org/10.1214/20-EJP484>); the current paper focuses on the global branching dynamics.

**Keywords** First-passage percolation · Stochastic mean-field model of distance · Strong disorder · Continuous-time branching processes · Invasion percolation

---

Communicated by Hal Tasaki.

Extended author information available on the last page of the article

## Contents

1	Model and Summary of Results	366
1.1	First Passage Percolation with Regularly-Varying Edge Weights	367
1.2	First Passage Percolation with $n$ -Dependent Edge Weights	369
2	The Universal Picture	370
2.1	Description of the Class of Edge Weights to Which Our Results Apply	370
2.2	Discussion of Our Results	373
2.2.1	The Universality Class in Terms of $s_n$	373
2.2.2	First Passage Percolation on Random Graphs	375
2.2.3	Extremal Functionals for FPP on the Complete Graph	375
3	Detailed Results, Overview and Classes of Edge Weights	376
3.1	FPP Exploration Processes	376
3.1.1	One-Source Exploration Process	376
3.1.2	Two-Source Exploration Process	377
3.2	Coupling FPP on $K_n$ to FPP on the Poisson-Weighted Infinite Tree	378
3.2.1	Order Statistics of Exponentials and Poisson Processes	379
3.2.2	The Poisson-Weighted Infinite Tree	379
3.2.3	Coupling One-Source Exploration to the PWIT	380
3.2.4	Coupling Two-Source Exploration to the PWIT	381
3.3	FPP on the PWIT as a CTBP	382
3.3.1	FPP on the PWIT as a CTBP: One-Vertex Characteristics	382
3.3.2	FPP on the PWIT as a CTBP: Two-Vertex Characteristics	384
3.4	CTBP Growth and the Need for Freezing: Medium Time Scales	385
3.4.1	Frozen FPP Exploration Process	386
3.4.2	FPP Exploration Process from Two Sources with Freezing and Collisions	388
3.5	Coupling FPP on $K_n$ from Two Sources to a CTBP	390
3.5.1	The Final Coupling Including Freezing	390
3.5.2	A Cox Process for the Collisions	392
3.6	The Collision Edge and Its Properties: Long Time Scales	393
3.7	Thinning and Completion of the Proof	395
4	Scaling Properties of $f_n$ and $\mu_n$	397
4.1	Growth and Density Bounds for $f_n$ and $\mu_n$	397
4.2	Analysis of Specific Edge-Weight Distributions	398
4.3	Equivalence of Conditions: Proof of Lemma 2.7	400
5	Coupling $K_n$ and the PWIT	401
5.1	Exploration Processes and the Definition of the Coupling	402
5.2	Minimal-Rule Exploration Processes	403
5.3	FPP and the Two Smallest-Weight Trees: Proof of Theorem 3.22	404
5.4	$\mathcal{B}$ and $\mathcal{S}$ as Exploration Processes: Proof of Theorem 3.26	406
5.5	Coupling and Cox Processes: Proof of Theorem 3.27	408
6	Branching Processes and Random Walks	411
6.1	Continuous-Time Branching Processes and Random Walks	411
6.2	Random Walk Convergence	413
6.3	Convergence of Random Walks: Proof of Theorem 6.3	414
6.4	Means of One-Vertex Characteristics: Proof of Theorem 3.13	417
7	Continuous-Time Branching Processes: Two-Vertex Characteristics	420
7.1	Truncated Two-Vertex Characteristic: Proof of Theorem 3.15	421
7.2	The Effect of Truncation: Proof of Theorem 3.16	423
8	First Points of Cox Processes: Proof of Theorem 3.29	429
8.1	Preparations: Weak Convergence of Cox Processes	429
8.2	Convergence of First Point of Cox Processes: Proof of Theorem 3.29	432
9	Moment Estimates and the Cluster After Unfreezing	434
9.1	Preparations: Frozen Intensity Measures	434
9.2	A First Moment Estimate: Proof of Lemma 3.32 (a)	435
9.3	Second Moment Estimates: Proof of Theorem 3.31	436
9.4	No Collisions from the Frozen Cluster: Proof of Lemma 3.32 (b)	442
9.5	Properties of the Freezing Time and Frozen Cluster: Proof of Theorem 3.18	445
	References	446

# 1 Model and Summary of Results

In this paper, we study first passage percolation on the complete graph equipped with independent and identically distributed positive and continuous edge weights. In contrast to earlier work [11,12,16,20,27], we consider the case where the extreme values of the edge weights are highly separated.

We start by introducing first passage percolation (FPP). Given a graph  $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$ , let  $(Y_e^{(\mathcal{G})})_{e \in E(\mathcal{G})}$  denote a collection of positive edge weights. Thinking of  $Y_e^{(\mathcal{G})}$  as the cost of crossing an edge  $e$ , we can define a metric on  $V(\mathcal{G})$  by setting

$$d_{\mathcal{G}, Y(\mathcal{G})}(i, j) = \inf_{\pi: i \rightarrow j} \sum_{e \in \pi} Y_e^{(\mathcal{G})}, \quad (1.1)$$

where the infimum is over all paths  $\pi$  in  $\mathcal{G}$  that join  $i$  to  $j$ , and  $Y^{(\mathcal{G})}$  represents the edge weights  $(Y_e^{(\mathcal{G})})_{e \in E(\mathcal{G})}$ . We will always assume that the infimum in (1.1) is attained uniquely, by some (finite) path  $\pi_{i,j}$ . We are interested in the situation where the edge weights  $Y_e^{(\mathcal{G})}$  are *random*, so that  $d_{\mathcal{G}, Y(\mathcal{G})}$  is a random metric. In particular, when the graph  $\mathcal{G}$  is very large, with  $|V(\mathcal{G})| = n$  say, we wish to understand the scaling behavior of the following quantities for fixed  $i, j \in V(\mathcal{G})$ :

- (a) The *distance*  $W_n = d_{\mathcal{G}, Y(\mathcal{G})}(i, j)$ —the total edge cost of the optimal path  $\pi_{i,j}$ ;
- (b) The *hopcount*  $H_n$ —the number of edges in the optimal path  $\pi_{i,j}$ ;
- (c) The *topological structure*—the shape of the random neighborhood of a point.

In this paper, we consider FPP on the complete graph, which acts as a mean-field model for FPP on finite graphs. In [11], the question was raised what the *universality classes* are for this model. We bring the discussion substantially further by describing a way to distinguish several universality classes and by identifying the limiting behavior of first passage percolation in one of these classes. The cost regime introduced in (1.1) uses the information from all edges along the path and is known as the *weak disorder* regime. By contrast, in the *strong disorder* regime the cost of a path  $\pi$  is given by  $\max_{e \in \pi} Y_e^{(\mathcal{G})}$ . We establish a firm connection between the weak and strong disorder regimes in first passage percolation. Interestingly, this connection also establishes a strong relation to invasion percolation (IP) on the Poisson-weighted infinite tree (PWIT), which is the local limit of IP on the complete graph, and also arises in the context of the minimal spanning tree on the complete graph (see e.g. [1]).

Our main interest is in the case  $\mathcal{G} = K_n$ , the complete graph on  $n$  vertices  $V(K_n) = [n] := \{1, \dots, n\}$ , equipped with independent and identically distributed (i.i.d.) edge weights  $(Y_e^{(K_n)})_{e \in E(K_n)}$ . We write  $Y$  for a random variable with  $Y \stackrel{d}{=} Y_e^{(\mathcal{G})}$ , and assume that the distribution function  $F_Y$  of  $Y$  is continuous. For definiteness, we study the optimal path  $\pi_{1,2}$  between vertices 1 and 2. First, we introduce some general notation:

**Notation.** All limits in this paper are taken as  $n$  tends to infinity unless stated otherwise. A sequence of events  $(\mathcal{A}_n)_n$  happens *with high probability (whp)* if  $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$ . For random variables  $(X_n)_n$ ,  $X$ , we write  $X_n \xrightarrow{d} X$ ,  $X_n \xrightarrow{\mathbb{P}} X$  and  $X_n \xrightarrow{a.s.} X$  to denote convergence in distribution, in probability and almost surely, respectively. For real-valued sequences  $(a_n)_n$ ,  $(b_n)_n$ , we write  $a_n = O(b_n)$  if the sequence  $(a_n/b_n)_n$  is bounded;  $a_n = o(b_n)$  if  $a_n/b_n \rightarrow 0$ ;  $a_n = \Theta(b_n)$  if the sequences  $(a_n/b_n)_n$  and  $(b_n/a_n)_n$  are both bounded; and  $a_n \sim b_n$  if  $a_n/b_n \rightarrow 1$ . Similarly, for sequences  $(X_n)_n$ ,  $(Y_n)_n$  of random variables, we write  $X_n = O_{\mathbb{P}}(Y_n)$  if the sequence  $(X_n/Y_n)_n$  is tight;  $X_n = o_{\mathbb{P}}(Y_n)$  if  $X_n/Y_n \xrightarrow{\mathbb{P}} 0$ ; and  $X_n = \Theta_{\mathbb{P}}(Y_n)$  if the sequences  $(X_n/Y_n)_n$  and  $(Y_n/X_n)_n$  are both tight. We denote by  $\lfloor x \rfloor$  the greatest integer not exceeding  $x$ . Moreover,  $E$  denotes an exponentially distributed random variable

with mean 1. We often need to refer to results from [21], and we will write, e.g., [Part I, Lemma 2.18] for [21, Lemma 2.18].

For a brief overview of notation particular to this paper, see p. 81.

## 1.1 First Passage Percolation with Regularly-Varying Edge Weights

In this paper, we will consider edge-weight distributions with a heavy tail near 0, in the sense that the distribution function  $F_Y(y)$  decays slowly to 0 as  $y \downarrow 0$ . It will prove more convenient to express this notion in terms of inverse  $F_Y^{-1}(u)$ , since we can write

$$Y_e^{(K_n)} \stackrel{d}{=} F_Y^{-1}(U), \quad (1.2)$$

where  $U$  is uniformly distributed on  $[0, 1]$ . Expressed in terms of  $F_Y^{-1}$ , saying that the edge-weight distribution is heavy-tailed near 0 means that  $F_Y^{-1}(u)$  decays rapidly to 0 as  $u \downarrow 0$ . We will quantify this notion in terms of the logarithmic derivative of  $F_Y^{-1}$ , which will become large as  $u \downarrow 0$ .

In this section, we will assume that

$$u \frac{d}{du} \log F_Y^{-1}(u) = u^{-\alpha} L(1/u), \quad (1.3)$$

where  $\alpha \geq 0$  and  $t \mapsto L(t)$  is slowly varying as  $t \rightarrow \infty$ . That is, for all  $a > 0$ ,  $\lim_{t \rightarrow \infty} L(at)/L(t) = 1$ . In other words, we assume that  $u \mapsto u \frac{d}{du} \log F_Y^{-1}(u) = \frac{d}{d(\log u)} \log F_Y^{-1}(u)$  is regularly varying as  $u \downarrow 0$ . Recall that a function  $\tilde{L} : (0, \infty) \rightarrow (0, \infty)$  is called regularly varying as  $u \downarrow 0$  if  $\lim_{u \downarrow 0} \tilde{L}(au)/\tilde{L}(u)$  is finite but nonzero for all  $a > 0$ .

Define a sequence  $s_n$  by setting  $u = 1/n$ :

$$s_n = u \frac{d}{du} \log F_Y^{-1}(u) \Big|_{u=1/n} = \frac{(F_Y^{-1})'(1/n)}{n F_Y^{-1}(1/n)}. \quad (1.4)$$

The asymptotics of the sequence  $(s_n)_n$  quantify how heavy-tailed the edge-weight distribution is. For instance, an identically constant sequence, say  $s_n = s$ , corresponds to a pure power law  $F_Y(y) = y^{1/s}$ ,  $F_Y^{-1}(u) = u^s$ ; larger values of  $s$  correspond to heavier-tailed distributions.

In this paper, we are interested in the regime where  $s_n \rightarrow \infty$ , which corresponds to a very heavy-tailed distribution function  $F_Y(y)$  that decays to 0 slower than any power of  $y$ , as  $y \downarrow 0$ .

To describe our scaling results, define

$$u_n(x) = F_Y^{-1}(x/n). \quad (1.5)$$

Then, for i.i.d. random variables  $(Y_i)_{i \in \mathbb{N}}$  with distribution function  $F_Y$ ,

$$\mathbb{P}\left(\min_{i \in [n]} Y_i \leq u_n(x)\right) \rightarrow 1 - e^{-x}. \quad (1.6)$$

In view of (1.6), the family  $(u_n(x))_{x \in (0, \infty)}$  are the *characteristic values* for  $\min_{i \in [n]} Y_i$ . See [22] for a detailed discussion of extreme value theory.

**Theorem 1.1** (Weight and hopcount—regularly-varying logarithmic derivatives) *Suppose that the edge weights  $(Y_e^{(K_n)})_{e \in E(K_n)}$  follow an  $n$ -independent distribution  $F_Y$  that satisfies (1.3). If the sequence  $(s_n)_n$  from (1.4) satisfies  $s_n / \log \log n \rightarrow \infty$  and  $s_n = o(n^{1/3})$ ,*

then there exist sequences  $(\lambda_n)_n$  and  $(\phi_n)_n$  with  $\phi_n/s_n \rightarrow 1$ ,  $\lambda_n u_n(1) \rightarrow e^{-\gamma}$ , where  $\gamma$  is Euler's constant, such that

$$nF_Y\left(W_n - \frac{1}{\lambda_n} \log(n/s_n^3)\right) \xrightarrow{d} M^{(1)} \vee M^{(2)}, \quad (1.7)$$

$$\frac{H_n - \phi_n \log(n/s_n^3)}{\sqrt{s_n^2 \log(n/s_n^3)}} \xrightarrow{d} Z. \quad (1.8)$$

Here  $Z$  is standard normal, and  $M^{(1)}, M^{(2)}$  are i.i.d. random variables for which  $\mathbb{P}(M^{(j)} \leq x)$  is the survival probability of a Poisson Galton–Watson branching process with mean  $x$ .

Let us discuss the result in Theorem 1.1 in more detail. Under the hypotheses of Theorem 1.1,  $u_n(x)$  varies heavily in  $x$  in the sense in that  $u_n(x + \delta)/u_n(x) \rightarrow \infty$  for every  $x, \delta > 0$ . Consequently, the extreme values are widely separated, which is characteristic of the *strong disorder regime*.

We see in (1.7) that  $W_n - \frac{1}{\lambda_n} \log(n/s_n^3) \approx u_n(M^{(1)} \vee M^{(2)})$ , which means that the weight of the smallest-weight path has a deterministic part  $\frac{1}{\lambda_n} \log(n/s_n^3)$ , while its random fluctuations are of the same order of magnitude as some of the typical values for the minimal edge weight adjacent to vertices 1 and 2. For  $j \in \{1, 2\}$ , one can think of  $M^{(j)}$  as the time needed to “escape” from the local neighborhood of vertex  $j$ . The sequences  $(\lambda_n)_n$  and  $(\phi_n)_n$  will be identified in (3.17)–(3.18), subject to slightly stronger assumptions.

The optimal paths in Theorem 1.1 are *long paths* because the asymptotic mean of the path length  $H_n$  in (1.8) is of larger order than  $\log n$ , the path length that arises in many random graph contexts. See Sect. 2.2 for a comprehensive literature overview. The following example collects some edge-weight distributions that are covered by Theorem 1.1:

**Example 1.2** (Examples of weight distributions)

- (a) Let  $a, \gamma > 0$ . Take  $Y_e^{(K_n)} \stackrel{d}{=} \exp(-aE^\gamma)$ , for which  $\log F_Y^{-1}(u) = -a(\log(1/u))^\gamma$  and

$$s_n = a\gamma(\log n)^{\gamma-1}. \quad (1.9)$$

The hypotheses of Theorem 1.1 are satisfied whenever  $\gamma > 1$ .

- (b) Let  $a, \gamma > 0$ . Take  $Y_e^{(K_n)} \stackrel{d}{=} U^{a(\log(1+\log(1/U)))^\gamma}$ , for which  $\log F_Y^{-1}(u) = a \log u (\log(1 + \log(1/u)))^\gamma$  and

$$s_n = a(\log(1 + \log n))^\gamma + a\gamma \frac{\log n}{1 + \log n} (\log(1 + \log n))^{\gamma-1}. \quad (1.10)$$

We note that  $s_n \sim a(\log \log n)^\gamma$  as  $n \rightarrow \infty$ . The hypotheses of Theorem 1.1 are satisfied whenever  $\gamma > 1$ . We shall see, however, that the conclusions of Theorem 1.1 also hold when  $0 < \gamma \leq 1$ ; see Sect. 2.1 and Lemma 4.8.

- (c) Let  $a, \beta > 0$ . Take  $Y_e^{(K_n)} \stackrel{d}{=} \exp(-aU^{-\beta}/\beta)$ , for which  $\log F_Y^{-1}(u) = -au^{-\beta}/\beta$  and

$$s_n = an^\beta. \quad (1.11)$$

The hypotheses of Theorem 1.1 are satisfied when  $0 < \beta < 1/3$ . When  $\beta \geq 1/3$ , we conjecture that the hopcount scaling (1.8) fails; see the discussion in Sect. 2.2. An analogue of the weight convergence (1.7) holds in a modified form; see [Part I, Theorem 1.1 and Example 1.4 (c)].

Notice that every sequence  $(s_n)_n$  of the form  $s_n = n^\alpha L(n)$ , for  $\alpha \geq 0$  and  $L$  slowly varying at infinity, can be obtained from a distribution by taking  $\log F_Y^{-1}(u) = \int u^{-1-\alpha} L(1/u) du$ , i.e., the indefinite integral of the function  $u \mapsto u^{-1-\alpha} L(1/u)$ . In Sect. 2.1 we will weaken the requirement  $s_n / \log \log n \rightarrow \infty$  to the requirement  $s_n \rightarrow \infty$  subject to an additional regularity assumption.

## 1.2 First Passage Percolation with $n$ -Dependent Edge Weights

In Theorem 1.1, we started with a fixed edge-weight distribution and extracted a specific sequence  $(s_n)_n$ . For an essentially arbitrary distribution [subject to the relatively modest regular variation assumption in (1.3)], its FPP properties are fully encoded, at least for the purposes of the conclusions of Theorem 1.1, by the scaling properties of this sequence  $(s_n)_n$ . Thus, Theorem 1.1 shows the common behaviour of a *universality class* of edge-weight distributions, and shows that this universality class is described in terms of a sequence of real numbers  $(s_n)_n$  and its scaling behaviour.

In this section, we reverse this setup. We take as input a sequence  $(s_n)_n$  and consider the  $n$ -dependent edge-weight distribution

$$Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}, \quad (1.12)$$

where  $E$  is exponentially distributed with mean 1. (For legibility, our notation will not indicate the implicit dependence of  $Y_e^{(K_n)}$  on  $n$ .) Then the conclusions of Theorem 1.1 hold verbatim:

**Theorem 1.3** (Weight and hopcount— $n$ -dependent edge weights) *Let  $Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}$ , where  $(s_n)_n$  is a positive sequence with  $s_n \rightarrow \infty$ ,  $s_n = o(n^{1/3})$ . Then*

$$n \left( W_n - \frac{1}{n^{s_n} \Gamma(1 + 1/s_n)^{s_n}} \log(n/s_n^3) \right)^{1/s_n} \xrightarrow{d} M^{(1)} \vee M^{(2)}, \quad (1.13)$$

and

$$\frac{H_n - s_n \log(n/s_n^3)}{\sqrt{s_n^2 \log(n/s_n^3)}} \xrightarrow{d} Z, \quad (1.14)$$

where  $Z$  is standard normal and  $M^{(1)}, M^{(2)}$  are i.i.d. random variables for which  $\mathbb{P}(M^{(j)} \leq x)$  is the survival probability of a Poisson Galton–Watson branching process with mean  $x$ .

We note that Theorem 1.3 resolves a conjecture in [11]. This problem is closely related to the problem of *strong disorder* on the complete graph, and has attracted considerable attention in the physics literature [18,23,33]. The convergence in (1.13) was proved in [Part I, Theorem 1.5 (a)] without the subtraction of the term  $\frac{1}{n^{s_n} \Gamma(1+1/s_n)^{s_n}} \log(n/s_n^3)$  in the argument, and under the stronger assumption that  $s_n / \log \log n \rightarrow \infty$ .<sup>1</sup>

The edge-weight distribution in Theorem 1.3 allows for a simpler intuitive explanation of (1.7), while the convergence (1.14) verifies the heuristics for the strong disorder regime in [11, Sect. 1.4]. See Remark 4.6 for a discussion of the relation between these two results. As mentioned in Sect. 1.1, strong disorder here refers to the fact that when  $s_n \rightarrow \infty$  the values of the random weights  $E_e^{s_n}$  depend strongly on the disorder  $(E_e)_{e \in E(G)}$ , making small

<sup>1</sup> [Part I, Theorem 1.5 (a)] in fact deals with powers of uniform random variables, but for this discussion that makes no difference.

values increasingly more, and large values increasingly less, favorable. Mathematically, the elementary limit

$$\lim_{s \rightarrow \infty} (x_1^s + x_2^s)^{1/s} = x_1 \vee x_2 \quad (1.15)$$

expresses the convergence of the  $\ell^s$  norm towards the  $\ell^\infty$  norm and establishes a relationship between the weak disorder regime and the strong disorder regime of FPP.

Remarkably, a similar argument actually also applies to Theorem 1.1, exemplifying that these settings are in the same universality class. Indeed, Theorem 1.3 shows that  $n$ -dependent distributions can be understood in the same framework as the  $n$ -independent distributions in Example 1.2. We next explain this comparison and generalize our results further by explaining the universal picture behind them.

## 2 The Universal Picture

In Sect. 2.1, we generalize the results in Theorems 1.1 and 1.3 to a larger class of edge weights and provide a common language that allows us to prove these results in one go. Having reached this higher level of abstraction, in Sect. 2.2, we will embed the results achieved here in the wider picture of universality classes of FPP, and provide conjectures or results how to describe all universality classes and what the scaling behaviour of each of them might be. Links to the relevant literature and existing results are provided. For a short guide to notation, see p. 81.

### 2.1 Description of the Class of Edge Weights to Which Our Results Apply

In this section, we describe a general framework containing both Theorem 1.1 as well as Theorem 1.3. This framework, which is in terms of i.i.d. exponential random variables, determines the precise conditions that the edge weights need to satisfy for the results in Theorems 1.1–1.3 to apply. Interestingly, due to the parametrization in terms of exponential random variables, this general framework also provides a clear link between the near-critical Erdős–Rényi random graph and our first passage percolation problem where the lower extremes of the edge-weight distribution are highly separated. Finally and conveniently, this framework allows us to prove these theorems simultaneously. In particular, both the  $n$ -independent edge weights in Theorem 1.1, as well as the  $n$ -dependent ones in Theorem 1.3, are key examples of the class of edge weights that we will study in this paper.

For fixed  $n$ , the edge weights  $(Y_e^{(K_n)})_{e \in E(K_n)}$  are independent for different  $e$ . However, there is no requirement that they are independent over  $n$ , and in fact in Sect. 5, we will produce  $Y_e^{(K_n)}$  using a fixed source of randomness not depending on  $n$ . Therefore, it will be useful to describe the randomness on the edge weights  $((Y_e^{(K_n)})_{e \in E(K_n)} : n \in \mathbb{N})$  uniformly across the sequence. It will be most useful to give this description in terms of exponential random variables. Fix independent exponential mean 1 variables  $(X_e^{(K_n)})_{e \in E(K_n)}$ , and define

$$Y_e^{(K_n)} = g(X_e^{(K_n)}), \quad (2.1)$$

where  $g : (0, \infty) \rightarrow (0, \infty)$  is a strictly increasing function. The relation between  $g$  and the distribution function  $F_Y$  is given by

$$F_Y(y) = 1 - e^{-g^{-1}(y)} \quad \text{and} \quad g(x) = F_Y^{-1}(1 - e^{-x}). \quad (2.2)$$

We define

$$f_n(x) = g(x/n) = F_Y^{-1}(1 - e^{-x/n}). \quad (2.3)$$

Let  $Y_1, \dots, Y_n$  be i.i.d. with  $Y_i = g(E_i)$  as in (2.1). Since  $g$  is increasing,

$$\min_{i \in [n]} Y_i = g\left(\min_{i \in [n]} E_i\right) \stackrel{d}{=} g(E/n) = f_n(E). \quad (2.4)$$

Because of this convenient relation between the edge weights  $Y_e^{(K_n)}$  and exponential random variables, we will express our hypotheses about the distribution of the edge weights in terms of conditions on the functions  $f_n(x)$  as  $n \rightarrow \infty$ .

Consider first the case  $Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}$  from Theorem 1.3. From (2.1), we have  $g(x) = g_n(x) = x^{s_n}$ , so that (2.3) yields

$$\text{for } Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}, \quad f_n(x) = (x/n)^{s_n} = f_n(1)x^{s_n}. \quad (2.5)$$

Thus, (2.4)–(2.5) show that the parameter  $s_n$  measures the *relative sensitivity* of  $\min_{i \in [n]} Y_i$  to fluctuations in the variable  $E$ . In general, we will have  $f_n(x) \approx f_n(1)x^{s_n}$  if  $x$  is appropriately close to 1 and  $s_n \approx f'_n(1)/f_n(1)$ . These observations motivate the following conditions on the functions  $(f_n)_n$ , which we will use to relate the distributions of the edge weights  $Y_e^{(K_n)}$ ,  $n \in \mathbb{N}$ , to a sequence  $(s_n)_n$ :

**Condition 2.1** (Scaling of  $f_n$ ) *For every  $x \geq 0$ ,*

$$\frac{f_n(x^{1/s_n})}{f_n(1)} \rightarrow x. \quad (2.6)$$

Even though we will rely on Condition 2.1 when  $s_n \rightarrow \infty$  and  $s_n = o(n^{1/3})$ , we strongly believe that the scaling of the sequence  $(s_n)_n$  actually *characterises* the universality classes, in the sense that the behaviour of  $H_n$  and  $W_n$  is similar for edge weights for  $(s_n)_n$  with similar scaling behaviour, and different for sequences that have different scaling. We elaborate on this in Sect. 2.2.1, where we identify eight different universality classes and the expected and/or proved results in them.

**Condition 2.2** (Density bound for small weights) *There exist  $\varepsilon_0 > 0$ ,  $\delta_0 \in (0, 1]$  and  $n_0 \in \mathbb{N}$  such that*

$$\varepsilon_0 s_n \leq \frac{x f'_n(x)}{f_n(x)} \leq s_n / \varepsilon_0, \quad \text{whenever } 1 - \delta_0 \leq x \leq 1 \quad \text{and } n \geq n_0. \quad (2.7)$$

**Condition 2.3** (Density bound for large weights)

(a) *For all  $R > 1$ , there exist  $\varepsilon > 0$  and  $n_0 \in \mathbb{N}$  such that for every  $1 \leq x \leq R$ , and  $n \geq n_0$ ,*

$$\frac{x f'_n(x)}{f_n(x)} \geq \varepsilon s_n. \quad (2.8)$$

(b) *For all  $C > 1$ , there exist  $\varepsilon > 0$  and  $n_0 \in \mathbb{N}$  such that (2.8) holds for every  $n \geq n_0$  and every  $x \geq 1$  satisfying  $f_n(x) \leq C f_n(1) \log n$ .*

Notice that Condition 2.1 implies that  $f_n(1) \sim u_n(1)$  [recall the definition of  $u_n(x)$  in (1.5)] whenever  $s_n = o(n)$ . Indeed, by (2.3) we can write  $u_n(1) = f_n(x_n^{1/s_n})$  for  $x_n = (-n \log(1 - 1/n))^{s_n}$ . Since  $s_n = o(n)$ , we have  $x_n = 1 - o(1)$  and the monotonicity of  $f_n$  implies that  $f_n(x_n^{1/s_n})/f_n(1) \rightarrow 1$ . We remark also that (1.6) remains valid if  $u_n(x)$  is replaced by  $f_n(x)$ .

We are now in a position to state our main theorem:



**Theorem 2.4** (Weight and hopcount—general edge weights) *Assume that Conditions 2.1–2.3 hold for a positive sequence  $(s_n)_n$  with  $s_n \rightarrow \infty$  and  $s_n = o(n^{1/3})$ . Then there exist sequences  $(\lambda_n)_n$  and  $(\phi_n)_n$  such that  $\phi_n/s_n \rightarrow 1$ ,  $\lambda_n f_n(1) \rightarrow e^{-\gamma}$ , where  $\gamma$  is Euler's constant, and*

$$f_n^{-1}\left(W_n - \frac{1}{\lambda_n} \log(n/s_n^3)\right) \xrightarrow{d} M^{(1)} \vee M^{(2)}, \quad (2.9)$$

$$\frac{H_n - \phi_n \log(n/s_n^3)}{\sqrt{s_n^2 \log(n/s_n^3)}} \xrightarrow{d} Z, \quad (2.10)$$

where  $Z$  is standard normal, and  $M^{(1)}, M^{(2)}$  are i.i.d. random variables for which  $\mathbb{P}(M^{(i)} \leq x)$  is the survival probability of a Poisson Galton–Watson branching process with mean  $x$ . The convergences in (2.9)–(2.10) hold jointly and the limiting random variables are independent.

The sequences  $(\lambda_n)_n$  and  $(\phi_n)_n$  are identified in (3.17)–(3.18), subject to the additional Condition 2.6. The proof of Theorem 2.4 is given in Sect. 3.7.

**Relation between Theorem 2.4 and Theorems 1.1 and 1.3.** Theorems 1.1 and 1.3 follow from Theorem 2.4: in the case  $Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}$  from Theorem 1.3, (2.6)–(2.8) hold identically with  $\varepsilon_0 = \varepsilon = 1$  and we explicitly compute  $\lambda_n = n^{s_n} \Gamma(1 + 1/s_n)^{s_n}$  and  $\phi_n = s_n$  in Example 6.1. We will prove in Lemma 4.8 that the distributions in Theorem 1.1 satisfy the assumptions of Theorem 2.4. The convergence (1.7) in Theorem 1.1 is equivalent to (2.9) in Theorem 2.4 by the observation that, for any non-negative random variables  $(T_n)_n$  and  $\mathcal{M}$ ,

$$nF_Y(T_n) \rightarrow \mathcal{M} \iff f_n^{-1}(T_n) \rightarrow \mathcal{M}, \quad (2.11)$$

where the convergence is in distribution, in probability or almost surely; see e.g. [Part I, Lemma 5.5] for an example.

The following example describes a generalization of Theorem 1.3:

**Example 2.5** Let  $(s_n)_n$  be a positive sequence with  $s_n \rightarrow \infty$ ,  $s_n = o(n^{1/3})$ . Let  $Z$  be a positive-valued continuous random variable with distribution function  $G$  such that  $G'(z)$  exists and is continuous at  $z = 0$  with  $G'(0) > 0$  [with  $G'(0)$  interpreted as a right-hand derivative]. Take  $Y_e^{(K_n)} \stackrel{d}{=} Z^{s_n}$ , i.e.,  $F_Y(y) = G(y^{1/s_n})$ . Then Conditions 2.1–2.3 hold and Theorem 2.4 applies.

For instance, we can take  $Z$  to be a uniform distribution on an interval  $(0, b)$ , for any  $b > 0$ . We give a proof of this assertion in Lemma 4.8.

Condition 2.3 can be strengthened to the following condition that will be equivalent for our purposes:

**Condition 2.6** (Extended density bound) *There exist  $\varepsilon_0 > 0$  and  $n_0 \in \mathbb{N}$  such that*

$$\frac{xf'_n(x)}{f_n(x)} \geq \varepsilon_0 s_n \quad \text{for every } x \geq 1, n \geq n_0. \quad (2.12)$$

**Lemma 2.7** *It suffices to prove Theorem 2.4 assuming Conditions 2.1, 2.2 and 2.6.*

Lemma 2.7, which is proved in Sect. 4.3, reflects the fact that the upper tail of the edge-weight distribution does not substantially influence the first passage percolation problem.

Henceforth, except where otherwise noted, we will assume Conditions 2.1, 2.2 and 2.6. We will reserve the notation  $\varepsilon_0, \delta_0$  for some fixed choice of the constants in Conditions 2.2 and 2.6, with  $\varepsilon_0$  chosen small enough to satisfy both conditions.

## 2.2 Discussion of Our Results

In this section we discuss our results and state open problems.

### 2.2.1 The Universality Class in Terms of $s_n$

In Sect. 2.1, we have described an edge-weight universality class in terms of  $s_n$ . In this paper, we investigate the case where  $s_n \rightarrow \infty$  with  $s_n = o(n^{1/3})$ . We conjecture that all universality classes can be described in terms of the scaling behaviour of the sequence  $(s_n)_n$  and below identify the eight universality classes that describe the different scaling behaviours. These eight cases are defined by how fast  $s_n \rightarrow 0$  (this gives rise to four cases), the case where  $s_n$  converges to a positive and finite constant, and by how  $s_n \rightarrow \infty$  (giving rise to three cases, including the one that is studied in this paper). We believe that this paper represents a major step forward in this direction in that it describes the scaling behaviour in a large regime of  $(s_n)_n$  sequences. We next describe the eight scaling regimes of  $s_n$  and the results proved and/or predicted for them. We conjecture that these eight cases describe *all* universality classes for FPP on the complete graph, and it would be of interest to make this universal picture complete. Let us now describe these eight cases.

**The regime  $s_n \rightarrow 0$ .** In view of (2.5), for  $s_n \rightarrow 0$ , the first passage percolation problem approximates the graph metric, where the approximation is stronger the faster  $s_n$  tends to zero. We distinguish four different scaling regimes according to how fast  $s_n \rightarrow 0$ :

- (i) Firstly,  $s_n \log n \rightarrow \gamma \in [0, \infty)$ : the case that  $Y \stackrel{d}{=} E^{-\gamma}$  for  $\gamma \in (0, \infty)$  falls into this class with  $s_n = \gamma / \log n$  (see [20, Sect. 1.2]) and was investigated in [16], where it is proved that  $H_n$  is concentrated on at most two values. For the case of  $n$ -dependent edge weights  $Y_e^{(\kappa_n)} \stackrel{d}{=} E^{s_n}$ , it was observed in [20] that when  $s_n \log n$  converges to  $\gamma$  fast enough, the methods of [16] can be imitated and the concentration result for the hopcount continues to hold.
- (ii) When  $s_n \log n \rightarrow \infty$  but  $s_n^2 \log n \rightarrow 0$ , the variance factor  $s_n^2 \log(n/s_n^3)$  from the central limit theorem (CLT) in (2.10) tends to zero. Since  $H_n$  is integer-valued, it follows that (2.10) must fail in this case. First order asymptotics are investigated in [20], and it is shown that  $H_n/(s_n \log n) \xrightarrow{\mathbb{P}} 1$ ,  $W_n/(u_n(1)s_n \log n) \xrightarrow{\mathbb{P}} e$ . It is tempting to conjecture that there exists an integer  $k = k_n \approx s_n \log n$  such that  $H_n \in \{k_n, k_n + 1\}$  whp.
- (iii) The regime where  $s_n \log n \rightarrow \infty$  but  $s_n^2 \log n \rightarrow \gamma \in (0, \infty)$  corresponds to a *critical window* between the one- or two-point concentration conjectured in (ii) and the CLT scaling conjectured in (iv). It is natural to expect that  $H_n - \lfloor \phi_n \log n \rfloor$  is tight for an appropriately chosen sequence  $\phi_n \sim s_n$ , although the distribution of  $H_n - \lfloor \phi_n \log n \rfloor$  might only have subsequential limits because of integer effects. Moreover, we would expect these subsequential limits in distribution to match with (ii) and (iv) in the limits  $\gamma \rightarrow 0$  or  $\gamma \rightarrow \infty$ , respectively.
- (iv) When  $s_n \rightarrow 0$ ,  $s_n^2 \log n \rightarrow \infty$ , we conjecture that the CLT for the hopcount in Theorem 2.4 remains true, and that  $u_n(1)^{-1}W_n - \frac{1}{\lambda_n} \log n$  converges to a Gumbel distribution for a suitable sequence  $\lambda_n$  and  $u_n(1)$ . Unlike in the fixed  $s$  case, we expect no martingale limit terms to appear in the limiting distribution.

**The fixed  $s$  regime.** The fixed  $s$  regime was investigated in [11] in the case where  $Y \stackrel{d}{=} E^s$ , and describes the boundary case between  $s_n \rightarrow 0$  and  $s_n \rightarrow \infty$ . We conjecture that for other sequences of random variables for which Condition 2.1 is valid for some  $s \in (0, \infty)$  the CLT for the hopcount remains valid, while there exist  $V$  and  $\lambda(s)$  depending on the distribution

$Y$  such that  $u_n(1)^{-1}W_n - \frac{1}{\lambda(s)} \log(n/s^3) \xrightarrow{d} V$ . In general,  $V$  will not be equal to  $(M^{(1)} \vee M^{(2)})^s$ , see for example [11]. Instead, it is described by a sum of different terms involving Gumbel distributions and the martingale limit of a certain continuous-time branching process depending on the distribution. Our proof is inspired by the methods developed in [11]. The CLT for  $H_n$  in the fixed  $s$  regime can be recovered from our proofs; in fact, the reasoning in that case simplifies considerably compared to our more general setup.

The results in [11] match up nicely with ours. Indeed, in [11], it was shown that

$$f_n^{-1}(W_n - \log n/\lambda(s)) \xrightarrow{d} \lambda(s)^{-1/s} (\Lambda_{1,2} - \log L_s^{(1)} - \log L_s^{(2)} - \log(1/s))^{1/s}, \quad (2.13)$$

where  $\lambda(s) = \Gamma(1 + 1/s)^s$ ,  $\Lambda_{1,2}$  is a Gumbel variable so that  $\mathbb{P}(\Lambda_{1,2} \leq x) = e^{-e^{-x}}$  and  $L_s^{(1)}, L_s^{(2)}$  are two independent copies of the random variable  $L_s$  with  $\mathbb{E}(L_s) = 1$  solving the distributional equation

$$L_s \stackrel{d}{=} \sum_{i \geq 1} e^{-\lambda(s)(E_1 + \dots + E_i)^s} L_{s,i}, \quad (2.14)$$

where  $(L_{s,i})_{i \geq 1}$  are i.i.d. copies of  $L_s$  and  $(E_i)_{i \geq 1}$  are i.i.d. exponentials with mean 1. We claim that the right-hand side of (2.13) converges to  $M^{(1)} \vee M^{(2)}$  as  $s \rightarrow \infty$ , where  $M^{(1)}, M^{(2)}$  are as in Theorem 2.4. This is equivalent to the statement that  $(-\log L_s^{(j)})^{1/s} \xrightarrow{d} M^{(j)}$  as  $s \rightarrow \infty$ . Assume that  $(-\log L_s^{(j)})^{1/s}$  converges in distribution to a random variable  $\mathcal{M}$ . Then

$$\begin{aligned} \lim_{s \rightarrow \infty} \left( -\log \left( \sum_{i \geq 1} e^{-\lambda(s)(E_1 + \dots + E_i)^s} L_{s,i} \right) \right)^{1/s} &= \min_{i \geq 1} \lim_{s \rightarrow \infty} \left( \lambda(s)(E_1 + \dots + E_i)^s - \log L_{s,i} \right)^{1/s} \\ &= \min_{i \geq 1} \left( (E_1 + \dots + E_i) \vee \left( \lim_{s \rightarrow \infty} (-\log L_{s,i})^{1/s} \right) \right), \end{aligned} \quad (2.15)$$

and using (2.14) we deduce that  $\mathcal{M}$  is the solution of the equation

$$\mathcal{M} \stackrel{d}{=} \min_{i \geq 1} (E_1 + \dots + E_i) \vee \mathcal{M}_i, \quad (2.16)$$

where  $(\mathcal{M}_i)_{i \geq 1}$  are i.i.d. copies of  $\mathcal{M}$  independent of  $(E_i)_{i \geq 1}$ . The unique solution to (2.16) is the random variable with  $\mathbb{P}(\mathcal{M} \leq x)$  being the survival probability of a Poisson Galton–Watson process with mean  $x$ , so that  $\mathcal{M} \stackrel{d}{=} M^{(1)}$ .

**The regime**  $s_n \rightarrow \infty$ . The regime  $s_n \rightarrow \infty$  can be further separated into three cases.

- (i) Firstly, the case where  $s_n \rightarrow \infty$  with  $s_n/n^{1/3} \rightarrow 0$  is the main topic of this paper.
- (ii) Secondly, the regime where  $s_n/n^{1/3} \rightarrow \gamma \in (0, \infty)$  corresponds to the *critical window* between the minimal spanning tree case discussed below and the case (i) studied here. It is natural to expect (see also Theorems 1.1 and 1.3) that  $H_n/n^{1/3}$  converges to a non-trivial limit that depends sensitively on  $\gamma$ , and that, when  $\gamma \rightarrow 0$  and  $\gamma \rightarrow \infty$  matches up with the cases (i) and (iii) discussed above and below, respectively.
- (iii) Finally, the regime  $s_n/n^{1/3} \rightarrow \infty$ . Several of our methods do not extend to the case where  $s_n/n^{1/3} \rightarrow \infty$ ; indeed, we conjecture that the CLT in Theorem 2.4 ceases to hold in this regime. In this case, our proof clearly suggests that first passage percolation (FPP) on the complete graph is closely approximated by invasion percolation (IP) on the Poisson-weighted infinite tree (PWIT), studied in [2], whenever  $s_n \rightarrow \infty$ , see also [21]. It is tempting to predict that  $H_n/n^{1/3}$  converges to the same limit as the graph distance between two vertices for the minimal spanning tree on the complete graph as identified in [3].

## 2.2.2 First Passage Percolation on Random Graphs

FPP on random graphs has attracted considerable attention in the past years, and our research was strongly inspired by its studies. In [17], the authors show that for the configuration model with finite-variance degrees (and related graphs) and edge weights with a continuous distribution not depending on  $n$ , there exists only a *single* universality class. Indeed, if we define  $W_n$  and  $H_n$  to be the weight of and the number of edges in the smallest-weight path between two uniform vertices in the graph, then there exist positive, finite constants  $\alpha, \beta, \lambda$  and sequences  $(\alpha_n)_n, (\lambda_n)_n$ , with  $\alpha_n \rightarrow \alpha, \lambda_n \rightarrow \lambda$ , such that  $W_n - (1/\lambda_n) \log n$  converges in distribution, while  $H_n$  satisfies a CLT with asymptotic mean  $\alpha_n \log n$  and asymptotic variance  $\beta \log n$ .

Related results for exponential edge weights appear for the Erdős–Rényi random graph in [15], to certain inhomogeneous random graphs in [28] and to the small-world model in [30]. The diameter of the weighted graph is studied in [6], and relations to competition on  $r$ -regular graphs are examined in [7]. Finally, the smallest-weight paths with most edges from a single source or between any pair in the graph are investigated in [5].

We conjecture that our results are closely related to FPP on random graphs with *infinite-variance degrees*. Such graphs, sometimes called *scale-free* random graphs, have been suggested in the networking community as appropriate models for various real-world networks. See [8,31] for extended surveys of real-world networks, and [19,24,32] for more details on random graph models of such real-world networks. FPP on infinite-variance random graphs with exponential weights was first studied in [13,14], of which the case of finite-mean degrees studied in [14] is most relevant for our discussion here. There, it was shown that a linear transformation of  $W_n$  converges in distribution, while  $H_n$  satisfies a CLT with asymptotic mean and variance  $\alpha \log n$ , where  $\alpha$  is a simple function of the power-law exponent of the degree distribution of the configuration model. Since the configuration model with infinite-variance degrees whp contains a complete graph of size a positive power of  $n$ , it can be expected that the universality classes on these random graphs are closely related to those on the complete graph  $K_n$ . In particular, the strong universality result for finite-variance random graphs is false, which can be easily seen by observing that for the weight distribution  $1 + E$ , where  $E$  is an exponential random variable, the hopcount  $H_n$  is of order  $\log \log n$  (as for the graph distance [26]), rather than  $\log n$  as it is for exponential weights. See [9] for two examples proving that strong universality indeed fails in the infinite-variance setting, and [4,10] for further results. The area has attracted substantial attention through the work of Komjáthy and collaborators, see also [25,29] for recent work in geometric contexts.

## 2.2.3 Extremal Functionals for FPP on the Complete Graph

Many more fine results are known for FPP on the complete graph with exponential edge weights. In [27], the weak limits of the rescaled path weight and flooding are determined, where the flooding is the maximal smallest weight between a source and all vertices in the graph. In [12] the same is performed for the diameter of the graph. It would be of interest to investigate the weak limits of the flooding and diameter in our setting.

### 3 Detailed Results, Overview and Classes of Edge Weights

In this section, we provide an overview of the proof of our main results.

This section is organised as follows. In Sect. 3.1, we explain how FPP clusters can be described in terms of an appropriate *exploration process*, both from one as well as from two sources. In Sect. 3.2, we discuss how this exploration process can be coupled to first passage percolation on the Poisson-weighted infinite tree (PWIT). In Sect. 3.3, we interpret the FPP dynamics on the PWIT as a continuous-time branching process, and study one- and two-vertex characteristics associated with it. The two-vertex characteristics are needed since we explore from two sources. Due to the near-critical behavior of the involved branching processes, the FPP clusters may run at rather different speeds, and we need to make sure that their sizes are comparable. This is achieved by *freezing* the fastest growing one, which is explained in detail in Sect. 3.4, both for the time at which this happens as well as the sizes of the FPP cluster at the freezing times. There, we also investigate the collision times between the two exploration processes, which correspond to (near-) shortest paths between the two sources. In Sect. 3.5, we couple FPP on the complete graph from two sources to a continuous-time branching process from which we can retrieve the FPP clusters by a *thinning* procedure. In Sect. 3.6, we use the explicit distribution of the collision edge (whether thinned or not) to derive its scaling properties, both for the time at which it occurs, as well as the generations of the vertices it consist of. Finally, in Sect. 3.7, we show that the first point of the Cox process that describes the collision edge is with high probability *unthinned* and complete the proof of our main results.

#### 3.1 FPP Exploration Processes

To understand smallest-weight paths in the complete graph, we study the first passage *exploration process* from one or two sources. Recall from (1.1) that  $d_{K_n, Y^{(K_n)}}(i, j)$  denotes the total cost of the optimal path  $\pi_{i,j}$  between vertices  $i$  and  $j$ .

##### 3.1.1 One-Source Exploration Process

For a vertex  $j \in V(K_n)$ , let the *one-source smallest-weight tree*  $SWT_t^{(j)}$  be the connected subgraph of  $K_n$  defined by

$$\begin{aligned} V(SWT_t^{(j)}) &= \{i \in V(K_n) : d_{K_n, Y^{(K_n)}}(i, j) \leq t\}, \\ E(SWT_t^{(j)}) &= \{e \in E(K_n) : e \in \pi_{j,i} \text{ for some } i \in V(SWT_t^{(j)})\}. \end{aligned} \quad (3.1)$$

Note that  $SWT_t^{(j)}$  is indeed a tree: if two optimal paths  $\pi_{j,k}, \pi_{j,k'}$  pass through a common vertex  $i$ , both paths must contain  $\pi_{j,i}$  since the minimizers of (1.1) are unique. Moreover, by construction, FPP distances from the source vertex  $j$  can be recovered from arrival times in the process  $SWT_t^{(j)}$ :

$$d_{K_n, Y^{(K_n)}}(i, j) = \inf \{t : i \in SWT_t^{(j)}\}. \quad (3.2)$$

To visualize the process  $(SWT_t^{(j)})_{t \geq 0}$ , think of the edge weight  $Y_e^{(K_n)}$  as the time required for fluid to flow across the edge  $e$ . Place a source of fluid at  $j$  and allow it to spread through the graph. Then  $V(SWT_t^{(j)})$  is precisely the set of vertices that have been wetted by time  $t$ , while  $E(SWT_t^{(j)})$  is the set of edges along which, at any time up to  $t$ , fluid has flowed from a wet vertex to a previously dry vertex. Equivalently, an edge is added to  $SWT_t^{(j)}$  whenever it

becomes completely wet, with the additional rule that an edge is not added if it would create a cycle.

Because fluid begins to flow across an edge only after one of its endpoints has been wetted, the *age* of a vertex—the length of time that a vertex has been wet—determines how far fluid has traveled along the adjoining edges. Given  $SWT_t^{(j)}$ , the future of the exploration process will therefore be influenced by the current ages of vertices in  $SWT_t^{(j)}$ , and the nature of this effect depends on the probability law of the edge weights  $(Y_e^{(K_n)})_e$ . In the sequel, for a subgraph  $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$  of  $K_n$ , we write  $\mathcal{G}$  instead of  $V(\mathcal{G})$  for the vertex set when there is no risk of ambiguity.

### 3.1.2 Two-Source Exploration Process

Consider now two vertices from  $K_n$ , which for simplicity we take to be vertices 1 and 2. The *two-source smallest-weight tree*  $SWT_t^{(1,2)}$  is the subgraph of  $K_n$  defined by

$$SWT_t^{(1,2)} = SWT_t^{(1,2;1)} \cup SWT_t^{(1,2;2)}, \quad (3.3)$$

where

$$\begin{aligned} V(SWT_t^{(1,2;1)}) &= \{i \in [n]: d_{K_n, Y(K_n)}(1, i) \leq t \text{ and } d_{K_n, Y(K_n)}(1, i) < d_{K_n, Y(K_n)}(2, i)\}, \\ V(SWT_t^{(1,2;2)}) &= \{i \in [n]: d_{K_n, Y(K_n)}(2, i) \leq t \text{ and } d_{K_n, Y(K_n)}(2, i) < d_{K_n, Y(K_n)}(1, i)\}, \\ E(SWT_t^{(1,2;j)}) &= \{e \in E(K_n): e \in \pi_{j,i} \text{ for some } i \in V(SWT_t^{(1,2;j)})\}. \end{aligned} \quad (3.4)$$

In other words,  $SWT_t^{(1,2)}$  is the union, over all vertices  $i$  within FPP distance  $t$  of vertex 1 or vertex 2, of an optimal path, either  $\pi_{1,i}$  or  $\pi_{2,i}$  whichever has smaller weight.

Because the edge weight distribution has no atoms, no two optimal paths have the same length. It follows that, a.s.,  $SWT_t^{(1,2)}$  is the union of two *vertex-disjoint* trees for all  $t$ . (To see this, suppose vertex  $i$  is closer to vertex  $j$  than to vertex  $j'$ , where  $\{j, j'\} = \{1, 2\}$ . Then, given another vertex  $i'$  and a path  $\pi$  passing from  $i'$  to  $i$  to  $j'$ , there must be a strictly shorter path passing from  $i'$  to  $i$  to  $j$ .) We note that

$$V(SWT_t^{(1,2)}) = V(SWT_t^{(1)}) \cup V(SWT_t^{(2)}), \quad E(SWT_t^{(1,2)}) \subset E(SWT_t^{(1)}) \cup E(SWT_t^{(2)}), \quad (3.5)$$

with strict containment for sufficiently large  $t$ .

To visualize the process  $(SWT_t^{(1,2)})_{t \geq 0}$ , place sources of fluid at vertices 1 and 2 and allow both fluids to spread through the graph. Then, as before,  $V(SWT_t^{(1,2)})$  is precisely the set of vertices that have been wetted by time  $t$ , while  $E(SWT_t^{(1,2)})$  is the set of edges along which, at any time up to  $t$ , fluid has flowed from a wet vertex to a previously dry vertex. Equivalently, an edge is added to  $SWT_t^{(1,2)}$  whenever it becomes completely wet, with the additional rules that an edge is not added if it would create a cycle *or* if it would connect the two connected components of  $SWT_t^{(1,2)}$ .

From the process  $SWT_t^{(1,2)}$ , we can partially recover FPP distances. Denote by  $T^{SWT^{(1,2)}}(i) = \inf \{t \geq 0: i \in SWT_t^{(1,2)}\}$  the arrival time of a vertex  $i \in [n]$ . Then, for  $j \in \{1, 2\}$ ,

$$d_{K_n, Y(K_n)}(i, j) = T^{SWT^{(1,2)}}(i) \quad \text{provided that } i \in \cup_{t \geq 0} SWT_t^{(1,2;j)}. \quad (3.6)$$

More generally, observing the process  $(SWT_t^{(1,2)})_{t \geq 0}$  allows us to recover the edge weights  $Y_e^{(K_n)}$  for all  $e \in \cup_{t \geq 0} E(SWT_t^{(1,2)})$ . However, in contrast to the one-source case, the FPP distance  $W_n = d_{K_n, Y(K_n)}(1, 2)$  cannot be determined by observing the process  $(SWT_t^{(1,2)})_{t \geq 0}$ .

Indeed, if vertices  $i_1, i_2$  satisfy  $i_1 \in SWT_t^{(1,2;1)}$  and  $i_2 \in SWT_t^{(1,2;2)}$  for some  $t$ , then by construction the edge  $\{i_1, i_2\}$  between them will never be added to  $SWT^{(1,2)}$  and there is no arrival time from which to determine the edge weight  $Y_{\{i_1, i_2\}}^{(K_n)}$ .

The optimal weight  $W_n$  is the minimum value

$$W_n = \min_{i_1 \in SWT_\infty^{(1,2;1)}, i_2 \in SWT_\infty^{(1,2;2)}} \left( T^{SWT^{(1,2)}}(i_1) + Y_{\{i_1, i_2\}}^{(K_n)} + T^{SWT^{(1,2)}}(i_2) \right), \quad (3.7)$$

which is uniquely attained a.s. by our assumptions on the edge weights.

**Definition 3.1** (*Collision time and edge*) The *SWT collision time* is  $T_{\text{coll}}^{SWT} = \frac{1}{2} W_n$ . Let  $I_1^{SWT}, I_2^{SWT}$  denote the (a.s. unique) minimizers in (3.7). The edge  $\{I_1^{SWT}, I_2^{SWT}\}$  is called the *SWT collision edge*.

In the fluid flow description above,  $T_{\text{coll}}^{SWT}$  is the time when the fluid from vertex 1 and the fluid from vertex 2 first collide, and this collision takes place inside the collision edge. Note that since fluid flows at rate 1 from both sides simultaneously, the overall distance is given by  $W_n = 2T_{\text{coll}}^{SWT}$ .

**Proposition 3.2** (*Smallest-weight path*) *The end points of the collision edge are explored before the collision time,  $T^{SWT^{(1,2)}}(I_j^{SWT}) < T_{\text{coll}}^{SWT}$  for  $j \in \{1, 2\}$ . The optimal path  $\pi_{1,2}$  from vertex 1 to vertex 2 is the union of the unique path in  $SWT_{T_{\text{coll}}^{SWT}}^{(1,2;1)}$  from 1 to  $I_1^{SWT}$ ; the collision edge  $\{I_1^{SWT}, I_2^{SWT}\}$ ; and the unique path in  $SWT_{T_{\text{coll}}^{SWT}}^{(1,2;2)}$  from  $I_2^{SWT}$  to 2. Furthermore*

$$\begin{aligned} & \mathbb{P} \left( T_{\text{coll}}^{SWT} > t \mid (SWT_u^{(1,2)})_{u \geq 0} \right) \\ &= \prod_{i_1 \in SWT_t^{(1,2;1)}} \prod_{i_2 \in SWT_t^{(1,2;2)}} \\ & \quad \mathbb{P} \left( Y > t - T^{SWT^{(1,2)}}(i_1) + t - T^{SWT^{(1,2)}}(i_2) \mid Y > |T^{SWT^{(1,2)}}(i_1) - T^{SWT^{(1,2)}}(i_2)| \right) \\ &= \exp \left( -\frac{1}{n} \sum_{i_1 \in SWT_t^{(1,2;1)}} \sum_{i_2 \in SWT_t^{(1,2;2)}} \left( f_n^{-1} \left( t - T^{SWT^{(1,2)}}(i_1) + t - T^{SWT^{(1,2)}}(i_2) \right) \right. \right. \\ & \quad \left. \left. - f_n^{-1} \left( |T^{SWT^{(1,2)}}(i_1) - T^{SWT^{(1,2)}}(i_2)| \right) \right) \right). \end{aligned} \quad (3.8)$$

We will not use Proposition 3.2 and the formula (3.7), which are a special case of Lemma 3.20 and Theorem 3.22. These generalizations deal with a *freezing procedure* that we will explain below. Note that the conditioning in (3.8) reflects the information about  $Y_{\{i_1, i_2\}}$  gained by knowing that  $i_1$  and  $i_2$  belong to different connected components of  $SWT_t^{(1,2)}$ : during the period of time when one vertex was explored but not the other, the fluid must not have had time to flow from the earlier-explored vertex to the later-explored vertex.

### 3.2 Coupling FPP on $K_n$ to FPP on the Poisson-Weighted Infinite Tree

In this section, we state results that couple FPP on  $K_n$  to FPP on the Poisson-weighted infinite tree (PWIT). We start by explaining the key idea, coupling of order statistics of exponentials to Poisson processes, in Sect. 3.2.1. We continue to define the PWIT in Sect. 3.2.2. We then



couple FPP on  $K_n$  to FPP on the PWIT, for one source in Sect. 3.2.3 and finally for two sources in Sect. 3.2.4.

### 3.2.1 Order Statistics of Exponentials and Poisson Processes

To study the smallest-weight tree from a vertex, say vertex 1, let us consider the time until the first vertex is added. By construction,  $\min_{i \in [n] \setminus \{1\}} Y_{\{1,i\}}^{(K_n)} \stackrel{d}{=} f_n \left( \frac{n}{n-1} E \right)$  [cf. (2.4)], where  $E$  is an exponential random variable of mean 1. We next extend this to describe the distribution of the order statistics  $Y_{(1)}^{(K_n)} < Y_{(2)}^{(K_n)} < \dots < Y_{(n-1)}^{(K_n)}$  of the weights of edges from vertex 1 to *all* other vertices.

By (2.1) and (2.3), we can write  $Y_{\{1,i\}}^{(K_n)} = f_n(E'_i)$ , where  $E'_i = nX_{\{1,i\}}^{(K_n)}$  are independent, exponential random variables with rate  $1/n$ . We can realize  $E'_i$  as the first point of a Poisson point process  $\mathcal{P}^{(i)}$  with rate  $1/n$ , with points  $X_1^{(i)} < X_2^{(i)} < \dots$ , chosen independently for different  $i = 2, \dots, n$ . We can also form the Poisson point process  $\mathcal{P}^{(1)}$  with rate  $1/n$ , corresponding to  $i = 1$ , although this Poisson point process is not needed to produce an edge weight. To each point of  $\mathcal{P}^{(i)}$ , associate the mark  $i$ .

Now amalgamate all  $n$  Poisson point processes to form a single Poisson point process of intensity 1, with points  $X_1 < X_2 < \dots$ . Each point  $X_k$  has an associated *mark*  $M_k$ . By properties of Poisson point processes, given the points  $X_1 < X_2 < \dots$ , the marks  $M_k$  are chosen uniformly at random from  $[n]$ , different marks being independent.

To complete the construction of the edge weights  $Y_{\{1,i\}}^{(K_n)}$ , we need to recover the first points  $X_1^{(i)}$ , for all  $i = 2, \dots, n$ , from the amalgamated points  $X_1 < X_2 < \dots$ . Thus we will *thin* a point  $X_k$  when  $M_k = 1$  (since  $i = 1$  is not used to form an edge weight) or when  $M_k = M_{k'}$  for some  $k' < k$  (since such a point is not the first point of its corresponding Poisson point process). Then

$$(Y_{(k)}^{(K_n)})_{k \in [n-1]} \stackrel{d}{=} (f_n(X_k))_{k \in \mathbb{N}, X_k \text{ unthinned}}. \quad (3.9)$$

In the next step, we extend this result to the smallest-weight tree  $SWT^{(1)}$  using a relation to FPP on the Poisson-weighted infinite tree.

### 3.2.2 The Poisson-Weighted Infinite Tree

The Poisson-weighted infinite tree is an infinite edge-weighted tree in which every vertex has infinitely many (ordered) children. Before giving the definitions, we recall the Ulam–Harris notation for describing trees.

Define the tree  $\mathcal{T}^{(1)}$  as follows. The vertices of  $\mathcal{T}^{(1)}$  are given by finite sequences of natural numbers headed by the symbol  $\emptyset_1$ , which we write as  $\emptyset_1 j_1 j_2 \dots j_k$ . The sequence  $\emptyset_1$  denotes the root vertex of  $\mathcal{T}^{(1)}$ . We concatenate sequences  $v = \emptyset_1 i_1 \dots i_k$  and  $w = \emptyset_1 j_1 \dots j_m$  to form the sequence  $vw = \emptyset_1 i_1 \dots i_k j_1 \dots j_m$  of length  $|vw| = |v| + |w| = k + m$ . Identifying a natural number  $j$  with the corresponding sequence of length 1, the  $j^{\text{th}}$  child of a vertex  $v$  is  $vj$ , and we say that  $v$  is the parent of  $vj$ . Write  $p(v)$  for the (unique) parent of  $v \neq \emptyset_1$ , and  $p^k(v)$  for the ancestor  $k$  generations before,  $k \leq |v|$ .

We can place an edge (which we could consider to be directed) between every  $v \neq \emptyset_1$  and its parent; this turns  $\mathcal{T}^{(1)}$  into a tree with root  $\emptyset_1$ . With a slight abuse of notation, we will use  $\mathcal{T}^{(1)}$  to mean both the set of vertices and the associated graph, with the edges given implicitly according to the above discussion, and we will extend this convention to any subset  $\tau \subset \mathcal{T}^{(1)}$ . We also write  $\partial\tau = \{v \notin \tau : p(v) \in \tau\}$  for the set of children one generation away from  $\tau$ .



To describe the PWIT formally, we associate weights to the edges of  $\mathcal{T}^{(1)}$ . By construction, we can index these edge weights by non-root vertices, writing the weights as  $X = (X_v)_{v \neq \emptyset_1}$ , where the weight  $X_v$  is associated to the edge between  $v$  and its parent  $p(v)$ . We make the convention that  $X_{v_0} = 0$ .

**Definition 3.3** (*Poisson-weighted infinite tree*) The *Poisson-weighted infinite tree* (PWIT) is the random tree  $(\mathcal{T}^{(1)}, X)$  for which  $X_{vk} - X_{v(k-1)}$  is exponentially distributed with mean 1, independently for each  $v \in \mathcal{T}^{(1)}$  and each  $k \in \mathbb{N}$ . Equivalently, the weights  $(X_{v1}, X_{v2}, \dots)$  are the (ordered) points of a Poisson point process of intensity 1 on  $(0, \infty)$ , independently for each  $v$ .

Motivated by (3.9), we study FPP on  $\mathcal{T}^{(1)}$  with edge weights  $(f_n(X_v))_v$ :

**Definition 3.4** (*First passage percolation on the Poisson-weighted infinite tree*) For FPP on  $\mathcal{T}^{(1)}$  with edge weights  $(f_n(X_v))_v$ , let the FPP edge weight between  $v \in \mathcal{T}^{(1)} \setminus \{\emptyset_1\}$  and  $p(v)$  be  $f_n(X_v)$ . The FPP distance from  $\emptyset_1$  to  $v \in \mathcal{T}^{(1)}$  is

$$T_v = \sum_{k=0}^{|v|-1} f_n(X_{p^k(v)}) \quad (3.10)$$

and the FPP exploration process  $BP^{(1)} = (BP_t^{(1)})_{t \geq 0}$  on  $\mathcal{T}^{(1)}$  is defined by  $BP_t^{(1)} = \{v \in \mathcal{T}^{(1)} : T_v \leq t\}$ .

Note that the FPP edge weights  $(f_n(X_{vk}))_{k \in \mathbb{N}}$  are themselves the points of a Poisson point process on  $(0, \infty)$ , independently for each  $v \in \mathcal{T}^{(1)}$ . The intensity measure of this Poisson point process, which we denote by  $\mu_n$ , is the image of Lebesgue measure on  $(0, \infty)$  under  $f_n$ . Since  $f_n$  is strictly increasing by assumption,  $\mu_n$  has no atoms and we may abbreviate  $\mu_n((a, b])$  as  $\mu_n(a, b)$  for simplicity. Thus  $\mu_n$  is characterized by

$$\mu_n(a, b) = f_n^{-1}(b) - f_n^{-1}(a), \quad \int_0^\infty h(y) d\mu_n(y) = \int_0^\infty h(f_n(x)) dx, \quad (3.11)$$

for any measurable function  $h: [0, \infty) \rightarrow [0, \infty)$ .

Clearly, and as suggested by the notation, the FPP exploration process  $BP$  is a continuous-time branching process:

**Proposition 3.5** (FPP on PWIT is CTBP) *The process  $BP^{(1)}$  is a continuous-time branching process (CTBP), started from a single individual  $\emptyset_1$ , where the ages at childbearing of an individual form a Poisson point process with intensity  $\mu_n$ , independently for each individual. The time  $T_v$  is the birth time  $T_v = \inf \{t \geq 0 : v \in BP_t^{(1)}\}$  of the individual  $v \in \mathcal{T}^{(1)}$ .*

### 3.2.3 Coupling One-Source Exploration to the PWIT

Similar to the analysis of the weights of the edges containing vertex 1, we now introduce a thinning procedure that allows us to couple  $BP^{(1)}$  and  $SWT^{(1)}$ . Define  $M_{\emptyset_1} = 1$  and to each other  $v \in \mathcal{T}^{(1)} \setminus \{\emptyset_1\}$  associate a mark  $M_v$  chosen independently and uniformly from  $[n]$ .

**Definition 3.6** (*Thinning—one CTBP*) The vertex  $v \in \mathcal{T}^{(1)} \setminus \{\emptyset_1\}$  is *thinned* if it has an ancestor  $v_0 = p^k(v)$  (possibly  $v$  itself) such that  $M_{v_0} = M_w$  for some unthinned vertex  $w \in \mathcal{T}^{(1)}$  with  $T_w < T_{v_0}$ .

This definition also appears as [Part I, Definition 2.8]. As explained there, this definition is not circular since whether or not a vertex  $v$  is thinned can be assessed recursively in terms of earlier-born vertices. Write  $\widetilde{BP}_t^{(1)}$  for the subgraph of  $BP_t^{(1)}$  consisting of unthinned vertices.

**Definition 3.7** Given a subset  $\tau \subset \mathcal{T}^{(1)}$  and marks  $M = (M_v : v \in \tau)$  with  $M_v \in [n]$ , define  $\pi_M(\tau)$  to be the subgraph of  $K_n$  induced by the mapping  $\tau \rightarrow [n], v \mapsto M_v$ . That is,  $\pi_M(\tau)$  has vertex set  $\{M_v : v \in \tau\}$ , with an edge between  $M_v$  and  $M_{p(v)}$  whenever  $v, p(v) \in \tau$ .

Note that if the marks  $(M_v)_{v \in \tau}$  are distinct then  $\pi_M(\tau)$  and  $\tau$  are isomorphic graphs.

The following theorem, taken from [Part I, Theorem 2.10] and proved in [Part I, Sect. 3.3], establishes a close connection between FPP on  $K_n$  and FPP on the PWIT with edge weights  $(f_n(X_v))_{v \in \tau}$ :

**Theorem 3.8** (Coupling to FPP on PWIT—one source) *The law of  $(SWT_t^{(1)})_{t \geq 0}$  is the same as the law of  $(\pi_M(\widetilde{BP}_t^{(1)}))_{t \geq 0}$ .*

Theorem 3.8 is based on an explicit coupling between the edge weights  $(Y_e^{(K_n)})_e$  on  $K_n$  and  $(X_v)_v$  on  $\mathcal{T}^{(1)}$ . We will describe a related coupling in Sect. 3.5. A general form of those couplings is given in Sect. 5.

### 3.2.4 Coupling Two-Source Exploration to the PWIT

Let  $\mathcal{T}^{(1,2)}$  be the disjoint union of two independent copies  $(\mathcal{T}^{(j)}, X^{(j)})$ ,  $j \in \{1, 2\}$ , of the PWIT. We shall assume that the copies  $\mathcal{T}^{(j)}$  are vertex-disjoint, with roots  $\emptyset_j$ , so that we can unambiguously write  $X_v$  instead of  $X_v^{(j)}$  for  $v \in \mathcal{T}^{(j)}$ ,  $v \neq \emptyset_j$ . We set  $M_{\emptyset_j} = j$  for  $j = 1, 2$ , and otherwise the notation introduced for  $\mathcal{T}^{(1)}$  is used on  $\mathcal{T}^{(1,2)}$ , verbatim. For example, for any subset  $\tau \subseteq \mathcal{T}^{(1,2)}$ , we write  $\partial\tau = \{v \notin \tau : p(v) \in \tau\}$  for the boundary vertices of  $\tau$ , and we define the subgraph  $\pi_M(\tau)$  for  $\tau \subset \mathcal{T}^{(1,2)}$  just as in Definition 3.7.

As in Proposition 3.5, the two-source FPP exploration process on  $\mathcal{T}^{(1,2)}$  with edge weights  $(f_n(X_v))_v$  starting from  $\emptyset_1$  and  $\emptyset_2$  is equivalent to the union  $BP = BP^{(1)} \cup BP^{(2)}$  of two CTBPs. (In the fluid-flow formulation, the additional rule—an edge is not explored if it would join the connected components containing the two sources—does not apply.)

**Definition 3.9** (Thinning—two CTBPs) The vertex  $v \in \mathcal{T}^{(1,2)} \setminus \{\emptyset_1, \emptyset_2\}$  is *thinned* if it has an ancestor  $v_0 = p^k(v)$  (possibly  $v$  itself) such that  $M_{v_0} = M_w$  for some unthinned vertex  $w \in \mathcal{T}^{(1,2)}$  with  $T_w < T_{v_0}$ .

Note that this two-CTBP thinning rule is applied simultaneously across both trees: for instance, a vertex  $v \in \mathcal{T}^{(1)}$  can be thinned due to an unthinned vertex  $w \in \mathcal{T}^{(2)}$ . Henceforth we will be concerned with the two-CTBP version of thinning. Write  $\widetilde{BP}_t$  for the subgraph of  $BP_t = BP_t^{(1)} \cup BP_t^{(2)}$  consisting of unthinned vertices.

The following theorem is a special case of Theorem 3.26:

**Theorem 3.10** (Coupling to FPP on PWIT—two sources) *The law of  $(SWT_t^{(1,2)})_{t \geq 0}$  is the same as the law of  $(\pi_M(\widetilde{BP}_t))_{t \geq 0}$ .*

We will not use Theorem 3.10, but instead rely on its generalization Theorem 3.26, since, in our setting,  $BP^{(1)}$  and  $BP^{(2)}$  can grow at rather different speeds. We will counteract this unbalance by an appropriate *freezing* procedure, as explained in more detail later on. Theorem 3.26 generalizes Theorem 3.10 to include this freezing.

We next state an equality in law for the collision time and collision edge. As a preliminary step, note that (3.8) can be rewritten in terms of the measure  $\mu_n$  as

$$\begin{aligned} & \mathbb{P}(T_{\text{coll}}^{\text{SWT}} > t \mid (SWT_u)_{u \geq 0}) \\ &= \exp \left( - \sum_{\substack{i_1 \in SWT_t^{(1,2;1)} \\ i_2 \in SWT_t^{(1,2;2)}}} \mu_n \left( |T^{\text{SWT}^{(1,2)}}(i_1) - T^{\text{SWT}^{(1,2)}}(i_2)|, t - T^{\text{SWT}^{(1,2)}}(i_1) + t - T^{\text{SWT}^{(1,2)}}(i_2) \right) / n \right). \end{aligned} \quad (3.12)$$

By a *Cox process* with random intensity measure  $Z$  (with respect to a  $\sigma$ -algebra  $\mathcal{F}$ ) we mean a random point measure  $\mathcal{P}$  such that  $Z$  is  $\mathcal{F}$ -measurable and, conditionally on  $\mathcal{F}$ ,  $\mathcal{P}$  has the distribution of a Poisson point process with intensity measure  $Z$ . For notational convenience, given a sequence of intensity measures  $Z_n$  on  $\mathbb{R} \times \mathcal{X}$ , for some measurable space  $\mathcal{X}$ , we write  $Z_{n,t}$  for the measures on  $\mathcal{X}$  defined by  $Z_{n,t}(\cdot) = Z_n((-\infty, t] \times \cdot)$ .

Thus (3.12) states that  $T_{\text{coll}}^{\text{SWT}}$  has the law of the first point of a Cox process on  $\mathbb{R}$ , where the intensity measure is given by a sum over  $SWT^{(1,2;1)} \times SWT^{(1,2;2)}$  (see also [11, Proposition 2.3]). Using Theorem 3.10, we can lift this equality in law to apply to the collision time and collision edge.

**Theorem 3.11** (Cox process for collision edge) *Let  $\mathcal{P}_n^{\text{SWT}}$  be a Cox process on  $[0, \infty) \times T^{(1)} \times T^{(2)}$  (with respect to the  $\sigma$ -algebra generated by  $BP$  and  $(M_v)_{v \in T^{(1,2)}}$ ) with random intensity  $Z_n^{\text{SWT}} = (Z_{n,t}^{\text{SWT}})_{t \geq 0}$  defined by*

$$Z_{n,t}^{\text{SWT}}(\{v_1\} \times \{v_2\}) = \mathbb{1}_{\{v_1 \in BP_t^{(1)}, v_2 \in BP_t^{(2)}\}} \frac{1}{n} \mu_n(|T_{v_1} - T_{v_2}|, t - T_{v_1} + t - T_{v_2}) \quad (3.13)$$

for all  $t \geq 0$ . Let  $(T_{\text{coll}}^{\mathcal{P}_n, \text{SWT}}, V_{\text{coll}}^{(1, \text{SWT})}, V_{\text{coll}}^{(2, \text{SWT})})$  denote the first point of  $\mathcal{P}_n^{\text{SWT}}$  for which  $V_{\text{coll}}^{(1, \text{SWT})}$  and  $V_{\text{coll}}^{(2, \text{SWT})}$  are unthinned. Then the law of  $(T_{\text{coll}}^{\mathcal{P}_n, \text{SWT}}, \pi_M(\widehat{BP}_{T_{\text{coll}}^{\mathcal{P}_n, \text{SWT}}}), M_{V_{\text{coll}}^{(1, \text{SWT})}}, M_{V_{\text{coll}}^{(2, \text{SWT})}})$  is the same as the joint law of  $T_{\text{coll}}^{\text{SWT}} = \frac{1}{2}W_n$ ; the smallest-weight tree  $SWT_{T_{\text{coll}}^{\text{SWT}}}$  at the time  $T_{\text{coll}}^{\text{SWT}}$ ; and the endpoints  $I_1^{\text{SWT}}, I_2^{\text{SWT}}$  of the SWT collision edge. In particular, the hopcount  $H_n$  has the same distribution as  $|V_{\text{coll}}^{(1, \text{SWT})}| + |V_{\text{coll}}^{(2, \text{SWT})}| + 1$ .

### 3.3 FPP on the PWIT as a CTBP

In this section, we relate FPP on the PWIT to a continuous-time branching process (CTBP). In Sect. 3.3.1, we investigate the exploration from one vertex and describe this in terms of one-vertex characteristics. In Sect. 3.3.2, we extend this to the exploration from two vertices and relate this to two-vertex characteristics of CTBPs, which will be crucial to analyse shortest-weight paths in FPP on  $K_n$  which we explore from two sources.

#### 3.3.1 FPP on the PWIT as a CTBP: One-Vertex Characteristics

In this section, we analyze the CTBP  $BP^{(1)}$  introduced in Sect. 3.2. Notice that  $(BP_t^{(1)})_{t \geq 0}$  depends on  $n$  through its offspring distribution. We have to understand the coupled double asymptotics of  $n$  and  $t$  tending to infinity simultaneously.

Recall that we write  $|v|$  for the generation of  $v$  (i.e., its graph distance from the root in the genealogical tree). To count particles in  $BP_t^{(1)}$ , we use a non-random characteristic

$\chi: [0, \infty) \rightarrow [0, \infty)$ . Following [11], define the *generation-weighted vertex characteristic* by

$$z_t^\chi(a) = z_t^{\chi, BP^{(1)}}(a) = \sum_{v \in BP_t^{(1)}} a^{|v|} \chi(t - T_v) \quad \text{for all } a, t \geq 0. \quad (3.14)$$

We make the convention that  $\chi(t) = z_t^\chi(a) = 0$  for  $t < 0$ . For characteristics  $\chi, \eta$  and for  $a, b, t, u \geq 0$ , write

$$m_t^\chi(a) = \mathbb{E}(z_t^\chi(a)) \quad \text{and} \quad M_{t,u}^{\chi,\eta}(a, b) = \mathbb{E}(z_t^\chi(a) z_u^\eta(b)). \quad (3.15)$$

Let  $\hat{\mu}_n(\lambda) = \int e^{-\lambda y} d\mu_n(y)$  denote the Laplace transform of  $\mu_n$ . For  $a > 0$ , define  $\lambda_n(a) > 0$  by

$$a \hat{\mu}_n(\lambda_n(a)) = 1 \quad (3.16)$$

whenever (3.16) has a unique solution. The parameters  $\lambda_n$  and  $\phi_n$  in Theorem 2.4 are given by

$$\lambda_n = \lambda_n(1), \quad (3.17)$$

$$\phi_n = \lambda'_n(1)/\lambda_n(1). \quad (3.18)$$

The asymptotics of  $\lambda_n$  and  $\phi_n$  stated in Theorem 2.4 is the content of the following lemma:

**Lemma 3.12** (Asymptotics of BP-parameters) *As  $n \rightarrow \infty$ ,  $\phi_n/s_n \rightarrow 1$  and  $\lambda_n f_n(1) \rightarrow e^{-\gamma}$ , where  $\gamma$  is Euler's constant.*

Lemma 3.12 is proved in Sect. 6.3.

Typically,  $z_t^\chi(a)$  grows exponentially in  $t$  at rate  $\lambda_n(a)$ . Therefore, we write

$$\begin{aligned} \bar{z}_t^\chi(a) &= e^{-\lambda_n(a)t} z_t^\chi(a), \\ \bar{m}_t^\chi(a) &= \mathbb{E}(\bar{z}_t^\chi(a)) = e^{-\lambda_n(a)t} m_t^\chi(a), \\ \bar{M}_{t,u}^{\chi,\eta}(a, b) &= \mathbb{E}(\bar{z}_t^\chi(a) \bar{z}_u^\eta(b)) = e^{-\lambda_n(a)t} e^{-\lambda_n(b)u} M_{t,u}^{\chi,\eta}(a, b). \end{aligned} \quad (3.19)$$

In the following theorem, we investigate the asymptotics of such generation-weighted one-vertex characteristics:

**Theorem 3.13** (Asymptotics of one-vertex characteristics) *Given  $\varepsilon > 0$  and a compact subset  $A \subset (0, 2)$ , there is a constant  $K < \infty$  such that for  $n$  sufficiently large, uniformly for  $a, b \in A$  and for  $\chi$  and  $\eta$  bounded, non-negative, non-decreasing functions,  $\lambda_n(1)[t \wedge u] \geq K$ ,*

$$\left| s_n^{-1} \bar{m}_t^\chi(a^{1/s_n}) - \int_0^\infty e^{-z} \chi(z/\lambda_n(a^{1/s_n})) dz \right| \leq \varepsilon \|\chi\|_\infty, \quad (3.20)$$

$$\left| s_n^{-3} \bar{M}_{t,u}^{\chi,\eta}(a^{1/s_n}, b^{1/s_n}) - \frac{\int_0^\infty e^{-z} \chi(z/\lambda_n(a^{1/s_n})) dz \int_0^\infty e^{-w} \eta(w/\lambda_n(b^{1/s_n})) dw}{\log(1/a + 1/b)} \right|$$

$$\leq \varepsilon \|\chi\|_\infty \|\eta\|_\infty. \quad (3.21)$$

Moreover, there is a constant  $K' < \infty$  independent of  $\varepsilon$  such that  $\bar{m}_t^\chi(a^{1/s_n}) \leq K' \|\chi\|_\infty s_n$  and  $\bar{M}_{t,u}^{\chi,\eta}(a^{1/s_n}, b^{1/s_n}) \leq K' \|\chi\|_\infty \|\eta\|_\infty s_n^3$  for all  $n$  sufficiently large, uniformly over  $u, t \geq 0$  and  $a, b \in A$ .

**Corollary 3.14** (Asymptotics of means and variance of population size) *The population size  $|BP_t^{(1)}|$  satisfies  $\mathbb{E}(|BP_t^{(1)}|) \sim s_n e^{\lambda_n(1)t}$  and  $\text{Var}(|BP_t^{(1)}|) \sim s_n^3 e^{2\lambda_n(1)t} / \log 2$  in the limit as  $\lambda_n(1)t \rightarrow \infty, n \rightarrow \infty$ .*

Theorem 3.13 is proved in Sect. 6.4. Generally, we will be interested in characteristics  $\chi = \chi_n$  for which  $\chi_n (\lambda_n(1)^{-1} \cdot)$  converges as  $n \rightarrow \infty$ , so that the integral in (3.20) acts as a limiting value. In particular, Corollary 3.14 is the special case  $\chi = \mathbb{1}_{[0,\infty)}$ ,  $a = 1$ .

Since  $s_n \rightarrow \infty$ , Theorem 3.13 and Corollary 3.14 show that the variance of  $\bar{z}_t^\chi (a^{1/s_n})$  is larger compared to the square of the mean, by a factor of order  $s_n$ . This suggests that  $BP_t^{(1)}$  is typically of order 1 when  $\lambda_n(1)t$  is of order 1 [i.e., when  $t$  is of order  $f_n(1)$ , see Lemma 3.12] but has probability of order  $1/s_n$  of being of size of order  $s_n^2$ . See also [Part I, Proposition 2.17] which confirms this behavior.

### 3.3.2 FPP on the PWIT as a CTBP: Two-Vertex Characteristics

Theorem 3.11 expresses the collision time  $T_{\text{coll}}^{\text{SWT}}$  as the first point of a Cox process whose cumulative intensity is given by a double sum over two branching processes. To study such intensities, we introduce *generation-weighted two-vertex characteristics*. Let  $\chi$  be a non-random, non-negative function on  $[0, \infty)^2$  and recall that  $T_v = \inf \{t \geq 0: v \in BP_t\}$  denotes the birth time of vertex  $v$  and  $|v|$  the generation of  $v$ . The generation-weighted two-vertex characteristic is given by

$$z_t^\chi(\vec{a}) = \sum_{v_1 \in BP_{t_1}^{(1)}} \sum_{v_2 \in BP_{t_2}^{(2)}} a_1^{|v_1|} a_2^{|v_2|} \chi(t_1 - T_{v_1}, t_2 - T_{v_2}), \quad (3.22)$$

for all  $t_1, t_2, a_1, a_2 \geq 0$ , where we use vector notation  $\vec{a} = (a_1, a_2)$ ,  $\vec{t} = (t_1, t_2)$ , and so on. We make the convention that  $\chi(t_1, t_2) = z_{t_1, t_2}^\chi(\vec{a}) = 0$  for  $t_1 \wedge t_2 < 0$ . As in (3.19), we rescale and write

$$\begin{aligned} \bar{z}_t^\chi(\vec{a}) &= e^{-\lambda_n(a_1)t_1} e^{-\lambda_n(a_2)t_2} z_{t_1, t_2}^\chi(\vec{a}), \\ \bar{m}_t^\chi(\vec{a}) &= \mathbb{E}(\bar{z}_t^\chi(\vec{a})), \\ \bar{M}_{t, \vec{a}}^{\chi, \eta}(\vec{a}, \vec{b}) &= \mathbb{E}(\bar{z}_t^\chi(\vec{a}) \bar{z}_t^\eta(\vec{b})). \end{aligned} \quad (3.23)$$

In (3.13), the cumulative intensity  $Z_{n,t}^{\text{SWT}}$  can be expressed in terms of a two-vertex characteristic. If we define

$$\chi_n(t_1, t_2) = \mu_n(|t_1 - t_2|, t_1 + t_2) \quad (3.24)$$

then the total cumulative intensity is given by

$$|Z_{n,t}^{\text{SWT}}| = Z_{n,t}^{\text{SWT}}(\mathcal{T}^{(1)} \times \mathcal{T}^{(2)}) = \frac{1}{n} z_{t,t}^{\chi_n}(1, 1). \quad (3.25)$$

We will use the parameters  $a_1, a_2$  to compute moment generating functions corresponding to  $Z_{n,t}^{\text{SWT}}$ .

The characteristic  $\chi_n$  will prove difficult to control directly, because its values fluctuate significantly in size: for instance,  $\chi_n(\frac{1}{2}f_n(1), \frac{1}{2}f_n(1)) = 1$  whereas  $\chi_n(\frac{1}{2}f_n(1), f_n(1)) = O(1/s_n)$ . Therefore, for  $K \in (0, \infty)$ , we define the truncated measure

$$\mu_n^{(K)} = \mu_n|_{(f_n(1-K/s_n), f_n(1+K/s_n)]}, \quad (3.26)$$

and again write  $\mu_n^{(K)}((a, b]) = \mu_n^{(K)}(a, b)$  to shorten notation. For convenience, we will always assume that  $n$  is large enough that  $s_n \geq K$ . By analogy with (3.24), define

$$\chi_n^{(K)}(t_1, t_2) = \mu_n^{(K)}(|t_1 - t_2|, t_1 + t_2). \quad (3.27)$$

By construction, the total mass of  $\mu_n^{(K)}$  is  $2K/s_n$ , so that  $s_n \chi_n^{(K)}$  is uniformly bounded.

The following results identify the asymptotic behavior of  $\bar{z}_{t_1, t_2}^{\chi_n^{(K)}}(\vec{a})$  and show that, for  $K \rightarrow \infty$ , the contribution due to  $\chi_n - \chi_n^{(K)}$  becomes negligible. These results are formulated in Theorem 3.15, which investigates the truncated two-vertex characteristic, and Theorem 3.16, which studies the effect of truncation:

**Theorem 3.15** (Convergence of truncated two-vertex characteristic) *For every  $\varepsilon > 0$  and every compact subset  $A \subset (0, 2)$ , there exists a constant  $K_0 < \infty$  such that for every  $K \geq K_0$  there are constants  $K' < \infty$  and  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$ ,  $a_1, a_2, b_1, b_2 \in A$  and  $\lambda_n(1)[t_1 \wedge t_2 \wedge u_1 \wedge u_2] \geq K'$ ,*

$$\left| s_n^{-1} \bar{m}_t^{\chi_n^{(K)}}(\vec{a}^{1/s_n}) - \zeta(a_2/a_1) \right| \leq \varepsilon, \quad (3.28)$$

and

$$\left| s_n^{-4} \bar{M}_{\vec{t}, \vec{u}}^{\chi_n^{(K)}, \chi_n^{(K)}}(\vec{a}^{1/s_n}, \vec{b}^{1/s_n}) - \frac{\zeta(a_2/a_1)\zeta(b_2/b_1)}{\log(1/a_1 + 1/b_1)\log(1/a_2 + 1/b_2)} \right| \leq \varepsilon, \quad (3.29)$$

where  $\zeta : (0, \infty) \rightarrow \mathbb{R}$  is the continuous function defined by

$$\zeta(a_1/a_2) = \begin{cases} \frac{2a_1a_2}{a_1+a_2} \frac{\log(a_2/a_1)}{a_2-a_1}, & \text{if } a_1 \neq a_2, \\ 1, & \text{if } a_1 = a_2. \end{cases} \quad (3.30)$$

Moreover, for every  $K < \infty$  there are constants  $K'' < \infty$  and  $n'_0 \in \mathbb{N}$  such that for all  $n \geq n'_0$ ,  $t_1, t_2, u_1, u_2 \geq 0$  and  $a_1, a_2, b_1, b_2 \in A$ ,  $\bar{m}_t^{\chi_n^{(K)}}(\vec{a}^{1/s_n}) \leq K''s_n$  and  $\bar{M}_{\vec{t}, \vec{u}}^{\chi_n^{(K)}, \chi_n^{(K)}}(\vec{a}^{1/s_n}, \vec{b}^{1/s_n}) \leq K''s_n^4$ .

The exponents in Theorem 3.15 can be understood as follows. By Theorem 3.13, the first and second moments of a bounded one-vertex characteristic are of order  $s_n$  and  $s_n^3$ , respectively. Therefore, for two-vertex characteristics, one can expect  $s_n^2$  and  $s_n^6$ . Since  $\chi_n^{(K)} = \frac{1}{s_n} \chi_n$  appears once in the first and twice in the second moment, we arrive at  $s_n$  and  $s_n^4$ , respectively.

**Theorem 3.16** (The effect of truncation) *For every  $K > 0$ ,  $\bar{m}_t^{\chi_n - \chi_n^{(K)}}(\vec{1}) = O(s_n)$ , uniformly over  $t_1, t_2$ . Furthermore, given  $\varepsilon > 0$ , there exists  $K < \infty$  such that, for all  $n$  sufficiently large,  $\bar{m}_t^{\chi_n - \chi_n^{(K)}}(\vec{1}) \leq \varepsilon s_n$  whenever  $\lambda_n(1)[t_1 \wedge t_2] \geq K$ .*

Theorems 3.15 and 3.16 are proved in Sect. 7.

### 3.4 CTBP Growth and the Need for Freezing: Medium Time Scales

Theorem 3.11 shows how to analyze the weight  $W_n$  and hopcount  $H_n$  in terms of a Cox process driven by two ( $n$ -dependent) branching processes. In this section, we describe how this analysis works when the branching processes grow normally (i.e., exponentially with a fixed prefactor). In the CTBP scaling results from Sect. 3.3, we have seen that the class of edge weights we consider gives rise to a more complicated scaling, with  $n$ -dependent prefactors that diverge to infinity. As we will explain, this causes a direct analysis to break down, and we define an appropriate *freezing* mechanism that we use to overcome this obstacle. In Sect. 3.4.1, we first explain what we mean with freezing and why we need it, and in Sect. 3.4.2 we explain how FPP from two sources can be frozen, and then later unfrozen, such that CTBP asymptotics can be used and collision times between the two FPP clusters can be analyzed.

### 3.4.1 Frozen FPP Exploration Process

Under reasonable hypotheses, a fixed CTBP grows exponentially under all measures of size. More precisely, there will be a single constant  $\lambda$  such that  $m_t^\eta(1) \sim A_\eta e^{\lambda t}$  and  $M_{t,t}^{\eta,\eta}(1,1) \sim B_\eta e^{2\lambda t}$  for constants  $A_\eta, B_\eta$ , over a wide class of one-vertex characteristics  $\eta: [0, \infty) \rightarrow \mathbb{R}$ , including the choice  $\eta = 1$  that encodes the population size. Similarly, if  $\chi$  is a two-vertex characteristic, we can expect that  $m_{t,t}^\chi(\tilde{1}) \sim C_\chi e^{2\lambda t}$  and  $M_{(t,t),(t,t)}^\chi \sim D_\chi e^{4\lambda t}$ . Taking  $|Z_{n,t}| = \frac{1}{n} z_{t,t}^\chi(1)$  as in (3.25), we would then expect the first point of the Cox process from Theorem 3.11 to appear at times  $t$  for which  $e^{2\lambda t} \approx n$ . For such  $t$ , we have  $e^{\lambda t} \approx \sqrt{n}$ , so that each branching process has of order  $\sqrt{n}$  individuals and a typical individual is of order  $\log n$  generations away from the root.

If these asymptotics hold, then a typical vertex  $v$  alive at such times  $t$  is unthinned whp. Indeed, for each of the  $\approx \log n$  ancestors, there are at most  $\approx \sqrt{n}$  other vertices that might have the same mark. Each pair of vertices has probability  $1/n$  of having the same mark, leading to an upper bound of  $\approx \frac{\sqrt{n} \log n}{n}$  on the probability that  $v$  is thinned.

In particular, the first point of  $\mathcal{P}_n^{SWT}$  coincides whp with  $(T_{\text{coll}}^{(\mathcal{P}_n, SWT)}, V_{\text{coll}}^{(1, SWT)}, V_{\text{coll}}^{(2, SWT)})$ , the first unthinned point of  $\mathcal{P}_n^{SWT}$ . Using Theorem 3.11, it is therefore possible to derive asymptotics of  $W_n$  and  $H_n$  by analysing the first point of  $\mathcal{P}_n^{SWT}$ , which in turn can be done by a first- and second-moment analysis of the intensity measure  $Z_{n,t}^{SWT}$ .

In the setting of this paper, however, this analysis breaks down. The branching processes now themselves depend on  $n$ , and their behaviour becomes irregular when  $n$  is large. One-vertex characteristics satisfy  $m_t^\eta(1) \sim A'_\eta s_n e^{\lambda_n(1)t}$  and  $M_{t,t}^{\eta,\eta}(1,1) \sim B'_\eta s_n^3 e^{2\lambda_n(1)t}$ . The mismatch of prefactors,  $s_n$  versus  $s_n^3$ , suggests that the branching process has probability of order  $1/s_n$  of growing to size  $s_n^2$  in a time of order  $1/\lambda_n(1) \approx f_n(1)$ , and that this unlikely event is important to the long-run growth of the branching process.

We can balance the mismatched first and second moments by aggregating  $s_n$  independent copies of the branching process. The sum of  $s_n$  independent copies of  $z_t^\eta(1)$  will have mean of order  $A'_\eta s_n^2 e^{\lambda_n(1)t}$  and second moment of order  $B'_\eta s_n^4 e^{2\lambda_n(1)t}$ , where now the second moment is on the order of the square of the mean. [With proper attention to correlations, it is also possible to show that the two-vertex characteristics  $z_{t,t}^{\chi_n}(1)$ , summed over two groups of  $s_n$  independent branching processes each, will have mean of order  $C' s_n^3 e^{2\lambda_n(1)t}$  and second moment  $D' s_n^6 e^{4\lambda_n(1)t}$ .] This balancing makes a first- and second-moment analysis possible.

To achieve the same effect starting from two branching processes, wait until each branching process is large enough that it has of order  $s_n$  new children in time of order 1. Then the collection of all individuals born after that time (and their descendants) will again have balanced first and second moments. However, as we will see, the time when each branching process becomes large enough is highly variable. In particular, by the time the slower-growing of the two branching processes is large enough, the faster-growing branching process will have become much too large. For this reason, we will need to *freeze* the faster-growing branching process to allow the other to catch up. In the following sections we explain how freezing affects the FPP exploration process, the coupling to the PWIT, the Cox process representation for the optimal path, and the effect of thinning. We now first explain precisely how we freeze our two branching processes.

The choice of the freezing times  $T_{\text{fr}}^{(j)}$  must attain two goals. First, we must ensure that, at the collision time  $T_{\text{coll}}$ , the two branching processes with freezing are of comparable size (see Theorem 3.33 and the discussion following it). Second, we must ensure that, after the freezing times, the branching processes grow predictably, with the relatively steady exponential growth

typical of supercritical branching processes (in spite of Theorem 3.13 and Corollary 3.14, where the mismatch between mean and variance shows that a branching process from a single initial individual has highly variable growth).

It has been argued in [Part I, Sect. 2.5] that the crossover to typical branching process behavior occurs when we begin to discover “lucky” vertices that have a large number of descendants—of order  $s_n^2$ —in a time of order  $f_n(1)$ . From [Part I, Theorem 2.15], we can see that this crossover coincides approximately with several other milestones: for instance, around the same time, the branching process also reaches height of order  $s_n$  and total size of order  $s_n^2$ . For our purposes, it will be most important to control *moments* of the branching process after unfreezing, which will involve exponentially discounting future births at rate  $\lambda_n(1)$ .

These considerations lead to the following definition of the freezing times:

**Definition 3.17** (*Freezing*) Define, for  $j = 1, 2$ , the *freezing times*

$$T_{\text{fr}}^{(j)} = \inf \left\{ t \geq 0 : \sum_{v \in BP_t^{(j)}} \int_{t-T_v}^{\infty} e^{-\lambda_n(1)(y-(t-T_v))} d\mu_n(y) \geq s_n \right\}, \quad (3.31)$$

and the *unfreezing time*  $T_{\text{unfr}} = T_{\text{fr}}^{(1)} \vee T_{\text{fr}}^{(2)}$ . The *frozen cluster* is given by

$$\mathcal{B}_{\text{fr}} = BP_{T_{\text{unfr}}} = \mathcal{B}_{\text{fr}}^{(1)} \cup \mathcal{B}_{\text{fr}}^{(2)}, \quad \text{where } \mathcal{B}_{\text{fr}}^{(j)} = BP_{T_{\text{fr}}^{(j)}}^{(j)}. \quad (3.32)$$

The variable  $\int_{t-T_v}^{\infty} e^{-\lambda_n(1)(y-(t-T_v))} d\mu_n(y)$  represents the expected number of future offspring of vertex  $v \in BP_t^{(j)}$ , exponentially time-discounted at rate  $\lambda_n(1)$ . In Definition 3.17, this expected discounted number (summed over all  $v \in BP_t^{(j)}$ ) is required to exceed  $s_n$ . This is the correct choice of scaling, because each newly born vertex has probability of order  $1/s_n$  of being “lucky”—i.e., having of order  $s_n^2$  descendants in time  $f_n(1)$ , see [Part I, Definition 2.14 and Proposition 2.17]—and the scaling  $s_n$  in Definition 3.17 ensures that such a lucky vertex will be born in time  $O_{\mathbb{P}}(f_n(1))$  after unfreezing.

Recall that  $M^{(1)}, M^{(2)}$  are i.i.d. random variables for which  $\mathbb{P}(M^{(j)} \leq x)$  is the survival probability of a Poisson Galton–Watson branching process with mean  $x$ . The asymptotics of the freezing times  $T_{\text{fr}}^{(j)}$  and the frozen cluster  $\mathcal{B}_{\text{fr}}$  are as follows.

**Theorem 3.18** (Properties of the freezing times and frozen cluster)

- (a) The freezing times satisfy  $f_n^{-1}(T_{\text{fr}}^{(j)}) \xrightarrow{\mathbb{P}} M^{(j)}$  for  $j = 1, 2$ .
- (b) The volume  $|\mathcal{B}_{\text{fr}}|$  of the frozen cluster is  $O_{\mathbb{P}}(s_n^2)$ .
- (c) The maximum height  $\max \{|v| : v \in \mathcal{B}_{\text{fr}}\}$  of the frozen cluster is  $O_{\mathbb{P}}(s_n)$ .

We expect, but do not prove, that the bounds in parts (b) and (c) are of the correct order, i.e., that the volume is  $\Theta_{\mathbb{P}}(s_n^2)$  and the diameter  $\Theta_{\mathbb{P}}(s_n)$ . The proof of Theorem 3.18 is based on [Part I, Theorem 2.15] and is given in Sect. 9.5.

Since  $M^{(1)} \neq M^{(2)}$  a.s., Theorem 3.18 (a) and the scaling properties of  $f_n$  confirm that the two CTBPs  $BP^{(1)}$  and  $BP^{(2)}$  require substantially different times to grow large enough. Theorem 3.18 (b) and (c) will allow us to ignore the elements coming from the frozen cluster in the proof of Theorem 2.4. For instance, part (c) shows that heights within the frozen cluster are negligible in the central limit theorem scaling of (2.10).

From the proof of Theorem 2.4, we will see that

$$W_n - \frac{1}{\lambda_n} \log(n/s_n^3) = T_{\text{fr}}^{(1)} + T_{\text{fr}}^{(2)} + O_{\mathbb{P}}(f_n(1)). \quad (3.33)$$



[The presence of a logarithm in (3.33) reflects the fact that, after  $T_{\text{unfr}}$ , the branching processes grow exponentially.] The effects of the three terms in (3.33) can be combined using the following lemma:

**Lemma 3.19** (Sums behave like maxima) *Let  $(T_n^{(1)})_n, (T_n^{(2)})_n, \mathcal{M}^{(1)}, \mathcal{M}^{(2)}$  be random variables such that  $\mathcal{M}^{(1)} \vee \mathcal{M}^{(2)} \geq 1$  a.s. and  $(f_n^{-1}(T_n^{(1)}), f_n^{-1}(T_n^{(2)})) \xrightarrow{\mathbb{P}} (\mathcal{M}^{(1)}, \mathcal{M}^{(2)})$ . Then  $f_n^{-1}(T_n^{(1)} + T_n^{(2)}) \xrightarrow{\mathbb{P}} \mathcal{M}^{(1)} \vee \mathcal{M}^{(2)}$ .*

Lemma 3.19 is proved in Sect. 4. Theorem 3.18 (a) and Lemma 3.19 yield that  $f_n^{-1}(T_{\text{fr}}^{(1)} + T_{\text{fr}}^{(2)}) \xrightarrow{\mathbb{P}} M^{(1)} \vee M^{(2)}$ . Because  $M^{(1)}, M^{(2)} > 1$  a.s., we will be able to ignore the term  $O_{\mathbb{P}}(f_n(1))$ , and the scaling of  $W_n$  in Theorem 2.4 will follow.

### 3.4.2 FPP Exploration Process from Two Sources with Freezing and Collisions

Let  $T_{\text{fr}}^{(j)}$  be the freezing times defined in Definition 3.17, which are stopping times with respect to the filtration induced by  $BP_t^{(j)}$ ,  $j \in \{1, 2\}$ . Define

$$T_{\text{unfr}} = T_{\text{fr}}^{(1)} \vee T_{\text{fr}}^{(2)}, \quad (3.34)$$

$$R_j(t) = (t \wedge T_{\text{fr}}^{(j)}) + ((t - T_{\text{unfr}}) \vee 0), \quad (3.35)$$

and

$$\mathcal{B}_t = \bigcup_{j=1}^2 \mathcal{B}_t^{(j)}, \quad \mathcal{B}_t^{(j)} = \{v \in \mathcal{T}^{(j)} : T_v \leq R_j(t)\} = BP_{R_j(t)}^{(j)} \quad \text{for all } t \geq 0. \quad (3.36)$$

In words, we run the two branching processes  $BP = BP^{(1)} \cup BP^{(2)}$  normally until the first freezing time,  $T_{\text{fr}}^{(1)} \wedge T_{\text{fr}}^{(2)}$ , when one of the two branching processes has become large enough. Then we freeze the larger CTBP and allow the smaller one to evolve normally until it is large enough, at time  $T_{\text{unfr}} = T_{\text{fr}}^{(1)} \vee T_{\text{fr}}^{(2)}$ . At this time, which we call the *unfreezing* time  $T_{\text{unfr}} = T_{\text{fr}}^{(1)} \vee T_{\text{fr}}^{(2)}$ , both CTBPs resume their usual evolution. The processes  $R_j(t)$  are the on-off processes that encode this behaviour:  $R_j(t)$  increases at constant rate 1, except for the interval between  $T_{\text{fr}}^{(1)} \wedge T_{\text{fr}}^{(2)}$  and  $T_{\text{unfr}}$ , where one of the two processes is constant. In the fluid flow picture,  $R_j(t)$  represents the distance traveled by fluid from vertex  $j \in \{1, 2\}$ . We call the process  $(\mathcal{B}_t)_{t \geq 0}$  the *two-source branching process with freezing*.

As with  $\mathcal{T}^{(1,2)}$ , we can consider  $\mathcal{B}_t$  to be the union of two trees by placing an edge between each non-root vertex  $v \notin \{\emptyset_1, \emptyset_2\}$  and its parent. We denote by  $T_v^{\mathcal{B}} = \inf \{t \geq 0 : v \in \mathcal{B}_t\}$  the arrival time of the individual  $v \in \mathcal{T}^{(1,2)}$  in  $\mathcal{B} = (\mathcal{B}_t)_{t \geq 0}$ . Using the left-continuous inverse of  $R_j(t)$ , defined by

$$R_j^{-1}(y) = \inf \{t \geq 0 : R_j(t) \geq y\} = \begin{cases} t & \text{if } t \leq T_{\text{fr}}^{(j)}, \\ T_{\text{unfr}} - T_{\text{fr}}^{(j)} + t & \text{if } t > T_{\text{fr}}^{(j)}, \end{cases} \quad (3.37)$$

we obtain

$$T_v^{\mathcal{B}} = R_j^{-1}(T_v) \quad \text{for } v \in \mathcal{T}^{(j)}. \quad (3.38)$$

We next define the *two-source FPP exploration process with freezing* on  $K_n$ , which we will denote  $(S_t)_{t \geq 0}$ . Intuitively,  $S_t$  is the analogue of  $SWT_t^{(1,2)}$  under the assumption that fluid from vertex  $j$  has flowed a distance  $R_j(t)$  by time  $t$ . As with  $SWT_t^{(1,2)}$ , fluid from one vertex blocks fluid from the other vertex, so that  $S_t$  will consist of two vertex-disjoint trees for all  $t$ . However, because fluid from one vertex may be frozen while still blocking fluid from the

other vertex, it will no longer be possible to directly specify the vertex set  $V(S_t)$  as in (3.4). Instead we will define  $S = (S_t)_{t \geq 0}$  inductively using  $R_j^{-1}$  such that at every time  $t \geq 0$ ,  $S_t = S_t^{(1)} \cup S_t^{(2)}$  is the disjoint union of two trees  $S_t^{(1)}$  and  $S_t^{(2)}$  with root 1 and 2, respectively.

At time  $t = 0$ , let  $S_0$  be the subgraph of  $K_n$  with vertex set  $\{1, 2\}$  and no edges. Suppose inductively that we have constructed  $(S_t)_{0 \leq t \leq \tau_{k-1}}$  up to the time  $\tau_{k-1}$  where the  $(k-1)$ st vertex (not including the vertices 1 and 2) was added, for  $1 \leq k \leq n-2$ , with the convention that  $\tau_0 = 0$ . Denote by  $T^S(i) = \inf \{t \geq 0 : i \in S_t\}$  the arrival time of a vertex  $i \in [n]$ .

Consider the set  $\partial S_{\tau_{k-1}}$  of edges  $e$  joining a previously explored vertex  $\underline{e} \in S_{\tau_{k-1}}$  to a new vertex  $\bar{e} \notin S_{\tau_{k-1}}$ . For such an edge, write  $j(e) \in \{1, 2\}$  for the index defined by  $\underline{e} \in S_{\tau_{k-1}}^{(j(e))}$ . At time

$$\tau_k = \min_{e \in \partial S_{\tau_{k-1}}} R_{j(e)}^{-1} \left( R_{j(e)}(T^S(\underline{e})) + Y_e^{(K_n)} \right), \quad (3.39)$$

we add the edge  $e_k$  that attains the minimum in (3.39). Our assumptions on the edge weights  $Y_e^{(K_n)}$  and the processes  $R_1, R_2$  will imply that this minimum, and in addition the minimum

$$\min_{j \in \{1, 2\}} \min_{e: \underline{e} \in S_{\tau_{k-1}}^{(j)}} R_j^{-1} \left( R_j(T^S(\underline{e})) + Y_e^{(K_n)} \right) \quad (3.40)$$

(with edges between  $S^{(1)}$  and  $S^{(2)}$  included), are uniquely attained a.s. We set  $S_t = S_{\tau_{k-1}}$  for  $\tau_{k-1} \leq t < \tau_k$ , and we define  $S_{\tau_k}$  to be the graph obtained by adjoining  $e_k$  to  $S_{\tau_{k-1}}$ .

In the case  $R_1(t) = R_2(t) = t$ ,  $S$  coincides with the two-source smallest-weight tree  $SWT^{(1,2)}$ . In general, because the processes  $R_1, R_2$  increase at variable speeds, the relationship between  $S$ ,  $T^S(i)$  and the FPP distances  $d_{K_n, Y^{(K_n)}}(i, j)$  is subtle. For instance, it need not hold that  $d_{K_n, Y^{(K_n)}}(1, i) = R_1(T^S(i))$  for  $i \in \cup_{t \geq 0} S_t^{(1)}$ . However, we have the following analogue of (3.7):

**Lemma 3.20** (Minimal-weight representation) *The weight of the optimal path  $\pi_{1,2}$  from vertex 1 to vertex 2 is given by*

$$W_n = \min_{i_1 \in S^{(1)}, i_2 \in S^{(2)}} \left( R_1(T^S(i_1)) + Y_{\{i_1, i_2\}}^{(K_n)} + R_2(T^S(i_2)) \right), \quad (3.41)$$

and the minimum is attained uniquely a.s.

The conclusion of Lemma 3.20 is easily seen when  $R_1(t) = t$ ,  $R_2(t) = 0$  (in which case  $S^{(1)}$  is the same as  $SWT^{(1)}$  with vertex 2 removed) or when  $R_1(t) = R_2(t) = t$  (in which case  $S$  reduces to  $SWT^{(1,2)}$ ). The proof of Lemma 3.20 in general requires some care, and is given in Sect. 5.3. The equality in (3.41) will be the basis of our analysis of  $W_n$ .

**Definition 3.21** The collision time is

$$T_{\text{coll}} = \inf \{t \geq 0 : R_1(t) + R_2(t) \geq W_n\}. \quad (3.42)$$

The collision edge is the edge between the vertices  $I_1 \in S^{(1)}$  and  $I_2 \in S^{(2)}$  that attain the minimum in (3.41). We denote by  $H(I_1)$ ,  $H(I_2)$  the graph distance between 1 and  $I_1$  in  $S^{(1)}$  and between 2 and  $I_2$  in  $S^{(2)}$ , respectively.

**Theorem 3.22** (Exploration process at the collision time) *The following statements hold almost surely: the endpoints  $I_1, I_2$  of the collision edge are explored before time  $T_{\text{coll}}$ . The optimal path  $\pi_{1,2}$  from vertex 1 to vertex 2 is the union of the unique path in  $S_{T_{\text{coll}}}^{(1)}$  from 1 to  $I_1$ ; the collision edge  $\{I_1, I_2\}$ ; and the unique path in  $S_{T_{\text{coll}}}^{(2)}$  from  $I_2$  to 2. The weight and hopcount satisfy*

$$W_n = R_1(T_{\text{coll}}) + R_2(T_{\text{coll}}), \quad H_n = H(I_1) + H(I_2) + 1. \quad (3.43)$$

Theorem 3.22 is proved in Sect. 5.3. The first equality in (3.43) is a simple consequence of continuity of  $t \mapsto R_j(t)$  and the definition of  $T_{\text{coll}}$  in Definition 3.21.

**Remark 3.23** The values of the process  $(S_t)_{t \geq 0}$  depend implicitly on the choice of the processes  $R_1, R_2$ , which we have defined in terms of  $\mathcal{B}$ . Similarly, the value of  $(T_{\text{coll}}, I_1, I_2)$  depends on the on-off processes  $R_1(t), R_2(t)$ , as well as on the edge weights  $Y_{\{i_1, i_2\}}^{(K_n)}$  for  $i_1 \in \mathcal{S}^{(1)}, i_2 \in \mathcal{S}^{(2)}$ . In particular, the law of  $(T_{\text{coll}}, I_1, I_2)$  will depend on the relationship between  $\mathcal{B}$  and the edge weights  $(Y_e^{(K_n)})$ , which we will specify in Sect. 3.5.

However, regardless of the laws of  $\mathcal{S}$  and of  $(T_{\text{coll}}, I_1, I_2)$ , the laws of  $W_n$  and  $H_n$  are the same.

### 3.5 Coupling FPP on $K_n$ from Two Sources to a CTBP

In this section, we revisit the coupling of FPP on  $K_n$  from two sources to a CTBP, which we initialized in Sect. 3.2, using the tools in Sect. 3.3 on vertex characteristics, as well as the freezing, unfreezing and collisions discussed in Sect. 3.4. In Sect. 3.5.1, we give the final conclusions of the coupling including freezing, and in Sect. 3.5.2, we relate the law of the weight of the smallest-weight path and its number of collision edges to a certain Cox process of collisions.

#### 3.5.1 The Final Coupling Including Freezing

Similarly to Theorem 3.8, we next couple the FPP process  $\mathcal{S}$  on  $K_n$  and the FPP process  $\mathcal{B}$  on  $\mathcal{T}^{(1,2)}$ . To this end, we introduce a thinning procedure for  $\mathcal{B} = (\mathcal{B}_t)_{t \geq 0}$ : define  $M_{\emptyset_j} = j$ , for  $j = 1, 2$ . To each other  $v \in \mathcal{T}^{(1,2)} \setminus \{\emptyset_1, \emptyset_2\}$ , we associate a mark  $M_v$  chosen uniformly and independently from  $[n]$ .

**Definition 3.24** (*Thinning*) The vertex  $v \in \mathcal{T} \setminus \{\emptyset_1, \emptyset_2\}$  is *thinned* if it has an ancestor  $v_0 = p^k(v)$  (possibly  $v$  itself) such that  $M_{v_0} = M_w$  for some unthinned vertex  $w$  with  $T_w^{\mathcal{B}} < T_{v_0}^{\mathcal{B}}$ .

The difference between Definition 3.24 and its closely related cousin Definition 3.9 is that Definition 3.24 includes freezing, which, as explained in Sect. 3.4, is crucial for our analysis. As with Definition 3.9, this definition is not circular, as vertices are investigated in their order of appearance. Write  $\tilde{\mathcal{B}}_t$  for the subgraph of  $\mathcal{B}_t$  consisting of unthinned vertices.

From here onwards, we will work on a probability space that contains

- the two independent PWITs  $(\mathcal{T}^{(j)}, X^{(j)})$ ,  $j \in \{1, 2\}$ ,
- the marks  $M_v$ ,  $v \in \mathcal{T}^{(1,2)}$ ,
- and a family of independent exponential random variables  $E_e$ ,  $e \in E(K_\infty)$ , with mean 1, independent of the PWITs and the marks.

Here  $E(K_\infty) = \{\{i, j\} : i, j \in \mathbb{N}, i < j\}$ .

On this probability space, we can construct the FPP edge weights on  $K_n$  as follows. Let

$$T^{\tilde{\mathcal{B}}}(i) = \inf \{t \geq 0 : M_v = i \text{ for some } v \in \tilde{\mathcal{B}}_t\} \quad (3.44)$$

be the first time that a vertex with mark  $i$  appears in  $\tilde{\mathcal{B}}$  and denote the corresponding vertex by  $V(i) \in \mathcal{T}^{(1,2)}$ . Note that  $T^{\tilde{\mathcal{B}}}(i)$  is finite for all  $i$  almost surely since the FPP exploration process eventually explores every edge. For every edge  $\{i, i'\} \in E(K_n)$ , we define

$$X(i, i') = \min \{X_v : M_v = i', p(v) = V(i)\}, \quad (3.45)$$

and

$$X_{\{i,i'\}}^{(K_n)} = \begin{cases} \frac{1}{n} X(i, i') & \text{if } T^{\tilde{\mathcal{B}}}(i) < T^{\tilde{\mathcal{B}}}(i'), \\ \frac{1}{n} X(i', i) & \text{if } T^{\tilde{\mathcal{B}}}(i') < T^{\tilde{\mathcal{B}}}(i), \\ E_{\{i,i'\}} & \text{if } T^{\tilde{\mathcal{B}}}(i) = T^{\tilde{\mathcal{B}}}(i') = 0. \end{cases} \quad (3.46)$$

The following proposition states that the random variables in (3.46) can be used to produce the correct edge weights on  $K_n$  for our FPP problem:

**Proposition 3.25** *Let  $(X_e^{(K_n)})_{e \in E(K_n)}$  be defined in (3.46), and write  $Y_e^{(K_n)} = g(X_e^{(K_n)})$ , where  $g$  is the strictly increasing function from (2.1)–(2.2). Then the edge weights  $(Y_e^{(K_n)})_{e \in E(K_n)}$  are i.i.d. with distribution function  $F_Y$ .*

Proposition 3.25 will be generalized in Theorem 5.3 and proved in Sect. 5.4.

In Theorem 3.22, we have explained the (deterministic) relationship between  $\mathcal{S}$  and the FPP problem with edge weights  $Y_e^{(K_n)}$ . Proposition 3.25 shows that, subject to (3.46), these edge weights have the desired distribution. We next explain the relationship between  $\mathcal{B}$  and  $\mathcal{S}$ . Recall the subgraph  $\pi_M(\tau)$  of  $K_n$  introduced in Definition 3.7, which we extend to the case where  $\tau \subset \mathcal{T}^{(1,2)}$ .

**Theorem 3.26** (The coupling with freezing) *Under the edge-weight coupling (3.46),  $\mathcal{S}_t = \pi_M(\tilde{\mathcal{B}}_t)$  for all  $t \geq 0$  almost surely. Moreover, when  $\tilde{\mathcal{B}}$  and  $\mathcal{S}$  are equipped with the FPP edge weights  $Y_v = f_n(X_v)$  and  $Y_e^{(K_n)} = g(X_e^{(K_n)})$ , respectively, the mapping  $\pi_M: \tilde{\mathcal{B}} \rightarrow \mathcal{S}$  is an isomorphism of edge-weighted graphs.*

The proof is given in Sect. 5.4. Even though Theorem 3.26 is closely related to Theorem 3.10, it will be the version in Theorem 3.26 that we rely upon in our technical proofs. We now discuss its importance in more detail.

Theorem 3.26 achieves two goals. First, it relates the exploration process  $\mathcal{B}$ , defined in terms of two infinite underlying trees, to the smallest-weight tree process  $\mathcal{S}$ , defined in terms of a single finite graph. Because thinning gives an explicit coupling between these two objects, we will be able to control its effect, even when the total number of thinned vertices is relatively large. Consequently we will be able to study the FPP problem by analyzing a corresponding problem expressed in terms of  $\mathcal{B}$  (see Theorems 3.27 and 3.28) and showing that whp thinning does not affect our conclusions (see Theorem 3.33).

Second, Theorem 3.26 allows us to relate FPP on the complete graph ( $n$ -independent dynamics run on an  $n$ -dependent weighted graph) with an exploration defined in terms of a pair of Poisson-weighted infinite trees ( $n$ -dependent dynamics run on an  $n$ -independent weighted graph). By analyzing the dynamics of  $\mathcal{B}$  when  $n$  and  $s_n$  are large, we obtain a fruitful *dual picture*: when the number of explored vertices is large, we find a *dynamic* rescaled branching process approximation that is essentially independent of  $n$ . When the number of explored vertices is small, we make use of a *static* approximation by invasion percolation found in [21]. In fact, under our scaling assumptions, FPP on the PWIT is closely related to *invasion percolation* (IP) on the PWIT which is defined as follows. Set  $\text{IP}^{(1)}(0)$  to be the subgraph consisting of  $\emptyset_1$  only. For  $k \in \mathbb{N}$ , form  $\text{IP}^{(1)}(k)$  inductively by adjoining to  $\text{IP}^{(1)}(k-1)$  the boundary vertex  $v \in \partial \text{IP}^{(1)}(k-1)$  of minimal weight. We note that, since we consider only the relative ordering of the various edge weights, we can use either the PWIT edge weights  $(X_v)_v$  or the FPP edge weights  $(f_n(X_v))_v$ .

Write  $\text{IP}^{(1)}(\infty) = \bigcup_{k=1}^{\infty} \text{IP}^{(1)}(k)$  for the limiting subgraph. We remark that  $\text{IP}^{(1)}(\infty)$  is a strict subgraph of  $\mathcal{T}^{(1)}$  a.s. (in contrast to FPP, which eventually explores every edge). Indeed, define

$$M^{(1)} = \sup \{X_v : v \in \text{IP}^{(1)}(\infty) \setminus \{\emptyset_1\}\}, \quad (3.47)$$

the largest weight of an invaded edge. Then  $\mathbb{P}(M^{(1)} < x)$  is the survival probability of a Poisson Galton–Watson branching process with mean  $x$ , as in Theorems 1.1, 1.3 and 2.4.

Consequently, (2.9) in Theorem 2.4 can be read as a decomposition of the weight  $W_n$  of the smallest-weight path into a deterministic part  $\frac{1}{\lambda_n} \log(n/s_n^3)$  coming from the branching process dynamics and the weight of the largest edge explored by invasion percolation starting from two sources  $f_n(M^{(1)} \vee M^{(2)})$ .

### 3.5.2 A Cox Process for the Collisions

Similarly to Theorem 3.11, we can relate the collision time and collision edge to a Cox process driven by  $\mathcal{B}$ . To state this result, we will use a slightly different coupling between  $\mathcal{B}$  and  $K_n$  than the one just described. More precisely, we alter the definition (3.36) of  $\mathcal{B}$  and work with a copy having the same law. This alteration only affects the thinned part  $\mathcal{B} \setminus \tilde{\mathcal{B}}$ , and the pointwise relationships in (3.46) and Theorem 3.26 continue to apply. As discussed in Remark 3.23, this change affects the law of  $(S, T_{\text{coll}}, I_1, I_2)$  but not of  $W_n$  and  $H_n$ . The full details can be found in Sect. 5.5.

**Theorem 3.27** (A Cox process for the collision edges with freezing) *Let  $\mathcal{P}_n$  be a Cox process on  $[0, \infty) \times T^{(1)} \times T^{(2)}$  [with respect to the  $\sigma$ -algebra generated by  $\mathcal{B}$ ,  $R_1$ ,  $R_2$  and  $(M_v)_{v \in \mathcal{B}}$ ] with random intensity measure  $Z_n = (Z_{n,t})_{t \geq 0}$  defined by*

$$Z_{n,t}(\{v_1\} \times \{v_2\}) = \mathbb{1}_{\{v_1 \in \mathcal{B}_t^{(1)}, v_2 \in \mathcal{B}_t^{(2)}\}} \frac{1}{n} \mu_n(\Delta R_{v_1, v_2}, R_1(t) - R_1(T_{v_1}^{\mathcal{B}}) + R_2(t) - R_2(T_{v_2}^{\mathcal{B}})) \quad (3.48)$$

for all  $t \geq 0$ , where

$$\Delta R_{v_1, v_2} = \begin{cases} R_1(T_{v_2}^{\mathcal{B}}) - R_1(T_{v_1}^{\mathcal{B}}), & \text{if } T_{v_1}^{\mathcal{B}} \leq T_{v_2}^{\mathcal{B}}, \\ R_2(T_{v_1}^{\mathcal{B}}) - R_2(T_{v_2}^{\mathcal{B}}), & \text{if } T_{v_2}^{\mathcal{B}} \leq T_{v_1}^{\mathcal{B}}. \end{cases} \quad (3.49)$$

Let  $(T_{\text{coll}}^{(\mathcal{P}_n)}, V_{\text{coll}}^{(1)}, V_{\text{coll}}^{(2)})$  denote the first point of  $\mathcal{P}_n$  for which  $V_{\text{coll}}^{(1)}$  and  $V_{\text{coll}}^{(2)}$  are unthinned. Then, for a suitable coupling, the law of  $(T_{\text{coll}}^{(\mathcal{P}_n)}, R_1(T_{\text{coll}}^{(\mathcal{P}_n)}) + R_2(T_{\text{coll}}^{(\mathcal{P}_n)}), \pi_M(\tilde{\mathcal{B}}_{T_{\text{coll}}^{(\mathcal{P}_n)}}), M_{V_{\text{coll}}^{(1)}}^{(1)}, M_{V_{\text{coll}}^{(2)}}^{(2)})$  is the same as the joint law of the collision time  $T_{\text{coll}}$ ; the optimal weight  $W_n$ ; the smallest-weight tree  $\mathcal{S}_{T_{\text{coll}}}$  at time  $T_{\text{coll}}$ ; and the endpoints  $I_1, I_2$  of the collision edge. In particular, the hopcount  $H_n$  has the same distribution as  $|V_{\text{coll}}^{(1)}| + |V_{\text{coll}}^{(2)}| + 1$ .

Theorem 3.27 is the version of Theorem 3.11 that includes freezing.

*Sketch of the proof* By Lemma 3.20, Theorem 3.22 and the fact that  $R_1 + R_2$  is strictly increasing,  $T_{\text{coll}} > t$  is equivalent to  $W_n > R_1(t) + R_2(t)$ , which is in turn equivalent to

$$Y_{\{i_1, i_2\}}^{(K_n)} > R_1(t) + R_2(t) - R_1(T^{\mathcal{S}}(i_1)) - R_2(T^{\mathcal{S}}(i_2)) \quad \text{for all } i_1 \in \mathcal{S}_t^{(1)}, \quad i_2 \in \mathcal{S}_t^{(2)}. \quad (3.50)$$

On the other hand, the fact that  $i_1 \in \mathcal{S}_t^{(1)}$  and  $i_2 \in \mathcal{S}_t^{(2)}$  implies that between the times  $T^{\mathcal{S}}(i_1) \wedge T^{\mathcal{S}}(i_2)$  and  $T^{\mathcal{S}}(i_1) \vee T^{\mathcal{S}}(i_2)$  when the first vertex and the second vertex of  $\{i_1, i_2\}$  were explored, respectively, the flow from the first explored vertex did not reach the other vertex. This translates to precisely the information that  $Y_{\{i_1, i_2\}}^{(K_n)} > \Delta R_{i_1, i_2}$ .

Because of the relation  $\mathcal{S}_t = \pi_M(\tilde{\mathcal{B}}_t)$ , the set of pairs  $(i_1, i_2) \in \mathcal{S}_t^{(1)} \times \mathcal{S}_t^{(2)}$  can be identified with the set of pairs of unthinned vertices  $(v_1, v_2) \in \tilde{\mathcal{B}}_t^{(1)} \times \tilde{\mathcal{B}}_t^{(2)}$  via  $i_j = M_{v_j}$ . Moreover, under the edge-weight coupling (3.46), the connecting edge weight  $Y_{\{i_1, i_2\}}^{(K_n)}$  is determined based on the birth times of children of  $v_1$  with mark  $M_{v_2}$  (if  $T_{v_1}^{\mathcal{B}} < T_{v_2}^{\mathcal{B}}$ ) or children of  $v_2$

with mark  $M_{v_2}$  (if  $T_{v_1}^{\mathcal{B}} < T_{v_2}^{\mathcal{B}}$ ). In either case, given  $\mathcal{B}_t$  and  $(M_v)_{v \in \mathcal{B}_t}$  for  $t \leq T_{v_1}^{\mathcal{B}} \vee T_{v_2}^{\mathcal{B}}$ , the conditional law of such birth times is Poisson with intensity given by (3.48).

To complete the proof, it remains to ensure that knowledge of  $\mathcal{B}_t$  and  $(M_v)_{v \in \mathcal{B}_t}$  for  $t > T_{v_1}^{\mathcal{B}} \vee T_{v_2}^{\mathcal{B}}$  does not reveal any other information about the birth times used to determine the connecting edge weights  $Y_{[i_1, i_2]}^{(K_n)}$ . We will accomplish this by redefining  $\mathcal{B}$  and  $R_1, R_2$  so that they use a conditionally independent copy of those birth times. See Sect. 5.5 for more details and the full proof of Theorem 3.27.  $\square$

Theorem 3.27 means that we can study first passage percolation on  $K_n$  by studying a CTBP problem and then controlling the effect of thinning. In fact, with Theorem 3.27 in hand, we will no longer need to refer to  $K_n$  at all.

The remainder of the proof of Theorem 2.4 will be to use Theorems 3.26 and 3.27. In Sect. 3.6, we will first study the properties of the collision edges ignoring the thinning, and in Sect. 3.7, we will show that whp the first collision edge is not thinned to conclude the proof.

### 3.6 The Collision Edge and Its Properties: Long Time Scales

Theorem 3.27 expresses the collision edge in terms of the first unthinned point of  $\mathcal{P}_n$ . We begin by stating the asymptotic behavior of the first point (whether thinned or not) of  $\mathcal{P}_n$ :

**Theorem 3.28** (The first point of the Cox process) *Let  $\mathcal{P}_n$  be the Cox process in Theorem 3.27, and let  $(T_{\text{first}}, V_{\text{first}}^{(1)}, V_{\text{first}}^{(2)})$  denote its first point. Then*

$$T_{\text{first}} = T_{\text{unfr}} + \frac{\log(n/s_n^3)}{2\lambda_n} + O_{\mathbb{P}}(1/\lambda_n). \quad (3.51)$$

Furthermore, recalling the sequence  $(\phi_n)_n$  from (3.18), the pair

$$\left( \frac{|V_{\text{first}}^{(1)}| - \frac{1}{2}\phi_n \log(n/s_n^3)}{\sqrt{s_n^2 \log(n/s_n^3)}}, \frac{|V_{\text{first}}^{(2)}| - \frac{1}{2}\phi_n \log(n/s_n^3)}{\sqrt{s_n^2 \log(n/s_n^3)}} \right) \quad (3.52)$$

converges in distribution to a pair of independent normal random variables of mean 0 and variance  $\frac{1}{2}$ , and is asymptotically independent of  $T_{\text{first}}$  and of  $\mathcal{B}$ .

The proof of Theorem 3.28, presented at the end of the current section, is based on a general convergence result for Cox processes, which we now describe. Consider a sequence of Cox processes  $(\mathcal{P}_n)_n$  on  $\mathbb{R} \times \mathbb{R}^2$  with random intensity measures  $(Z_n^*)_n$ , with respect to  $\sigma$ -fields  $(\mathcal{F}_n)_n$ . We will write  $\mathcal{P}_{n,t}^*$  for the measure defined by  $\mathcal{P}_{n,t}^*(\cdot) = \mathcal{P}_n^*([-\infty, t) \times \cdot)$ . Define

$$T_{n,k}^* = \inf \{t: |\mathcal{P}_{n,t}^*| \geq k\} \quad (3.53)$$

and let  $A_{n,k}$  be the event that  $T_{n,j}^* \notin \{\pm\infty\}$  and  $|\mathcal{P}_{n,T_{n,j}^*}^*| = j$ , for  $j = 1, \dots, k$ . That is,  $A_{n,k}$  is the event that the points of  $\mathcal{P}_n^*$  with the  $k$  smallest  $t$ -values are uniquely defined. On  $A_{n,k}$ , let  $X_{n,k}$  denote the unique point for which  $\mathcal{P}_n^*(\{T_{n,k}^*\} \times \{X_{n,k}\}) = 1$ , and otherwise set  $X_{n,k} = \dagger$ , an isolated cemetery point.

The following theorem gives a sufficient condition for the first points of such a Cox process to converge towards independent realizations of a probability measure  $\mathcal{Q}$ . To state it, we write

$$\hat{R}(\vec{\xi}) = \int_{\mathbb{R}^d} e^{\vec{\xi} \cdot \vec{x}} dR(\vec{x}) \quad (3.54)$$

for the moment generating function of a measure  $R$  on  $\mathbb{R}^d$  and  $\vec{\xi} \in \mathbb{R}^d$ .

**Theorem 3.29** Fix a probability measure  $Q$  on  $\mathbb{R}^2$  with  $\hat{Q}(\vec{\xi}) < \infty$  for all  $\vec{\xi} \in \mathbb{R}^2$ , a non-decreasing continuous function  $q: \mathbb{R} \rightarrow (0, \infty)$  satisfying  $\lim_{t \rightarrow -\infty} q(t) = 0$ . Suppose that we can find a decomposition  $Z_n^* = Z_n^{(K)} + Z_n^{\prime(K)}$  for each  $K > 0$ , and sub- $\sigma$ -fields  $\mathcal{F}'_n \subset \mathcal{F}_n$ , such that

(a) for each fixed  $\varepsilon > 0$ ,  $t, u \in \mathbb{R}$ ,  $\vec{\xi} \in \mathbb{R}^2$ , there exists  $K_0 < \infty$  such that, for all  $K \geq K_0$ ,

$$(1 - \varepsilon)q(t)\hat{Q}(\vec{\xi}) \leq \mathbb{E}\left(\hat{Z}_{n,t}^{\prime(K)}(\vec{\xi}) \mid \mathcal{F}'_n\right) \leq (1 + \varepsilon)q(t)\hat{Q}(\vec{\xi}), \quad (3.55)$$

$$\mathbb{E}\left(\left(\frac{\hat{Z}_{n,t}^{\prime(K)}(\vec{\xi})}{q(t)\hat{Q}(\vec{\xi})} - \frac{|Z_{n,u}^{\prime(K)}|}{q(u)}\right)^2 \mid \mathcal{F}'_n\right) \leq \varepsilon, \quad \text{and} \quad (3.56)$$

$$\mathbb{E}\left(|Z_{n,t}^{\prime(K)}| \mid \mathcal{F}'_n\right) < \varepsilon q(t), \quad (3.57)$$

all this with probability at least  $1 - \varepsilon$  for  $n$  sufficiently large; and

(b) for each  $\varepsilon > 0$ , there exists  $\bar{t}$  such that

$$\liminf_{n \rightarrow \infty} \mathbb{P}\left(|Z_{n,\bar{t}}^*| > 1/\varepsilon\right) \geq 1 - \varepsilon. \quad (3.58)$$

Then the random sequence  $(X_{n,j})_{j=1}^\infty$  converges in distribution to an i.i.d. random sequence  $(X_j)_{j=1}^\infty$  where  $X_j$  has law  $Q$ . Moreover  $\{(T_{n,j})_{j=1}^k: n \in \mathbb{N}\}$  is tight,  $(X_{n,j})_{j=1}^\infty$  is asymptotically independent of  $\mathcal{F}_n$  and, if  $(T_j, X_j)_{j=1}^\infty$  is any subsequential limit of  $(T_{n,j}, X_{n,j})_{j=1}^\infty$ , then  $(T_j)_{j=1}^\infty$  and  $(X_j)_{j=1}^\infty$  are independent.

Theorem 3.29 is proved in Sect. 8.

To apply Theorem 3.29, we will rescale and recentre both time and the heights of vertices. Furthermore, we will remove the effect of the frozen cluster  $\mathcal{B}_{\text{fr}}$ .

**Definition 3.30** (Rescaling and recentering) For  $v \in \mathcal{T}^{(1,2)} \setminus \mathcal{B}_{\text{fr}}$ , define  $p^{\text{unfr}}(v)$  to be the unique ancestor  $v'$  of  $v$  for which  $v' \in \partial\mathcal{B}_{\text{fr}}$  (with  $p^{\text{unfr}}(v) = v$  if  $p(v) \in \mathcal{B}_{\text{fr}}$ ). Write

$$|v|^* = \frac{|v| - |p^{\text{unfr}}(v)| - \frac{1}{2}\phi_n \log(n/s_n^3)}{\sqrt{s_n^2 \log(n/s_n^3)}}, \quad (3.59)$$

$$t^* = \lambda_n(1)(t - T_{\text{unfr}}) - \frac{1}{2} \log(n/s_n^3). \quad (3.60)$$

Define  $\mathcal{P}_n^*$  to be the image under the mapping  $(t, v_1, v_2) \mapsto (t^*, |v_1|^*, |v_2|^*)$  of the restriction of  $\mathcal{P}_n$  to  $[0, \infty) \times (\mathcal{T}^{(1)} \setminus \mathcal{B}_{\text{fr}}^{(1)}) \times (\mathcal{T}^{(2)} \setminus \mathcal{B}_{\text{fr}}^{(2)})$ .

**Theorem 3.31** (Our collision Cox process is nice) The point measures  $(\mathcal{P}_n^*)_n$  are Cox processes and satisfy the hypotheses of Theorem 3.29 when  $Q$  is the law of a pair of independent  $N(0, \frac{1}{2})$  random variables,  $q(t^*) = e^{2t^*}$ , and  $\mathcal{F}'_n$  is the  $\sigma$ -field generated by the frozen cluster  $\mathcal{B}_{\text{fr}}$ .

We prove Theorem 3.31 in Sect. 9.3. All the vertices relevant to  $\mathcal{P}_n^*$  are born after the unfreezing time  $T_{\text{unfr}}$ , and therefore appear according to certain CTBPs. Theorem 3.31 will therefore be proved by a first and second moment analysis of the two-vertex characteristics from Sect. 3.3.2.

To use Theorem 3.31 in the proof of Theorem 3.28, we will show that the first point  $(T_{\text{first}}^*, H_1^*, H_2^*)$  of  $\mathcal{P}_n^*$  and the first point  $(T_{\text{first}}, V_{\text{first}}^{(1)}, V_{\text{first}}^{(2)})$  of  $\mathcal{P}_n$  are whp related as in (3.59)–(3.60). This will follow from part (b) of the following lemma, which we will prove in Sects. 9.2 and 9.4:



**Lemma 3.32** Let  $K < \infty$  and  $\bar{t} = T_{\text{unfr}} + \lambda_n(1)^{-1} \left( \frac{1}{2} \log(n/s_n^3) + K \right)$ . Then

- (a)  $|\mathcal{B}_{\bar{t}}| = O_{\mathbb{P}}(\sqrt{ns_n})$ ; and  
 (b)  $\mathcal{P}_n([0, \bar{t}] \times \mathcal{B}_{\text{fr}}^{(1)} \times \mathcal{T}^{(2)}) = \mathcal{P}_n([0, \bar{t}] \times \mathcal{T}^{(1)} \times \mathcal{B}_{\text{fr}}^{(2)}) = 0$  whp.

Assuming Lemma 3.32 (b) and Theorem 3.31, we can now prove Theorem 3.28:

**Proof of Theorem 3.28** By construction, the first point  $(T_{\text{first}}^*, H_1^*, H_2^*)$  of  $\mathcal{P}_n^*$  is the image of some point  $(T, V_1, V_2)$  of  $\mathcal{P}_n$  under the mapping  $(t, v_1, v_2) \mapsto (t^*, |v_1|^*, |v_2|^*)$ . Theorems 3.29 and 3.31 imply that  $T_{\text{first}}^* = O_{\mathbb{P}}(1)$ , so that  $T = T_{\text{unfr}} + \lambda_n(1)^{-1} \left( \frac{1}{2} \log(n/s_n^3) + O_{\mathbb{P}}(1) \right)$  by (3.60). We may therefore apply Lemma 3.32 (b) to conclude that  $\mathcal{P}_n([0, T] \times \mathcal{B}_{\text{fr}}^{(1)} \times \mathcal{T}^{(2)}) = \mathcal{P}_n([0, T] \times \mathcal{T}^{(1)} \times \mathcal{B}_{\text{fr}}^{(2)}) = 0$  whp.

In particular, whp,  $(T, V_1, V_2)$  equals the first point  $(T_{\text{first}}, V_{\text{first}}^{(1)}, V_{\text{first}}^{(2)})$  of  $\mathcal{P}_n$ , and therefore  $H_j^* = |V_{\text{first}}^{(j)}|^*$ . In Theorem 3.28, the heights are to be rescaled as in (3.52) rather than (3.59). However, these differ only by the term  $|p^{\text{unfr}}(V_{\text{first}}^{(j)})|/s_n \sqrt{\log(n/s_n^3)}$ . By Theorem 3.18 (c), we have  $|p^{\text{unfr}}(V_{\text{first}}^{(j)})| = 1 + O_{\mathbb{P}}(s_n)$ , since  $p(p^{\text{unfr}}(V_{\text{first}}^{(j)})) \in \mathcal{B}_{\text{fr}}^{(j)}$  by construction. Hence the term  $|p^{\text{unfr}}(V_{\text{first}}^{(j)})|/s_n \sqrt{\log(n/s_n^3)}$  is  $o_{\mathbb{P}}(1)$ . Finally, the asymptotic independence statements follow from those in Theorem 5.3 and (3.51) follows from the tightness of  $T_{\text{first}}^*$ .  $\square$

### 3.7 Thinning and Completion of the Proof

In this section, we explain that the first point of the Cox process is whp unthinned and conclude our main results:

**Theorem 3.33** (First point of Cox process is whp unthinned) Let  $\mathcal{P}_n$  be the Cox process in Theorem 3.27, and let  $(T_{\text{first}}, V_{\text{first}}^{(1)}, V_{\text{first}}^{(2)})$  denote its first point. Then  $V_{\text{first}}^{(1)}$  and  $V_{\text{first}}^{(2)}$  are whp unthinned. Consequently, whp  $T_{\text{coll}}^{(\mathcal{P}_n)} = T_{\text{first}}, V_{\text{coll}}^{(1)} = V_{\text{first}}^{(1)}, V_{\text{coll}}^{(2)} = V_{\text{first}}^{(2)}$ .

**Proof** According to Definition 3.24, the vertex  $V_{\text{first}}^{(j)}$ ,  $j \in \{1, 2\}$ , will be thinned if and only if some non-root ancestor  $v_0$  of  $V_{\text{first}}^{(j)}$  has  $M_{v_0} = M_w$ , where  $w \in \mathcal{B}_{T_{\text{first}}}$  is unthinned and  $T_w^{\mathcal{B}} < T_{v_0}^{\mathcal{B}}$ . We obtain an upper bound by dropping the requirement that  $w$  should be unthinned and relaxing the condition  $T_w^{\mathcal{B}} < T_{v_0}^{\mathcal{B}}$  to  $T_w^{\mathcal{B}} \leq T_{\text{first}}^{\mathcal{B}}$  and  $w \neq v_0$ . Each such pair of vertices  $(v_0, w)$  has conditional probability  $1/n$  of having the same mark, so, by a union bound,

$$\mathbb{P}(V_{\text{first}}^{(1)} \text{ or } V_{\text{first}}^{(2)} \text{ is thinned} \mid V_{\text{first}}^{(1)}, V_{\text{first}}^{(2)}, |\mathcal{B}_{T_{\text{first}}}|) \leq \frac{1}{n}(|V_{\text{first}}^{(1)}| + |V_{\text{first}}^{(2)}|) |\mathcal{B}_{T_{\text{first}}}|. \quad (3.61)$$

By Theorem 3.28,  $|V_{\text{first}}^{(j)}| = O_{\mathbb{P}}(s_n \log(n/s_n^3))$ . Moreover  $T_{\text{first}} = T_{\text{unfr}} + \lambda_n(1)^{-1} \left( \frac{1}{2} \log(n/s_n^3) + O_{\mathbb{P}}(1) \right)$ , so that  $|\mathcal{B}_{T_{\text{first}}}| = O_{\mathbb{P}}(\sqrt{ns_n})$  by Lemma 3.32 (a). Hence

$$\mathbb{P}(V_{\text{first}}^{(1)} \text{ or } V_{\text{first}}^{(2)} \text{ is thinned} \mid V_{\text{first}}^{(1)}, V_{\text{first}}^{(2)}, |\mathcal{B}_{T_{\text{first}}}|) \leq O_{\mathbb{P}}\left(\frac{\log(n/s_n^3)}{\sqrt{n/s_n^3}}\right), \quad (3.62)$$

and this upper bound is  $o_{\mathbb{P}}(1)$  since  $n/s_n^3 \rightarrow \infty$ .  $\square$

Note that other choices of  $R_1(t)$ ,  $R_2(t)$  would make Theorem 3.33 false. For the first point of  $\mathcal{P}_n$  to appear, the intensity measure  $Z_{n,t}$ , which is given by  $1/n$  times a sum over  $\mathcal{B}_t^{(1)} \times \mathcal{B}_t^{(2)}$ , must be of order 1. If  $R_1(t) = t$ ,  $R_2(t) = 0$ , for instance, then  $\mathcal{B}_t^{(2)}$  is small and it follows that  $\mathcal{B}_t^{(1)}$  must be large (of size at least of order  $n$ ) at time  $t = T_{\text{first}}$ . In this case thinning would have a very strong effect. We note that this argument applies even to relatively well-behaved



edge distributions such as the  $E^s$  edge weights considered in [11], where the exploration must proceed simultaneously from both endpoints with  $R_1(t) = R_2(t) = t$ .

In the heavy-tailed case that we consider, even the symmetric choice  $R_1(t) = R_2(t) = t$  is in effect unbalanced. Indeed, at the earlier of the two freezing times,  $t = \min \{T_{\text{fr}}^{(1)}, T_{\text{fr}}^{(2)}\}$ , the faster-growing cluster has reached size  $O_{\mathbb{P}}(s_n^2)$ , whereas

$$\min \{T_{\text{fr}}^{(1)}, T_{\text{fr}}^{(2)}\} \approx f_n(\min \{M^{(1)}, M^{(2)}\}) < f_n(\max \{M^{(1)}, M^{(2)}\}) \quad (3.63)$$

(see Theorem 3.18) implies that the slower-growing cluster has not yet explored the unique edge of weight  $\max \{M^{(1)}, M^{(2)}\}$  and therefore has size  $O_{\mathbb{P}}(1)$ . This is a crucial reason for introducing the freezing procedure of Sect. 3.4.

We are now ready to complete the proof of Theorem 2.4:

**Proof of Theorem 2.4** According to Lemma 2.7, we can assume Conditions 2.1, 2.2 and 2.6.

We begin with the hopcount result (2.10). By Theorem 3.27,  $H_n \stackrel{d}{=} |V_{\text{coll}}^{(1)}| + |V_{\text{coll}}^{(2)}| + 1$ , where  $(T_{\text{coll}}^{(\mathcal{P}_n)}, V_{\text{coll}}^{(1)}, V_{\text{coll}}^{(2)})$  is the first point of the Cox process  $\mathcal{P}_n$  for which  $V_{\text{coll}}^{(1)}$  and  $V_{\text{coll}}^{(2)}$  are unthinned. By Theorem 3.33,  $T_{\text{coll}}^{(\mathcal{P}_n)} = T_{\text{first}}$  whp, so that the pairs  $(V_{\text{coll}}^{(1)}, V_{\text{coll}}^{(2)})$  and  $(V_{\text{first}}^{(1)}, V_{\text{first}}^{(2)})$  in Theorems 3.27 and 3.28 are the same whp. Hence, whp,

$$\frac{H_n - \phi_n \log(n/s_n^3)}{\sqrt{s_n^2 \log(n/s_n^3)}} \stackrel{d}{=} \frac{|V_{\text{first}}^{(1)}| - \frac{1}{2}\phi_n \log(n/s_n^3)}{\sqrt{s_n^2 \log(n/s_n^3)}} + \frac{|V_{\text{first}}^{(2)}| - \frac{1}{2}\phi_n \log(n/s_n^3)}{\sqrt{s_n^2 \log(n/s_n^3)}} + o_{\mathbb{P}}(1), \quad (3.64)$$

so that Theorem 3.28 implies the CLT for  $H_n$  in (2.10).

For the weight result (2.9), Theorem 3.27 states that  $W_n \stackrel{d}{=} R_1(T_{\text{coll}}^{(\mathcal{P}_n)}) + R_2(T_{\text{coll}}^{(\mathcal{P}_n)})$ . On the event  $\{T_{\text{first}} \geq T_{\text{unfr}}\} \cap \{T_{\text{first}} = T_{\text{coll}}^{(\mathcal{P}_n)}\}$  (which, by Theorem 3.28 and the argument above, occurs whp), the definition (3.35) of  $R_1(t)$ ,  $R_2(t)$  leads to  $R_1(T_{\text{coll}}^{(\mathcal{P}_n)}) + R_2(T_{\text{coll}}^{(\mathcal{P}_n)}) = T_{\text{fr}}^{(1)} + T_{\text{fr}}^{(2)} + 2(T_{\text{first}} - T_{\text{unfr}})$ . Using again Theorem 3.28, we obtain

$$W_n \stackrel{d}{=} T_{\text{fr}}^{(1)} + T_{\text{fr}}^{(2)} + \frac{\log(n/s_n^3)}{\lambda_n} + O_{\mathbb{P}}(1/\lambda_n). \quad (3.65)$$

Therefore, (2.9) follows from Theorem 3.18 (a) and Lemmas 3.12 and 3.19. Finally, the independence of the limiting variables follows from the asymptotic independence in Theorem 3.28.  $\square$

**Organisation of this paper.** In the remainder of the paper, we use the outline given in Sects. 2.1–3.7 and give the details and proofs omitted there. We have already mentioned how these results complete the proof of our main results, as well as where these results are proved, we thus now restrict to the relation to the local behavior as described in the companion paper [21]. While [21] focusses on the local behaviour of FPP on  $K_n$ , in this paper we extend the analysis to the *global* behavior. We will rely on some results from [21], but in a highly localized way. Indeed, in Sect. 5 we rely on some coupling results from [21], in particular [Part I, Theorem 3.4] and some of its extensions. In Sect. 9, we mainly rely on [Part I, Theorem 2.15] and some related results. The other sections do not rely on [21].

## 4 Scaling Properties of $f_n$ and $\mu_n$

In this section, we derive several useful consequences of Conditions 2.1, 2.2 and 2.6, including Lemma 3.19. We also verify that the edge-weight distributions in Examples 1.2 and 2.5 satisfy the hypotheses of Theorem 2.4, and we prove Lemma 2.7, showing that we may without loss of generality assume the stronger condition Condition 2.6 in place of Condition 2.3.

### 4.1 Growth and Density Bounds for $f_n$ and $\mu_n$

In this section, we explore the key implications of Conditions 2.1–2.2 and 2.6 on  $f_n$  and on the intensity measure  $\mu_n$ .

**Lemma 4.1** *Assume Conditions 2.2 and 2.6. Then there exists  $n_0 \in \mathbb{N}$  such that*

$$f_n(x) \leq \left(\frac{x}{x'}\right)^{\varepsilon_0 s_n} f_n(x') \quad \text{whenever } 1 - \delta_0 \leq x \leq x', \quad n \geq n_0. \quad (4.1)$$

**Proof** Divide (2.7) or (2.12) by  $x$  and integrate between  $x$  and  $x'$  to obtain  $\log f_n(x') - \log f_n(x) \geq \varepsilon_0 s_n (\log x' - \log x)$  whenever  $1 - \delta_0 \leq x \leq x'$ ,  $n \geq n_0$ , as claimed.  $\square$

We call Condition 2.6 a density bound because it implies the following lemma, which will also be useful in the study of two-vertex characteristics in Sect. 7:

**Lemma 4.2** *Assume Conditions 2.2 and 2.6. Then, for  $n$  sufficiently large, on the interval  $(f_n(1 - \delta_0), \infty)$ , the measure  $\mu_n$  is absolutely continuous with respect to Lebesgue measure and*

$$\mathbb{1}_{\{y > f_n(1 - \delta_0)\}} d\mu_n(y) \leq \frac{1}{\varepsilon_0 s_n} \frac{f_n^{-1}(y)}{y} dy. \quad (4.2)$$

**Proof.** By Conditions 2.2 and 2.6,  $f_n$  is strictly increasing on  $(1 - \delta_0, \infty)$ , so  $y = f_n(\mu_n(0, y))$  for  $y > f_n(1 - \delta_0)$ . Differentiating and again applying Conditions 2.2 and 2.6, we get

$$1 = f'_n(\mu_n(0, y)) \frac{d}{dy} \mu_n(0, y) \geq \varepsilon_0 s_n \frac{f_n(\mu_n(0, y))}{\mu_n(0, y)} \frac{d}{dy} \mu_n(0, y), \quad y > f_n(1 - \delta_0). \quad \square$$

**Lemma 4.3** *Assume Conditions 2.2 and 2.6. Then, for  $n$  sufficiently large, the density of  $\mu_n$  with respect to Lebesgue measure is at most  $1/(\varepsilon_0 s_n f_n(1))$  on the interval  $(f_n(1), \infty)$ .*

**Proof** From Lemma 4.1 it follows immediately that  $f_n^{-1}(y) \leq (y/f_n(1))^{1/\varepsilon_0 s_n} \leq y/f_n(1)$  for all  $y > f_n(1)$  and sufficiently large  $n$ . The result now follows from Lemma 4.2.  $\square$

**Lemma 4.4** *Assume Conditions 2.2 and 2.6. Then, given  $\varepsilon, \bar{\varepsilon} > 0$ , there exist  $n_0 \in \mathbb{N}$  and  $K < \infty$  such that, for all  $n \geq n_0$  and  $t \geq 0$ ,*

$$\int e^{-\varepsilon y/f_n(1)} \mathbb{1}_{\{y \geq K f_n(1)\}} \mu_n(t + dy) \leq \bar{\varepsilon}/s_n. \quad (4.3)$$

**Proof** By Lemma 4.3, for large  $n$ , the density of  $\mu_n$  with respect to Lebesgue measure is bounded from above by  $1/(\varepsilon_0 s_n f_n(1))$  on  $(f_n(1), \infty)$ . Hence, for  $K > 1$ ,

$$\int e^{-\varepsilon y/f_n(1)} \mathbb{1}_{\{y \geq K f_n(1)\}} \mu_n(t + dy) \leq \int_t^\infty e^{-\varepsilon(y-t)/f_n(1)} \mathbb{1}_{\{y-t \geq K f_n(1)\}} \frac{dy}{\varepsilon_0 s_n f_n(1)} = \frac{e^{-\varepsilon K}}{\varepsilon_0 s_n \varepsilon}.$$

Taking  $K$  sufficiently large proves the claim.  $\square$

**Lemma 4.5** Assume Conditions 2.1, 2.2 and 2.6. Then, given  $K < \infty$ , there exist  $\varepsilon_K > 0$  and  $n_0 \in \mathbb{N}$  such that, for  $0 \leq t \leq Kf_n(1)$  and  $n \geq n_0$ ,

$$\int e^{-\lambda_n(1)y} \mu_n(t+dy) \geq \varepsilon_K/s_n. \quad (4.4)$$

**Proof** For any  $0 \leq t \leq Kf_n(1)$ ,

$$\int e^{-\lambda_n(1)y} \mu_n(t+dy) = \int e^{-\lambda_n(1)(y-t)} \mathbb{1}_{\{y \geq t\}} d\mu_n(y) \geq e^{-2\lambda_n(1)Kf_n(1)} \mu_n(Kf_n(1), 2Kf_n(1)). \quad (4.5)$$

By Lemma 3.12,  $\lambda_n f_n(1)$  converges to a finite constant. Since  $f_n$  is strictly increasing on  $(1 - \delta_0, \infty)$ ,  $y = f_n(f_n^{-1}(y))$  for all  $y > f_n(1 - \delta_0)$ . Writing  $x_n = f_n^{-1}(y)$  and using Conditions 2.1 and 2.6, we get

$$\frac{y(f_n^{-1})'(y)}{f_n^{-1}(y)} = \left[ \frac{x_n f'_n(x_n)}{f_n(x_n)} \right]^{-1} \leq (\varepsilon_0 s_n)^{-1} \quad \forall y > f_n(1 - \delta_0).$$

By Condition 2.1,  $f_n(1 + x/s_n)/f_n(1) \rightarrow e^x$  for every given  $x$ , and it follows with Taylor's theorem that  $f_n^{-1}(f_n(1)K) \sim f_n^{-1}\left(f_n\left(1 + \frac{\log K}{s_n}\right)\right)$  and, therefore,  $\mu_n(Kf_n(1), 2Kf_n(1)) = f_n^{-1}(2Kf_n(1)) - f_n^{-1}(Kf_n(1)) \sim (\log 2)/s_n$ .  $\square$

We are now in the position to prove Lemma 3.19:

**Proof of Lemma 3.19** By monotonicity,  $f_n^{-1}(T_n^{(1)} + T_n^{(2)}) \geq \max\{f_n^{-1}(T_n^{(1)}), f_n^{-1}(T_n^{(2)})\} \xrightarrow{\mathbb{P}} \mathcal{M}^{(1)} \vee \mathcal{M}^{(2)}$ . For the matching upper bound, let  $T_n = T_n^{(1)} \vee T_n^{(2)}$  and  $\mathcal{M} = \mathcal{M}^{(1)} \vee \mathcal{M}^{(2)}$ , so that  $f_n^{-1}(T_n) \xrightarrow{\mathbb{P}} \mathcal{M}$ . Noting that  $f_n^{-1}(T_n^{(1)} + T_n^{(2)}) \leq f_n^{-1}(2T_n)$ , it suffices to show that  $f_n^{-1}(2T_n) \xrightarrow{\mathbb{P}} \mathcal{M}$ . But for  $\delta > 0$ , Lemma 4.1 implies that  $f_n(x + \delta)/f_n(x)$  tends to infinity, and is in particular larger than 2 for  $n$  sufficiently large, uniformly over  $x \in [1, R]$  for any  $R < \infty$ . It follows that  $f_n^{-1}(2T_n) \leq (1 \vee f_n^{-1}(T_n)) + \delta$  with high probability, for any  $\delta > 0$ . Since  $\mathcal{M} \geq 1$  a.s. and  $\delta > 0$  was arbitrary, this completes the proof.  $\square$

We conclude the section with a remark on the connection between Theorem 2.4 above and [Part I, Theorem 2.1] which states that, if  $s_n/\log \log n \rightarrow \infty$ , then

$$f_n^{-1}(W_n) \xrightarrow{d} M^{(1)} \vee M^{(2)}. \quad (4.6)$$

**Remark 4.6** The statements for  $W_n$  in Theorem 2.4 and (4.6) are consistent. Indeed, if  $s_n/\log \log n \rightarrow \infty$ , then Lemmas 4.1 and 3.12 imply

$$f_n^{-1}\left(\frac{1}{\lambda_n} \log(n/s_n^3)\right) \leq \left(\frac{\log(n/s_n^3)}{\lambda_n f_n(1)}\right)^{1/\varepsilon_0 s_n} = (1+o(1)) \exp\left(\frac{1}{\varepsilon_0 s_n} \log \log(n/s_n^3)\right) = 1+o(1). \quad (4.7)$$

Hence, Lemma 3.19 gives that (2.9) and (4.6) agree if  $s_n/\log \log n \rightarrow \infty$ .

## 4.2 Analysis of Specific Edge-Weight Distributions

**Lemma 4.7** With the notation of (2.2)–(2.3), the relation (1.3) [with  $t \mapsto L(t)$  slowly varying as  $t \rightarrow \infty$ ] holds if and only if

$$x \frac{d}{dx} \log g(x) = x^{-\alpha} \tilde{L}(1/x) \quad (4.8)$$

with  $t \mapsto \tilde{L}(t)$  slowly varying as  $t \rightarrow \infty$ . If either of these two equivalent conditions hold, then  $L(t) \sim \tilde{L}(t)$  as  $t \rightarrow \infty$  and the sequences  $(s_n)_n$ ,  $(\tilde{s}_n)_n$  defined by (1.4) and

$$\tilde{s}_n = \frac{f'_n(1)}{f_n(1)} \quad (4.9)$$

satisfy  $s_n \sim \tilde{s}_n$  as  $n \rightarrow \infty$ . Moreover if in addition  $s_n \rightarrow \infty$  (or equivalently  $\tilde{s}_n \rightarrow \infty$ ) then Conditions 2.1, 2.2 and 2.3 (a) hold for the sequences  $(f_n(x), s_n)$  or  $(f_n(x), \tilde{s}_n)$ . If the stronger statement  $s_n / \log \log n \rightarrow \infty$  (or equivalently  $\tilde{s}_n / \log \log n \rightarrow \infty$ ) holds, then Condition 2.3 (b) holds.

**Proof.** The proof is virtually identical to the proof of [Part I, Lemma 5.4], and we indicate only those parts of the proof that differ.

We prove Condition 2.1 for  $f_n$  and the sequence  $\tilde{s}_n$  defined in (4.9). We compute

$$\frac{1}{\tilde{s}_n} \frac{x f'_n(x)}{f_n(x)} = \frac{g(1/n)}{(1/n)g'(1/n)} \frac{\frac{x}{n} g'(\frac{x}{n})}{g(\frac{x}{n})} = x^{-\alpha} \frac{\tilde{L}(n/x)}{\tilde{L}(n)}. \quad (4.10)$$

Noting that  $x f'_n(x) / f_n(x)$  is the derivative of  $\log f_n(x)$  with respect to  $q = \log x$ , we find

$$\frac{f_n(x^{1/s_n})}{f_n(1)} = \exp \left( \int_0^{\log(x^{1/s_n})} s_n e^{-\alpha q} \frac{\tilde{L}(n e^{-q})}{\tilde{L}(n)} dq \right) = \exp \left( \int_0^{\log x} e^{-\alpha \theta / s_n} \frac{\tilde{L}(n e^{-\theta / s_n})}{\tilde{L}(n)} d\theta \right), \quad (4.11)$$

after the substitution  $\theta = s_n q$ . As  $n \rightarrow \infty$ , we have  $\tilde{s}_n \rightarrow \infty$  by assumption, so that the last integrand converges to 1 pointwise. For each fixed  $x$ , the convergence is uniform over  $\theta$  by properties of slowly varying functions, so that  $f_n(x^{1/s_n}) / f_n(1) \rightarrow e^{\log x} = x$  as required. (This argument remains valid, interchanging the limits of integration as necessary, when  $x < 1$ .)

For Condition 2.3 (b), we note that if  $s_n / \log \log n \rightarrow \infty$  then Condition 2.3 (a) implies Condition 2.3 (b) because, for any  $\varepsilon > 0$ ,  $R > 1$ ,

$$\frac{f_n(R)}{f_n(1) \log n} \geq \frac{R^{\varepsilon s_n}}{\log n} = \exp(\varepsilon \log R s_n - \log \log n) \rightarrow \infty. \quad \square$$

**Lemma 4.8** *The edge-weight distributions and the associated sequences  $(s_n)_n$  from Example 1.2 (for all values of the parameters  $a$ ,  $\gamma$  and  $\beta$ ) and Example 2.5 satisfy Conditions 2.1–2.3.*

**Proof** For Example 1.2, it is readily verified that the regular-variation condition (1.3) holds. By Lemma 4.7 it remains to show that Condition 2.3 (b) holds when  $s_n / \log \log n \rightarrow \infty$ , i.e., for Example 1.2 (b) with  $0 < \gamma \leq 1$ . It suffices to find a sequence  $x_n$  with  $g(x_n/n) / g(1/n) \log n \rightarrow \infty$  such that  $s_n^{-1} x g'(x) / g(x) \rightarrow 1$  uniformly over  $1/n \leq x \leq x_n/n$ ; in fact we will take  $x_n = \log n$ .

Since  $u \mapsto u^{a(\log(1+\log(1/u)))^\gamma}$  is increasing, it follows that, with  $q = \log(1/u)$ ,

$$F_Y^{-1}(u) = u^{a(\log(1+\log(1/u)))^\gamma} = \exp(-aq(\log(1+q))^\gamma). \quad (4.12)$$

Setting  $u = 1 - e^{-x}$  and  $q = \log(1/u) = -\log(1 - e^{-x})$ , (2.2) and the chain rule lead to

$$\frac{x g'(x)}{g(x)} = \frac{x e^{-x}}{1 - e^{-x}} \left( a(\log(1+q))^\gamma + a\gamma \frac{q}{1+q} (\log(1+q))^{\gamma-1} \right). \quad (4.13)$$

In particular, we see that  $x g'(x) / g(x)$  is a slowly-varying function of  $q = -\log(1 - e^{-x})$ , say  $x g'(x) / g(x) = h(q)$  where  $q \mapsto h(q)$  is slowly varying as  $x \downarrow 0$ ,  $q \rightarrow \infty$ . Since

$q = \log(1/x) + O(x)$  as  $x \downarrow 0$ , we find that  $q = \log n + O(\log \log n) + O((\log n)/n) = (\log n)(1 + o(1))$  uniformly on  $1/n \leq x \leq (\log n)/n$ , and consequently

$$\frac{1}{s_n} \frac{xg'(x)}{g(x)} = \frac{h(q)}{h(-\log(1 - e^{-1/n}))} = 1 + o(1) \quad (4.14)$$

uniformly on  $1/n \leq x \leq (\log n)/n$  by properties of slowly varying functions, as required. On the other hand  $xg'(x)/g(x) \rightarrow \infty$  as  $x \downarrow 0$  implies in particular that  $xg'(x)/g(x) \geq 2$  for  $x$  sufficiently small, so that

$$\frac{g((\log n)/n)}{g(1/n)} \geq (\log n)^2 \quad (4.15)$$

for  $n$  sufficiently large and we have shown that  $g(x_n/n)/g(1/n) \log n \rightarrow \infty$ , as required.

For Example 2.5, we compute  $f_n(x) = (G^{-1}(1 - e^{-x/n}))^{s_n}$  and

$$\frac{1}{s_n} \frac{xf'_n(x)}{f_n(x)} = \frac{(x/n)(G^{-1})'(1 - e^{-x/n})}{G^{-1}(1 - e^{-x/n})}. \quad (4.16)$$

Write  $a = G'(0) > 0$ . Since  $Z$  is positive-valued, we have  $G(0) = 0$  and therefore  $u/G^{-1}(u) \rightarrow a$  as  $u \downarrow 0$ , whereas  $(G^{-1})'(u) \rightarrow 1/a$  as  $u \rightarrow 0$ . This implies that the quantity in (4.16) tends to 1 whenever  $x/n \rightarrow 0$ , which allows us to conclude Conditions 2.2 and 2.3 (a). Similarly, from  $s_n \rightarrow \infty$  we can infer that  $xf'_n(x)/f_n(x) \geq 2$ , uniformly over  $x \leq \log n$ , for  $n$  sufficiently large, whence  $f_n(\log n)/f_n(1) \log n \rightarrow \infty$  and Condition 2.3 (b) holds. Finally, as in (4.11), we find

$$\frac{f_n(x^{1/s_n})}{f_n(1)} = \exp \left( \int_0^{\log x} \frac{1}{s_n} \frac{\tilde{x} f'_n(\tilde{x})}{f_n(\tilde{x})} \Big|_{\tilde{x}=e^{\theta/s_n}} d\theta \right) \quad (4.17)$$

and comparing with (4.16), the integrand converges to 1 as  $n \rightarrow \infty$ , uniformly over  $\theta$  for any fixed  $x$ , and Condition 2.1 follows.  $\square$

### 4.3 Equivalence of Conditions: Proof of Lemma 2.7

The proof of Lemma 2.7 is based on the observation that if the functions  $f_n, \tilde{f}_n$  agree on the interval  $[0, \bar{x}]$  then, for the FPP problems with edge weights  $f_n(nX_e^{(K_n)})$  and  $\tilde{f}_n(nX_e^{(K_n)})$ , respectively, the optimal paths and their corresponding edge weights are identical whenever either optimal path has weight less than  $f_n(\bar{x}) = \tilde{f}_n(\bar{x})$ .

**Proof of Lemma 2.7** Let  $\delta_0$  be the constant from Condition 2.2, let  $R > 1$ , and define

$$x_{n,R} = R \vee \left( \inf \{x \geq 1 : f_n(x) \geq 4e^\gamma f_n(1) \log n\} \right),$$

$$\varepsilon_0^{(R)} = \frac{1}{2} \liminf_{n \rightarrow \infty} \inf_{1 - \delta_0 \leq x \leq x_{n,R}} \frac{x}{s_n} \frac{d}{dx} \log f_n(x). \quad (4.18)$$

Conditions 2.2 and 2.3 imply that  $\varepsilon_0^{(R)} > 0$  for any  $R > 1$ , and there exists  $n_0^{(R)} \in \mathbb{N}$  such that  $x \frac{d}{dx} \log f_n(x) \geq \varepsilon_0^{(R)} s_n$  for  $x \in [1 - \delta_0, x_{n,R}]$  whenever  $n \geq n_0^{(R)}$ . For definiteness, we take  $n_0^{(R)}$  minimal with this property. We may uniquely define  $f_{n,R} : [0, \infty) \rightarrow [0, \infty)$  by requiring that  $f_{n,R} = f_n$  if  $n < n_0^{(R)}$  and if  $n \geq n_0^{(R)}$  then (a)  $f_{n,R}(x) = f_n(x)$  for all  $x \leq x_{n,R}$ , and (b)  $\frac{x}{s_n} \frac{d}{dx} \log f_{n,R}(x)$  is constant on  $[x_{n,R}, \infty)$ . By construction, the sequence  $(f_{n,R})_n$  satisfies Condition 2.6 for any fixed  $R > 1$ . Furthermore, given any  $x > 0$ ,  $R > 1$  implies that  $x^{1/s_n} \leq R \leq x_{n,R}$  for  $n$  sufficiently large, and it follows that  $f_{n,R}$  satisfies Condition 2.1. Since  $x_{n,R} \geq R > 1$  it follows that Condition 2.2 holds for  $(f_{n,R})_n$ , too.

Let  $\mu_{n,R}$  and  $\lambda_{n,R}$  denote the analogues of  $\mu_n$  and  $\lambda_n$  when  $f_n$  is replaced by  $f_{n,R}$ , and let  $\lambda_{n,R} = \lambda_{n,R}(1)$  and  $\phi_{n,R} = \lambda'_{n,R}(1)/\lambda_{n,R}(1)$  denote the corresponding parameters [see (3.11) and (3.16)–(3.18)]. Let  $W_{n,R}$ ,  $H_{n,R}$  denote the weight and hopcount, respectively, associated to the FPP problem on  $K_n$  with edge weights  $f_{n,R}(nX_e^{(K_n)})$ . Abbreviate  $w_{n,R} = W_{n,R} - \log(n/s_n^3)/\lambda_{n,R}$ ,  $h_{n,R} = (H_{n,R} - \phi_{n,R} \log(n/s_n^3))/\sqrt{s_n^2 \log(n/s_n^3)}$ .

By assumption, Theorem 2.4 holds assuming Conditions 2.1, 2.2 and 2.6. Therefore, it applies to  $f_{n,R}$ . Using Theorem 2.4, we conclude that for any  $k \in \mathbb{N}$ , we may find  $n_0^{(R,k)} \in \mathbb{N}$  such that  $n_0^{(R,k)} \geq n_0^{(R)}$  and

$$\sup_{x,y \in \mathbb{R}} \left| \mathbb{P}\left(f_{n,R}^{-1}(w_{n,R}) \leq x, h_{n,R} \leq y\right) - \mathbb{P}(M^{(1)} \vee M^{(2)} \leq x) \mathbb{P}(Z \leq y) \right| \leq \frac{1}{k}, \quad (4.19a)$$

$$|\lambda_{n,R} f_{n,R}(1) - e^{-\gamma}| \leq \frac{1}{k}, \quad \left| \frac{\phi_{n,R}}{s_n} - 1 \right| \leq \frac{1}{k}, \quad (4.19b)$$

$$f_{n,R}(R-1) + \frac{\log(n/s_n^3)}{\lambda_{n,R}} \leq f_{n,R}(R) \vee 4e^\gamma f_n(1) \log n \quad (4.19c)$$

whenever  $n \geq n_0^{(R,k)}$ , and for definiteness we take  $n_0^{(R,k)}$  minimal with these properties. Indeed, using the continuity of  $M^{(1)} \vee M^{(2)}$  and  $Z$ , the uniform convergence in (4.19a) follows from the pointwise convergence at a finite grid  $((x_i, y_i))_i$  depending on  $k$  and monotonicity of the distribution functions. For (4.19c), use the inequality  $a + b \leq 2(a \vee b)$ , Lemma 3.12, and note that  $2f_{n,R}(R-1) \leq f_{n,R}(R)$  for  $n$  sufficiently large by Lemma 4.1. Set

$$R_n = (2 \vee \max \{k \in \mathbb{N} : n \geq n_0^{(k,k)}\}) \wedge n, \quad \text{and} \quad \lambda_n = \lambda_{n,R_n}, \quad \phi_n = \phi_{n,R_n}. \quad (4.20)$$

Since  $n_0^{(k,k)}$  is finite for each  $k \in \mathbb{N}$ , it follows that  $R_n \rightarrow \infty$ . Moreover, as soon as  $n \geq n_0^{(2,2)}$ , we have  $n \geq n_0^{(R_n, R_n)}$ , so that (4.19a)–(4.19c) hold with  $(R, k) = (R_n, R_n)$ . By construction,  $f_{n,R}(1) = f_n(1)$ , and we conclude in particular that  $\phi_n/s_n \rightarrow 1$  and  $\lambda_n f_n(1) \rightarrow e^{-\gamma}$ .

Given two functions  $f_n$  and  $\tilde{f}_n$ , we can couple the corresponding FPP problems by choosing edge weights  $f_n(nX_e^{(K_n)})$  and  $\tilde{f}_n(nX_e^{(K_n)})$ , respectively. Let  $\bar{x} > 0$ . On the event  $\{W_n \leq f_n(\bar{x})\}$ , the optimal path  $\pi_{1,2}$  uses only edges of weight at most  $f_n(\bar{x})$ . If  $f_n$  and  $\tilde{f}_n$  agree on the interval  $[0, \bar{x}]$ , then the edges along that path have the same weights in the two FPP problems and we deduce that  $W_n = \tilde{W}_n$  and  $H_n = \tilde{H}_n$ , where  $\tilde{W}_n$  and  $\tilde{H}_n$  are the weight and the hopcount of the optimal path in the problem corresponding to  $\tilde{f}_n$ .

Consequently, on the event  $\{W_{n,R_n} \leq f_{n,R_n}(x_{n,R_n})\}$ ,  $W_n = W_{n,R_n}$  and  $H_n = H_{n,R_n}$ . By (4.19a), it remains to show that this event occurs whp. Since  $R_n \rightarrow \infty$ , we conclude from (4.19a) that  $W_{n,R_n} \leq f_{n,R_n}(R_n - 1) + \log(n/s_n^3)/\lambda_{n,R_n}$  whp. But from the definition of  $x_{n,R_n}$  and  $f_{n,R_n}$  it follows that  $f_{n,R_n}(x_{n,R_n}) \geq f_{n,R_n}(R_n) \vee 4e^\gamma f_{n,R_n}(1) \log n$ , so (4.19c) completes the proof.  $\square$

## 5 Coupling $K_n$ and the PWIT

In Theorem 3.26, we indicated that two random processes, the first passage exploration processes  $\mathcal{S}$  and  $\mathcal{B}$  on  $K_n$  and  $\mathcal{T}^{(1,2)}$ , respectively, could be coupled. In this section we explain how this coupling arises as a special case of a general family of couplings between  $K_n$ , understood as a random edge-weighted graph with i.i.d. exponential edge weights, and the PWIT. We rely on some results from the companion paper [21], in particular [Part I, Theorem 3.4, Lemma 3.6 and Proposition 3.7].

## 5.1 Exploration Processes and the Definition of the Coupling

As in Sect. 3.5, we define  $M_{\emptyset_j} = j$ , for  $j = 1, 2$ , and to each  $v \in \mathcal{T}^{(1,2)} \setminus \{\emptyset_1, \emptyset_2\}$ , we associate a mark  $M_v$  chosen uniformly and independently from  $[n]$ . We next define what an exploration process is:

**Definition 5.1** (*Exploration process on two PWITs*) Let  $\mathcal{F}_0$  be a  $\sigma$ -field containing all null sets, and let  $(\mathcal{T}^{(1,2)}, X)$  be independent of  $\mathcal{F}_0$ . We call a sequence  $\mathcal{E} = (\mathcal{E}_k)_{k \in \mathbb{N}_0}$  of subsets of  $\mathcal{T}^{(1,2)}$  an *exploration process* if, with probability 1,  $\mathcal{E}_0 = \{\emptyset_1, \emptyset_2\}$  and, for every  $k \in \mathbb{N}$ , either  $\mathcal{E}_k = \mathcal{E}_{k-1}$  or else  $\mathcal{E}_k$  is formed by adjoining to  $\mathcal{E}_{k-1}$  a previously unexplored child  $v_k \in \partial \mathcal{E}_{k-1}$ , where the choice of  $v_k$  depends only on the weights  $X_w$  and marks  $M_w$  for vertices  $w \in \mathcal{E}_{k-1} \cup \partial \mathcal{E}_{k-1}$  and on events in  $\mathcal{F}_0$ .

Examples for exploration processes are given by FPP and IP on  $\mathcal{T}^{(1,2)}$ . For FPP, as defined in Definition 3.4, it is necessary to convert to discrete time by observing the branching process at those moments when a new vertex is added. The standard IP on  $\mathcal{T}^{(1,2)}$  is defined as follows. Set  $\text{IP}(0) = \{\emptyset_1, \emptyset_2\}$ . For  $k \in \mathbb{N}$ , form  $\text{IP}(k)$  inductively by adjoining to  $\text{IP}(k-1)$  the boundary vertex  $v \in \partial \text{IP}(k-1)$  of minimal weight. However, an exploration process is also obtained when we specify at each step (in any suitably measurable way) whether to perform an invasion step in  $\mathcal{T}^{(1)}$  or  $\mathcal{T}^{(2)}$ .

For  $k \in \mathbb{N}$ , let  $\mathcal{F}_k$  be the  $\sigma$ -field generated by  $\mathcal{F}_0$  together with the weights  $X_w$  and marks  $M_w$  for vertices  $w \in \mathcal{E}_{k-1} \cup \partial \mathcal{E}_{k-1}$ . Note that the requirement on the choice of  $v_k$  in Definition 5.1 can be expressed as the requirement that  $\mathcal{E}$  is  $(\mathcal{F}_k)_k$ -adapted.

For  $v \in \mathcal{T}^{(1,2)}$ , define the exploration time of  $v$  by

$$N_v = \inf \{k \in \mathbb{N}_0 : v \in \mathcal{E}_k\}. \quad (5.1)$$

**Definition 5.2** (*Thinning of two PWITs*) The vertex  $v \in \mathcal{T}^{(1,2)} \setminus \{\emptyset_1, \emptyset_2\}$  is *thinned* if it has an ancestor  $v_0 = p^k(v)$  (possibly  $v$  itself) such that  $M_{v_0} = M_w$  for some unthinned vertex  $w$  with  $N_w < N_{v_0}$ . Write  $\tilde{\mathcal{E}}_k$  for the subgraph of  $\mathcal{E}_k$  consisting of unthinned vertices.

We define the stopping times

$$N(i) = \inf \{k \in \mathbb{N}_0 : M_v = i \text{ for some } v \in \tilde{\mathcal{E}}_k\} \quad (5.2)$$

at which  $i \in [n]$  first appears as a mark in the unthinned exploration process. Note that, on the event  $\{N(i) < \infty\}$ ,  $\tilde{\mathcal{E}}_k$  contains a *unique* vertex in  $\mathcal{T}^{(1,2)}$  whose mark is  $i$ , for any  $k \geq N(i)$ ; call that vertex  $V(i)$ . On this event, we define

$$X(i, i') = \min \{X_w : M_w = i', p(w) = V(i)\}. \quad (5.3)$$

We define, for an edge  $\{i, i'\} \in E(K_n)$ ,

$$X_{\{i, i'\}}^{(K_n)} = \begin{cases} \frac{1}{n} X(i, i') & \text{if } N(i) < N(i'), \\ \frac{1}{n} X(i', i) & \text{if } N(i') < N(i), \\ E_{\{i, i'\}} & \text{if } N(i) = N(i') = \infty \text{ or } N(i) = N(i') = 0, \end{cases} \quad (5.4)$$

where  $(E_e)_{e \in E(K_n)}$  are exponential variables with mean 1, independent of each other and of  $(X_v)_{v \in \mathcal{T}^{(1,2)}}$ .

**Theorem 5.3** *If  $\mathcal{E}$  is an exploration process on the union  $\mathcal{T}^{(1,2)}$  of two PWITs, then the edge weights  $X_e^{(K_n)}$  defined in (5.4) are exponential with mean 1, independently for each  $e \in E(K_n)$ .*

The idea underlying Theorem 5.3 is that each variable  $\frac{1}{n}X(i, i')$  is exponentially distributed conditionally on the past up to the moment  $N(i)$  when it may be used to set the value of  $X_{\{i, i'\}}^{(K_n)}$ . Theorem 5.3 restates [Part I, Theorem 3.4] and is proved in that paper.

## 5.2 Minimal-Rule Exploration Processes

An important class of exploration processes, which includes both FPP and IP, are those exploration processes determined by a minimal rule in the following sense:

**Definition 5.4** A *minimal rule* for an exploration process  $\mathcal{E}$  on  $\mathcal{T}^{(1,2)}$  is an  $(\mathcal{F}_k)_k$ -adapted sequence  $(S_k, \prec_k)_{k=1}^\infty$ , where  $S_k \subset \partial\mathcal{E}_{k-1}$  is a (possibly empty) subset of the boundary vertices of  $\mathcal{E}_{k-1}$  and  $\prec_k$  is a strict total ordering of the elements of  $S_k$  (if any) such that the implication

$$w \in S_k, \quad p(v) = p(w), \quad M_v = M_w, \quad X_v < X_w \implies v \in S_k, v \prec_k w \quad (5.5)$$

holds. An exploration process is *determined by the minimal rule*  $(S_k, \prec_k)_{k=1}^\infty$  if  $\mathcal{E}_k = \mathcal{E}_{k-1}$  whenever  $S_k = \emptyset$  and otherwise  $\mathcal{E}_k$  is formed by adjoining to  $\mathcal{E}_{k-1}$  the unique vertex  $v_k \in S_k$  that is minimal with respect to  $\prec_k$ .

In words, in every step  $k$  there is a set of boundary vertices  $S_k$  from which we can select for the next exploration step. The content of (5.5) is that, whenever a vertex  $w \in S_k$  is available for selection, then all siblings of  $w$  with the same mark but smaller weight are also available for selection and are preferred over  $w$ .

For FPP without freezing on  $\mathcal{T}^{(1,2)}$  with edge weights  $f_n(X_v)$ , we take  $v \prec_k w$  if and only if  $T_v < T_w$  [recall (3.10)] and take  $S_k = \partial\mathcal{E}_{k-1}$ . For IP on  $\mathcal{T}^{(1,2)}$ , we have  $v \prec_k w$  if and only if  $X_v < X_w$ ; the choice of subset  $S_k$  can be used to enforce, for instance, whether the  $k$ th step is taken in  $\mathcal{T}^{(1)}$  or  $\mathcal{T}^{(2)}$ .

Recall the subtree  $\tilde{\mathcal{E}}_k$  of unthinned vertices from Definition 5.2 and the subgraph  $\pi_M(\tilde{\mathcal{E}}_k)$  from Definition 3.7. That is,  $\pi_M(\tilde{\mathcal{E}}_k)$  is the union of two trees with roots 1 and 2, respectively, and for  $v \in \tilde{\mathcal{E}}_k \setminus \{\emptyset_1, \emptyset_2\}$ ,  $\pi_M(\tilde{\mathcal{E}}_k)$  contains vertices  $M_v$  and  $M_{p(v)}$  and the edge  $\{M_v, M_{p(v)}\}$ .

For any  $i \in [n]$  for which  $N(i) < \infty$ , recall that  $V(i)$  is the unique vertex of  $\tilde{\mathcal{E}}_k$  ( $k \geq N(i)$ ) for which  $M_{V(i)} = i$ . Define  $V(i, i')$  to be the first child of  $V(i)$  with mark  $i'$ .

Recalling (5.3), an equivalent characterization of  $V(i, i')$  is

$$X(i, i') = X_{V(i, i')}. \quad (5.6)$$

The following lemma shows that, for an exploration process determined by a minimal rule, unthinned vertices must have the form  $V(i, i')$ :

**Lemma 5.5** Suppose  $\mathcal{E}$  is an exploration process determined by a minimal rule  $(S_k, \prec_k)_{k=1}^\infty$  and  $k \in \mathbb{N}$  is such that  $\tilde{\mathcal{E}}_k \neq \tilde{\mathcal{E}}_{k-1}$ . Let  $i_k = M_{p(v_k)}$  and  $i'_k = M_{v_k}$ . Then  $v_k = V(i_k, i'_k)$ .

See the proof of [Part I, Lemma 3.6].

If  $\mathcal{E}$  is an exploration process determined by a minimal rule, then we define

$$S_k^{(K_n)} = \{\{i, i'\} \in E(K_n) : i \in \pi_M(\tilde{\mathcal{E}}_{k-1}), i' \notin \pi_M(\tilde{\mathcal{E}}_{k-1}), V(i, i') \in S_k\} \quad (5.7)$$

and

$$e_1 \preceq_k e_2 \iff V(i_1, i'_1) \prec_k V(i_2, i'_2), \quad e_1, e_2 \in S_k^{(K_n)}, \quad (5.8)$$

where  $e_j = \{i_j, i'_j\}$  and  $i_j \in \pi_M(\tilde{\mathcal{E}}_{k-1})$ ,  $i'_j \notin \pi_M(\tilde{\mathcal{E}}_{k-1})$  as in (5.7).



**Proposition 5.6** (Thinned minimal rule) *Suppose  $\mathcal{E}$  is an exploration process determined by a minimal rule  $(S_k, \prec_k)_{k=1}^\infty$ . Then, under the edge-weight coupling (5.4), the edge weights of  $\pi_M(\tilde{\mathcal{E}}_k)$  are determined by*

$$X_{\{M_v, M_{p(v)}\}}^{(K_n)} = \frac{1}{n} X_v \quad \text{for any } v \in \cup_{k=1}^\infty \tilde{\mathcal{E}}_k \setminus \{\emptyset_1, \emptyset_2\} \quad (5.9)$$

and generally

$$X_{\{i, i'\}}^{(K_n)} = \frac{1}{n} X_{V(i, i')} \quad \text{whenever } i \in \pi_M(\tilde{\mathcal{E}}_{k-1}), \quad i' \notin \pi_M(\tilde{\mathcal{E}}_{k-1}) \quad \text{for some } k \in \mathbb{N}. \quad (5.10)$$

Moreover, for any  $k \in \mathbb{N}$  for which  $\tilde{\mathcal{E}}_k \neq \tilde{\mathcal{E}}_{k-1}$ ,  $\pi_M(\tilde{\mathcal{E}}_k)$  is formed by adjoining to  $\pi_M(\tilde{\mathcal{E}}_{k-1})$  the unique edge  $e_k \in S_k^{(K_n)}$  that is minimal with respect to  $\tilde{\prec}_k$ .

Proposition 5.6 asserts that the subgraph  $\pi_M(\tilde{\mathcal{E}}_k)$  of  $K_n$ , equipped with the edge weights  $(X_e^{(K_n)})_{e \in E(\pi_M(\tilde{\mathcal{E}}_k))}$ , is isomorphic as an edge-weighted graph to the subgraph  $\tilde{\mathcal{E}}_k$  of  $\mathcal{T}^{(1,2)}$ , equipped with the edge weights  $(\frac{1}{n} X_v)_{v \in \tilde{\mathcal{E}}_k \setminus \{\emptyset_1, \emptyset_2\}}$ . Furthermore, the subgraphs  $\pi_M(\tilde{\mathcal{E}}_k)$  can be grown by an inductive rule. Thus the induced subgraphs  $(\pi_M(\tilde{\mathcal{E}}_k))_{k=0}^\infty$  themselves form a minimal-rule exploration process on  $K_n$ , with a minimal rule derived from that of  $\mathcal{E}$ , with the caveat that  $\tilde{\prec}_k$  may depend on edge weights from  $\mathcal{E}_{k-1} \setminus \tilde{\mathcal{E}}_{k-1}$  as well as from  $\pi_M(\tilde{\mathcal{E}}_{k-1})$ .

See the proof of [Part I, Proposition 3.7] for the proof of Proposition 5.6.

### 5.3 FPP and the Two Smallest-Weight Trees: Proof of Theorem 3.22

In this section, we discuss the relationship between  $\mathcal{S}$  and FPP distances, and we prove Lemma 3.20 and Theorem 3.22.

**Proof of Lemma 3.20** Define

$$\tilde{\mathcal{S}}_u = \tilde{\mathcal{S}}_u^{(1)} \cup \tilde{\mathcal{S}}_u^{(2)}, \quad \tilde{\mathcal{S}}_u^{(j)} = SWT_{R_j(u)}^{(j)}. \quad (5.11)$$

To describe the discrete-time evolution of  $\tilde{\mathcal{S}} = (\tilde{\mathcal{S}}_u)_{u \geq 0}$ , denote by  $\bar{\tau}_{k-1}$  the time where the  $(k-1)$ th vertex (not including vertices 1 and 2) was added to  $\tilde{\mathcal{S}}$ . At time  $u = 0$ ,  $\tilde{\mathcal{S}}_0$  is equal to  $\mathcal{S}_0$ , and contains vertex 1 and 2 and no edges. Having constructed the process until time  $\bar{\tau}_{k-1}$ , at time

$$\bar{\tau}_k = \min_{j \in \{1,2\}} \min_{e \in \partial \tilde{\mathcal{S}}_{\bar{\tau}_{k-1}}^{(j)}} R_j^{-1} (d_{K_n, Y^{(K_n)}}(j, e) + Y_e^{(K_n)}) \quad (5.12)$$

adjoin the boundary edge  $e_k$  to  $\tilde{\mathcal{S}}_{\bar{\tau}_{k-1}}^{(j)}$ , where  $(j_k, e_k)$  is the minimizer in (5.12). [As in (3.39),  $e$  is an edge from  $\underline{e} \in \tilde{\mathcal{S}}_{\bar{\tau}_{k-1}}^{(j)}$  to  $\bar{e} \notin \tilde{\mathcal{S}}_{\bar{\tau}_{k-1}}^{(j)}$ ]. Note that the arrival times in  $\tilde{\mathcal{S}}^{(j)}$  are given by

$$T^{\mathcal{S}^{(j)}}(e) = R_j^{-1} (d_{K_n, Y^{(K_n)}}(j, e)) \quad (5.13)$$

by construction, so that we may rewrite (5.12) as

$$\bar{\tau}_k = \min_{j \in \{1,2\}} \min_{e \in \partial \tilde{\mathcal{S}}_{\bar{\tau}_{k-1}}^{(j)}} R_j^{-1} (R_j(T^{\mathcal{S}^{(j)}}(e)) + Y_e^{(K_n)}). \quad (5.14)$$

Comparing (5.14) with (3.39),  $\mathcal{S}$  and  $\tilde{\mathcal{S}}$  will evolve in the same way until the time

$$\bar{\tau} = \min \{t : \tilde{\mathcal{S}}_t^{(1)} \cap \tilde{\mathcal{S}}_t^{(2)} \neq \emptyset\} \quad (5.15)$$

when  $\bar{S}$  first accepts an edge between  $\bar{S}^{(1)}$  and  $\bar{S}^{(2)}$ . In particular, the minimization problem in (5.14) will be the same as in (3.40), and the minimizer will be a.s. unique, as long as  $\bar{\tau}_k \leq \bar{\tau}$ . Therefore we can choose  $J, J'$  with  $\{J, J'\} = \{1, 2\}$  and  $I \in \bar{S}_{\bar{\tau}-}^{(J)}, I' \in \bar{S}_{\bar{\tau}-}^{(J')}$  such that, at time  $\bar{\tau}$ , the edge between  $I$  and  $I'$  is adjoined to  $\bar{S}^{(J)}$ . [In other words,  $j = J$  and  $e = \{I, I'\}$  is the minimizer in (5.14)].

Because the minimizer in (5.14) is unique, no vertex is added to  $S$  at time  $\bar{\tau}$ . In particular,  $T^S(i) < \bar{\tau}$  for every  $i \in S_{\bar{\tau}}$ . Since  $S_t^{(j)}$  and  $\bar{S}_t^{(j)}$  agree for  $t < \bar{\tau}$ , the arrival times before  $\bar{\tau}$  must coincide. Recalling (5.13),

$$\begin{aligned} T^S(i) &= T^{\bar{S}^{(j)}}(i) = R_j^{-1}(d_{K_n, Y^{(K_n)}}(j, i)) \quad \text{if } i \in S_{\bar{\tau}}^{(j)}. \\ d_{K_n, Y^{(K_n)}}(j, i) &= R_j(T^S(i)) \end{aligned} \quad (5.16)$$

In addition,  $S_{\bar{\tau}}^{(j')}$  and  $\bar{S}_{\bar{\tau}}^{(j')}$  have the same vertex set while  $S_{\bar{\tau}}^{(j)}$  and  $\bar{S}_{\bar{\tau}}^{(j)}$  differ only by the vertex  $I'$ . It follows that

$$d_{K_n, Y^{(K_n)}}(j, i) \geq R_j(\tau) \quad \text{if } i \notin S_{\bar{\tau}}^{(j)}. \quad (5.17)$$

Consider the optimal path from vertex  $J$  to vertex  $I'$ . Since  $I'$  is adjoined to  $\bar{S}^{(j)}$  at time  $\bar{\tau}$ , it follows from (5.13) that  $d_{K_n, Y^{(K_n)}}(J, I') = R_J(\bar{\tau})$ . Moreover, since  $I$  is the parent of  $I'$  in  $\bar{S}_{\bar{\tau}}^{(j)} = SWT_{R_J(\bar{\tau})}^{(j)}$ , we have  $d_{K_n, Y^{(K_n)}}(J, I') = d_{K_n, Y^{(K_n)}}(J, I) + Y_{\{I, I'\}}^{(K_n)}$ . Applying (5.16) to the path from  $J$  to  $I$  to  $I'$  to  $J'$ ,

$$\begin{aligned} W_n &\leq d_{K_n, Y^{(K_n)}}(J, I) + Y_{\{I, I'\}}^{(K_n)} + d_{K_n, Y^{(K_n)}}(J', I') \\ &= R_J(T^S(I)) + Y_{\{I, I'\}}^{(K_n)} + R_{J'}(T^S(I')) \\ &= R_J(\tau) + R_{J'}(T^S(I')) \leq R_J(\bar{\tau}) + R_{J'}(\bar{\tau}). \end{aligned} \quad (5.18)$$

In particular, both sides of (3.41) are bounded above by  $R_1(\bar{\tau}) + R_2(\bar{\tau})$ .

The bound (5.18) will allow us to exclude vertices that arrive after time  $\bar{\tau}$ . To this end, we will show that (a) if a path  $\pi$  from vertex 1 to vertex 2 contains a vertex not belonging to  $S_{\bar{\tau}}$ , then the weight of  $\pi$  is greater than  $R_1(\bar{\tau}) + R_2(\bar{\tau})$ ; and (b) if  $i_1 \in S^{(1)} \setminus S_{\bar{\tau}}^{(1)}$  or  $i_2 \in S^{(2)} \setminus S_{\bar{\tau}}^{(2)}$ , then the term  $R_1(T^S(i_1)) + Y_{\{i_1, i_2\}}^{(K_n)} + R_2(T^S(i_2))$  in the minimum (3.41) is greater than  $R_1(\bar{\tau}) + R_2(\bar{\tau})$ .

For (a), suppose that  $\pi$  contains a vertex  $i \notin S_{\bar{\tau}}$ . Since the vertex sets of  $S_{\bar{\tau}}$  and  $\bar{S}_{\bar{\tau}}$  coincide, it follows that  $i \notin \bar{S}_{\bar{\tau}}$ , and by right continuity  $i \notin \bar{S}_t$  for some  $t > \bar{\tau}$ . Since  $R_1 + R_2$  is strictly increasing, (5.13) shows that the weight of  $\pi$  is at least

$$d_{K_n, Y^{(K_n)}}(1, i) + d_{K_n, Y^{(K_n)}}(2, i) \geq R_1(t) + R_2(t) > R_1(\bar{\tau}) + R_2(\bar{\tau}). \quad (5.19)$$

For (b), suppose for specificity that  $i_1 \in S^{(1)} \setminus S_{\bar{\tau}}^{(1)}$ . If in addition  $i_2 \in S^{(2)} \setminus S_{\bar{\tau}}^{(2)}$  then  $T^S(i_1), T^S(i_2) > \bar{\tau}$  and the strict monotonicity of  $R_1 + R_2$  gives the desired result. We may therefore suppose  $i_2 \in S_{\bar{\tau}}^{(2)}$ . Since  $S^{(1)}$  and  $S^{(2)}$  are disjoint, we must have  $i_1 \notin S_{\bar{\tau}}^{(2)}$ , so that  $d_{K_n, Y^{(K_n)}}(2, i_1) \geq R_2(\bar{\tau})$ . In particular, by considering the optimal path from 2 to  $i_2$  together with the edge from  $i_2$  to  $i_1$ , we conclude that  $d_{K_n, Y^{(K_n)}}(2, i_2) + Y_{\{i_1, i_2\}}^{(K_n)} \geq R_2(\bar{\tau})$ . By the assumption  $i_2 \in S_{\bar{\tau}}^{(2)}$ , we may rewrite this as  $R_2(T^S(i_2)) + Y_{\{i_1, i_2\}}^{(K_n)} \geq R_2(\bar{\tau})$ . Together with  $R_1(T^S(i_1)) > R_1(\bar{\tau})$ , this proves (b).

To complete the proof, consider a path  $\pi$  from vertex 1 to vertex 2. By statement (a) and (5.18),  $\pi$  must contain only vertices from  $S_{\bar{\tau}}$  if it is to be optimal. Since  $1 \in S_{\bar{\tau}}^{(1)}$  but  $2 \in S_{\bar{\tau}}^{(2)}$ , it follows that  $\pi$  must contain an edge between some pair of vertices  $i_1 \in S_{\bar{\tau}}^{(1)}$  and  $i_2 \in S_{\bar{\tau}}^{(2)}$ . The minimum possible weight of such a path is  $d_{K_n, Y^{(K_n)}}(1, i_1) + Y_{\{i_1, i_2\}}^{(K_n)} + d_{K_n, Y^{(K_n)}}(2, i_2)$ ,

which agrees with the corresponding term in (3.41) by (5.16). Therefore (3.41) is verified if the minimum is taken only over  $i_1 \in \mathcal{S}_{\bar{\tau}}^{(1)}$ ,  $i_2 \in \mathcal{S}_{\bar{\tau}}^{(2)}$ . But statement (b) and (5.18) shows that the remaining terms must be strictly greater.  $\square$

**Proof of Theorem 3.22** Since  $R_1 + R_2$  is strictly increasing, the relation  $W_n = R_1(T_{\text{coll}}) + R_2(T_{\text{coll}})$  is a reformulation of Definition 3.21.

Recall the time  $\bar{\tau}$  from the proof of Lemma 3.20. We showed there that the minimizer of (3.41) must come from vertices  $i_1 \in \mathcal{S}_{\bar{\tau}}^{(1)}$ ,  $i_2 \in \mathcal{S}_{\bar{\tau}}^{(2)}$ . In particular, by (5.16),

$$W_n = R_1(T^{\mathcal{S}}(I_1)) + Y_{\{I_1, I_2\}}^{(\kappa_n)} + R_2(T^{\mathcal{S}}(I_2)) \quad (5.20)$$

expresses  $W_n$  as the weight of the path formed as the union of the optimal path from 1 to  $I_1$ ; the edge from  $I_1$  to  $I_2$ ; and the optimal path from  $I_2$  to 2. In particular,  $\pi_{1,2}$  is the same as this path. Since  $T^{\mathcal{S}}(I_j) < \tau$  and  $\mathcal{S}_t^{(j)} = SWT_{R_j(t)}^{(j)}$  for  $t < \bar{\tau}$ , it follows that the optimal paths from  $j$  to  $I_j$  coincide with the unique paths in  $\mathcal{S}_{\bar{\tau}}^{(j)}$  between these vertices. The relation  $H_n = H(I_1) + H(I_2) + 1$  follows by counting the edges in these subpaths.

It remains to show that  $T^{\mathcal{S}}(I_j) < T_{\text{coll}}$ . Define  $t_1 = R_1^{-1}(R_1(T^{\mathcal{S}}(I_1)) + Y_{\{I_1, I_2\}}^{(\kappa_n)})$ . Recalling (3.39), we see that  $t_1$  is the time at which the edge from  $I_1$  to  $I_2$  is adjoined to  $\mathcal{S}^{(1)}$ , provided that  $I_2$  has not already been added to  $\mathcal{S}$  at some point strictly before time  $t_1$ . By construction,  $I_2$  is added to  $\mathcal{S}^{(2)}$ , not  $\mathcal{S}^{(1)}$ , so it must be that  $T^{\mathcal{S}}(I_2) < t_1$ . [Equality is not possible because of our assumption that the minimizers of (3.39) are unique.] Aiming for a contradiction, suppose that  $T^{\mathcal{S}}(I_2) \geq T_{\text{coll}}$ . Comparing the relation  $W_n = R_1(T_{\text{coll}}) + R_2(T_{\text{coll}})$  to (5.20) gives  $R_1(T^{\mathcal{S}}(I_2)) + Y_{\{I_1, I_2\}}^{(\kappa_n)} \leq R_1(T_{\text{coll}})$ , so that  $t_1 \leq T_{\text{coll}}$ . This is a contradiction since  $t_1 > T^{\mathcal{S}}(I_2) \geq T_{\text{coll}}$ . Similarly we must have  $T^{\mathcal{S}}(I_1) < T_{\text{coll}}$ . This shows that the unique paths in  $\mathcal{S}_{\bar{\tau}}^{(j)}$  from  $j$  to  $I_j$  are actually paths in  $\mathcal{S}_{T_{\text{coll}}}^{(j)}$ , as claimed.  $\square$

## 5.4 $\mathcal{B}$ and $\mathcal{S}$ as Exploration Processes: Proof of Theorem 3.26

Before proving Theorem 3.26, we show that the discrete-time analogue of  $\mathcal{B}$  is an exploration process determined by a minimal rule:

**Lemma 5.7** *Let  $v_k$  denote the  $k$ th vertex added to  $\mathcal{B}$ , excluding the root vertices  $\emptyset_1, \emptyset_2$ , and set  $\mathcal{E}_k = \mathcal{B}_{T_{v_k}^{\mathcal{B}}}$  for  $k \geq 1$ ,  $\mathcal{E}_0 = \mathcal{B}_0 = \{\emptyset_1, \emptyset_2\}$ . Then  $\mathcal{E}$  is an exploration process determined by a minimal rule.*

**Proof** Consider the  $k$ th step and define

$$\tau_{\text{next}}^{(j)}(k) = \min_{v \in \partial \mathcal{E}_{k-1} \cap \mathcal{T}^{(j)}} T_v, \quad (5.21)$$

i.e., the next birth time of a vertex in  $\mathcal{T}^{(j)}$  in the absence of freezing, and let  $v_k^{(j)}$  denote the a.s. unique vertex attaining the minimum in (5.21). Recalling the definition of the filtration  $\mathcal{F}_k$ , we see that  $\tau_{\text{next}}^{(j)}(k)$  and  $v_k^{(j)}$  are  $\mathcal{F}_k$ -measurable.

The variable  $T_{\text{fr}}^{(j)}$  is not  $\mathcal{F}_k$ -measurable. However, the event  $\{T_{\text{fr}}^{(j)} < \tau_{\text{next}}^{(j)}(k)\}$  is  $\mathcal{F}_k$ -measurable. To see this, define

$$\tau_{\text{fr}}^{(j)}(k) = \inf \left\{ t \geq 0 : \sum_{v \in \mathcal{E}_{k-1} \cap \mathcal{T}^{(j)}} \mathbb{1}_{\{T_v \leq t\}} \int_{t-T_v}^{\infty} e^{-\lambda_n(1)(y-(t-T_v))} d\mu_n(y) \geq s_n \right\}, \quad (5.22)$$

so that  $\tau_{\text{fr}}^{(j)}(k)$  is  $\mathcal{F}_k$ -measurable, and abbreviate  $\tau_{\text{unfr}}(k) = \tau_{\text{fr}}^{(1)}(k) \vee \tau_{\text{fr}}^{(2)}(k)$ . We will use  $\tau_{\text{fr}}^{(j)}(k)$  as an approximation to  $T_{\text{fr}}^{(j)}$  based on the information available in  $\mathcal{F}_k$ . By analogy

with (3.35) and (3.37), we also define

$$\begin{aligned} R_{j,k}(t) &= (t \wedge \tau_{\text{fr}}^{(j)}(k)) + ((t - \tau_{\text{unfr}}(k)) \vee 0), \\ R_{j,k}^{-1}(t) &= \begin{cases} t & \text{if } t \leq \tau_{\text{fr}}^{(j)}(k), \\ \tau_{\text{unfr}}(k) - \tau_{\text{fr}}^{(j)}(k) + t & \text{if } t > \tau_{\text{fr}}^{(j)}(k). \end{cases} \end{aligned} \quad (5.23)$$

We note the following:

- (i)  $\mathcal{E}_{k-1} \cap \mathcal{T}^{(j)} = BP_{t'}^{(j)}$  and  $\tau_{\text{next}}^{(j)}(k) = \min \{T_v : v \in \partial BP_{t'}^{(j)}\}$ , where  $t' = T_{v_{k-1}}$ .
- (ii) The sum in (5.22) agrees with the sum in the definition (3.31) of  $T_{\text{fr}}^{(j)}$  whenever  $t < \tau_{\text{next}}^{(j)}(k)$ .
- (iii)  $T_{\text{fr}}^{(j)} < \tau_{\text{next}}^{(j)}(k)$  if and only if  $\tau_{\text{fr}}^{(j)}(k) < \tau_{\text{next}}^{(j)}(k)$ , and in this case  $T_{\text{fr}}^{(j)} = \tau_{\text{fr}}^{(j)}(k)$ .
- (iv) For  $j \in \{1, 2\}$ , let  $I_k^{(j)} \subseteq [0, \tau_{\text{next}}^{(j)}(k)]$  be nonempty. The minimizers of  $R_{j,k}^{-1}(t)$  and  $R_j^{-1}(t)$  over all pairs  $(t, j)$  with  $t \in I_k^{(j)}$  and  $j \in \{1, 2\}$  agree and return the same value.

Statement (i) follows by induction and (3.36). For (ii), note that the sums in (3.31) and (5.22) agree whenever  $BP_{t'}^{(j)} \subset \mathcal{E}_{k-1} \cap \mathcal{T}^{(j)}$ , so using (i) it suffices to show that  $BP_{t'}^{(j)} \subset BP_{t'}^{(j)}$  for  $t < \tau_{\text{next}}^{(j)}(k)$ . But for any  $t'$  and any  $t < \min \{T_v : v \in \partial BP_{t'}^{(j)}\}$  we have  $BP_t^{(j)} \subset BP_{t'}^{(j)}$  by definition, so the second part of (i) completes the proof.

Statement (iii) follows from (ii): if one of the sums in (3.31) or (5.22) exceeds  $s_n$  before time  $\tau_{\text{next}}^{(j)}(k)$ , then so does the other, and the first times where they do so are the same. In particular,  $\{T_{\text{fr}}^{(j)} < \tau_{\text{next}}^{(j)}(k)\}$  is  $\mathcal{F}_k$ -measurable because  $\tau_{\text{fr}}^{(j)}(k)$  and  $\tau_{\text{next}}^{(j)}(k)$  are.

To prove (iv), we distinguish three cases. If  $\tau_{\text{fr}}^{(j)}(k) \geq \tau_{\text{next}}^{(j)}(k)$ , then  $R_{j,k}^{-1}$  and  $R_j^{-1}$  reduce to the identity on  $[0, \tau_{\text{next}}^{(j)}(k)]$  by (iii). Hence, (iv) holds if  $\tau_{\text{fr}}^{(j)}(k) \geq \tau_{\text{next}}^{(j)}(k)$  for both  $j \in \{1, 2\}$ . If  $\tau_{\text{fr}}^{(j)}(k) < \tau_{\text{next}}^{(j)}(k)$  for both  $j \in \{1, 2\}$ , then  $R_{j,k}^{-1}$  and  $R_j^{-1}$  agree everywhere according to (iii). Finally, consider the case that  $\tau_{\text{fr}}^{(j)}(k) \geq \tau_{\text{next}}^{(j)}(k)$  and  $\tau_{\text{fr}}^{(j')}(k) < \tau_{\text{next}}^{(j')}(k)$  for  $\{j, j'\} = \{1, 2\}$ . Then  $T_{\text{fr}}^{(j)} \geq \tau_{\text{next}}^{(j)}(k)$  and, therefore,  $R_{j,k}^{-1}(t) = R_j^{-1}(t) = t$  on  $I_k^{(j)}$ . Moreover,  $T_{\text{fr}}^{(j')} = \tau_{\text{fr}}^{(j')}(k)$  implying that  $R_{j',k}^{-1}(t) = R_{j'}^{-1}(t) = t$  on  $[0, \tau_{\text{fr}}^{(j')}(k)]$  and for  $t \in (\tau_{\text{fr}}^{(j')}(k), \tau_{\text{next}}^{(j')}(k)]$ ,  $R_{j',k}^{-1}(t) \geq \tau_{\text{unfr}}(k) \geq \tau_{\text{next}}^{(j)}(k)$  and  $R_j^{-1}(t) \geq T_{\text{unfr}} \geq \tau_{\text{next}}^{(j)}(k)$ . Hence, in all three cases, the functions agree on the relevant domain and we have proved (iv).

Set  $(v_k^\mathcal{E}, j_k^\mathcal{E})$  to be the pair that minimizes  $R_{j,k}^{-1}(T_v)$  among all pairs  $(v, j)$  with  $v \in \partial \mathcal{E}_{k-1} \cap \mathcal{T}^{(j)}$ ,  $j \in \{1, 2\}$ . Note that  $R_{j,k}^{-1}(t)$  can be infinite when  $\tau_{\text{fr}}^{(j)}(k) < \infty$  and  $\tau_{\text{fr}}^{(j')}(k) = \infty$ , where  $\{j, j'\} = \{1, 2\}$ , but in this case  $R_{j',k}^{-1}(t)$  must be finite for all  $t$ . Furthermore  $R_{j,k}^{-1}$  is strictly increasing whenever it is finite. Recalling that the times  $T_v$  are formed from variables with continuous distributions, it follows that the minimizing pair  $(v_k^\mathcal{E}, j_k^\mathcal{E})$  is well defined a.s.

Since  $R_{j,k}^{-1}$  is  $\mathcal{F}_k$ -measurable, the choice of  $v_k^\mathcal{E}$  is determined by a minimal rule with  $S_k = \partial \mathcal{E}_{k-1}$ . To complete the proof, we must show that  $v_k^\mathcal{E} = v_k$ .

The vertex  $v_k$  is the first coordinate of the pair  $(v_k, j_k)$  that minimizes  $T_v^\mathcal{B} = R_j^{-1}(T_v)$  over all pairs  $(v, j)$  with  $v \in \partial \mathcal{B}_{T_{v_{k-1}}}^\mathcal{B}$ ,  $j \in \{1, 2\}$ . (Once again, this minimizing pair is well defined a.s., reflecting the fact that, a.s.,  $\mathcal{B}$  never adds more than one vertex at a time.) Since  $\mathcal{E}_{k-1} = \mathcal{B}_{T_{v_{k-1}}}^\mathcal{B}$ , the pairs  $(v_k^\mathcal{E}, j_k^\mathcal{E})$  and  $(v_k, j_k)$  are both minimizers over the set  $\{(v, j) : v \in \partial \mathcal{E}_{k-1} \cap \mathcal{T}^{(j)}\}$ . Moreover, since  $R_j^{-1}$  and  $R_{j,k}^{-1}$  are both strictly increasing (when finite), both minimizations can be restricted to the set  $\{(v_k^{(j)}, j) : j = 1, 2\}$ , where  $v_k^{(j)}$  is the minimizer in (5.21). Since  $T_{v_k^{(j)}} = \tau_{\text{next}}^{(j)}(k)$ , statement (iv) with  $I_{j,k} = \{\tau_{\text{next}}^{(j)}(k)\}$  implies  $v_k^\mathcal{E} = v_k$ , as claimed.  $\square$

**Proof of Proposition 3.25** By Theorem 5.3, the edge weights  $X_e^{(K_n)}$  associated via (5.4) to the exploration process in Lemma 5.7 are independent exponential random variables with mean 1. Recalling (2.1)–(2.2), the corresponding edge weights  $Y_e^{(K_n)} = g(X_e^{(K_n)})$  are independent with distribution function  $F_Y$ . To complete the proof it suffices to observe that  $N(i) < N(i')$  if and only if  $T^{\tilde{B}}(i) < T^{\tilde{B}}(i')$  and that  $T^{\tilde{B}}(i)$  is finite for all  $i \in [n]$  since the FPP process explores every edge in  $\mathcal{T}^{(1,2)}$  eventually. Hence, definitions (3.46) and (5.4) agree.  $\square$

**Proof of Theorem 3.26** We resume the notation from the proof of Lemma 5.7. Create the edge weights on  $K_n$  according to (5.4). Denote by  $\tau_{k'-1}$  the time where the  $(k' - 1)$ th vertex (not including the vertices 1 and 2) was added to  $\mathcal{S}$  [see (3.39)]. As in the proof of Theorem 3.8, both  $\tilde{B}$  and  $\mathcal{S}$  are increasing jump processes and  $\pi_M(\tilde{B}_0) = \mathcal{S}_0$ . By an inductive argument we can suppose that  $k, k' \in \mathbb{N}$  are such that  $\tilde{\mathcal{E}}_k \neq \tilde{\mathcal{E}}_{k-1}$  and  $\pi_M(\tilde{\mathcal{E}}_{k-1}) = \mathcal{S}_{\tau_{k'-1}}$ . The proof will be complete if we can prove that (a) the edge  $e'_{k'}$  adjoined to  $\mathcal{S}_{\tau_{k'-1}}$  to form  $\mathcal{S}_{\tau_{k'}}$  is the same as the edge  $e_k = \{i_k, i'_k\}$  adjoined to  $\pi_M(\tilde{\mathcal{E}}_{k-1})$  to form  $\pi_M(\tilde{\mathcal{E}}_k)$ ; and (b)  $\tau_{k'} = T_{v_k}^{\mathcal{B}}$ .

Let  $i \in \mathcal{S}_{\tau_{k'-1}}$ , and let  $j \in \{1, 2\}$  be such that  $i \in \mathcal{S}_{\tau_{k'-1}}^{(j)}$ . By the inductive hypothesis,  $V(i) \in \tilde{\mathcal{E}}_{k-1} \cap \mathcal{T}^{(j)}$  and the unique path in  $\mathcal{S}_{\tau_{k'-1}}$  from  $i$  to  $j$  is the image of the unique path in  $\tilde{\mathcal{E}}_{k-1}$  from  $V(i)$  to  $\emptyset_j$  under the mapping  $v \mapsto M_v$  (recall Definition 3.7). According to (5.9), (2.1) and (2.3), the edge weights along this path are  $Y_{\{M_{p^{m-1}(V(i))}, M_{p^m(V(i))}\}}^{(K_n)} = g(X_{\{M_{p^{m-1}(V(i))}, M_{p^m(V(i))}\}}^{(K_n)}) = g\left(\frac{1}{n} X_{p^{m-1}(V(i))}\right)$  and  $f_n(X_{p^{m-1}(V(i))})$ , for  $m = 1, \dots, |V(i)|$ . Summing gives  $d_{\mathcal{S}_{\tau_{k'-1}}, Y^{(K_n)}}(j, i) = T_{V(i)}$ .

In addition, let  $i' \notin \mathcal{S}_{\tau_{k'-1}}$  and write  $e = \{i, i'\}$ . By (5.10),  $X_e^{(K_n)} = \frac{1}{n} X_{V(i, i')}$ , so that  $Y_e^{(K_n)} = f_n(X_{V(i, i')})$ . Thus the expression in the right-hand side of (3.39) reduces to  $R_j^{-1}(T_{V(i)} + f_n(X_{V(i, i')}))$ , i.e.,  $R_j^{-1}(T_{V(i, i')})$ . The edge  $e'_{k'}$  minimizes this expression over all  $i \in \mathcal{S}_{\tau_{k'-1}}$ ,  $i' \notin \mathcal{S}_{\tau_{k'-1}}$ . By Proposition 5.6, the edge  $e_k = \{i_k, i'_k\}$  minimizes  $R_{j,k}^{-1}(T_{V(i, i')})$  over all  $i \in \pi_M(\tilde{\mathcal{E}}_{k-1})$ ,  $i' \notin \pi_M(\tilde{\mathcal{E}}_{k-1})$  (with  $j$  such that  $V(i) \in \mathcal{E}_{k-1}^{(j)}$ ). By the induction hypothesis, statement (iv) in the proof of Lemma 5.7 and monotonicity, these two minimization problems have the same minimizers, proving (a), and return the same value, which completes the proof of (b).  $\square$

## 5.5 Coupling and Cox Processes: Proof of Theorem 3.27

In this section we explain the modified coupling in, and give the proof of Theorem 3.27.

**Proof of Theorem 3.27** The edge-weight coupling (3.46) selects the edge weights  $Y_{\{i_1, i_2\}}^{(K_n)}$  for  $i_1 \in \mathcal{S}^{(1)}$ ,  $i_2 \in \mathcal{S}^{(2)}$  based on values  $f_n(X_w)$  for which  $p(w) \in \widetilde{B}\widetilde{P}$  and  $\{M_{p(w)}, M_w\} = \{i_1, i_2\}$ . Under the present definition of  $\mathcal{B}$  [see (3.36)], such vertices  $w$  are eventually explored, and consequently the values  $f_n(X_w)$  can be recovered by observing  $(\mathcal{B}_t)_{t \geq 0}$ . On the other hand, in the context of Theorem 3.27, we want the values  $f_n(X_w)$  to behave as a Cox process [with respect to  $\mathcal{B}$ ,  $R_1$ ,  $R_2$ ,  $(M_v)_{v \in \mathcal{B}}$ ]. For this reason, we will modify the definition of  $\mathcal{B}$  so that it does not explore vertices  $w$  of this kind, and we will replace the contribution of those vertices using an additional source of randomness.<sup>2</sup>

As always, we have two independent PWITs  $(\mathcal{T}^{(j)}, X^{(j)})$ ,  $j \in \{1, 2\}$ , the marks  $M_v$ ,  $v \in \mathcal{T}^{(1,2)}$ , and a family of independent exponential random variables  $E_e$ ,  $e \in E(K_\infty)$ , with

<sup>2</sup> Alternatively, we could retain the definition of  $\mathcal{B}$  and set the edge weights  $Y_{\{i_1, i_2\}}^{(K_n)}$  for  $i_1 \in \mathcal{S}^{(1)}$ ,  $i_2 \in \mathcal{S}^{(2)}$  using an additional source of randomness. We prefer to avoid this option because it would entail proving a more complicated version of Theorem 5.3.

mean 1, independent of the PWITs and the marks. In addition, from each vertex  $v$  we initialise an independent PWIT with vertices  $(v, w')$ , edge weights  $X'_{(v,w')}$  [such that  $(X'_{(v,w'k)})_{k=1}^\infty$  forms a Poisson point process with rate 1 on  $(0, \infty)$ ] and marks  $M'_{(v,w')}$  uniform on  $[n]$ , all independent of each other and of the original variables  $X_v, M_v$ .

First consider  $(\mathcal{B}, \tilde{\mathcal{B}}, R_1, R_2)$  as normally constructed, without using the auxiliary PWITs with vertices  $(v, w')$ . Fix  $i, i' \in [n], i \neq i'$ , and suppose for definiteness that  $T^{\tilde{\mathcal{B}}}(i) < T^{\tilde{\mathcal{B}}}(i')$ . [If instead  $T^{\tilde{\mathcal{B}}}(i') < T^{\tilde{\mathcal{B}}}(i)$ , interchange the roles of  $i$  and  $i'$  in the following discussion.] According to (3.46) or (5.4), the edge weight  $X_{\{i,i'\}}^{(K_n)}$  is set to be  $\frac{1}{n}X(i, i')$ , where  $X(i, i')$  is the first point in the Poisson point process

$$\mathcal{P}^{(i,i')} = \sum_{w: p(w)=V(i), M_w=i'} \delta_{X_w} \quad (5.24)$$

of intensity  $1/n$ .

Now condition on  $V(i)$  and  $V(i')$  belonging to different trees, say  $V(i) \in \mathcal{T}^{(J)}, V(i') \in \mathcal{T}^{(J')}$  where  $\{J, J'\} = \{1, 2\}$ . For this to happen, the children of  $V(i)$  having mark  $i'$  must *not* have been explored by time  $T_{V(i')}^{\tilde{\mathcal{B}}}$ . A child  $w$  of  $V(i)$  is explored at time  $R_J^{-1}(T_w) = R_J^{-1}(T_{V(i)} + f_n(X_w))$ , so we can reformulate this by saying that the Poisson point process  $\mathcal{P}^{(i,i')}$  must contain no point  $x$  with  $R_J^{-1}(T_{V(i)} + f_n(x)) \leq T_{V(i')}^{\tilde{\mathcal{B}}}$ . Using the relation  $T_v = R_J(T_v^{\tilde{\mathcal{B}}})$  for  $v \in \mathcal{T}^{(J)}$ , we can rewrite this as the condition that  $\mathcal{P}^{(i,i')}$  contains no point  $x$  with  $f_n(x) \leq R_J(T_{V(i')}^{\tilde{\mathcal{B}}}) - R_J(T_{V(i)}^{\tilde{\mathcal{B}}}) = \Delta R_{V(i), V(i')}$ . However, the condition gives no information about points of larger value. It follows that, conditionally on  $(\mathcal{B}_t, \tilde{\mathcal{B}}_t, R_1(t), R_2(t))_{t \leq T^{\tilde{\mathcal{B}}}(i')}$ ,  $\mathcal{P}^{(i,i')}$  is a Poisson point process of intensity  $1/n$  on  $(f_n^{-1}(\Delta R_{V(i), V(i')}), \infty)$ .

To preserve this property when conditioning on  $(\mathcal{B}_t, R_1(t), R_2(t), (M_v)_{v \in \mathcal{B}_t})$  for  $t > T^{\tilde{\mathcal{B}}}(i')$ , replace the edge weights and marks of all vertices  $w$  (and their descendants) for which  $p(w) = V(i), M_w = i'$  and  $f_n(X_w) > \Delta R_{V(i), V(i')}$ , by the edge weights and marks of the auxiliary vertices  $(V(i), w')$  (and their descendants) for which  $p(w') = V(i), M'_{(V(i), w')} = i'$  and  $f_n(X'_{(V(i), w')}) > \Delta R_{V(i), V(i')}$ . Modify the evolution of  $\mathcal{B}, R_1, R_2$  for  $t > T^{\tilde{\mathcal{B}}}(i')$  so as to use the replacement edge weights  $X'_{v, w'}$ , but continue to use the original edge weights  $X_w$  in the edge-weight coupling (5.4). [Formally, modify the minimal rule from Lemma 5.7 so that vertices are ineligible for exploring once they have been replaced, but add to the sum in (5.22) the contribution from any replacement vertices from the auxiliary PWITs that would have been explored at time  $t$ .]

These replacements do not change the law of  $\mathcal{B}, R_1, R_2, (M_v)_{v \in \mathcal{B}}$ , or the edge weights  $X_e^{(K_n)}$ . The pointwise equality between  $\pi_M(\tilde{\mathcal{B}})$  and  $\mathcal{S}$  is unaffected: the replaced vertices are thinned and therefore do not affect  $\tilde{\mathcal{B}}$ . Finally, the evolution of  $\tilde{\mathcal{B}}$  for  $t > T^{\tilde{\mathcal{B}}}(i_2)$  now gives no additional information about the edge weights  $\{X_w: p(w) = V(i_1), M_w = i_2, f_n(X_w) > \Delta R_{V(i), V(i')}\}$ . In particular, conditionally on  $\mathcal{B}, R_1, R_2, (M_v)_{v \in \mathcal{B}}$  and the event that  $V(i) \in \mathcal{T}^{(J)}, V(i') \in \mathcal{T}^{(J')}$  for some choice of  $\{J, J'\} = \{1, 2\}$ , the law of  $\mathcal{P}^{(i,i')}$  is that of a Poisson point process with intensity measure  $1/n$  on  $(f_n^{-1}(\Delta R_{V(i), V(i')}), \infty)$ . Furthermore, the Poisson point processes corresponding to different  $i, i'$  will be conditionally independent.

We can now give an explicit construction of  $\mathcal{P}_n$ . We begin by defining  $\mathcal{P}_n$  on the subspace given by unthinned pairs of vertices:

for  $i_1, i_2 \in [n]$  such that  $V(i_1) \in \mathcal{T}^{(1)}$ ,  $V(i_2) \in \mathcal{T}^{(2)}$ ,

and for  $\{j, j'\} = \{1, 2\}$  with  $T_{V(i_j)}^{\mathcal{B}} < T_{V(i_{j'})}^{\mathcal{B}}$ ,

$$\mathcal{P}_n|_{[0, \infty) \times \{V(i_1)\} \times \{V(i_2)\}} = \sum_{\substack{w: p(w)=V(i_j), \\ M_w=i_{j'}}} \delta_{((R_1+R_2)^{-1}(T_{V(i_1)}+f_n(X_w)+T_{V(i_2)}), V(i_1), V(i_2))}. \quad (5.25)$$

In the notation above,  $\mathcal{P}_n|_{[0, \infty) \times \{V(i_1)\} \times \{V(i_2)\}}$  is the image of  $\mathcal{P}^{(i_j, i_{j'})}$  under the mapping

$$x \mapsto (R_1 + R_2)^{-1}(T_{V(i_1)} + f_n(x) + T_{V(i_2)}). \quad (5.26)$$

In particular, by the remarks above,  $\mathcal{P}_n|_{[0, \infty) \times \tilde{\mathcal{B}}^{(1)} \times \tilde{\mathcal{B}}^{(2)}}$  has the conditional law of a Poisson point process conditionally on  $\mathcal{B}, R_1, R_2, (M_v)_{v \in \mathcal{B}}$ . To compute its intensity measure, note that the mapping  $x \mapsto y = f_n(x)$  sends  $1/n$  times Lebesgue measure on  $(f_n^{-1}(\Delta R_{V(i), V(i')}), \infty)$  to the measure  $\frac{1}{n} \mu_n|_{(\Delta R_{V(i), V(i')}, \infty)}$ . It follows that the further mapping  $y \mapsto (R_1 + R_2)^{-1}(T_{V(i_1)} + y + T_{V(i_2)})$  leads to the intensity measure specified by (3.48), where we have again used the relation  $T_v = R_j(T_v^{\mathcal{B}})$  for  $v \in \mathcal{T}^{(j)}$ . Thus  $\mathcal{P}_n|_{[0, \infty) \times \tilde{\mathcal{B}}^{(1)} \times \tilde{\mathcal{B}}^{(2)}}$  is a Cox process of the correct intensity.

Finally, we may extend  $\mathcal{P}_n$  to be a Cox process on  $[0, \infty) \times \mathcal{T}^{(1)} \times \mathcal{T}^{(2)}$  with the specified intensity, by defining  $\mathcal{P}_n|_{[0, \infty) \times \{v_1\} \times \{v_2\}}$  using an independent source of randomness for any pair of vertices  $v_1, v_2$  for which  $v_1 \in \mathcal{B}^{(1)} \setminus \tilde{\mathcal{B}}^{(1)}$  or  $v_2 \in \mathcal{B}^{(2)} \setminus \tilde{\mathcal{B}}^{(2)}$ . Note that the details of this extension are unimportant since such pairs  $(v_1, v_2)$  are not considered in the definition of  $(T_{\text{coll}}^{(\mathcal{P}_n)}, V_{\text{coll}}^{(1)}, V_{\text{coll}}^{(2)})$ .

Observe that under this construction of  $\mathcal{P}_n$  and under the edge-weight coupling (3.46),

$$(T_{\text{coll}}^{(\mathcal{P}_n)}, V_{\text{coll}}^{(1)}, V_{\text{coll}}^{(2)}) = (T_{\text{coll}}, V(I_1), V(I_2)). \quad (5.27)$$

Indeed, consider any  $i_1, i_2 \in [n]$  with  $V(i_1) \in \mathcal{T}^{(1)}$  and  $V(i_2) \in \mathcal{T}^{(2)}$ . We note that this assumption is equivalent to  $i_1 \in \mathcal{S}^{(1)}$  and  $i_2 \in \mathcal{S}^{(2)}$ . Taking  $\{J, J'\} = \{1, 2\}$  with  $T_{V(i_J)}^{\mathcal{B}} < T_{V(i_{J'})}^{\mathcal{B}}$ , (3.46) gives

$$Y_{\{i_1, i_2\}}^{(K_n)} = g\left(X_{\{i_1, i_2\}}^{(K_n)}\right) = g\left(\frac{1}{n}X(i_J, i_{J'})\right) = f_n(X(i_J, i_{J'})). \quad (5.28)$$

The value  $X(i_J, i_{J'})$  coincides with the first point of  $\mathcal{P}^{(i_J, i_{J'})}$ , and applying the increasing mapping (5.26) it follows that the first point of  $\mathcal{P}_n|_{[0, \infty) \times \{V(i_1)\} \times \{V(i_2)\}}$  has time coordinate

$$(R_1 + R_2)^{-1}\left(T_{V(i_1)} + Y_{\{i_1, i_2\}}^{(K_n)} + T_{V(i_2)}\right). \quad (5.29)$$

Using Lemma 3.20, the strict monotonicity of  $R_1 + R_2$ , and the relation  $T_{V(i)} = R_j(T_{V(i)}^{\mathcal{B}}) = R_j(T^{\mathcal{S}}(i))$  for  $V(i) \in \mathcal{T}^{(j)}$ ,  $i \in \mathcal{S}^{(j)}$ , we see that

$$\begin{aligned} T_{\text{coll}} &= (R_1 + R_2)^{-1}(W_n) = (R_1 + R_2)^{-1}\left(\min_{i_1 \in \mathcal{S}^{(1)}, i_2 \in \mathcal{S}^{(2)}} T_{V(i_1)} + Y_{\{i_1, i_2\}}^{(K_n)} + T_{V(i_2)}\right) \\ &= \min_{i_1 \in \mathcal{S}^{(1)}, i_2 \in \mathcal{S}^{(2)}} (R_1 + R_2)^{-1}\left(T_{V(i_1)} + Y_{\{i_1, i_2\}}^{(K_n)} + T_{V(i_2)}\right) \end{aligned} \quad (5.30)$$

is the result of minimizing (5.29) over all choices of  $i_1, i_2$ , and  $I_1, I_2$  are the corresponding minimizers. On the other hand,  $T_{\text{coll}}^{(\mathcal{P}_n)}$  is the result of minimizing the first point of  $\mathcal{P}_n|_{[0, \infty) \times \{v_1\} \times \{v_2\}}$  over all choices of unthinned vertices  $v_1 \in \mathcal{T}^{(1)}$ ,  $v_2 \in \mathcal{T}^{(2)}$ , and  $V_{\text{coll}}^{(1)}, V_{\text{coll}}^{(2)}$



are the corresponding minimizers. Every such pair  $(v_1, v_2)$  can be written as  $v_j = V(i_j)$  for some  $i_j \in \mathcal{S}^{(j)}$ ,  $j = 1, 2$ , and in fact  $i_j = M_{v_j}$  in this correspondence. Hence these two minimizations problems are equivalent and their unique minimizers coincide, and we have proved (5.27).

The remaining statements in Theorem 3.27 follow from (5.27) and the relations  $W_n = R_1(T_{\text{coll}}) + R_2(T_{\text{coll}})$ ,  $\mathcal{S}_t = \pi_M(\tilde{\mathcal{B}}_t)$  for all  $t$ , and  $M_{V(i)} = i$ ,  $V(M_v) = v$  for  $i \in \mathcal{S}$ ,  $v \in \tilde{\mathcal{B}}$ .  $\square$

In the remainder of the paper, we will be concerned only with the equality in law from Theorem 3.27. We can therefore continue to define  $\mathcal{B}$  as in (3.36), ignoring the modified construction given in the proof of Theorem 3.27. The edge-weight coupling (3.46) between  $\mathcal{T}^{(1,2)}$  and  $K_n$ , and indeed the edge weights on  $K_n$  generally, will play no further direct role in the analysis.

## 6 Branching Processes and Random Walks

In this section, we prove Theorem 3.13 by continuing the analysis of the branching process  $BP^{(1)}$  introduced in Sect. 3. In Sect. 6.1 we identify a random walk which facilitates moment computations of the one-vertex characteristics. Section 6.2 contains precise results about the scaling behavior of the random walk and the parameter  $\lambda_n(a)$ . The results are proved in Sect. 6.3. Section 6.4 identifies the asymptotics of the first two moments of one-vertex characteristics. Having understood these, we investigate two-vertex characteristics in Sect. 7 and prove Theorem 3.15.

### 6.1 Continuous-Time Branching Processes and Random Walks

Recall that  $BP^{(1)} = (BP_t^{(1)})_{t \geq 0}$  denotes a CTBP with original ancestor  $\emptyset_1$ . Using Ulam–Harris notation, the children of the root are the vertices  $v$  with  $p(v) = \emptyset_1$ , and their birth times  $(T_v)_{p(v)=\emptyset_1}$  form a Poisson point process with intensity  $\mu_n$ . For  $v \in \mathcal{T}^{(1)}$ , write  $BP^{(v)}$  for the branching process of descendants of such a  $v$ , re-rooted and time-shifted to start at  $t = 0$ . Formally,

$$BP_t^{(v)} = \left\{ w \in \mathcal{T}^{(1)} : vw \in BP_{T_v+t}^{(1)} \right\}. \quad (6.1)$$

In particular,  $BP^{(1)} = BP^{(\emptyset_1)}$ , and the processes  $(BP^{(v)})_{p(v)=\emptyset_1}$  are independent of each other and of  $(T_v)_{p(v)=\emptyset_1}$ . We may express this compactly by saying that the sum of point masses  $\mathcal{Q} = \sum_{p(v)=\emptyset_1} \delta_{(T_v, BP^{(v)})}$  forms a Poisson point process with intensity  $d\mu_n \otimes d\mathbb{P}(BP^{(1)} \in \cdot)$ , where  $\mathbb{P}(BP^{(1)} \in \cdot)$  is the law of the entire branching process. Recalling the definition of the one-vertex characteristic from (3.14), we deduce that

$$\begin{aligned} z_t^\chi(a) &= \chi(t) + a \sum_{v: p(v)=\emptyset} \mathbb{1}_{\{T_v \leq t\}} z_{t-T_v}^{\chi, BP^{(v)}}(a) \\ &= \chi(t) + a \int d\mathcal{Q}(y, bp) \mathbb{1}_{\{y \leq t\}} z_{t-y}^{\chi, bp}(a). \end{aligned} \quad (6.2)$$

Note that (6.2) holds jointly for all  $a, t, \chi$ . To draw conclusions for  $m_t^\chi(a)$  and  $M_{t,u}^{\chi,\eta}(a, b)$ , the expectation of  $z_t^\chi(a)$  and  $z_t^\chi(a)z_u^\eta(b)$ , respectively, defined in (3.15), we will use the

formulas

$$\begin{aligned}\mathbb{E}\left(\sum_{p \in \tilde{\mathcal{Q}}} f(p)\right) &= \int f(p) d\tilde{\mu}(p), \\ \text{Cov}\left(\sum_{p \in \tilde{\mathcal{Q}}} f_1(p), \sum_{p \in \tilde{\mathcal{Q}}} f_2(p)\right) &= \int f_1(p) f_2(p) d\tilde{\mu}(p),\end{aligned}\quad (6.3)$$

where  $\tilde{\mathcal{Q}}$  is a Poisson point process with some intensity  $\tilde{\mu}$  (and assuming the integrals exist). Apply (6.3) to (6.2) with  $f(y, bp) = \mathbb{1}_{\{y \leq t\}} z_{t-y}^{\chi, bp}(a)$  to get

$$\begin{aligned}m_t^\chi(a) &= \chi(t) + a \int d\mu_n(y) \mathbb{1}_{\{y \leq t\}} \int d\mathbb{P}(BP^{(1)} = bp) z_{t-y}^{\chi, bp}(a) \\ &= \chi(t) + a \int_0^t d\mu_n(y) m_{t-y}^\chi(a).\end{aligned}\quad (6.4)$$

Similarly

$$\begin{aligned}M_{t,u}^{\chi, \eta}(a, b) - m_t^\chi(a) m_u^\eta(b) &= ab \int d\mu_n(y) \mathbb{1}_{\{y \leq t\}} \mathbb{1}_{\{y \leq u\}} \int d\mathbb{P}(BP^{(1)} = bp) z_{t-y}^{\chi, bp}(a) z_{u-y}^{\eta, bp}(b) \\ &= ab \int_0^{t \wedge u} d\mu_n(y) M_{t-y, u-y}^{\chi, \eta}(a, b).\end{aligned}\quad (6.5)$$

Recall from (3.16) that  $\hat{\mu}_n(\lambda) = \int e^{-\lambda t} d\mu_n(t)$  denotes the Laplace transform of  $\mu_n$  and that for  $a > 0$  the parameter  $\lambda_n(a) > 0$  is the unique solution to  $a\hat{\mu}_n(\lambda_n(a)) = 1$ . [In general, if  $\hat{\mu}_n(\lambda_0)$  is finite for some  $\lambda_0 \geq 0$ , then the equation has a solution whenever  $1/\hat{\mu}_n(\lambda_0) \leq a < 1/\mu_n(\{0\})$ , and this solution is unique if  $\mu_n$  assigns positive mass to  $(0, \infty)$ . Our assumptions imply that, for  $n$  sufficiently large,  $\lambda_n(a)$  exists uniquely for any  $a > 0$ .] Since  $z_t^\chi(a)$  typically grows exponentially at rate  $\lambda_n(a)$ , we study the rescaled versions  $\bar{m}_t^\chi(a)$ ,  $\bar{M}_{t,u}^{\chi, \eta}(a, b)$  defined in (3.19) and let

$$dv_a(y) = ae^{-\lambda_n(a)y} d\mu_n(y). \quad (6.6)$$

Then (6.4) becomes  $\bar{m}_t^\chi(a) = e^{-\lambda(a)t} \chi(t) + \int_0^t dv_a(y) \bar{m}_{t-y}^\chi(a)$ . Since  $v_a$  is a probability measure by construction, this recursion can be solved in terms of a random walk:

$$\bar{m}_t^\chi(a) = \mathbb{E}_a\left(\sum_{j=0}^{\infty} e^{-\lambda_n(a)(t-S_j)} \chi(t-S_j)\right), \quad \text{where } S_j = \sum_{i=1}^j D_i, \quad \mathbb{P}_a(D_i \in \cdot) = v_a(\cdot). \quad (6.7)$$

From (6.5), we obtain similarly

$$\bar{M}_{t,u}^{\chi, \eta}(a, b) = \mathbb{E}_{ab}\left(\sum_{j=0}^{\infty} e^{-S_j[\lambda_n(a) + \lambda_n(b) - \lambda_n(ab)]} \bar{m}_{t-S_j}^\chi(a) \bar{m}_{u-S_j}^\eta(b)\right), \quad (6.8)$$

where  $S_j = \sum_{i=1}^j D_i$  now has distribution  $\mathbb{P}_{ab}(D_i \in \cdot) = v_{ab}(\cdot)$ .

Note that for a random variable  $D$  with law  $v_a$ , for every  $h \geq 0$  measurable,

$$\mathbb{E}_a(h(D)) = \int h(y) ae^{-\lambda_n(a)y} d\mu_n(y). \quad (6.9)$$

Moreover, let  $v_a^*$  denote the size-biasing of  $v_a$ , i.e.,

$$dv_a^*(y) = \frac{y dv_a(y)}{\int y dv_a(y)} \quad \text{so that} \quad \mathbb{E}_a(h(D^*)) = \frac{\mathbb{E}_a(Dh(D))}{\mathbb{E}_a(D)} \quad (6.10)$$

for  $h \geq 0$  measurable. Here and in all of the following we assume that  $D$  and  $D^*$  have laws  $v_a$  and  $v_a^*$  respectively under  $\mathbb{E}_a$ . Let  $U$  be uniform on  $[0, 1]$ , and let  $(D_i)_{i \geq 1}$  be independent with law  $v_a$ , and independent of  $U$  and  $D^*$ . Besides the random walk  $(S_j)_j$  from (6.7)–(6.8), it is useful to study the random walk  $(S_j^*)_j$  with

$$S_0^* = UD^* \quad \text{and} \quad S_j^* = S_0^* + \sum_{i=1}^j D_i \quad \text{for all } j \geq 1. \quad (6.11)$$

## 6.2 Random Walk Convergence

In this section, we investigate asymptotics of the random walks in (6.7)–(6.8) and (6.11). Recall that the branching process  $BP^{(1)}$  is derived from the intensity measure  $\mu_n$ , where for  $h: [0, \infty) \rightarrow [0, \infty)$  measurable,

$$\int h(y) d\mu_n(y) = \int_0^\infty h(f_n(x)) dx. \quad (6.12)$$

In particular, all the quantities  $z$ ,  $m$ ,  $M$ ,  $v_a$ ,  $D_i$ ,  $\mathbb{P}_a$  and  $\mathbb{E}_a$  as well as the random walks  $(S_j)_j$  and  $(S_j^*)_j$  depend implicitly on  $n$ . We will give sufficient conditions for the sequence of walks  $(S_j)_j$ ,  $(S_j^*)_j$  to have scaling limits; in all cases the scaling limit is the Gamma process. This is the content of Theorem 6.3 below.

As a motivation, we first look at the key example  $Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}$ :

**Example 6.1** Let  $Y_e^{(K_n)} \stackrel{d}{=} E^{s_n}$ . Then

$$\begin{aligned} F_Y(x) &= 1 - e^{-x^{1/s_n}} \quad \text{and} \quad F_Y^{-1}(x) = (-\log(1-x))^{s_n}, \\ f_n(x) &= (x/n)^{s_n} = f_n(1)x^{s_n} \quad \text{and} \quad f_n^{-1}(x) = nx^{1/s_n}. \end{aligned} \quad (6.13)$$

One easily checks that for all  $a > 0$ ,  $\beta > 0$ ,

$$\begin{aligned} f_n(1)\lambda_n(a^{1/s_n}) &= a\Gamma(1 + 1/s_n)^{s_n}, \quad s_n\lambda_n(a^{1/s_n})\mathbb{E}_{a^{1/s_n}}(D) = 1, \\ s_n\lambda_n(a^{1/s_n})^2\mathbb{E}_{a^{1/s_n}}(D^2) &= 1 + 1/s_n, \quad a(\hat{\mu}_n(\lambda_n(a^{1/s_n})\beta))^{s_n} = 1/\beta. \end{aligned} \quad (6.14)$$

Notice that  $\Gamma(1 + 1/s_n)^{s_n} \rightarrow e^{-\gamma}$  for  $n \rightarrow \infty$ .

Theorem 6.3 will show that in general the same identities hold asymptotically under our conditions on  $f_n$ . In fact, we will prove Theorem 6.3 under weaker assumptions on  $f_n$ :

**Condition 6.2** There exist  $\varepsilon_0 > 0$  and a sequence  $(\delta_n)_{n \in \mathbb{N}} \in (0, 1]^{\mathbb{N}}$  such that  $s_n f_n(1 - \delta_n)/f_n(1) = o(1)$ ,  $f_n(x^{1/s_n}) \geq f_n(1)x^{\varepsilon_0}$  for  $x \geq 1$  and  $f_n(x^{1/s_n}) \leq f_n(1)x^{\varepsilon_0}$  for  $(1 - \delta_n)^{s_n} \leq x \leq 1$ .

Conditions 2.2 and 2.6 together imply Condition 6.2: we may set  $\delta_n = \delta_0$ , with  $\varepsilon_0$  chosen as for Conditions 2.2 and 2.6, and replacing  $(x, x')$  in Lemma 4.1 by  $(1, x^{1/s_n})$  or  $(x^{1/s_n}, 1)$  verifies the inequalities in Condition 6.2.

**Theorem 6.3** (Random walk convergence) Suppose that Conditions 2.1 and 6.2 hold. Then, for any  $a \in (0, \infty)$ ,

(a) the parameter  $\lambda_n(a)$  exists for  $n$  sufficiently large, and, for all  $\beta > 0$ ,

$$\lim_{n \rightarrow \infty} f_n(1) \lambda_n(a^{1/s_n}) = a e^{-\gamma}, \quad (6.15)$$

$$\lim_{n \rightarrow \infty} s_n \lambda_n(a^{1/s_n}) \mathbb{E}_{a^{1/s_n}}(D) = 1, \quad (6.16)$$

$$\lim_{n \rightarrow \infty} s_n \lambda_n(a^{1/s_n})^2 \mathbb{E}_{a^{1/s_n}}(D^2) = 1, \quad (6.17)$$

$$\lim_{n \rightarrow \infty} a (\hat{\mu}_n(\lambda_n(a^{1/s_n})\beta))^{s_n} = 1/\beta, \quad (6.18)$$

where  $\gamma$  is Euler's constant;

- (b) under  $\mathbb{E}_{a^{1/s_n}}$ , the process  $(\lambda_n(a^{1/s_n})S_{\lfloor s_n t \rfloor})_{t \geq 0}$  converges in distribution (in the Skorohod topology on compact subsets) to a Gamma process  $(\Gamma_t)_{t \geq 0}$ , i.e., the Lévy process such that  $\Gamma_t$  has the Gamma( $t, 1$ ) distribution;
- (c) under  $\mathbb{E}_{a^{1/s_n}}$ , the variable  $\lambda_n(a^{1/s_n})D^*$  converges in distribution to an exponential random variable  $E$  with mean 1, and the process  $(\lambda_n(a^{1/s_n})S_{\lfloor s_n t \rfloor}^*)_{t \geq 0}$  converges in distribution to the sum  $(UE + \Gamma_t)_{t \geq 0}$  where  $U$  is Uniform on  $[0, 1]$  and  $U, E, (\Gamma_t)_{t \geq 0}$  are independent.

Moreover, given a compact subset  $A \subset (0, \infty)$ , all of the convergences occur uniformly for  $a \in A$  and, for (6.18), for  $\beta \in A$ .

Theorem 6.3 will be proved in Sect. 6.3. We stress that the proof of Theorem 6.3 uses only Conditions 2.1 and 6.2 and the relevant definitions but no other results stated so far.

### 6.3 Convergence of Random Walks: Proof of Theorem 6.3

For notational convenience, we make the abbreviations

$$\tilde{x} = x^{1/s_n}, \quad \tilde{a} = a^{1/s_n}, \quad \tilde{b} = b^{1/s_n}, \text{ etc.}, \quad (6.19)$$

which we will use extensively in this section and in Sects. 7 and 9.

**Proof of Theorem 6.3** We begin by proving

$$\hat{\mu}_n\left(\frac{ae^{-\gamma}}{f_n(1)}\right) = 1 - \frac{\log a}{s_n} + o(1/s_n). \quad (6.20)$$

Recalling (6.12), we have

$$\hat{\mu}_n(\lambda) = \int_0^\infty e^{-\lambda f_n(\tilde{x})} d\tilde{x}. \quad (6.21)$$

Write  $\tilde{f}_n(\tilde{x}) = f_n(\tilde{x}) \mathbb{1}_{\{\tilde{x} \geq 1 - \delta_n\}}$ . Then

$$\hat{\mu}_n(\lambda) = \int_0^\infty e^{-\lambda \tilde{f}_n(\tilde{x})} d\tilde{x} - \int_0^{1-\delta_n} (1 - e^{-\lambda f_n(\tilde{x})}) d\tilde{x}, \quad (6.22)$$

and take  $\lambda = ae^{-\gamma}/f_n(1)$  to estimate  $\int_0^{1-\delta_n} (1 - e^{-ae^{-\gamma} f_n(\tilde{x})/f_n(1)}) d\tilde{x} = O(f_n(1 - \delta_n)/f_n(1)) = o(1/s_n)$  by Condition 6.2. Hence, for the purposes of proving (6.20), it is no loss of generality to assume that  $f_n(x^{1/s_n}) \leq f_n(1)x^{\varepsilon_0}$  for all  $x \leq 1$ .

Inspired by (2.5), where  $f_n(\tilde{x}) = f_n(1)\tilde{x}^{s_n}$ , we compute

$$\int_0^\infty e^{-\lambda f_n(1)\tilde{x}^{s_n}} d\tilde{x} = \int_0^\infty \frac{1}{s_n} x^{1/s_n-1} e^{-\lambda f_n(1)x} dx = \left( \frac{\Gamma(1 + 1/s_n)^{s_n}}{\lambda f_n(1)} \right)^{1/s_n}. \quad (6.23)$$

In particular, setting  $\lambda = a\Gamma(1 + 1/s_n)^{s_n}/f_n(1)$  gives  $\int_0^\infty \exp(-a\Gamma(1 + 1/s_n)^{s_n} \tilde{x}^{s_n}) d\tilde{x} = a^{-1/s_n}$ , which is  $1 - (\log a)/s_n + o(1/s_n)$ . Subtracting this from (6.21), we can therefore prove (6.20) if we show that

$$s_n \int_0^\infty \left( e^{-ae^{-\gamma} f_n(\tilde{x})/f_n(1)} - e^{-a\Gamma(1+1/s_n)^{s_n} \tilde{x}^{s_n}} \right) d\tilde{x} \rightarrow 0, \quad (6.24)$$

or equivalently, by the substitution  $\tilde{x} = x^{1/s_n}$ , if we show that

$$\int_0^\infty x^{1/s_n-1} \left( e^{-ae^{-\gamma} f_n(x^{1/s_n})/f_n(1)} - e^{-a\Gamma(1+1/s_n)^{s_n} x} \right) dx \rightarrow 0. \quad (6.25)$$

Note that  $\Gamma(1 + 1/s)^s \rightarrow e^{-\gamma}$  as  $s \rightarrow \infty$ . Together with Condition 2.1, this implies that the integrand in (6.25) converges pointwise to 0. For  $x \leq 1$ ,  $f_n(x^{1/s_n}) \leq f_n(1)x^{\varepsilon_0}$  means that the integrand is bounded by  $O(x^{\varepsilon_0-1} + 1)$ . For  $x \geq 1$ , Condition 6.2 implies that the integrand is bounded by  $e^{-\delta x^{\varepsilon_0}}$  for some  $\delta > 0$ . Dominated convergence therefore completes the proof of (6.20). It is easy to see that the proof of (6.25), and hence (6.20), holds uniformly in  $a \in A$ , where  $A \subseteq (0, \infty)$  is a fixed but arbitrary compact set.

To conclude (6.15) from (6.20), we use the monotonicity of  $\mu_n$  and  $\lambda_n$ : given  $\delta > 0$ , we have  $\hat{\mu}_n(ae^{-\gamma}/f_n(1))^{s_n} \leq a^{-1} + \delta$  for all sufficiently large  $n$ , uniformly in  $a \in A$ . Replacing  $a$  by  $a' = (1/a - \delta)^{-1}$  shows that  $\lambda_n(\tilde{a}) \leq a'e^{-\gamma}/f_n(1)$  for all  $n$  large enough, uniformly in  $a$ . A lower bound holds similarly; take  $\delta \rightarrow 0$  to conclude (6.15).

The proof of (6.16) is similar to that of (6.15). Using (6.9), we compute

$$\begin{aligned} s_n \lambda_n(\tilde{a}) \mathbb{E}_{\tilde{a}}(D) &= s_n \lambda_n(\tilde{a}) \int y \tilde{a} e^{-\lambda_n(\tilde{a})y} d\mu_n(y) \\ &= \lambda_n(\tilde{a}) f_n(1) \tilde{a} \int_0^\infty x^{1/s_n-1} \frac{f_n(\tilde{x})}{f_n(1)} e^{-\lambda_n(\tilde{a})f_n(\tilde{x})} dx. \end{aligned} \quad (6.26)$$

By (2.6) and (6.15), the integrand in (6.26) converges pointwise to  $e^{-ae^{-\gamma}x}$ , and satisfies a similar dominated convergence estimates as (6.25) by Condition 6.2. Hence (6.26) converges to  $ae^{-\gamma} \int_0^\infty e^{-ae^{-\gamma}x} dx = 1$  as claimed. The proof of (6.17) is similar.

To prove (6.18), let  $b_n$  be defined by  $\lambda_n(\tilde{b}_n) = \beta \lambda_n(\tilde{a})$ . By (6.15) and monotonicity, it follows that  $b_n \rightarrow \beta a$  [for if  $\limsup_{n \rightarrow \infty} b_n \geq (1+\varepsilon)\beta a$  then  $\limsup_{n \rightarrow \infty} f_n(1)\lambda_n(\tilde{b}_n) \geq (1+\varepsilon)\beta ae^{-\gamma} = (1+\varepsilon) \lim_{n \rightarrow \infty} f_n(1)\beta \lambda_n(\tilde{a})$ , a contradiction, and similarly if  $\liminf_{n \rightarrow \infty} b_n \leq (1-\varepsilon)\beta a$ ]. But  $\hat{\mu}_n(\lambda_n(\tilde{b}_n))^{s_n} = b_n^{-1}$ , giving the result.

Since  $\lambda_n(\tilde{a})S_{[s_n t]}$  is non-decreasing and right-continuous, and has i.i.d. increments if  $s_n t$  is restricted to integer values, it suffices to show that its limiting distribution is  $\Gamma(t, 1)$  for a fixed  $t$ , where  $\Gamma(t, 1)$  denotes a standard Gamma variable with parameter  $t$ . For this, we note that its Laplace transform is

$$\mathbb{E}_{\tilde{a}} \left( e^{-\tau \lambda_n(\tilde{a}) S_{[s_n t]}} \right) = \left( \tilde{a} \hat{\mu}_n(\lambda_n(\tilde{a})(1+\tau)) \right)^{[s_n t]}. \quad (6.27)$$

Since  $s_n \rightarrow \infty$ , (6.18) yields that the right-hand side tends to  $(1+\tau)^{-t}$ . This is the Laplace transform of a  $\Gamma(t, 1)$  variable, and thus completes the proof of (b).

For the proof of part (c) define  $b_n$  by  $\lambda_n(\tilde{b}_n) = (1+\tau)\lambda_n(\tilde{a})$  for a given  $\tau \geq 0$ . Then (6.10) and (6.9) yield

$$\mathbb{E}_{\tilde{a}}(e^{-\tau \lambda_n(\tilde{a}) D^*}) = \frac{\mathbb{E}_{\tilde{a}}(D e^{-\tau \lambda_n(\tilde{a}) D})}{\mathbb{E}_{\tilde{a}}(D)} = \frac{\tilde{a} \int y e^{-\lambda_n(\tilde{a})y(1+\tau)} d\mu_n(y)}{\mathbb{E}_{\tilde{a}}(D)} = \frac{\tilde{a} \mathbb{E}_{\tilde{b}_n}(D)}{\tilde{b}_n \mathbb{E}_{\tilde{a}}(D)}. \quad (6.28)$$

By the same argument as in the proof of (6.18),  $b_n \rightarrow (1 + \tau)a$ . Combining with (6.16), we conclude that  $\mathbb{E}_{\tilde{a}}(e^{-\tau\lambda_n(\tilde{a})D^*}) \rightarrow (1 + \tau)^{-1}$  and  $\lambda_n(\tilde{a})D^*$  converges to an exponential variable with mean 1. So the rest of part (c) follows from part (b).

The remaining uniformity claims follow from the uniformity in (6.15). The uniformity statements in parts (b) and (c) follow from the observation that the Radon–Nikodym derivatives  $d\mathbb{P}_{\tilde{a}}((\lambda_n(\tilde{a})S_{\lfloor s_n t \rfloor})_{0 \leq t \leq K} \in \cdot) / d\mathbb{P}_{\tilde{a}'}((\lambda_n(\tilde{a}')S_{\lfloor s_n t \rfloor})_{0 \leq t \leq K} \in \cdot)$  (and similarly for  $D^*$  and  $S^*$ ) are tight for  $K < \infty$ , uniformly over  $a, a' \in A$ .  $\square$

**Lemma 6.4** For all  $a \in (0, \infty)$ ,  $\lambda'_n(a) = 1/(\mathbb{E}_a(D))$ .

**Proof** Denote  $\hat{\mu}'_n(\lambda) = \frac{d}{d\lambda} \hat{\mu}_n(\lambda)$  and  $\hat{\mu}''_n(\lambda) = \frac{d^2}{d\lambda^2} \hat{\mu}_n(\lambda)$ . Using the definitions of  $\hat{\mu}_n$  and  $v_a$ , an elementary computation shows that  $-a\hat{\mu}'_n(\lambda_n(a)) = \mathbb{E}_a(D)$ . Moreover, by (3.16),  $a\hat{\mu}_n(\lambda_n(a)) = 1$  and the claim follows.

**Proof of Lemma 3.12** By (6.15),  $\lambda_n f_n(1) \rightarrow e^{-\gamma}$  and Lemma 6.4 gives

$$\phi_n = \frac{\lambda'_n(1)}{\lambda_n(1)} = \frac{1}{\lambda_n(1)\mathbb{E}_1(D)}. \quad (6.29)$$

Now (6.16) implies that  $\phi_n/s_n \rightarrow 1$ , as required.  $\square$

**Corollary 6.5** Uniformly for  $a$  in a compact subset of  $(0, \infty)$ ,

$$\lambda_n(a^{1/s_n}) = \lambda_n(1) \left( 1 + \frac{\phi_n}{s_n} (a - 1) + o((a - 1)^2) \right). \quad (6.30)$$

**Proof** By the same arguments as in the proof of Lemma 6.4, the function  $F(a) = \lambda_n(a^{1/s_n})$  satisfies  $-\tilde{a}\hat{\mu}'_n(F(a)) = \mathbb{E}_{\tilde{a}}(D)$  and  $\tilde{a}\hat{\mu}''_n(F(a)) = \mathbb{E}_{\tilde{a}}(D^2)$ . By (3.16),  $a\hat{\mu}_n(F(a))^{s_n} = 1$  and we deduce that  $s_n\mathbb{E}_{\tilde{a}}(D)F'(a) = 1/a$  and

$$s_n\mathbb{E}_{\tilde{a}}(D)F''(a) = \frac{s_n\mathbb{E}_{\tilde{a}}(D^2)}{(as_n\mathbb{E}_{\tilde{a}}(D))^2} - \frac{1 + \frac{1}{s_n}}{a^2}. \quad (6.31)$$

The right-hand side of (6.31) converges uniformly to 0 by (6.16)–(6.17). Applying (6.16) again and noting from (6.15) that  $\lambda_n(\tilde{a})/\lambda_n(1)$  is bounded, it follows that  $F''(a)/\lambda_n(1)$  converges uniformly to 0. A Taylor expansion of  $F$  around the point 1 and (6.29) complete the proof.  $\square$

The following lemma is a consequence of Theorem 6.3 and will be used in the proof of the second moment statements of Theorem 3.13:

**Lemma 6.6** Assume the hypotheses of Theorem 6.3, and let  $A \subset (0, 2)$  be compact. For any measurable, bounded function  $h \geq 0$ , set

$$\Xi(h) = \mathbb{E}_{a^{1/s_n}b^{1/s_n}} \left( \sum_{j=0}^{\infty} e^{-[\lambda_n(a^{1/s_n}) + \lambda_n(b^{1/s_n}) - \lambda_n(a^{1/s_n}b^{1/s_n})]S_j} h(S_j) \right). \quad (6.32)$$

There are constants  $K < \infty$  and  $n_0 \in \mathbb{N}$  independent of  $h$  such that  $\Xi(h) \leq K s_n \|h\|_{\infty}$  for all  $n \geq n_0$  and  $a, b \in A$ . Moreover, for any  $\varepsilon > 0$  there are constants  $K' < \infty$ ,  $n'_0 \in \mathbb{N}$  independent of  $h$  such that for all  $a, b \in A$ ,  $n \geq n'_0$ ,

$$-\varepsilon \|h\|_{\infty} + \frac{\inf \{h(y) : \lambda_n(1)y \leq K'\}}{\log(1/a + 1/b)} \leq \frac{\Xi(h)}{s_n} \leq \varepsilon \|h\|_{\infty} + \frac{\sup \{h(y) : \lambda_n(1)y \leq K'\}}{\log(1/a + 1/b)}. \quad (6.33)$$

Note that  $\log(1/a + 1/b)$  is positive and bounded away from 0 by our assumption on  $A$ .

**Proof of Lemma 6.6** We rewrite  $\Xi(h)$  in integral form, bound  $h$  by its maximum, and use (6.9) and the definition of  $\hat{\mu}_n$ , to obtain

$$\Xi(h) = s_n \int_0^\infty \mathbb{E}_{\tilde{a}\tilde{b}} \left( e^{-[\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}) - \lambda_n(\tilde{a}\tilde{b})]S_{\lfloor s_n t \rfloor}} h(S_{\lfloor s_n t \rfloor}) \right) dt \quad (6.34)$$

$$\leq s_n \|h\|_\infty \int_0^\infty \left[ \tilde{a}\tilde{b}\hat{\mu}_n(\lambda_n(\tilde{a}) + \lambda_n(\tilde{b})) \right]^{\lfloor s_n t \rfloor} dt. \quad (6.35)$$

By (6.15) and (6.18), we deduce  $(\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}))/\lambda_n(1) \rightarrow (a + b)$  and

$$[\tilde{a}\tilde{b}\hat{\mu}_n(\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}))]^{s_n} \rightarrow ab/(a + b). \quad (6.36)$$

Since  $\log((a + b)/ab) = \log(1/a + 1/b)$  is positive and uniformly bounded away from 0, the integral in (6.35) is uniformly bounded for sufficiently large  $n$  and we conclude that there is a constant  $K < \infty$  with  $\Xi(h) \leq K s_n \|h\|_\infty$ .

For (6.33), let  $\varepsilon > 0$  be given. Since  $A \subset (0, 2)$  is compact, (6.15) implies that there exists  $\delta > 0$  and  $n'_0 \in \mathbb{N}$  such that  $\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}) - \lambda_n(\tilde{a}\tilde{b}) \geq \delta f_n(1)$  for all  $n \geq n'_0$  and  $a, b \in A$ . Using again (6.36), we may take  $t_0$  and  $n'_0$  sufficiently large that  $\int_{t_0}^\infty [\tilde{a}\tilde{b}\hat{\mu}_n(\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}))]^{\lfloor s_n t \rfloor} dt \leq \frac{1}{3}\varepsilon$  and  $\left| \int_0^{t_0} [\tilde{a}\tilde{b}\hat{\mu}_n(\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}))]^{\lfloor s_n t \rfloor} dt - 1/\log(1/a + 1/b) \right| \leq \frac{1}{3}\varepsilon$  for all  $n \geq n'_0$ ,  $a, b \in A$ . Furthermore, Theorem 6.3(b) implies that the family of laws  $\mathbb{P}_{\tilde{a}\tilde{b}}(\lambda_n(1)S_{\lfloor s_n t \rfloor} \in \cdot)$ ,  $t \leq t_0$ ,  $a, b \in A$ , is tight. Hence we may take  $K'$  large enough that  $t_0 e^{-\delta K'} \leq 1$  and  $\mathbb{P}_{\tilde{a}\tilde{b}}(\lambda_n(1)S_{\lfloor s_n t \rfloor} > K') \leq \frac{1}{3}\varepsilon$ , uniformly for  $t \leq t_0$ . We conclude from (6.34) that

$$\begin{aligned} & \int_0^{t_0} \left[ \inf \{h(y) : \lambda_n(1)y \leq K'\} \mathbb{E}_{\tilde{a}\tilde{b}} \left( e^{-[\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}) - \lambda_n(\tilde{a}\tilde{b})]S_{\lfloor s_n t \rfloor}} \right) - \frac{1}{3}\varepsilon e^{-\delta K'} \|h\|_\infty \right] dt \\ & \leq \frac{\Xi(h)}{s_n} \leq \int_0^{t_0} \left[ \sup \{h(y) : \lambda_n(1)y \leq K'\} \mathbb{E}_{\tilde{a}\tilde{b}} \left( e^{-[\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}) - \lambda_n(\tilde{a}\tilde{b})]S_{\lfloor s_n t \rfloor}} \right) + \frac{1}{3}\varepsilon e^{-\delta K'} \|h\|_\infty \right] dt \\ & \quad + \int_{t_0}^\infty \|h\|_\infty \mathbb{E}_{\tilde{a}\tilde{b}} \left( e^{-(\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}) - \lambda_n(\tilde{a}\tilde{b}))S_{\lfloor s_n t \rfloor}} \right) dt \end{aligned} \quad (6.37)$$

for  $n \geq n'_0$ . Using again  $\mathbb{E}_{\tilde{a}\tilde{b}}(e^{-S_{\lfloor s_n t \rfloor}(\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}) - \lambda_n(\tilde{a}\tilde{b}))}) = [\tilde{a}\tilde{b}\hat{\mu}_n(\lambda_n(\tilde{a}) + \lambda_n(\tilde{b}))]^{\lfloor s_n t \rfloor}$ , we see that the hypotheses on  $t_0$  and  $n'_0$  imply (6.33).  $\square$

## 6.4 Means of One-Vertex Characteristics: Proof of Theorem 3.13

In this section, we prove Theorem 3.13. Further, we set the stage for the proofs of Theorems 3.15 and 3.16 in Sect. 7.

Recall from (6.7) that

$$\bar{m}_t^\chi(a) = \mathbb{E}_a \left( \sum_{j=0}^\infty e^{-\lambda_n(a)(t-S_j)} \chi(t - S_j) \right).$$

Thus,  $\bar{m}_t^\chi(a)$  can be understood as the expected integral of  $e^{-\lambda_n(a)t} \chi(t)$  against the counting measure on the random set  $\{t - S_j : j \in \mathbb{N}_0\}$ . When  $t \rightarrow \infty$ , this measure approaches its stationary equivalent, which is the counting measure on the point-stationary set  $\{t - S_j^* : j \in \mathbb{N}_0\}$  (see [34]). Since the expected value of this stationary measure is a



multiple of the Lebesgue measure,  $\bar{m}_t^\chi(a)$  will approach (as  $t \rightarrow \infty$ ) the same multiple of  $\int_0^\infty e^{-\lambda_n(a)t} \chi(t) dt$ . In the following we make this more precise. We begin with general estimates that apply to any CTBP with any intensity measure  $\mu$ , and we will write simply  $\lambda(a)$  for the parameter defined by the analogue of (3.16). Similarly, all other notation introduced for  $BP^{(1)}$  will be used for a general CTBP.

**Proposition 6.7** *Let  $(S_j^*)_j$  be the random walk defined in (6.11). Let  $\chi$  be a non-negative characteristic. Then, for all  $a, t > 0$ ,*

$$\mathbb{E}_a(\bar{m}_{t-UD^*}^\chi(a)) = \mathbb{E}_a \left( \sum_{j=0}^{\infty} e^{-\lambda(a)(t-S_j^*)} \chi(t-S_j^*) \right) = \frac{\int_0^t e^{-\lambda(a)u} \chi(u) du}{\mathbb{E}_a(D)}. \quad (6.38)$$

**Proof** The first equality is (6.7); the second follows because the set  $\{t-S_j^*: j \in \mathbb{N}_0\}$  is point-stationary in the sense of [34]. Alternatively, the equality may be verified by taking Laplace transforms with respect to  $t$ .  $\square$

In (6.7) and (6.38), we may collapse the tail of the sum into a single value of  $\bar{m}_u^\chi(a)$ . Namely, if  $J$  is a stopping time for  $(S_j)_j$  or  $(S_j^*)_j$ , respectively, then by the strong Markov property

$$\begin{aligned} \bar{m}_t^\chi(a) &= \mathbb{E}_a \left( \bar{m}_{t-S_J}^\chi(a) + \sum_{j=0}^{J-1} e^{-\lambda(a)(t-S_j)} \chi(t-S_j) \right), \\ \frac{\int_0^t e^{-\lambda(a)u} \chi(u) du}{\mathbb{E}_a(D)} &= \mathbb{E}_a \left( \bar{m}_{t-S_J^*}^\chi(a) + \sum_{j=0}^{J-1} e^{-\lambda(a)(t-S_j^*)} \chi(t-S_j^*) \right). \end{aligned} \quad (6.39)$$

The following lemmas provide bounds on  $\bar{m}_t^\chi(a)$  when  $m_t^\chi(a)$  is non-decreasing in  $t$ :

**Lemma 6.8** *Suppose  $\chi$  is a non-negative characteristic such that  $m_t^\chi(a)$  is non-decreasing in  $t$  (in particular, this holds if  $\chi$  is non-decreasing). Let  $(S_j)_j$  be as in (6.7) and  $(S_j^*)_j$  as in (6.38), and suppose that  $(S_j)_j$  and  $(S_j^*)_j$  are independent. Let  $\varepsilon > 0$  and set  $J = \inf \{j: |S_j - S_j^*| \leq \varepsilon\}$ . Then, for  $a, t > 0$ ,*

$$\begin{aligned} e^{-2\lambda(a)\varepsilon} \frac{\int_0^{t-\varepsilon} e^{-\lambda(a)u} \chi(u) du}{\mathbb{E}_a(D)} - \mathbb{E}_a \left( \sum_{j=0}^{J-1} e^{-\lambda(a)(t-\varepsilon-S_j^*)} \chi(t-\varepsilon-S_j^*) \right) \\ \leq \bar{m}_t^\chi(a) \leq e^{2\lambda(a)\varepsilon} \frac{\int_0^{t+\varepsilon} e^{-\lambda(a)u} \chi(u) du}{\mathbb{E}_a(D)} + \mathbb{E}_a \left( \sum_{j=0}^{J-1} e^{-\lambda(a)(t-S_j)} \chi(t-S_j) \right). \end{aligned} \quad (6.40)$$

**Proof** The hypotheses imply  $t-\varepsilon-S_J^* \leq t-S_J \leq t+\varepsilon-S_J^*$  and therefore  $e^{-2\lambda(a)\varepsilon} \bar{m}_{t-\varepsilon-S_J^*}^\chi(a) \leq \bar{m}_{t-S_J}^\chi(a) \leq e^{2\lambda(a)\varepsilon} \bar{m}_{t+\varepsilon-S_J^*}^\chi(a)$ . Combining with (6.39) gives the result.  $\square$

**Lemma 6.9** *Suppose  $\chi$  is a non-negative characteristic such that  $m_t^\chi(a)$  is non-decreasing in  $t$ . Then, for all  $a, t > 0$  and  $K > 0$ ,*

$$\bar{m}_t^\chi(a) \leq \frac{e^K}{\mathbb{E}_a \left( e^{\lambda(a)S_0^*} \mathbb{1}_{\{\lambda(a)S_0^* \leq K\}} \right)} \frac{\int_0^\infty e^{-\lambda(a)u} \chi(u) du}{\mathbb{E}_a(D)}. \quad (6.41)$$

**Proof** On  $\{\lambda(a)S_0^* \leq K\}$  we have  $\bar{m}_{t+K/\lambda(a)-S_0^*}^\chi(a) \geq e^{-K} e^{\lambda(a)S_0^*} \bar{m}_t^\chi(a)$ . Apply (6.38) and replace the limit of integration by  $\infty$  to obtain the result.  $\square$

**Lemma 6.10** Let  $\chi$  be a non-negative, non-decreasing characteristic such that  $\int_0^\infty e^{-\lambda(a)u} \chi(u) du < \infty$ , and fix  $a, K > 0$ . Then, for all  $t > 0$ ,  $\sum_{j=0}^\infty e^{-\lambda(a)(t-S_j)} \chi(t-S_j)$  is square-integrable under  $\mathbb{E}_a$  and, abbreviating  $C_{a,K} = e^K / \mathbb{E}_a(e^{\lambda(a)S_0^*} \mathbb{1}_{\{\lambda(a)S_0^* \leq K\}})$ ,

$$\mathbb{E}_a \left( \left( \sum_{j=0}^\infty e^{-\lambda(a)(t-S_j)} \chi(t-S_j) \right)^2 \right) \leq C_{a,K} \frac{\int_0^\infty e^{-2\lambda(a)u} \chi(u)^2 du}{\mathbb{E}_a(D)} + 2C_{a,K}^2 \frac{(\int_0^\infty e^{-\lambda(a)u} \chi(u) du)^2}{\mathbb{E}_a(D)^2}. \quad (6.42)$$

The same bound holds with  $(S_j)_j$  replaced by  $(S_j^*)_j$ .

**Proof** Since  $\chi$  is non-decreasing,  $\int_0^\infty e^{-\lambda(a)u} \chi(u) du < \infty$  implies that  $e^{-\lambda(a)u} \chi(u)$  must be bounded. Hence  $\int_0^\infty e^{-2\lambda(a)u} \chi(u)^2 du < \infty$  also. Applying Lemma 6.9 to  $\chi$  and  $\chi^2$ , we deduce

$$\begin{aligned} & \mathbb{E}_a \left( \left( \sum_{j=0}^\infty e^{-\lambda(a)(t-S_j)} \chi(t-S_j) \right)^2 \right) \\ &= \mathbb{E}_a \left( \sum_{j=0}^\infty e^{-2\lambda(a)(t-S_j)} \chi(t-S_j)^2 \right) \\ & \quad + 2\mathbb{E}_a \left( \sum_{j=0}^\infty e^{-\lambda(a)(t-S_j)} \chi(t-S_j) \sum_{k=j+1}^\infty e^{-\lambda(a)(t-S_k)} \chi(t-S_k) \right) \\ &= \bar{m}_t^{\chi^2}(a) + 2\mathbb{E}_a \left( \sum_{j=0}^\infty e^{-\lambda(a)(t-S_j)} \chi(t-S_j) \bar{m}_{t-S_{j+1}}^\chi(a) \right) \\ &\leq C_{a,K} \frac{\int_0^\infty e^{-\lambda(a)u} \chi(u)^2 du}{\mathbb{E}_a(D)} + 2C_{a,K} \frac{\int_0^\infty e^{-\lambda(a)u} \chi(u) du}{\mathbb{E}_a(D)} \mathbb{E}_a \left( \sum_{j=0}^\infty e^{-\lambda(a)(t-S_j)} \chi(t-S_j) \right). \end{aligned} \quad (6.43)$$

Another application of Lemma 6.9 gives (6.42). Finally replacing  $(S_j)_j$  by  $(S_j^*)_j$  is equivalent to replacing  $t$  by  $t - UD^*$ . Since the upper bound in (6.42) does not depend on  $t$ , the result follows.  $\square$

We now specialise to the offspring distribution  $\mu_n$  and apply the convergence results of Theorem 6.3 to prove Theorem 3.13:

**Proof of Theorem 3.13** By Lemma 6.9, for all  $a, t > 0$ ,

$$\bar{m}_t^\chi(\tilde{a}) \leq s_n \frac{e^1}{\mathbb{P}_{\tilde{a}}(\lambda_n(\tilde{a})S_0^* \leq 1)} \frac{\int_0^\infty \lambda_n(\tilde{a}) e^{-\lambda_n(\tilde{a})u} \|\chi\|_\infty du}{s_n \lambda_n(\tilde{a}) \mathbb{E}_{\tilde{a}}(D)}, \quad (6.44)$$

and, by Theorem 6.3,  $\mathbb{P}_{\tilde{a}}(\lambda_n(\tilde{a})S_0^* \leq 1) \rightarrow \mathbb{P}(UE \leq 1)$  and  $s_n \lambda_n(\tilde{a}) \mathbb{E}_{\tilde{a}}(D) \rightarrow 1$  uniformly in  $a \in A$ . Hence, the uniform bound for  $\bar{m}_t^\chi(\tilde{a})$  follows. By the same reasoning, Lemma 6.10 yields a constant  $C < \infty$  such that

$$\mathbb{E}_{\tilde{a}} \left( \left( \sum_{j=0}^\infty e^{-\lambda_n(\tilde{a})(t-S_j)} \chi(t-S_j) \right)^2 \right) \leq C s_n^2 \|\chi\|_\infty^2, \quad (6.45)$$

an estimate that will be needed shortly.

For (3.20), fix  $\varepsilon > 0$ . Apply Lemma 6.8 with  $\varepsilon$  and  $a$  replaced by  $\tilde{\varepsilon} = \lambda_n(\tilde{a})^{-1}\varepsilon$  and  $\tilde{a}$ , with the stopping time  $J_n = \inf \left\{ j: \left| S_j - S_j^* \right| \leq \tilde{\varepsilon} \right\} = \inf \left\{ j: \lambda_n(\tilde{a}) \left| S_j - S_j^* \right| \leq \varepsilon \right\}$ . By (6.15), we may choose  $K$  large enough that  $\int_{t-\tilde{\varepsilon}}^{\infty} \lambda_n(\tilde{a}) e^{-\lambda_n(\tilde{a})z} \chi(z) dz \leq \|\chi\|_{\infty} \varepsilon$  whenever  $\lambda_n(1)t \geq K$ . By (6.16), it follows that the first terms in the upper and lower bounds of Lemma 6.8 are  $s_n \int_0^{\infty} \lambda_n(\tilde{a}) e^{-\lambda_n(\tilde{a})z} \chi(z) dz + O(\varepsilon) s_n \|\chi\|_{\infty}$ . Therefore it is enough to show that the error term  $\mathbb{E}_{\tilde{a}} \left( \sum_{j=0}^{J_n-1} e^{-\lambda_n(\tilde{a})(t-\tilde{\varepsilon}-S_j^*)} \chi(t-\tilde{\varepsilon}-S_j^*) \right)$  is also  $O(\varepsilon) s_n \|\chi\|_{\infty}$  for  $\lambda_n(1)t$  sufficiently large, uniformly in  $a \in A$  (the same proof will work for the term with  $S_j$ ).

To prove this, observe that the variables  $(J_n/s_n, \lambda_n(\tilde{a})S_{j_n}^*)_{n \in \mathbb{N}, a \in A}$  are tight. Indeed, the rescaled processes  $(\lambda_n(\tilde{a})S_{\lfloor s_n t \rfloor})_{t \geq 0}$ ,  $(\lambda_n(\tilde{a})S_{\lfloor s_n t \rfloor}^*)_{t \geq 0}$  converge by Theorem 6.3 to independent Gamma processes (with different initial conditions). These limiting processes approach to within  $\varepsilon/2$  at some random but finite time, and tightness follows.

Thus we may choose  $C'$  large enough that the event  $\mathcal{A} = \{J_n \leq C's_n\} \cup \{\lambda_n(\tilde{a})S_{j_n}^* \leq C'\}$  satisfies  $\mathbb{P}_{\tilde{a}}(\mathcal{A}^c) \leq \varepsilon^2$ . Using the Cauchy–Schwarz inequality and (6.45),

$$\begin{aligned} \mathbb{E}_{\tilde{a}} \left( \mathbb{1}_{\mathcal{A}^c} \sum_{j=0}^{J_n-1} e^{-\lambda_n(\tilde{a})(t-\tilde{\varepsilon}-S_j^*)} \chi(t-\tilde{\varepsilon}-S_j^*) \right) \\ \leq \left[ \mathbb{P}_{\tilde{a}}(\mathcal{A}^c) \mathbb{E}_{\tilde{a}} \left( \left( \sum_{j=0}^{\infty} e^{-\lambda_n(\tilde{a})(t-\tilde{\varepsilon}-S_j^*)} \chi(t-\tilde{\varepsilon}-S_j^*) \right)^2 \right) \right]^{1/2} \leq \sqrt{C'} \varepsilon s_n \|\chi\|_{\infty}, \end{aligned} \quad (6.46)$$

whereas

$$\mathbb{E}_{\tilde{a}} \left( \mathbb{1}_{\mathcal{A}} \sum_{j=0}^{J_n-1} e^{-\lambda_n(\tilde{a})(t-\tilde{\varepsilon}-S_j^*)} \chi(t-\tilde{\varepsilon}-S_j^*) \right) \leq C' s_n \|\chi\|_{\infty} e^{-\lambda_n(\tilde{a})(t-\tilde{\varepsilon})+C'}. \quad (6.47)$$

By (6.15), the right-hand side is at most  $\varepsilon s_n \|\chi\|_{\infty}$ , uniformly over  $a \in A$ , if  $\lambda_n(1)t \geq K$  with  $K$  sufficiently large. This completes the proof of (3.20).

We turn to the estimates for  $\bar{M}_{t,u}^{\chi,\eta}(\tilde{a}, \tilde{b})$ . In view of (6.8), apply Lemma 6.6 to  $h(y) = \bar{m}_{t-y}^{\chi}(\tilde{a}) \bar{m}_{u-y}^{\eta}(\tilde{b})$ . By the first part of the current proof,  $\|h\|_{\infty} = O(s_n^2) \|\chi\|_{\infty} \|\eta\|_{\infty}$  and for any  $\varepsilon > 0$  we can make the infimum and supremum in (6.33) differ from  $s_n^2 \int_0^{\infty} e^{-z} \chi(z/\lambda_n(\tilde{a})) dz \cdot \int_0^{\infty} e^{-w} \eta(w/\lambda_n(\tilde{b})) dw$  by at most  $\varepsilon s_n^2$ , by taking  $\lambda_n(1)[t \wedge u]$  large enough.  $\square$

## 7 Continuous-Time Branching Processes: Two-Vertex Characteristics

In view of Theorems 3.11 and 3.27, we wish to consider generation-weighted two-vertex characteristics of the form

$$z_t^{\chi}(\vec{a}) = \sum_{v_1 \in BP_{t_1}^{(1)}} \sum_{v_2 \in BP_{t_2}^{(2)}} a_1^{|v_1|} a_2^{|v_2|} \chi(t_1 - T_{v_1}, t_2 - T_{v_2}) \quad (7.1)$$

for  $\chi(\vec{t}) = \chi_n(\vec{t}) = \mu_n(|t_1 - t_2|, t_1 + t_2)$  defined in (3.24). As discussed in Sect. 3.3.2, we split  $\chi_n$  into  $\chi_n^{(K)}$  and  $\chi_n - \chi_n^{(K)}$  for some  $K \in (0, \infty)$ .

Regarding  $t_1$  as fixed,

$$s_n \chi_n^{(K)}(t_1, t_2) = s_n \mu_n^{(K)}(t_1 - t_2, t_1 + t_2) - \mathbb{1}_{\{t_2 \geq t_1\}} s_n \mu_n^{(K)}(t_1 - t_2, t_2 - t_1) \quad (7.2)$$

expresses the one-vertex characteristic  $s_n \chi_n^{(K)}(t_1, \cdot)$  as the difference of two uniformly bounded, non-negative, non-decreasing functions.

We extend our abbreviations from (6.19) to vectors and write

$$\tilde{\vec{a}} = \vec{a}^{1/s_n} = (\tilde{a}_1, \tilde{a}_2) = (a_1^{1/s_n}, a_2^{1/s_n}) \text{ etc.} \quad (7.3)$$

## 7.1 Truncated Two-Vertex Characteristic: Proof of Theorem 3.15

In this section, we prove Theorem 3.15. For any two-vertex characteristic  $\chi$ , note that  $z_t^\chi(\vec{a})$  can be written in terms of two one-vertex characteristics as follows:

$$z_t^\chi(\vec{a}) = z_{t_1}^{\rho', BP^{(1)}}(a_1), \quad \text{where} \quad \rho'(t'_1) = \sum_{v_2 \in BP_{t_2}^{(2)}} a_2^{|v_2|} \chi(t'_1, t_2 - T_{v_2}) = z_{t_2}^{\chi(t'_1, \cdot), BP^{(2)}}(a_2). \quad (7.4)$$

Similarly, we may evaluate the two-vertex mean  $\tilde{m}_t^{\chi_n^{(K)}}(\vec{a})$  via two one-vertex means:

$$\tilde{m}_t^{\chi_n^{(K)}}(\vec{a}) = \tilde{m}_{t_1}^{\rho_{t_2, \vec{a}_2}}(\tilde{a}_1), \quad \text{where} \quad \rho_{t_2, \vec{a}_2}(t'_1) = \tilde{m}_{t_2}^{\chi_n^{(K)}(t'_1, \cdot)}(\tilde{a}_2). \quad (7.5)$$

For this reason we first give estimates for the one-vertex characteristic  $\chi_n^{(K)}(t_1, \cdot)$  uniformly in  $t_1$ . In their statements, we rely on the function  $\zeta$  given in (3.30) and on the following definition:

$$I(z) = \int_0^\infty \left( e^{-|y-z|} - e^{-(y+z)} \right) \frac{dy}{y}. \quad (7.6)$$

**Proposition 7.1** *Assume the hypotheses of Theorem 6.3. For every  $\varepsilon > 0$  and for every compact subset  $A \subset (0, 2)$ , there is a constant  $K_0 < \infty$  such that for every  $K \geq K_0$  there exist constants  $K' < \infty$  and  $n_0 \in \mathbb{N}$  such that for  $n \geq n_0$ ,  $a_1, a_2, b_1, b_2 \in A$  and  $t'_1 \geq 0$ ,*

$$\left| \tilde{m}_{t_2}^{\chi_n^{(K)}(t'_1, \cdot)}(\tilde{a}_2) - I(\lambda_n(\tilde{a}_2)t'_1) \right| \leq \varepsilon, \quad \text{if } \lambda_n(1)[t_1 \wedge t_2 \wedge u_1 \wedge u_2] \geq K'. \\ \left| s_n^{-3} \tilde{M}_{t_1, u_1}^{\rho_{t_2, \vec{a}_2}, \rho_{u_2, \vec{b}_2}}(\tilde{a}_1, \tilde{b}_1) - \frac{\zeta(a_2/a_1)\zeta(b_2/b_1)}{\log(1/a_1 + 1/b_1)} \right| \leq \varepsilon, \quad (7.7)$$

Moreover, for every  $K < \infty$ , there are constants  $K'' < \infty$  and  $n'_0 \in \mathbb{N}$  such that

$$\tilde{m}_{t_2}^{\chi_n^{(K)}(t_1, \cdot)}(\tilde{a}_2) \leq K'', \quad \tilde{M}_{t_1, u_1}^{\rho_{t_2, \vec{a}_2}, \rho_{u_2, \vec{b}_2}}(\tilde{a}_1, \tilde{b}_1) \leq K'' s_n^3 \quad (7.8)$$

for all  $n \geq n'_0$ ,  $t_1, t_2, u_1, u_2 \geq 0$  and  $a_1, a_2, b_1, b_2 \in A$ .

Note that  $\tilde{m}_{t_2}^{\chi_n^{(K)}(t_1, \cdot)}(\tilde{a}_2)$  is asymptotically constant, instead of growing like  $s_n$  as in Theorem 3.13. This reflects the fact that  $\chi_n^{(K)}$  itself is typically of scale  $1/s_n$ .

**Proof of Proposition 7.1** The integrand in (7.6) can be bounded by  $2e^{-y+1}$  if  $z \leq 1$  and by  $\mathbb{1}_{\{y \leq 1\}} 2e^{-z+1} + \mathbb{1}_{\{y > 1\}} e^{-|y-z|}/y$  if  $z \geq 1$ . It follows that we may choose  $K_0 < \infty$  sufficiently large that  $\left| \int_{ae^{-K-\gamma}}^{ae^{K-\gamma}} (e^{-|y-z|} - e^{-(y+z)}) dy/y - I(z) \right| < \frac{1}{3}\varepsilon$ , for all  $z \geq 0$ ,  $a \in A$ ,  $K \geq K_0$ . Here,  $\gamma$  again denotes Euler's constant.

Applying Theorem 3.13 to each of the uniformly bounded, non-negative, non-decreasing functions in (7.2), we conclude that for every  $K < \infty$ ,  $\bar{m}_{t_2}^{\chi_n^{(K)}(t'_1, \cdot)}(\tilde{a}_2)$  is uniformly bounded and is within  $\frac{1}{3}\varepsilon$  of  $\int_0^\infty s_n \lambda_n(\tilde{a}_2) e^{-\lambda_n(\tilde{a}_2)t} \chi_n^{(K)}(t'_1, t) dt$  if  $\lambda_n(1)t_2$  is sufficiently large, uniformly over  $a_2 \in A$ . Use Fubini's Theorem and (6.12), write  $z = \lambda_n(\tilde{a}_2)t'_1$  and substitute  $\tilde{x} = x^{1/s_n}$  to compute

$$\begin{aligned} & \int_0^\infty s_n \lambda_n(\tilde{a}_2) e^{-\lambda_n(\tilde{a}_2)t} \chi_n^{(K)}(t'_1, t) dt \\ &= s_n \int d\mu_n^{(K)}(y) \int_0^\infty \mathbb{1}_{\{|y-t'_1| \leq t \leq y+t'_1\}} \lambda_n(\tilde{a}_2) e^{-\lambda_n(\tilde{a}_2)t} dt \\ &= s_n \int_{1-K/s_n}^{1+K/s_n} \left( e^{-\lambda_n(\tilde{a}_2)|f_n(\tilde{x})-t'_1|} - e^{-\lambda_n(\tilde{a}_2)(f_n(\tilde{x})+t'_1)} \right) d\tilde{x} \\ &= \int_{(1-K/s_n)^{s_n}}^{(1+K/s_n)^{s_n}} \left( e^{-|\lambda_n(\tilde{a}_2)f_n(\tilde{x})-z|} - e^{-(\lambda_n(\tilde{a}_2)f_n(\tilde{x})+z)} \right) x^{1/s_n-1} dx. \end{aligned} \quad (7.9)$$

Monotonicity, Condition 2.1 and (6.15) imply that  $\lambda_n(\tilde{a}_2)f_n(\tilde{x}) \rightarrow a_2 e^{-\gamma}x$  as  $n \rightarrow \infty$ , uniformly over  $x \in [(1-K/s_n)^{s_n}, (1+K/s_n)^{s_n}]$  and  $a_2 \in A$ . Hence the integral in (7.9) is within  $\frac{1}{3}\varepsilon$  of  $\int_{e^{-K}}^{e^K} (e^{-|a_2 e^{-\gamma}x-z|} - e^{-(a_2 e^{-\gamma}x+z)}) dx/x$  for  $n$  sufficiently large, uniformly in  $z \geq 0$  and  $a_2 \in A$ . Substituting  $y = a_2 e^{-\gamma}x$  and using  $K \geq K_0$ , we obtain the desired statements for  $\bar{m}_{t_2}^{\chi_n^{(K)}(t'_1, \cdot)}(\tilde{a}_2)$ .

For  $\bar{M}_{t_1, u_1}^{\rho_{t_2, \tilde{a}_2}, \rho_{u_2, \tilde{b}_2}}(\tilde{a}_1, \tilde{b}_1)$ , the statements for  $\bar{m}_{t_2}^{\chi_n^{(K)}(t'_1, \cdot)}(\tilde{a}_2)$  can be interpreted to say that the characteristics  $\rho_{t_2, \tilde{a}_2}(\cdot)$ ,  $\rho_{u_2, \tilde{b}_2}(\cdot)$  are uniformly bounded and lie within  $\frac{1}{2}\varepsilon$  of the characteristics  $I(\lambda_n(\tilde{a}_2) \cdot)$ ,  $I(\lambda_n(\tilde{b}_2) \cdot)$  if  $\lambda_n(1)t_2$ ,  $\lambda_n(1)u_2$  are sufficiently large. It is readily verified that  $I(z)$  can be written as the difference of two bounded, non-negative, non-decreasing functions. We may therefore apply Theorem 3.13 to these characteristics. A calculation shows that

$$\int_0^\infty e^{-z} I(rz) dz = \zeta(r), \quad r = \lambda_n(\tilde{a}_2)/\lambda_n(\tilde{a}_1), \quad (7.10)$$

where  $\zeta$  is defined in (3.30). By (6.15), we have  $r \rightarrow a_2/a_1$  uniformly over  $a_1, a_2 \in A$ ; since  $\zeta$  is continuous, this completes the proof.  $\square$

**Proof of Theorem 3.15** Interchanging the roles of  $t_1$  and  $t_2$  in (7.2) and using Proposition 7.1, we can write  $\rho_{t_2, \tilde{a}_2}$  in (7.5) as the difference of two bounded, non-negative, non-decreasing functions and Theorem 3.13 yields that  $\frac{1}{s_n} \bar{m}_{t_1}^{\chi_n^{(K)}}(\tilde{a})$  is bounded. To show (3.28), Proposition 7.1 allows us to replace  $\rho_{t_2, \tilde{a}_2}$  in (7.5) by  $I(\lambda_n(\tilde{a}_2) \cdot)$ , making an error of at most  $\varepsilon s_n$ . Since  $I$  can be written as the difference of two bounded, non-negative, non-decreasing functions, Theorem 3.13, (7.10) and the fact that  $\zeta(r) \rightarrow \zeta(a_2/a_1)$  uniformly, yield the claim.

For  $\bar{M}_{t, \tilde{u}}^{\chi_n^{(K)}, \chi_n^{(K)}}(\tilde{a}, \tilde{b})$ , use (7.4) to obtain [similarly to (6.8)]

$$\begin{aligned} \bar{M}_{t, \tilde{u}}^{\chi, \eta}(\tilde{a}, \tilde{b}) &= \mathbb{E}_{a_1 b_1, a_2 b_2} \left( \sum_{j_1=0}^\infty \sum_{j_2=0}^\infty e^{-[\lambda_n(a_1)+\lambda_n(b_1)-\lambda_n(a_1 b_1)]S_{j_1}^{(1)}} e^{-[\lambda_n(a_2)+\lambda_n(b_2)-\lambda_n(a_2 b_2)]S_{j_2}^{(2)}} \right. \\ &\quad \cdot \bar{m}_{t_1-S_{j_1}^{(1)}, t_2-S_{j_2}^{(2)}}^{\chi}(\tilde{a}) \bar{m}_{u_1-S_{j_1}^{(1)}, u_2-S_{j_2}^{(2)}}^{\eta}(\tilde{b}) \Big), \end{aligned} \quad (7.11)$$

where now  $(S_j^{(1)})_j$  and  $(S_j^{(2)})_j$  are independent random walks and  $(S_j^{(i)})_j$  has step distribution  $\nu_{a_i b_i}$ ,  $i = 1, 2$ . Applying Lemma 6.6 twice and using the results from the first part of the proof, we obtain the desired conclusions.  $\square$

## 7.2 The Effect of Truncation: Proof of Theorem 3.16

In this section, we control the effect of truncation and prove Theorem 3.16, by showing that the remainder  $\chi_n - \chi_n^{(K)}$  has a negligible first moment.

We will write  $\chi_n = \int d\mu_n(y) \Psi_y$ , where

$$\Psi_y(t_1, t_2) = \mathbb{1}_{\{|t_1 - t_2| \leq y, t_1 + t_2 \geq y\}}. \quad (7.12)$$

The same is true for  $\chi_n^{(K)}$  and  $\mu_n^{(K)}$ , so that, by (6.12) and the substitution  $\tilde{x} = x^{1/s_n}$ ,

$$\begin{aligned} s_n^{-1} \bar{m}_t^{\chi_n - \chi_n^{(K)}}(\vec{1}) &= \int_0^{1-\delta_0} s_n^{-1} \bar{m}_t^{\Psi_{f_n(\tilde{x})}}(\vec{1}) d\tilde{x} \\ &\quad + \left( \int_{(1-\delta_0)^{s_n}}^{(1-K/s_n)^{s_n}} + \int_{(1+K/s_n)^{s_n}}^{\infty} \right) s_n^{-2} \bar{m}_t^{\Psi_{f_n(\tilde{x})}}(\vec{1}) x^{1/s_n-1} dx \\ &= I_0 + I_1 + I_2. \end{aligned} \quad (7.13)$$

We must therefore show that  $I_0, I_1, I_2$  are uniformly bounded and can be made small by making  $K$  and  $\lambda_n(1)[t_1 \wedge t_2]$  large. To this end, we will bound the two-vertex mean  $\bar{m}_t^{\Psi_y}(\vec{1})$  in terms of one-vertex means. Abbreviate

$$\eta^{(q)}(t) = \mathbb{1}_{\{0 \leq t \leq q\}}. \quad (7.14)$$

**Lemma 7.2** *For any  $y \in (0, \infty)$ ,*

$$\begin{aligned} \bar{m}_t^{\Psi_y}(\vec{1}) &\leq \frac{1}{y} \left[ \int_0^\infty e^{-2\lambda_n(1)r} \bar{m}_{t_1-r}^{\eta^{(2y)}}(1) \bar{m}_{t_2-r}^{\eta^{(2y)}}(1) dr \right. \\ &\quad \left. + \int_0^y e^{-\lambda_n(1)(y-r)} \left( \bar{m}_{t_1-y+r}^{\eta^{(2r)}}(1) \bar{m}_{t_2}^{\eta^{(r)}}(1) + \bar{m}_{t_2-y+r}^{\eta^{(2r)}}(1) \bar{m}_{t_1}^{\eta^{(r)}}(1) \right) dr \right]. \end{aligned} \quad (7.15)$$

**Proof** Note that

$$\Psi_y \leq \frac{1}{y} \left[ \int_0^\infty \mathbb{1}_{[r, r+2y]^2} dr + \int_0^y (\mathbb{1}_{[y-r, y+r] \times [0, r]} + \mathbb{1}_{[0, r] \times [y-r, y+r]}) dr \right] \quad (7.16)$$

since, for any  $\vec{t}$  for which  $\Psi_y(\vec{t}) > 0$ , the measure of the sets of parameter values  $r$  for which  $\vec{t}$  belongs to the relevant rectangles is at least  $y$  in total. Then the identities  $\bar{m}_{t_1, t_2}^{\mathbb{1}_{[a, b] \times [c, d]}}(\vec{a}) = \bar{m}_{t_1}^{\mathbb{1}_{[a, b]}}(a_1) \bar{m}_{t_2}^{\mathbb{1}_{[c, d]}}(a_2)$  and

$$\bar{m}_t^{\mathbb{1}_{[c, d]}}(a) = e^{-\lambda_n(a)c} \bar{m}_{t-c}^{\eta^{(d-c)}}(a) \quad (7.17)$$

complete the proof.  $\square$

Using Lemma 7.2, it will suffice to bound the one-vertex means  $\bar{m}_t^{\eta^{(q)}}(1)$ . We will use different bounds depending on the relative sizes of  $q, t$  and  $f_n(1)$ , as in the following lemma:

**Lemma 7.3** *There is a constant  $C < \infty$  such that, for  $n$  sufficiently large,*

$$\bar{m}_t^{\eta^{(q)}}(1) \leq C s_n \text{ for all } t, q \geq 0, \quad (7.18a)$$

$$\bar{m}_t^{\eta^{(q)}}(1) \leq C s_n \frac{q + s_n f_n(1 - \delta_0)}{f_n(1)} \text{ if } t \geq f_n(1 - 1/s_n), \quad q \leq \frac{1}{2}t, \quad (7.18b)$$

$$\bar{m}_t^{\eta^{(q)}}(1) \leq C \frac{q + s_n f_n(1 - \delta_0)}{s_n t (1 - f_n^{-1}(t))^2} \text{ if } t < f_n(1), \quad q \leq \frac{1}{2}t, \quad (7.18c)$$

$$\bar{m}_t^{\eta^{(2q)}}(1) \leq \frac{2}{1 - f_n^{-1}(q)}, \text{ if } q < f_n(1), \quad t \geq 0. \quad (7.18d)$$

**Proof** Theorem 3.13 and  $\|\eta^{(q)}\|_\infty = 1$  imply (7.18a). For (7.18d), use the representation (6.7) and note that, starting from the first index  $J$  for which  $\eta^{(2q)}(t - S_J) \neq 0$  (if one exists), the total number of indices  $j$  for which  $\eta^{(2q)}(t - S_j) \neq 0$  is stochastically bounded by the waiting time (starting from  $J$ ) until the second step where  $Y_j > q$ . Then  $\mathbb{P}_1(D_j \leq q) \leq f_n^{-1}(q)$  proves (7.18d).

For (7.18b)–(7.18c), we employ a size-biasing argument on the jump sizes  $D_i$ . For  $i \leq j$ , write  $S'_{j,i} = \sum_{1 \leq k \leq j, k \neq i} D_k$ . We can therefore rewrite (6.7) (noting that the term  $j = 0$  vanishes) as

$$\bar{m}_t^{\eta^{(q)}}(1) = \sum_{j=1}^{\infty} \sum_{i=1}^j \mathbb{E}_1 \left( e^{-\lambda_n(1)(t-S'_{j,i})} \mathbb{E}_1 \left( e^{\lambda_n(1)D_i} \frac{D_i}{S'_{j,i} + D_i} \mathbb{1}_{\{t-q-S'_{j,i} \leq D_i \leq t-S'_{j,i}\}} \middle| S'_{j,i} \right) \right). \quad (7.19)$$

We split according to whether  $D_i > f_n(1 - \delta_0)$  or  $D_i \leq f_n(1 - \delta_0)$ . For any measurable function  $h \geq 0$ , (6.9) and Lemma 4.2 imply

$$\mathbb{E}_1 \left( e^{\lambda_n(1)D_i} D_i h(D_i) \mathbb{1}_{\{D_i > f_n(1-\delta_0)\}} \right) \leq \frac{1}{\varepsilon_0 s_n} \int h(y) f_n^{-1}(y) dy. \quad (7.20)$$

On the other hand,  $\mathbb{E}_1 \left( e^{\lambda_n(1)D_i} D_i h(D_i) \mathbb{1}_{\{D_i \leq f_n(1-\delta_0)\}} \right) \leq \max \{yh(y) : y \leq f_n(1 - \delta_0)\}$  by (6.9). Consequently, writing  $x^+ := x \vee 0$ ,

$$\begin{aligned} \bar{m}_t^{\eta^{(q)}}(1) &\leq \sum_{j=1}^{\infty} \sum_{i=1}^j \mathbb{E}_1 \left( e^{-\lambda_n(1)(t-S'_{j,i})} \mathbb{1}_{\{t-S'_{j,i} \geq 0\}} \left[ \frac{1}{\varepsilon_0 s_n} \int_{(t-q-S'_{j,i})^+}^{t-S'_{j,i}} \frac{f_n^{-1}(y)}{S'_{j,i} + y} dy + \frac{f_n(1-\delta_0)}{t-q} \right] \right) \\ &\leq \sum_{j=1}^{\infty} \sum_{i=1}^j \mathbb{E}_1 \left( e^{-\lambda_n(1)(t-S'_{j,i})} \mathbb{1}_{\{t-S'_{j,i} \geq 0\}} \left[ \frac{f_n^{-1}(t-S'_{j,i})}{\varepsilon_0 s_n} \log \left( \frac{t}{t-q} \right) + \frac{f_n(1-\delta_0)}{t-q} \right] \right) \\ &\leq \frac{2}{t} \sum_{j=1}^{\infty} \sum_{i=1}^j \mathbb{E}_1 \left( e^{-\lambda_n(1)(t-S'_{j,i})} \mathbb{1}_{\{t-S'_{j,i} \geq 0\}} \left[ \frac{q f_n^{-1}(t-S'_{j,i})}{\varepsilon_0 s_n} + f_n(1-\delta_0) \right] \right) \end{aligned} \quad (7.21)$$

since  $-\log(1-x) \leq 2x$  for  $x \in [0, 1/2]$  and  $q \leq \frac{1}{2}t$ . To obtain (7.18c), we note that  $\mathbb{P}_1(S'_{j,i} \leq t) \leq f_n^{-1}(t)^{j-1}$ , as in the proof of (7.18d), and therefore

$$\bar{m}_t^{\eta^{(q)}}(1) \leq C' \frac{q f_n^{-1}(t) + s_n f_n(1 - \delta_0)}{s_n t} \sum_{j=1}^{\infty} j f_n^{-1}(t)^{j-1}, \quad (7.22)$$

and  $f_n^{-1}(t) < f_n^{-1}(f_n(1)) = 1$  completes the proof of (7.18c).



Finally, to prove (7.18b) we now reverse the argument that led to (7.21) by reintroducing a term  $D_i$ . By (6.12),

$$\int y_i \mathbb{1}_{\{y_i \leq f_n(1)\}} d\mu_n(y_i) = \int_0^1 f_n(\tilde{x}) d\tilde{x} \geq \int_{1-1/s_n}^1 f_n(1-1/s_n) d\tilde{x} = \frac{f_n(1-1/s_n)}{s_n}, \quad (7.23)$$

so that  $\mathbb{E}_1(e^{\lambda_n(1)D_i} D_i \mathbb{1}_{\{D_i \leq f_n(1)\}} / f_n(1-1/s_n)) \geq 1/s_n$ . Abbreviate  $\rho(u) = \mathbb{1}_{\{u \geq 0\}} [q f_n^{-1}(u)/\varepsilon_0 + s_n f_n(1-\delta_0)]$ . Note that  $\rho$  is increasing and that on  $\{D_i \leq f_n(1)\}$  we have  $t - S'_{j,i} \leq t - S_j + f_n(1)$ . Continuing from (7.21), we estimate

$$\begin{aligned} \bar{m}_t^{\eta^{(q)}}(1) &\leq \frac{2}{t} \sum_{j=1}^{\infty} \sum_{i=1}^j \mathbb{E}_1 \left( e^{-\lambda_n(1)(t-S'_{j,i})} \rho(t-S'_{j,i}) \mathbb{E}_1 \left( e^{\lambda_n(1)D_i} \frac{D_i}{f_n(1-1/s_n)} \mathbb{1}_{\{D_i \leq f_n(1)\}} \right) \right) \\ &\leq \frac{2}{t} \sum_{j=1}^{\infty} \sum_{i=1}^j \mathbb{E}_1 \left( e^{-\lambda_n(1)(t-S_j)} \rho(t+f_n(1)-S_j) \frac{D_i}{f_n(1-1/s_n)} \right) \\ &= \frac{2e^{\lambda_n(1)f_n(1)}}{f_n(1-1/s_n)} \sum_{j=0}^{\infty} \mathbb{E}_1 \left( e^{-\lambda_n(1)(t+f_n(1)-S_j)} \rho(t+f_n(1)-S_j) \frac{S_j}{t} \right) \\ &\leq \frac{2e^{\lambda_n(1)f_n(1)}}{f_n(1-1/s_n)} \frac{t+f_n(1)}{t} \bar{m}_{t+f_n(1)}^{\rho}(1), \end{aligned} \quad (7.24)$$

where in the last inequality we have used that  $\rho(t+f_n(1)-S_j) = 0$  if  $S_j > t+f_n(1)$ . As in the proof of Theorem 3.13, we use Lemma 6.9, Theorem 6.3 and the definition of  $\rho$  to obtain

$$\begin{aligned} \bar{m}_{t+f_n(1)}^{\rho}(1) &\leq O(1) \int_0^{\infty} s_n \lambda_n(1) e^{-\lambda_n(1)u} \rho(u) du \\ &= O(s_n) \left[ \frac{q}{\varepsilon_0} \int_0^{\infty} \lambda_n(1) e^{-\lambda_n(1)u} f_n^{-1}(u) du + s_n f_n(1-\delta_0) \right]. \end{aligned} \quad (7.25)$$

By Condition 6.2, we have  $f_n^{-1}(u) \leq (u/f_n(1))^{1/\varepsilon_0 s_n}$  for  $u \geq f_n(1)$ . Changing variables and using that  $\varepsilon_0 s_n \geq 1$  for large  $n$ , we obtain

$$\begin{aligned} \int_0^{\infty} \lambda_n(1) e^{-\lambda_n(1)u} f_n^{-1}(u) du &\leq 1 + \int_0^{\infty} \lambda_n(1) f_n(1) e^{-\lambda_n(1)f_n(1)u} u \\ du &= 1 + \frac{1}{\lambda_n(1)f_n(1)} = O(1) \end{aligned} \quad (7.26)$$

according to (6.15). Hence  $\bar{m}_{t+f_n(1)}^{\rho}(1) = O(s_n)(q + s_n f_n(1-\delta_0))$ . The other factors in (7.24) are  $O(1/f_n(1))$  because of (6.15), Condition 2.1, and the assumption  $t \geq f_n(1-1/s_n)$ . This completes the proof of (7.18b).  $\square$

To make use of the bounds (7.18c)–(7.18d), we note the following consequence of Condition 2.2:

**Lemma 7.4** *Suppose without loss of generality that the constant  $\delta_0$  from Condition 2.2 satisfies  $\delta_0 < 1$ . Then there exists  $C < \infty$  such that, uniformly over  $f_n(1-\delta_0) \leq u \leq f_n(1)$ ,*

$$f_n^{-1}(u) \leq 1 - \frac{1}{C s_n} \log(f_n(1)/u). \quad (7.27)$$

**Proof** It suffices to show that  $x \leq 1 - (Cs_n)^{-1} \log(f_n(1)/f_n(x))$  for  $1 - \delta_0 \leq x \leq 1$ , i.e., that  $\log(f_n(1)/f_n(x)) \leq Cs_n(1 - x)$ . But Condition 2.2 implies that  $\log(f_n(1)/f_n(x)) \leq \varepsilon_0^{-1} s_n \log(1/x)$ , as in the proof of Lemma 4.1, so a Taylor expansion gives the result.

**Proof of Theorem 3.16** We will show that each of the terms  $I_0, I_1, I_2$  in (7.13) is uniformly bounded, and furthermore can be made arbitrarily small by taking  $K$  large enough (for  $I_1$  and  $I_2$ ) and  $\lambda_n(1)[t_1 \wedge t_2]$  large enough (for  $I_0$ ). We begin with the term  $I_2$  [i.e.,  $x \geq (1 + K/s_n)^{s_n}$ ]. Lemma 7.2, (7.18a) and (6.15) give

$$\bar{m}_t^{\Psi_{f_n(\tilde{x})}}(\vec{1}) \leq \frac{O(s_n^2)}{f_n(\tilde{x})} \left[ \int_0^\infty e^{-2\lambda_n(1)r} dr + 2 \int_0^{f_n(\tilde{x})} e^{-\lambda_n(1)(f_n(\tilde{x})-r)} dr \right] = \frac{O(s_n^2)f_n(1)}{f_n(\tilde{x})}. \quad (7.28)$$

By Lemma 4.1,  $f_n(\tilde{x}) \geq f_n(1)x^{\varepsilon_0}$  for  $x \geq 1$ , so that  $I_2 \leq O(1) \int_{(1+K/s_n)^{s_n}}^\infty x^{1/s_n - \varepsilon_0 - 1} dx$ . Since  $\int_1^\infty x^{-\varepsilon_0 - 1} dx < \infty$ , it follows that  $I_2$  is uniformly bounded and can be made arbitrarily small by taking  $K$ , and hence  $(1 + K/s_n)^{s_n}$ , large enough, uniformly over  $t_1, t_2$ .

For  $I_1$ , we again start by estimating  $\bar{m}_t^{\Psi_y}(\vec{1})$  where  $y = f_n(\tilde{x})$  with  $\tilde{x} \in [1 - \delta_0, 1 - K/s_n]$ . Suppose for definiteness, and without loss of generality by symmetry, that  $t_1 \leq t_2$ . Split the first integral from Lemma 7.2 into the intervals  $[0, t_1 - f_n(1 - 1/s_n)]$ ,  $[t_1 - f_n(1 - 1/s_n), t_1 - 4y]$  and  $[t_1 - 4y, t_1]$  (noting that the integrand vanishes for  $r > t_1$ ) and denote the summand by  $\theta_n^{11}(y)$ ,  $\theta_n^{12}(y)$  and  $\theta_n^{13}(y)$ . The second summand in Lemma 7.2 is called  $\theta_n^{14}(y)$ . The corresponding parts of  $I_1$  are denoted by  $I_{11}, \dots, I_{14}$ .

We first estimate  $\theta_n^{13}(y)$  and  $\theta_n^{14}(y)$ . Since  $f_n(1 - \delta_0) \leq y \leq f_n(1 - K/s_n) < f_n(1)$ , (7.18d) and Lemma 7.4 give

$$\begin{aligned} \theta_n^{13}(y) + \theta_n^{14}(y) &\leq \frac{1}{y} \left[ \int_{t_1 - 4y}^{t_1} \left( \frac{2}{1 - f_n^{-1}(y)} \right)^2 dr + \int_0^y 2 \frac{2}{1 - f_n^{-1}(r)} \frac{2}{1 - f_n^{-1}(r/2)} dr \right] \\ &\leq \frac{O(1)}{(1 - f_n^{-1}(y))^2} \leq \frac{O(s_n^2)}{(\log(f_n(1)/y))^2}. \end{aligned} \quad (7.29)$$

According to Lemma 4.1,  $y = f_n(\tilde{x}) \leq f_n(1)x^{\varepsilon_0}$  for all  $1 - \delta_0 \leq \tilde{x} \leq 1$ . Substitute  $x = e^{-u}$  to obtain

$$I_{13} + I_{14} \leq O(1) \int_{(1-\delta_0)^{s_n}}^{(1-K/s_n)^{s_n}} \frac{1}{(\log(1/x^{\varepsilon_0}))^2} x^{1/s_n - 1} dx \leq O(1) \int_K^\infty \frac{1}{u^2} du. \quad (7.30)$$

Hence  $I_{13} + I_{14}$  is uniformly bounded and can be made arbitrarily small by taking  $K$  large.

For  $\theta_n^{11}(y)$ ,  $r \in [0, t_1 - f_n(1 - 1/s_n)]$  implies  $t_2 - r \geq t_1 - r \geq f_n(1 - 1/s_n)$  and since  $2y \leq 2f_n(1 - K/s_n) \leq \frac{1}{2}f_n(1 - 1/s_n)$  for large  $n$  by Condition 2.1, we can apply first (7.18b) and then (6.15) to obtain

$$\begin{aligned} \theta_n^{11}(y) &\leq \frac{1}{y} \int_0^{t_1 - f_n(1 - 1/s_n)} e^{-2\lambda_n(1)r} O(s_n^2) \left( \frac{2y + s_n f_n(1 - \delta_0)}{f_n(1)} \right)^2 dr \\ &= O(s_n^2) \frac{(y + s_n f_n(1 - \delta_0))^2}{y f_n(1)}. \end{aligned} \quad (7.31)$$

Using that  $(a + b)^2 \leq 2(a^2 + b^2)$  for all  $a, b \in \mathbb{R}$  and that  $f_n(\tilde{x}) \leq f_n(1)x^{\varepsilon_0}$  and, for  $x \in (1 - \delta_0, 1 - K/s_n)$ ,  $f_n(\tilde{x}) \geq f_n(1 - \delta_0)$ , we obtain

$$\begin{aligned} I_{11} &\leq \int_{(1-\delta_0)^{s_n}}^{(1-K/s_n)^{s_n}} O(1) \left( \frac{f_n(\tilde{x})}{f_n(1)} + \frac{s_n^2 f_n(1 - \delta_0)^2}{f_n(\tilde{x}) f_n(1)} \right) x^{1/s_n - 1} dx \\ &\leq O(1) \left[ \int_{(1-\delta_0)^{s_n}}^{(1-K/s_n)^{s_n}} x^{\varepsilon_0 + 1/s_n - 1} dx + s_n^2 \frac{f_n(1 - \delta_0)}{f_n(1)} \int_{(1-\delta_0)^{s_n}}^{(1-K/s_n)^{s_n}} x^{1/s_n - 1} dx \right]. \end{aligned} \quad (7.32)$$

The first summand is bounded and can be made arbitrarily small by choosing  $K$  large. The second summand is arbitrarily small for large  $n$  uniformly in  $K$  since  $f_n(1 - \delta_0)/f_n(1) \leq (1 - \delta_0)^{\varepsilon_0 s_n} = o(s_n^{-3})$  according to Lemma 4.1.

For  $\theta_n^{12}(y)$  we substitute  $u = t_1 - r$  to obtain

$$\theta_n^{12}(y) = \frac{1}{y} \int_{4y}^{f_n(1-1/s_n)} e^{-2\lambda_n(1)(t_1-u)} \bar{m}_u^{\eta(2y)}(1) \bar{m}_{t_2-t_1+u}^{\eta(2y)}(1) du. \quad (7.33)$$

We consider the two cases  $t_2 - t_1 \geq f_n(1)/2$  and  $0 \leq t_2 - t_1 < f_n(1)/2$  separately. First  $t_2 - t_1 \geq f_n(1)/2$ . Then  $t_2 - t_1 + u \geq f_n(1)/2 + 4f_n(1 - \delta_0) \geq f_n(1 - 1/s_n)$  for sufficiently large  $n$ . Hence (7.18b), (7.18c) and Lemma 7.4 yield

$$\begin{aligned} \theta_n^{12}(y) &\leq \frac{1}{y} \int_{4y}^{f_n(1-1/s_n)} O(1) \frac{2y + s_n f_n(1 - \delta_0)}{s_n u (1 - f_n^{-1}(u))^2} s_n \frac{2y + s_n f_n(1 - \delta_0)}{f_n(1)} du \\ &\leq O(s_n^2) \frac{(y + s_n f_n(1 - \delta_0))^2}{y f_n(1)} \int_{4y}^{f_n(1-1/s_n)} \frac{1}{u (\log(f_n(1)/u))^2} du \\ &\leq O(s_n^2) \frac{(y + s_n f_n(1 - \delta_0))^2}{y f_n(1)} \int_{\log(f_n(1)/f_n(1-1/s_n))}^{\infty} \frac{1}{\xi^2} d\xi, \end{aligned} \quad (7.34)$$

where the last integral [in which  $\xi = \log(f_n(1)/u)$ ] is  $O(1)$  since  $\log(f_n(1)/f_n(1-1/s_n)) \rightarrow 1$  as  $n \rightarrow \infty$ . Hence, in the case  $t_2 - t_1 \geq f_n(1)/2$ , we have the same bound for  $\theta_n^{12}(y)$  as for  $\theta_n^{11}(y)$ .

Now let  $t_2 - t_1 < f_n(1)/2$ , and abbreviate  $u' = t_2 - t_1 + u$ . Recalling that  $f_n(1 - 1/s_n)/f_n(1) \rightarrow e^{-1}$ , we have  $u' \leq f_n(1)/2 + f_n(1 - 1/s_n) < \frac{9}{10} f_n(1)$  for large  $n$ , uniformly over  $u \leq f_n(1 - 1/s_n)$ . We first apply (7.18c) to both factors in (7.33) and then use Lemma 7.4 to obtain

$$\begin{aligned} \theta_n^{12}(y) &\leq \frac{1}{y} \int_{4y}^{f_n(1-1/s_n)} O(1) \frac{2y + s_n f_n(1 - \delta_0)}{s_n u (1 - f_n^{-1}(u))^2} \frac{2y + s_n f_n(1 - \delta_0)}{s_n u' (1 - f_n^{-1}(u'))^2} du \\ &\leq O(1) \frac{(y + s_n f_n(1 - \delta_0))^2}{y s_n^2} \int_{4y}^{f_n(1-1/s_n)} \frac{s_n^2}{u (\log(f_n(1)/u))^2} \frac{s_n^2}{u' (\log(f_n(1)/u'))^2} du. \end{aligned} \quad (7.35)$$

In (7.35), we have  $u' \geq u$ , and we are particularly concerned with the case  $u' = u$ . The function  $u' \mapsto (u'/f_n(1))(\log(f_n(1)/u'))^2$  is not monotone over  $u' \in [u, \frac{9}{10} f_n(1)]$ , but it is increasing over  $(0, \frac{1}{10} f_n(1))$  and bounded from zero and infinity over  $[\frac{1}{10} f_n(1), \frac{9}{10} f_n(1)]$ . We may therefore find a constant  $c > 0$  such that  $u'(\log(f_n(1)/u'))^2 \geq cu(\log(f_n(1)/u))^2$  whenever  $u \leq u' \leq \frac{9}{10} f_n(1)$ . The bound (7.35) may therefore be simplified to

$$\theta_n^{12}(y) \leq O(s_n^2) \frac{(y + s_n f_n(1 - \delta_0))^2}{y} \int_{4y}^{f_n(1-1/s_n)} \frac{1}{u^2 \log(f_n(1)/u)^4} du, \quad (7.36)$$

and an integration by parts shows that

$$\theta_n^{12}(y) \leq O(s_n^2) \frac{(y + s_n f_n(1 - \delta_0))^2}{y^2 (\log(f_n(1)/y))^4}. \quad (7.37)$$

Inserting the bound (7.37) into  $I_{12}$  and using again  $(a + b)^2 \leq 2(a^2 + b^2)$ , we conclude that

$$I_{12} \leq \int_{(1-\delta_0)^{s_n}}^{(1-K/s_n)^{s_n}} O(1) \frac{f_n(\tilde{x})^2 + s_n^2 f_n(1 - \delta_0)^2}{f_n(\tilde{x})^2 (\log(f_n(1)/f_n(\tilde{x})))^4} x^{1/s_n-1} dx. \quad (7.38)$$

Recall from Lemma 4.1 the bounds  $f_n(\tilde{x}) \leq f_n(1)x^{\varepsilon_0}$  and  $f_n(1 - \delta_0) \leq f_n(\tilde{x})((1 - \delta_0)/\tilde{x})^{\varepsilon_0 s_n}$ . We split the integral for the second summand in (7.38) at  $(1 - \delta_0/2)^{s_n}$ . For  $x \geq (1 - \delta_0/2)^{s_n}$ ,  $f_n(1 - \delta_0) \leq f_n(\tilde{x})(1 - \delta')^{\varepsilon_0 s_n}$  for some  $\delta' \in (0, 1)$ . For  $x \leq (1 - \delta_0/2)^{s_n}$ ,  $\log(f_n(1)/f_n(\tilde{x})) \geq \varepsilon_0 \log(1/x) \geq c s_n$  for some  $c > 0$ . Hence

$$\begin{aligned} I_{12} &\leq O(1) \left( 1 + s_n^2 (1 - \delta')^{\varepsilon_0 s_n} \right) \int_{(1-\delta_0)^{s_n}}^{(1-K/s_n)^{s_n}} \frac{1}{x (\log(1/x))^4} dx \\ &\quad + O(1) \int_{(1-\delta_0)^{s_n}}^{(1-\delta_0/2)^{s_n}} \frac{s_n^2 f_n(1 - \delta_0)^2}{f_n(\tilde{x})^2 s_n^4} x^{1/s_n-1} dx \\ &\leq O(1) \int_0^{e^{-K}} \frac{1}{x (\log(1/x))^4} dx + O(s_n^{-2}) \int_{(1-\delta_0)^{s_n}}^{(1-\delta_0/2)^{s_n}} (1 - \delta_0)^{2\varepsilon_0 s_n} x^{-1-2\varepsilon_0} dx. \end{aligned} \quad (7.39)$$

The first summand in (7.39) is bounded and can be made arbitrarily small by choosing  $K$  large, while the second summand is arbitrarily small for large  $n$  uniformly in  $K$ . We have now handled all four contributions to  $I_1$ .

Finally, for  $I_0$ , let  $y \leq f_n(1 - \delta_0)$  and note that  $\Psi_y(t'_1, \cdot) = \mathbb{1}_{[|t'_1 - y|, t'_1 + y]} \leq \mathbb{1}_{[(t'_1 - y)^+, t'_1 + y]}$ , where  $(t'_1 - y)^+ := (t'_1 - y) \vee 0$ . From (7.17) and (7.18d) it follows that

$$\begin{aligned} \bar{m}_{t_2}^{\Psi_y(t'_1, \cdot)}(1) &\leq \mathbb{1}_{\{t_2 \geq t'_1 - f_n(1 - \delta_0)\}} e^{-\lambda_n(1)(t'_1 - y)^+} \bar{m}_{t_2 - (t'_1 - y)^+}^{\eta_{t'_1 + y - (t'_1 - y)^+}}(1) \\ &\leq \mathbb{1}_{\{t_2 \geq t'_1 - f_n(1 - \delta_0)\}} \frac{2}{1 - f_n^{-1}(2f_n(1 - \delta_0))} = O(1) \mathbb{1}_{[0, t_2 + f_n(1 - \delta_0)]}(t'_1), \end{aligned} \quad (7.40)$$

and therefore  $\bar{m}_t^{\Psi_y}(\vec{1}) \leq O(s_n)$  by (7.4)–(7.5) and (7.18a). We conclude that  $I_0$  is uniformly bounded. To show the smallness, we sharpen the bound (7.40) when  $t'_1$  is far from  $t_2$  by using (7.18b) instead of (7.18d), to obtain for  $t_2 \geq t'_1 + f_n(1 - \delta_0) + f_n(1)$ ,

$$\bar{m}_{t_2}^{\Psi_y(t'_1, \cdot)}(1) \leq \bar{m}_{t_2 - (t'_1 - y)^+}^{\eta_{t'_1 + y - (t'_1 - y)^+}}(1) \leq O(s_n) \frac{2y + s_n f_n(1 - \delta_0)}{f_n(1)} = O(s_n^2) \frac{f_n(1 - \delta_0)}{f_n(1)}. \quad (7.41)$$

Combining (7.40)–(7.41), we have  $\bar{m}_{t_2}^{\Psi_y(t'_1, \cdot)}(1) \leq O(s_n^2) \frac{f_n(1 - \delta_0)}{f_n(1)} + O(1) \mathbb{1}_{\{t_2 - 2f_n(1) \leq t'_1 \leq t_2 + f_n(1)\}}$ . Applying (7.4)–(7.5), (7.17), (7.18a) and (6.15) we conclude that

$$\bar{m}_t^{\Psi_y}(\vec{1}) \leq O(s_n^3) \frac{f_n(1 - \delta_0)}{f_n(1)} + O(s_n) e^{-\lambda_n(1)t_2} \quad (7.42)$$

and consequently  $I_0$  may be made small by taking  $t_2$  large.  $\square$

## 8 First Points of Cox Processes: Proof of Theorem 3.29

Let  $\mathcal{X}$  denote a topological space equipped with its Borel  $\sigma$ -field and let  $(\mathcal{P}_n)_{n \geq 1}$  be a sequence of Cox processes on  $\mathbb{R} \times \mathcal{X}$  with random intensity measures  $(Z_n)_{n \geq 1}$ . That is, there exist  $\sigma$ -fields  $\mathcal{F}_n$  such that  $Z_n$  is  $\mathcal{F}_n$ -measurable and, conditionally on  $\mathcal{F}_n$ ,  $\mathcal{P}_n$  is a Poisson point process with (random) intensity  $Z_n$ . For instance, Theorem 3.27 expresses the first passage distance and hopcount in terms of the first point of a Cox process. In this section, we determine sufficient conditions to identify the limiting distribution of the first points of  $\mathcal{P}_n$  based on the intensity measure at fixed times  $t$ .

This section is organised as follows. We start in Sect. 8.1 with preparations of convergence of Cox processes. In Sect. 8.2, we use these results to prove Theorem 3.29.

### 8.1 Preparations: Weak Convergence of Cox Processes

We will write  $\mathcal{P}_{n,t}$  for the measure defined by  $\mathcal{P}_{n,t}(\cdot) = \mathcal{P}_n((-\infty, t] \times \cdot)$ , and given a partition  $t_0 < \dots < t_N$  we abbreviate  $\Delta \mathcal{P}_{n,i} = \mathcal{P}_{n,t_i} - \mathcal{P}_{n,t_{i-1}}$ ; similarly for  $Z_{n,t}$ ,  $\Delta Z_{n,i}$ . Write  $|\mu|$  for the total mass of a measure  $\mu$ .

Define

$$T_{n,k} = \inf \{t : |\mathcal{P}_{n,t}| \geq k\}, \quad (8.1)$$

and let  $A_{n,k}$  be the event that  $T_{n,j} \notin \{\pm\infty\}$  and  $|\mathcal{P}_{n,T_{n,j}}| = j$ , for  $j = 1, \dots, k$ . That is,  $A_{n,k}$  is the event that the points of  $\text{supp } \mathcal{P}_n$  with the  $k$  smallest  $t$ -values are uniquely defined. On  $A_{n,k}$ , let  $X_{n,k}$  denote the unique point for which  $\mathcal{P}_n(\{T_{n,k}\} \times \{X_{n,k}\}) = 1$ , and otherwise set  $X_{n,k} = \dagger$ , an isolated cemetery point.

We will impose the following conditions on the intensity measures  $(Z_n)_n$ , expressed in terms of a probability measure  $Q$  on  $\mathcal{X}$  and a family  $\mathcal{H}$  of measurable functions  $h : \mathcal{X} \rightarrow \mathbb{R}$ .

**Condition 8.1** (Regularity of Cox process intensities)

(a) For any  $t \in \mathbb{R}$  and for any  $h \in \mathcal{H}$ ,

$$\int_{\mathcal{X}} h dZ_{n,t} - |Z_{n,t}| \int_{\mathcal{X}} h dQ \xrightarrow{\mathbb{P}} 0. \quad (8.2)$$

(b) For each  $\varepsilon > 0$ , there exists  $\underline{t} \in \mathbb{R}$  such that

$$\liminf_{n \rightarrow \infty} \mathbb{P}(|Z_{n,\underline{t}}| < \varepsilon) \geq 1 - \varepsilon. \quad (8.3)$$

(c) For each  $\varepsilon > 0$ , there exists  $\bar{t} \in \mathbb{R}$  such that

$$\liminf_{n \rightarrow \infty} \mathbb{P}(|Z_{n,\bar{t}}| > 1/\varepsilon) \geq 1 - \varepsilon. \quad (8.4)$$

(d) For each  $\varepsilon > 0$  and each  $\underline{t} < \bar{t}$ , there exists a partition  $t_0 = \underline{t} < t_1 < \dots < t_N = \bar{t}$  of  $[\underline{t}, \bar{t}]$  such that

$$\liminf_{n \rightarrow \infty} \mathbb{P}\left(\sum_{i=1}^N |\Delta Z_{n,i}|^2 \leq \varepsilon\right) \geq 1 - \varepsilon. \quad (8.5)$$

We make the convention that any function  $h$  on  $\mathcal{X}$  is extended to  $\mathcal{X} \cup \{\dagger\}$  by  $h(\dagger) = 0$ .

**Proposition 8.2** Suppose that Condition 8.1 holds for a probability measure  $Q$  on  $\mathcal{X}$  and a family  $\mathcal{H}$  of bounded measurable functions  $h : \mathcal{X} \rightarrow \mathbb{R}$ . Then, for each fixed  $k \in \mathbb{N}$ ,  $\mathbb{P}(A_{n,k}) \rightarrow 1$ , the collection  $\{(T_{n,j})_{j=1}^k : n \in \mathbb{N}\}$  of random vectors is tight, and

$$\mathbb{E} \left( \prod_{j=1}^k g_j(T_{n,j}) h_j(X_{n,j}) \mid \mathcal{F}_n \right) - \mathbb{E} \left( \prod_{j=1}^k g_j(T_{n,j}) \mid \mathcal{F}_n \right) \prod_{j=1}^k \int_{\mathcal{X}} h_j dQ \xrightarrow{\mathbb{P}} 0 \quad (8.6)$$

for all bounded continuous functions  $g_1, \dots, g_k: \mathbb{R} \rightarrow \mathbb{R}$  and all  $h_1, \dots, h_k \in \mathcal{H}$ .

**Theorem 8.3** Suppose Condition 8.1 holds when either

$\mathcal{H}$  is the family of all bounded continuous functions on  $\mathcal{X}$ ;  
 $\mathcal{X} = \mathbb{R}^d$  and  $\mathcal{H}$  is the family of functions  $h(\vec{x}) = e^{i\vec{\xi} \cdot \vec{x}}$  for  $\vec{\xi} \in \mathbb{R}^d$ ; or  
 $\mathcal{X} = [0, \infty)^d$  and  $\mathcal{H}$  is the family of functions  $h(\vec{x}) = e^{-\vec{\xi} \cdot \vec{x}}$  for  $\vec{\xi} \in [0, \infty)^d$ .

Then

- (a) the random sequence  $(X_{n,j})_{j=1}^\infty$  converges in distribution [with respect to the product topology on  $(\mathcal{X} \cup \{\dagger\})^\mathbb{N}$ ] to a random sequence  $(X_j)_{j=1}^\infty$ , where the  $X_j$  are independent with law  $Q$ ;
- (b) the sequence  $(X_{n,j})_{j=1}^\infty$  is asymptotically independent of  $\mathcal{F}_n$ ;
- (c) the collection  $\{(T_{n,j})_{j=1}^k : n \in \mathbb{N}\}$  of random vectors is tight; and
- (d) if  $(T_j, X_j)_{j=1}^\infty$  is any subsequential limit in distribution of  $(T_{n,j}, X_{n,j})_{j=1}^\infty$ , then  $(T_j)_{j=1}^\infty$  and  $(X_j)_{j=1}^\infty$  are independent.

**Proof of Theorem 8.3 assuming Proposition 8.2** Because of the product topology, it suffices to consider finite sequences  $(T_{n,j}, X_{n,j})_{j=1}^k$  for a fixed  $k \in \mathbb{N}$ . Applying (8.6) with  $g_j(t) = 1$  gives the convergence of  $(X_{n,j})_{j=1}^k$ . The independence of  $(T_j)_{j=1}^k$  and  $(X_j)_{j=1}^k$  follows from the product form of (8.6), and the asymptotic independence of  $X_j$  from  $\mathcal{F}_n$  follows because of the conditional expectations in (8.6).  $\square$

We first prove the following lemma. Given  $t_0 < \dots < t_N$ , write  $B_{n,k}$  for the event that there exist (random) integers  $1 \leq I_1 < \dots < I_k \leq N$  with  $|\Delta \mathcal{P}_{n,I_j}| = 1$  for  $j = 1, \dots, k$  and  $|\mathcal{P}_{n,t_k}| = k$ . (That is,  $B_{n,k}$  is the event that each of the first  $k$  points of  $\mathcal{P}_n$  is the unique point in some interval  $(t_{i-1}, t_i]$ . In particular  $B_{n,k} \subset A_{n,k}$ .)

**Lemma 8.4** Assume Conditions 8.1 (b)–(c). Then, given  $\varepsilon > 0$  and  $k \in \mathbb{N}$ , there exists  $[\underline{t}, \bar{t}]$  and a partition  $\underline{t} = t_0 < \dots < t_N = \bar{t}$  of  $[\underline{t}, \bar{t}]$  such that  $\liminf_{n \rightarrow \infty} \mathbb{P}(B_{n,k}) \geq 1 - \varepsilon$ . In particular,  $\mathbb{P}(A_{n,k}) \rightarrow 1$ .

**Proof** Given a partition  $\underline{t} = t_0 < \dots < t_N = \bar{t}$ , the complement  $B_{n,k}^c$  is the event that  $\mathcal{P}_n$  contains a point in  $(-\infty, \underline{t}]$ , fewer than  $k$  points in  $(-\infty, \bar{t}]$ , or more than one point in some interval  $(t_{i-1}, t_i]$ . By Conditions 8.1 (b)–(c), we may choose  $\underline{t}, \bar{t}$  such that the first two events each have probability at most  $\varepsilon/3$  for  $n$  large. Since  $\mathbb{P}(|\Delta \mathcal{P}_{n,i}| \geq 2 \mid Z_n) = 1 - e^{-|\Delta Z_{n,i}|}(1 + |\Delta Z_{n,i}|) \leq |\Delta Z_{n,i}|^2$ , Condition 8.1 (c) gives a partition of  $[\underline{t}, \bar{t}]$  such that the third event also has probability at most  $\varepsilon/3$  for  $n$  large.  $\square$

**Proof of Proposition 8.2** Fix any  $\varepsilon > 0$  and bounded, continuous functions  $g_1, \dots, g_k$ . Choose  $t_0 < \dots < t_N$  as in Lemma 8.4. By taking a refinement, we may assume that  $|g_j(t) - g_j(t_i)| \leq \varepsilon$  for each  $t \in (t_{i-1}, t_i]$  and each  $i, j$ . Define  $\psi(t) = t_i$  if  $t_{i-1} < t \leq t_i$  and  $\psi(t) = t_N$  otherwise, and set  $\tilde{g}_j = g_j \circ \psi$ . Partitioning according to the integers  $I_j$ ,

$$\mathbb{1}_{B_{n,k}} \prod_{j=1}^k \tilde{g}_j(T_{n,j}) h_j(X_{n,j}) = \sum_{\vec{i}} \mathbb{1}_{B_{n,k}} \mathbb{1}_{\{\vec{i}=\vec{i}\}} \prod_{j=1}^k g_j(t_{i_j}) \int_{\mathcal{X}} h_j \Delta \mathcal{P}_{n,i_j}, \quad (8.7)$$

where the sum is over  $\vec{i} \in \mathbb{N}^k$  with  $1 \leq i_1 < \dots < i_k \leq N$ , and we write  $\vec{I} = (I_1, \dots, I_k)$ . Observe that a Poisson point process  $\mathcal{P}$  with intensity  $\mu$  satisfies  $\mathbb{E}(\mathbb{1}_{\{|\mathcal{P}|=1\}} \int h d\mathcal{P}) = e^{-|\mu|} \int h d\mu$ . Consequently,

$$\mathbb{E} \left( \mathbb{1}_{B_{n,k}} \prod_{j=1}^k \tilde{g}_j(T_{n,j}) h_j(X_{n,j}) \middle| \mathcal{F}_n \right) = \sum_{\vec{i}} e^{-|Z_{n,t_{i_k}}|} \prod_{j=1}^k g_j(t_{i_j}) \int_{\mathcal{X}} h_j \Delta Z_{n,i_j}. \quad (8.8)$$

Apply (8.8) twice, with the original  $h_j$ 's and with the constant functions  $\tilde{h}_j(x) = \int h_j dQ$ , to get

$$\begin{aligned} & \mathbb{E} \left( \mathbb{1}_{B_{n,k}} \left[ \prod_{j=1}^k \tilde{g}_j(T_{n,j}) h_j(X_{n,j}) - \prod_{j=1}^k \tilde{g}_j(T_{n,j}) \int_{\mathcal{X}} h_j dQ \right] \middle| \mathcal{F}_n \right) \\ &= \sum_{\vec{i}} e^{-|Z_{n,t_{i_k}}|} \left[ \prod_{j=1}^k g_j(t_{i_j}) \int_{\mathcal{X}} h_j \Delta Z_{n,i_j} - \prod_{j=1}^k g_j(t_{i_j}) |\Delta Z_{n,i_j}| \int_{\mathcal{X}} h_j dQ \right]. \end{aligned} \quad (8.9)$$

The right-hand side of (8.9) is bounded (since  $g_j$ ,  $h_j$ , and  $|\Delta Z_{n,i_j}| e^{-|\Delta Z_{n,i_j}|}$  are bounded) and, by Condition 8.1 (a), converges to 0 in probability, and hence also in expectation. By the choice of the partition,  $|\tilde{g}_j(T_{n,j}) - g_j(T_{n,j})| \leq \varepsilon$  on  $B_{n,k}$  and  $\limsup_{n \rightarrow \infty} \mathbb{P}(B_{n,k}^c) \leq \varepsilon$ . Now let  $(F_n)_n$  be a uniformly bounded sequence of  $\mathbb{R}$ -valued random variables such that  $F_n$  is  $\mathcal{F}_n$ -measurable. Since all the functions involved are bounded, there exists  $C < \infty$  such that

$$\limsup_{n \rightarrow \infty} \left| \mathbb{E} \left( F_n \prod_{j=1}^k g_j(T_{n,j}) h_j(X_{n,j}) \right) - \mathbb{E} \left( F_n \prod_{j=1}^k g_j(T_{n,j}) \right) \prod_{j=1}^k \int_{\mathcal{X}} h_j dQ \right| \leq C\varepsilon, \quad (8.10)$$

which completes the proof.  $\square$

When  $\mathcal{X} = \mathbb{R}^d$ , another natural family is  $\mathcal{H} = \{h(\vec{x}) = e^{\vec{\xi} \cdot \vec{x}} : \vec{\xi} \in \mathbb{R}^d\}$ . However, these functions are *not* bounded, so it is necessary to modify the argument of Proposition 8.2 and Theorem 8.3. Recall from (3.54) that we write  $\hat{R}$  for the moment generating function of a measure  $R$  on  $\mathbb{R}^d$ .

**Proposition 8.5** *Let  $\mathcal{X} = \mathbb{R}^d$ . Suppose Condition 8.1 holds when  $\mathcal{H}$  is the family of functions  $h(x) = e^{\vec{\xi} \cdot \vec{x}}$  for  $\vec{\xi} \in \mathbb{R}^d$ ,  $|\vec{\xi}| \leq \delta$ , where  $\delta > 0$  and  $\hat{Q}(\vec{\xi}) < \infty$  for all  $|\vec{\xi}| \leq \delta$ . Then the conclusions (a)–(d) of Theorem 8.3 hold.*

**Proof** Fix any  $\varepsilon > 0$ ,  $k \in \mathbb{N}$ ,  $g_1, \dots, g_k$  bounded, continuous functions, and choose  $t_0 < \dots < t_N$  as in Lemma 8.4. By taking a refinement, we may assume that  $t_i - t_{i-1} \leq \varepsilon$ . Let  $C_n$  be the event that  $\hat{Z}_{n,t_i}(\vec{\xi}_0) \leq |Z_{n,t_i}| \hat{Q}(\vec{\xi}_0) + \varepsilon$  for each  $i = 1, \dots, N$  and for each  $\vec{\xi}_0 \in \{\delta/\sqrt{d}, -\delta/\sqrt{d}\}^d$ . By Condition 8.1 (a),  $\mathbb{P}(C_n) \rightarrow 1$ . Let  $X_j$  be independent random variables with law  $Q$ , and define  $\tilde{X}_{n,j} = X_{n,j}$  on  $B_{n,k} \cap C_n$  and  $\tilde{X}_{n,j} = X_j$  otherwise. Recall the notations  $\psi(t)$ ,  $\tilde{g}_j(t)$  from the proof of Proposition 8.2 and set  $\tilde{T}_{n,j} = \psi(T_{n,j})$ . Set  $h_j(\vec{x}) = e^{\vec{\xi}_j \cdot \vec{x}}$  for  $\|\vec{\xi}_j\|_\infty \leq \delta/\sqrt{d}$ . By the argument of the previous proof, this time using that the  $\tilde{X}_{n,j}$  have law  $Q$  on  $(B_{n,k} \cap C_n)^c$ , we find

$$\begin{aligned} & \mathbb{E} \left( \left| \prod_{j=1}^k g_j(\tilde{T}_{n,j}) h_j(\tilde{X}_{n,j}) - \prod_{j=1}^k g_j(\tilde{T}_{n,j}) \hat{Q}(\tilde{\xi}_j) \right| \middle| \mathcal{F}_n \right) \\ &= \mathbb{1}_{C_n} \sum_{\vec{i}} e^{-|Z_{n,i_k}|} \left[ \prod_{j=1}^k g_j(t_{i_j}) \widehat{\Delta Z}_{n,i_j}(\tilde{\xi}_j) - \prod_{j=1}^k g_j(t_{i_j}) |\Delta Z_{n,i_j}| \hat{Q}(\tilde{\xi}_j) \right]. \quad (8.11) \end{aligned}$$

By Condition 8.1 (a), the right-hand side of (8.11) converges to 0 in probability. Moreover, by the bound  $e^{\tilde{\xi}_j \cdot \tilde{x}} \leq \sum_{\tilde{\xi}_0 \in \{\pm \delta / \sqrt{d}\}^d} e^{\tilde{\xi}_0 \cdot \tilde{x}}$  and the choice of  $C_n$ , it is bounded as well. Hence we may repeat the argument from the proof of Theorem 8.3 to find that  $(\tilde{T}_{n,j}, \tilde{X}_{n,j})_j$  satisfy the desired conclusions. But by construction,  $\liminf_{n \rightarrow \infty} \mathbb{P}(X_{n,j} = \tilde{X}_{n,j}, |T_{n,j} - \tilde{T}_{n,j}| \leq \varepsilon) \geq 1 - \varepsilon$ . Since  $\varepsilon > 0$  was arbitrary, it follows that  $X_{n,j}$  and  $T_{n,j}$  themselves have the same convergence properties.  $\square$

## 8.2 Convergence of First Point of Cox Processes: Proof of Theorem 3.29

In this section, we use the above conditions for convergence of Cox processes to prove Theorem 3.29:

**Proof of Theorem 3.29** For any  $\varepsilon > 0$  we may define

$$\underline{t}(\varepsilon) = \max \left\{ t \in \mathbb{Z}, t < -1/\varepsilon : q(t) < \varepsilon^2 \right\}, \quad (8.12)$$

$$\bar{t}(\varepsilon) = \min \left\{ t \in \mathbb{Z}, t > 1/\varepsilon : \liminf_{n \rightarrow \infty} \mathbb{P}(|Z_{n,t}^*| > 1/\varepsilon) > 1 - \varepsilon \right\}, \quad (8.13)$$

$$\begin{aligned} K_1(\varepsilon) = \min & \left\{ k_1 \in \mathbb{N} : \liminf_{n \rightarrow \infty} \mathbb{P} \left( \mathbb{E} \left( |Z_{n,\bar{t}(\varepsilon)}^{(K)}| \middle| \mathcal{F}'_n \right) \leq 2q(\bar{t}(\varepsilon)) \right) \geq 1 - \varepsilon \text{ for all } K \geq k_1 \right\} \\ & \vee \min \left\{ k_1 \in \mathbb{N} : \liminf_{n \rightarrow \infty} \mathbb{P} \left( \mathbb{E} \left( |Z_{n,\bar{t}(\varepsilon)}^{(K)}| \middle| \mathcal{F}'_n \right) \leq \varepsilon^2 \right) \geq 1 - \varepsilon \text{ for all } K \geq k_1 \right\}, \end{aligned} \quad (8.14)$$

by  $\lim_{t \rightarrow -\infty} q(t) = 0$ , assumption (b), (3.55) and (3.57), respectively. Given any  $\varepsilon > 0$ , let  $\varepsilon' \in (0, \varepsilon)$  be arbitrary. By construction,  $\bar{t}(\varepsilon') \leq \bar{t}(\varepsilon)$ , so taking  $K = K_1(\varepsilon')$  shows that  $\mathbb{E} \left( |Z_{n,\bar{t}(\varepsilon)}^*| \middle| \mathcal{F}'_n \right)$  is uniformly bounded apart from an event of probability at most  $2\varepsilon' + o(1)$ . Since  $\varepsilon'$  was arbitrary, it follows that  $|Z_{n,\bar{t}(\varepsilon)}^*|$  is tight as  $n \rightarrow \infty$ , so we may define

$$z_0(\varepsilon) = \min \left\{ z \in \mathbb{N} : \limsup_{n \rightarrow \infty} \mathbb{P}(|Z_{n,\bar{t}(\varepsilon)}^*| \leq z) \geq 1 - \varepsilon \right\}. \quad (8.15)$$

For  $\varepsilon > 0, t, u \in \mathbb{R}, \tilde{\xi} \in \mathbb{R}^2$ , let  $K_0(\varepsilon, t, u, \tilde{\xi})$  denote the smallest integer exceeding  $K_1(\varepsilon)$  such that (3.55)–(3.57) hold with probability at least  $1 - \varepsilon$  for  $K \geq K_0(\varepsilon)$  and  $n \geq n_0(K, \varepsilon, t, u, \tilde{\xi})$ . Let  $K_0(\varepsilon)$  and  $n_0(\varepsilon)$  denote the maxima of  $K_0(\varepsilon, t, u, \tilde{\xi})$  and  $n_0(K_0(\varepsilon), \varepsilon, t, u, \tilde{\xi})$ , respectively, over all numbers  $t, u \in [\underline{t}(\varepsilon), \bar{t}(\varepsilon)]$  and  $\tilde{\xi}_1, \tilde{\xi}_2 \in [-1/\varepsilon, 1/\varepsilon]$  that are dyadic rationals of the form  $i2^{-\ell}$  ( $i \in \mathbb{Z}, \ell \in \mathbb{N}$ ) with  $\ell < 1/\varepsilon$ .

The hypotheses imply that  $K_0(\varepsilon)$  and  $n_0(\varepsilon)$  are finite for each  $\varepsilon > 0$ . Moreover, by construction,  $\bar{t}(\varepsilon) \rightarrow \infty, \underline{t}(\varepsilon) \rightarrow -\infty$  as  $\varepsilon \downarrow 0$ . Therefore, by letting  $\varepsilon = \varepsilon_n$  decrease to 0 sufficiently slowly as  $n \rightarrow \infty$  and setting  $K = K_0(\varepsilon_n)$ , we can assume that  $Z_n^* = Z_n^{(K)} + Z_n^{(K)}$  where



$$\mathbb{E} \left( \hat{Z}_{n,t}^{(K)}(\vec{\xi}) \mid \mathcal{F}'_n \right) \xrightarrow{\mathbb{P}} q(t) \hat{Q}(\vec{\xi}), \quad (8.16)$$

$$\mathbb{E} \left( \left( \frac{\hat{Z}_{n,t}^{(K)}(\vec{\xi})}{q(t) \hat{Q}(\vec{\xi})} - \frac{|Z_{n,u}^{(K)}|}{q(u)} \right)^2 \mid \mathcal{F}'_n \right) \xrightarrow{\mathbb{P}} 0, \quad (8.17)$$

$$\mathbb{E} (|Z_{n,t}^{(K)}| \mid \mathcal{F}'_n) \xrightarrow{\mathbb{P}} 0, \quad (8.18)$$

whenever  $t, u, \xi_1, \xi_2$  are dyadic rationals.

Let  $\mathcal{P}'_n, \mathcal{P}''_n$  denote the Cox processes with intensity measures  $Z_n^{(K)}, Z_n^{\prime(K)}$ , respectively. For any fixed  $t \in \mathbb{R}$ , (8.18) implies that the first point of  $\mathcal{P}''_n$  does not occur by time  $t$  with high probability. By Proposition 8.5, it therefore suffices to show that  $Z_n^{(K)}$  satisfies Condition 8.1 for the family of functions  $h(\vec{x}) = e^{\vec{\xi} \cdot \vec{x}}, \vec{\xi} \in \mathbb{R}^2$ .

Applying Chebyshev's inequality to (8.17),

$$\hat{Z}_{n,t}^{(K)}(\vec{\xi}) - \frac{q(t)}{q(u)} \hat{Q}(\vec{\xi}) |Z_{n,u}^{(K)}| \xrightarrow{\mathbb{P}} 0 \quad (8.19)$$

(a continuity argument extends the convergence from dyadic rationals to all  $t, u, \vec{\xi}$ ). Taking  $t = u$  verifies Condition 8.1 (a). For Condition 8.1 (b), fix  $\varepsilon > 0$  and choose  $\underline{t} \in \mathbb{R}$  such that  $q(\underline{t}) < \frac{1}{4}\varepsilon^2$ . Then (8.16) implies that  $\mathbb{E} \left( |Z_{n,\underline{t}}^{(K)}| \mid \mathcal{F}_n \right) \leq \frac{1}{2}\varepsilon^2$  whp, and Markov's inequality implies that  $\mathbb{P} \left( |Z_{n,\underline{t}}^{(K)}| \geq \varepsilon \mid \mathcal{F}_n \right) \leq \frac{1}{2}\varepsilon$  whp. Condition 8.1 (c) follows from (8.18) and assumption (b) in Theorem 3.29.

Finally, let  $\varepsilon > 0$  and a compact interval  $[\underline{t}, \bar{t}]$  be given. Expanding the interval if necessary, we may assume that  $\underline{t}, \bar{t}$  are dyadic rationals, and, decreasing  $\varepsilon$  if necessary, we may assume that  $\bar{t} \leq \bar{t}(\varepsilon)$ . Since  $q$  is continuous and non-decreasing, we may choose a partition  $t_0 < \dots < t_N$  of  $[\underline{t}, \bar{t}]$  consisting of dyadic rationals such that

$$\sum_{i=1}^N \left( \frac{q(t_i)}{q(t_{i-1})} - 1 \right)^2 \leq \frac{\varepsilon}{4z_0(\varepsilon)^2}. \quad (8.20)$$

[it is enough to choose the partition finely enough that  $\max_i (q(t_i) - q(t_{i-1})) \leq q(\underline{t})/4z_0(\varepsilon)^2 q(\bar{t})$ ], and bound

$$\begin{aligned} \sum_{i=1}^N \left| \Delta Z_{n,i}^{(K)} \right|^2 &= \sum_{i=1}^N \left( |Z_{n,t_i}^{(K)}| - \frac{q(t_i)}{q(t_{i-1})} |Z_{n,t_{i-1}}^{(K)}| + \left( \frac{q(t_i)}{q(t_{i-1})} - 1 \right) |Z_{n,t_{i-1}}^{(K)}| \right)^2 \\ &\leq \sum_{i=1}^N 2 \left( |Z_{n,t_i}^{(K)}| - \frac{q(t_i)}{q(t_{i-1})} |Z_{n,t_{i-1}}^{(K)}| \right)^2 + 2 \left( \frac{q(t_i)}{q(t_{i-1})} - 1 \right)^2 |Z_{n,t_{i-1}}^{(K)}|^2 \\ &\leq \frac{\varepsilon}{2z_0(\varepsilon)^2} |Z_{n,\bar{t}(\varepsilon)}^*|^2 + 2 \sum_{i=1}^N \left( |Z_{n,t_i}^{(K)}| - \frac{q(t_i)}{q(t_{i-1})} |Z_{n,t_{i-1}}^{(K)}| \right)^2. \end{aligned} \quad (8.21)$$

The latter sum is  $o_{\mathbb{P}}(1)$  by (8.19) with  $\vec{\xi} = \vec{0}$ , and the remaining term is at most  $\varepsilon/2$  on the event  $\{|Z_{n,\bar{t}(\varepsilon)}^*| \leq z_0(\varepsilon)\}$ . This event has probability at least  $1 - \varepsilon - o(1)$  by (8.15), which completes the proof.  $\square$

## 9 Moment Estimates and the Cluster After Unfreezing

While most other sections did not rely on the companion paper [21], in this section we *do* heavily rely on it. In particular, we make use of [Part I, Theorem 2.15], [Part I, Lemma 2.18], [Part I, Proposition 2.17] and [Part I, Lemma 6.4].

In this section we study  $\mathcal{B}_t$  for  $t \geq T_{\text{unfr}}$ , when the cluster resumes its CTBP behaviour. We will use moment methods to prove Lemma 3.32 and Theorem 3.31, completing the proof of our results.

This section is organised as follows. We start in Sect. 9.1 with some preparations concerning frozen intensity measures. In Sect. 9.2, we investigate the volume of the cluster after unfreezing and prove Lemma 3.32 (a). In Sect. 9.3 we use second moment methods on the collision edges to prove Theorem 3.31. In Sect. 9.4 we show that the collision edge whp does not originate from the frozen cluster, but rather from one of its descendants, to prove Lemma 3.32 (b). Finally, in Sect. 9.5, we study the freezing time and frozen cluster and use this to prove Theorem 3.18.

### 9.1 Preparations: Frozen Intensity Measures

We introduce the frozen intensity measures

$$d\mu_{n,\text{fr}}^{(j)}(y) = \sum_{v \in \mathcal{B}_{\text{fr}}^{(j)}} \mathbb{1}_{\{y \geq 0\}} \mu_n(T_{\text{fr}}^{(j)} - T_v + dy). \quad (9.1)$$

Recall that the notation  $\mu(t_0 + dy)$  denotes the translation of the measure  $\mu$  by  $t_0$ ; thus (9.1) means that, for a test function  $h \geq 0$ ,

$$\int h(y) d\mu_{n,\text{fr}}^{(j)}(y) = \sum_{v \in \mathcal{B}_{\text{fr}}^{(j)}} \int_{T_{\text{fr}}^{(j)} - T_v}^{\infty} h(y - (T_{\text{fr}}^{(j)} - T_v)) d\mu_n(y). \quad (9.2)$$

**Lemma 9.1** *Almost surely, for  $j = 1, 2$ ,*

$$s_n \leq \int e^{-\lambda_n(1)y} d\mu_{n,\text{fr}}^{(j)}(y) \leq s_n + 1. \quad (9.3)$$

**Proof** By (9.2),

$$\int e^{-\lambda_n y} d\mu_{n,\text{fr}}^{(j)}(y) = \sum_{v \in \mathcal{B}_{\text{fr}}^{(j)}} \int_{T_{\text{fr}}^{(j)} - T_v}^{\infty} e^{-\lambda_n(y - (T_{\text{fr}}^{(j)} - T_v))} d\mu_n(y). \quad (9.4)$$

The expression in (9.4) is the value of the process  $\sum_{v \in \mathcal{B}_t^{(j)}} \int_{t - T_v}^{\infty} e^{-\lambda_n(y - (t - T_v))} d\mu_n(y)$  from Definition 3.17, stopped at  $t = T_{\text{fr}}^{(j)}$  (recall that  $\mathcal{B}_{\text{fr}}^{(j)} = \mathcal{B}P_{T_{\text{fr}}^{(j)}}^{(j)}$ ). Since  $\mu_n$  has no atoms, this process is continuous in  $t$  except for jumps at the birth times, and since the birth times are distinct a.s., the corresponding jump has size  $\int_0^{\infty} e^{-\lambda_n y} d\mu_n(y) = 1$ . By definition,  $T_{\text{fr}}^{(j)}$  is the first time the process in (9.4) exceeds  $s_n$ , so it can have value at most  $s_n + 1$  at that time.  $\square$

For future reference, we now state a lemma, to be used in Sect. 9.3, showing that most of the mass of the frozen intensity measures  $\mu_{n,\text{fr}}^{(j)}$  comes from small times:

**Lemma 9.2** *Let  $\delta, \delta' > 0$  and  $a_0 > 0$  be given. Then there exists  $K < \infty$  and  $n_0 \in \mathbb{N}$  such that, for all  $a \geq a_0$  and  $n \geq n_0$ ,*

$$\mathbb{P} \left( \int e^{-\lambda_n(a^{1/s_n})y} \mathbb{1}_{\{\lambda_n(1)y \geq K\}} d\mu_{n,\text{fr}}^{(j)}(y) > \delta s_n \right) \leq \delta'. \quad (9.5)$$

**Proof** Let  $\varepsilon = a_0 e^{-\gamma}/2$ , where  $\gamma$  denotes Euler's constant. Using the definition of  $\mu_{n,\text{fr}}^{(j)}$  from (9.1), the monotonicity of  $\lambda_n(\cdot)$  and (6.15), we obtain  $n_0 \in \mathbb{N}$  such that for all  $K < \infty$ ,  $a \geq a_0$  and  $n \geq n_0$ ,

$$\int e^{-\lambda_n(\bar{a})y} \mathbb{1}_{\{\lambda_n(1)y \geq K\}} d\mu_{n,\text{fr}}^{(j)}(y) \leq \sum_{v \in \mathcal{B}_{\text{fr}}^{(j)}} \int e^{-\varepsilon y/f_n(1)} \mathbb{1}_{\{y \geq K f_n(1)\}} \mu_n(T_{\text{fr}}^{(j)} - T_v + dy). \quad (9.6)$$

According to Lemma 4.4, for any  $\varepsilon' > 0$  we can choose some  $K < \infty$  such that, after possibly increasing  $n_0$ , the right-hand side of (9.6) is bounded from above by  $|\mathcal{B}_{\text{fr}}^{(j)}| \varepsilon'/s_n$ . Since  $|\mathcal{B}_{\text{fr}}^{(j)}| = O_{\mathbb{P}}(s_n^2)$  by Theorem 3.18 (b), the proof is complete.  $\square$

## 9.2 A First Moment Estimate: Proof of Lemma 3.32 (a)

In this section we show how to express  $\mathcal{B}_t \setminus \mathcal{B}_{\text{fr}}$ ,  $t \geq T_{\text{unfr}}$ , as a suitable union of branching processes. This representation leads to a simple proof of Lemma 3.32 (a). We will also use it in Sect. 9.3 to prove Theorem 3.31.

Consider the immediate children  $v \in \partial \mathcal{B}_{\text{fr}}$  of individuals in the frozen cluster  $\mathcal{B}_{\text{fr}}$ . Then, for  $t' \geq 0$ ,

$$\mathcal{B}_{T_{\text{unfr}}+t'} \setminus \mathcal{B}_{\text{fr}} = \bigcup_{v \in \partial \mathcal{B}_{\text{fr}}: T_v^{\mathcal{B}} \leq T_{\text{unfr}}+t'} \left\{ vw : w \in BP_{t'+T_{\text{unfr}}-T_v^{\mathcal{B}}}^{(v)} \right\}, \quad (9.7)$$

where  $BP^{(v)}$  denotes the branching process of descendants of  $v$ , re-rooted and time-shifted as in (6.1). Furthermore, conditionally on  $\mathcal{B}_{\text{fr}}$ , the children  $v \in \partial \mathcal{B}_{\text{fr}}$  appear according to a Cox process. Formally, the point measures

$$\mathcal{P}_{n,\text{unfr}}^{(j)} = \sum_{v \in \partial \mathcal{B}_{\text{fr}}^{(j)}} \delta_{(T_v^{\mathcal{B}} - T_{\text{unfr}}, BP^{(v)})} \quad (9.8)$$

form Cox processes with intensities  $d\mu_{n,\text{fr}}^{(j)} \otimes d\mathbb{P}(BP^{(1)} \in \cdot)$ ,  $j = 1, 2$ , where the frozen intensity measures  $\mu_{n,\text{fr}}^{(j)}$  were introduced in (9.1).

**Proof of Lemma 3.32 (a)** By Theorem 3.18 (b), the volume  $|\mathcal{B}_{\text{fr}}|$  of the frozen cluster is  $O_{\mathbb{P}}(s_n^2)$ , and this is  $o_{\mathbb{P}}(\sqrt{n s_n})$  since  $n/s_n^3 \rightarrow \infty$ . It therefore suffices to show that  $|\mathcal{B}_{\bar{t}} \setminus \mathcal{B}_{\text{fr}}| = O_{\mathbb{P}}(\sqrt{n s_n})$  when  $\bar{t} = T_{\text{unfr}} + \lambda_n(1)^{-1} (\frac{1}{2} \log(n/s_n^3) + K)$ .

Abbreviate  $t' = \bar{t} - T_{\text{unfr}} = \lambda_n(1)^{-1} (\frac{1}{2} \log(n/s_n^3) + K)$ . By (9.7)–(9.8),

$$|\mathcal{B}_{\bar{t}} \setminus \mathcal{B}_{\text{fr}}| = \sum_{v \in \partial \mathcal{B}_{\text{fr}}: T_v^{\mathcal{B}} \leq T_{\text{unfr}}+t'} z_{t'+T_{\text{unfr}}-T_v^{\mathcal{B}}}^{1, BP^{(v)}}(1) = \sum_{j=1}^2 \int \mathbb{1}_{\{t \leq t'\}} z_{t'-t}^{1, bp}(1) d\mathcal{P}_{n,\text{unfr}}^{(j)}(t, bp), \quad (9.9)$$

so that, by Theorem 3.13 and Lemma 9.1, there exists a  $K' < \infty$  such that for sufficiently large  $n$ ,

$$\begin{aligned} \mathbb{E}(|\mathcal{B}_t \setminus \mathcal{B}_{\text{fr}}| \mid \mathcal{B}_{\text{fr}}) &= \sum_{j=1}^2 \int \mathbb{1}_{\{t \leq t'\}} e^{\lambda_n(1)(t'-t)} \bar{m}_{t'-t}^1(1) d\mu_{n,\text{fr}}^{(j)}(t) \\ &\leq \sqrt{\frac{n}{s_n^3}} e^K \sum_{j=1}^2 \int \mathbb{1}_{\{t \leq t'\}} K' s_n e^{-\lambda_n(1)t} d\mu_{n,\text{fr}}^{(j)}(t) \leq K' e^K \sqrt{\frac{n}{s_n^3}} s_n (s_n + 1). \end{aligned} \quad (9.10)$$

Markov's inequality completes the proof.  $\square$

### 9.3 Second Moment Estimates: Proof of Theorem 3.31

In this section we prove that  $\mathcal{P}_n^*$  satisfies the assumptions of Theorem 3.29. Namely, we will split  $\mu_n = \mu_n^{(K)} + (\mu_n - \mu_n^{(K)})$  into the truncated measure and a remainder, as in Sect. 7. This induces a splitting of the intensity measure into  $Z_n^{(K)} + Z_n^{(K)}$ , and the hypothesis (a) will be verified using the estimates for the two-vertex characteristics  $\chi_n^{(K)}$  and  $\chi_n - \chi_n^{(K)}$  in Theorems 3.15 and 3.16. The remaining hypothesis (b) will be proved using a separate argument.

Throughout the proof, the times  $t$  and  $t^*$  are related as in (3.60), and we recall from (3.54) that, for a measure  $Q$  on  $\mathbb{R}^d$ , we write  $\hat{Q}$  for its moment generating function.

**Proof of Theorem 3.31** Since  $\mathcal{P}_n^*$  is the image of a Cox process under a mapping that is measurable with respect to  $\mathcal{B}_{\text{fr}}$ , it is immediate that it itself is a Cox process, and its intensity measure is

$$Z_{n,t^*}^* = \sum_{v_1 \in \mathcal{B}_t^{(1)} \setminus \mathcal{B}_{\text{fr}}^{(1)}} \sum_{v_2 \in \mathcal{B}_t^{(2)} \setminus \mathcal{B}_{\text{fr}}^{(2)}} \frac{1}{n} \mu_n(\Delta R_{v_1, v_2}, R_1(t) - R_1(T_{v_1}^{\mathcal{B}}) + R_2(t) - R_2(T_{v_2}^{\mathcal{B}})) \delta_{(|v_1|^*, |v_2|^*)}, \quad (9.11)$$

where we recall  $|v|^*$  from (3.59) and  $|v_j|^*$  denotes  $|v|^*$  for  $\mathcal{B}^{(j)}$ . In the above sum,  $T_{v_j}^{\mathcal{B}} \geq T_{\text{unfr}}$ , so that  $R_{j'}(t) - R_{j'}(T_{v_j}^{\mathcal{B}}) = t - T_{v_j}^{\mathcal{B}}$  whenever  $t \geq T_{v_j}^{\mathcal{B}}$ ,  $j, j' \in \{1, 2\}$ , and, recalling (3.49),  $\Delta R_{v_1, v_2} = |T_{v_1}^{\mathcal{B}} - T_{v_2}^{\mathcal{B}}|$ .

We begin by expressing  $\hat{Z}_{n,t^*}^*(\vec{\xi})$  as a sum of two-vertex characteristics. As in (9.7), any vertex  $v' \in \mathcal{B}_t \setminus \mathcal{B}_{\text{fr}}$  is descended from a unique vertex  $v = p^{\text{unfr}}(v') \in \partial \mathcal{B}_{\text{fr}}$  and can therefore be written as  $v' = vw$  for some  $w \in B P_{t-T_{v'}^{\mathcal{B}}}^{(v)}$ . Hence  $|v'| - |p^{\text{unfr}}(v')| = |vw| - |v| = |w|$ . Thus

$$\begin{aligned} \hat{Z}_{n,t^*}^*(\vec{\xi}) &= \sum_{v_1 \in \partial \mathcal{B}_{\text{fr}}^{(1)}} \sum_{w_1 \in B P_{t-T_{v_1}^{\mathcal{B}}}^{(v_1)}} \sum_{v_2 \in \partial \mathcal{B}_{\text{fr}}^{(2)}} \sum_{w_2 \in B P_{t-T_{v_2}^{\mathcal{B}}}^{(v_2)}} \\ &\quad \exp \left( \frac{\xi_1 |w_1| + \xi_2 |w_2|}{s_n \sqrt{\log(n/s_n^3)}} - (\xi_1 + \xi_2) \frac{\phi_n}{2s_n} \sqrt{\log(n/s_n^3)} \right) \\ &\quad \times \frac{1}{n} \mu_n(|T_{v_1}^{\mathcal{B}} - T_{v_2}^{\mathcal{B}}|, t - T_{v_1}^{\mathcal{B}} + t - T_{v_2}^{\mathcal{B}}). \end{aligned} \quad (9.12)$$

We note that  $T_{v_j w_j}^{\mathcal{B}} = T_{v_j}^{\mathcal{B}} + T_{w_j}^{(v_j)}$ , where  $T_{w_j}^{(v_j)}$  denotes the birth time of  $w_j$  in the branching process  $BP^{(v_j)}$  defined in (6.1). (Note that  $T_{v_j}^{\mathcal{B}} \geq T_{\text{unfr}}$ , so that freezing plays no role after  $T_{v_j}^{\mathcal{B}}$  and we need not consider  $T_{w_j}^{\mathcal{B}}$ .) It follows that

$$\mu_n(|T_{v_1 w_1}^{\mathcal{B}} - T_{v_2 w_2}^{\mathcal{B}}|, t - T_{v_1 w_1}^{\mathcal{B}} + t - T_{v_2 w_2}^{\mathcal{B}}) = \chi_n(t - T_{v_1}^{\mathcal{B}} - T_{w_1}^{(v_1)}, t - T_{v_2}^{\mathcal{B}} - T_{w_2}^{(v_2)}), \quad (9.13)$$

where  $\chi_n$  is the two-vertex characteristic from (3.24). Recalling the notation from (3.22),

$$\hat{Z}_{n,t^*}^*(\vec{\xi}) = \frac{1}{n} \exp\left(-(\xi_1 + \xi_2) \frac{\phi_n}{2s_n} \sqrt{\log(n/s_n^3)}\right) \sum_{v_1 \in \partial \mathcal{B}_{\text{fr}}^{(1)}} \sum_{v_2 \in \partial \mathcal{B}_{\text{fr}}^{(2)}} z_{t-T_{v_1}^{\mathcal{B}}, t-T_{v_2}^{\mathcal{B}}}^{\chi_n, BP^{(v_1)}, BP^{(v_2)}}(\vec{a}), \quad (9.14)$$

where  $\vec{a} = (a_1^{1/s_n}, a_2^{1/s_n})$  as in (7.3) and

$$a_j = \exp\left(\frac{\xi_j}{\sqrt{\log(n/s_n^3)}}\right). \quad (9.15)$$

Note that  $a_1, a_2$  depend implicitly on  $n$  and  $a_j \rightarrow 1$  as  $n \rightarrow \infty$ .

As in Sect. 9.2, we express the sums over  $\partial \mathcal{B}_{\text{fr}}^{(1)}, \partial \mathcal{B}_{\text{fr}}^{(2)}$  in (9.14) in terms of the point measures  $\mathcal{P}_{n,\text{unfr}}^{(j)} = \sum_{v \in \partial \mathcal{B}_{\text{fr}}^{(j)}} \delta_{(T_v^{\mathcal{B}} - T_{\text{unfr}}, BP^{(v)})}$  from (9.8):

$$\begin{aligned} \hat{Z}_{n,t^*}^*(\vec{\xi}) &= \frac{1}{n} \exp\left(-(\xi_1 + \xi_2) \frac{\phi_n}{2s_n} \sqrt{\log(n/s_n^3)}\right) \int d\mathcal{P}_{n,\text{unfr}}^{(1)}(t_1, bp^{(1)}) \\ &\quad \times \int d\mathcal{P}_{n,\text{unfr}}^{(2)}(t_2, bp^{(2)}) z_{t-T_{\text{unfr}}-t_1, t-T_{\text{unfr}}-t_2}^{\chi_n, bp^{(1)}, bp^{(2)}}(\vec{a}). \end{aligned} \quad (9.16)$$

Recalling the notation from (3.23) and (6.19),

$$\begin{aligned} \hat{Z}_{n,t^*}^*(\vec{\xi}) &= \frac{1}{n} \exp\left(-(\xi_1 + \xi_2) \frac{\phi_n}{2s_n} \sqrt{\log(n/s_n^3)} + (t - T_{\text{unfr}})(\lambda_n(\tilde{a}_1) + \lambda_n(\tilde{a}_2))\right) \\ &\quad \times \int d\mathcal{P}_{n,\text{unfr}}^{(1)}(t_1, bp^{(1)}) \int d\mathcal{P}_{n,\text{unfr}}^{(2)}(t_2, bp^{(2)}) e^{-\lambda_n(\tilde{a}_1)t_1 - \lambda_n(\tilde{a}_2)t_2} z_{t-T_{\text{unfr}}-t_1, t-T_{\text{unfr}}-t_2}^{\chi_n, bp^{(1)}, bp^{(2)}}(\vec{a}) \\ &= \exp\left(\frac{\lambda_n(\tilde{a}_1) + \lambda_n(\tilde{a}_2)}{\lambda_n(1)} \left(t^* + \frac{1}{2} \log(n/s_n^3)\right) - (\xi_1 + \xi_2) \frac{\phi_n}{2s_n} \sqrt{\log(n/s_n^3)} - \log(n/s_n^3)\right) \\ &\quad \times \frac{1}{s_n^3} \int d\mathcal{P}_{n,\text{unfr}}^{(1)}(t_1, bp^{(1)}) \int d\mathcal{P}_{n,\text{unfr}}^{(2)}(t_2, bp^{(2)}) e^{-\lambda_n(\tilde{a}_1)t_1 - \lambda_n(\tilde{a}_2)t_2} z_{t-T_{\text{unfr}}-t_1, t-T_{\text{unfr}}-t_2}^{\chi_n, bp^{(1)}, bp^{(2)}}(\vec{a}). \end{aligned} \quad (9.17)$$

For the non-random factors on the right-hand side of (9.17), we use the asymptotics from Corollary 6.5:

$$\begin{aligned} \frac{\lambda_n(\tilde{a}_j)}{\lambda_n(1)} &= 1 + \frac{\phi_n}{s_n} \left( \exp\left(\xi_j / \sqrt{\log(n/s_n^3)}\right) - 1 \right) + o\left(\exp\left(\xi_j / \sqrt{\log(n/s_n^3)}\right) - 1\right)^2 \\ &= 1 + \frac{\phi_n}{s_n} \left( \xi_j / \sqrt{\log(n/s_n^3)} + \frac{1}{2} \xi_j^2 / \log(n/s_n^3) \right) + o(1 / \log(n/s_n^3)). \end{aligned} \quad (9.18)$$

Combining with the asymptotics  $\lambda_n(\tilde{a}_j)/\lambda_n(1) = 1 + o(1)$  and  $s_n/\phi_n = 1 + o(1)$  [see (6.15) and Lemma 3.12], we conclude that, for any fixed  $t^* \in \mathbb{R}$ ,

$$\frac{\lambda_n(\tilde{a}_j)}{\lambda_n(1)} \left(t^* + \frac{1}{2} \log(n/s_n^3)\right) - \xi_j \frac{\phi_n}{2s_n} \sqrt{\log(n/s_n^3)} - \frac{1}{2} \log(n/s_n^3) = t^* + \frac{1}{4} \xi_j^2 + o(1). \quad (9.19)$$

[When  $\xi_j = 0$ ,  $a_j = 1$ , the term  $o(1)$  in (9.19) is absent.] Combining (9.17) and (9.19),

$$\hat{Z}_{n,t^*}^*(\vec{\xi}) = \frac{e^{2t^* + \frac{1}{4}\|\vec{\xi}\|^2 + o(1)}}{s_n^3} \int d\mathcal{P}_{n,\text{unfr}}^{(1)}(t_1, bp^{(1)}) \int d\mathcal{P}_{n,\text{unfr}}^{(2)}(t_2, bp^{(2)}) \\ \cdot e^{-\lambda_n(\vec{a}_1)t_1} e^{-\lambda_n(\vec{a}_2)t_2} \bar{z}_{t-T_{\text{unfr}}-t_1, t-T_{\text{unfr}}-t_2}^{(K)}(\vec{a}). \quad (9.20)$$

This is the desired representation in terms of two-vertex characteristics.

Given  $K < \infty$ , define

$$Z_{n,t^*}^{(K)} = \sum_{v_1 \in \mathcal{B}_t^{(1)} \setminus \mathcal{B}_{\text{fr}}^{(1)}} \sum_{v_2 \in \mathcal{B}_t^{(2)} \setminus \mathcal{B}_{\text{fr}}^{(2)}} \frac{1}{n} \mu_n^{(K)}(|T_{v_1}^{\mathcal{B}} - T_{v_2}^{\mathcal{B}}|, t - T_{v_1}^{\mathcal{B}} + t - T_{v_2}^{\mathcal{B}}) \delta_{(|v_1|^*, |v_2|^*)}, \quad (9.21)$$

similarly to Definition 3.30, and set  $Z_{n,t^*}^{(K)} = Z_{n,t^*}^* - Z_{n,t^*}^{(K)}$ . (We suppress the dependence of  $Z_{n,t^*}^{(K)}$ ,  $Z_n^{(K)}$  on  $K$ .) Clearly, (9.20) remains true when  $Z_{n,t^*}^*$  and  $\chi_n$  are replaced by  $Z_{n,t^*}^{(K)}$  and  $\chi_n^{(K)}$ , or by  $Z_{n,t^*}^{(K)}$  and  $\chi_n - \chi_n^{(K)}$ , respectively. We will use (9.20) to control first and second moments of  $Z_{n,t^*}^{(K)}$ , for arbitrary but fixed values  $t^*$ ,  $\xi_1, \xi_2 \in \mathbb{R}$ , and first moments of  $Z_{n,t^*}^{(K)}$  with  $\xi_1 = \xi_2 = 0$ .

**Verification of (3.55):** Writing  $\lambda_n(\vec{a}) = (\lambda_n(\vec{a}_1), \lambda_n(\vec{a}_2))$ , we obtain from (9.20),

$$\mathbb{E}(\hat{Z}_{n,t^*}'(\vec{\xi}) | \mathcal{B}_{\text{fr}}) = \frac{e^{2t^* + \frac{1}{4}\|\vec{\xi}\|^2 + o(1)}}{s_n^2} \iint e^{-\lambda_n(\vec{a}) \cdot \vec{t}} \frac{\bar{m}_{t-T_{\text{unfr}}-\vec{t}}^{\chi_n^{(K)}}(\vec{a})}{s_n} d\mu_{n,\text{fr}}^{(1)}(t_1) d\mu_{n,\text{fr}}^{(2)}(t_2). \quad (9.22)$$

Since  $\lambda_n(1)(t - T_{\text{unfr}}) = \frac{1}{2} \log(n/s_n^3) + t^* \rightarrow \infty$  and  $a_j \rightarrow 1$ , Theorem 3.15 and Corollary 6.5 imply that for  $\varepsilon \in (0, 1)$  and  $K < \infty$  sufficiently large, there exists  $K' < \infty$  and  $n_0 \in \mathbb{N}$  such that the integrand of (9.22) lies between  $e^{-\lambda_n(1)(t_1+t_2)-\varepsilon}(1-\varepsilon)$  and  $e^{-\lambda_n(1)(t_1+t_2)+\varepsilon}(1+\varepsilon)$  for  $\lambda_n(1)[t_1 \vee t_2] \leq K'$  and  $n \geq n_0$ , and there are constants  $K'' < \infty$ ,  $n'_0 \in \mathbb{N}$  such that it is bounded by  $K''e^{-\lambda_n(\vec{a}) \cdot \vec{t}}$  for all  $t_1, t_2 \geq 0$ ,  $n \geq n'_0$ . Using Lemmas 9.1 and 9.2, it easily follows that

$$e^{2t^* + \frac{1}{4}\|\vec{\xi}\|^2} (1 - 2\varepsilon) \leq \mathbb{E}(\hat{Z}_{n,t^*}'(\vec{\xi}) | \mathcal{B}_{\text{fr}}) \leq e^{2t^* + \frac{1}{4}\|\vec{\xi}\|^2} (1 + 2\varepsilon) \quad (9.23)$$

whp. Since  $e^{\frac{1}{4}\|\vec{\xi}\|^2} = \hat{Q}(\vec{\xi})$ , where  $Q$  is the law of two independent  $N(0, \frac{1}{2})$  variables, we have therefore verified the first moment condition (3.55) of Theorem 3.29 with  $q(t^*) = e^{2t^*}$ .

**Verification of (3.57):** For  $Z_{n,t^*}^{(K)}$ , set  $\vec{\xi} = \vec{0} := (0, 0)$ . In (9.22), we can replace  $Z_{n,t^*}^{(K)}$  and  $\chi_n^{(K)}$  by  $Z_{n,t^*}^{(K)}$  and  $\chi_n - \chi_n^{(K)}$ , and Theorem 3.16 implies that, for any  $\varepsilon \in (0, 1)$  and large  $K$ , there is  $K' < \infty$  such that the integrand of the resulting equation is at most  $\varepsilon e^{-\lambda_n(1)(t_1+t_2)}$ , uniformly for  $\lambda_n(1)[t_1 \vee t_2] \leq K'$  for large  $n$ , and there is  $K'' < \infty$  such that it is bounded by  $K''e^{-\lambda_n(1)(t_1+t_2)}$  for all  $t_1, t_2 \geq 0$ ,  $n$  large. This verifies the first moment condition (3.57), as in the previous case.

**Verification of (3.56):** The second moment estimates for  $Z_{n,t^*}^{(K)}$ , though somewhat more complicated, are similar in spirit. Note the importance of freezing, which is not so far apparent from the first moment calculations only: the freezing times  $T_{\text{fr}}^{(j)}$  and the rescaled times  $t \approx T_{\text{unfr}} + \frac{1}{2}\lambda_n(1)^{-1} \log(n/s_n^3)$  have exactly the scaling needed so that both the first and second moments of  $Z_{n,t^*}^*$  will have order 1, even though  $\bar{m}_t^{\chi_n^{(K)}}(\vec{a}) \approx s_n$ ,  $\bar{M}_{t,\vec{u}}^{\chi_n^{(K)}, \chi_n^{(K)}}(\vec{a}, \vec{b}) \approx s_n^4$  by Theorem 3.15.

Equation (9.20) for  $Z'_{n,t^*}$  expresses  $\hat{Z}'_{n,t^*}(\vec{\xi})$  in terms of a double integral with respect to a pair of Cox processes, whose intensity is measurable with respect to  $\mathcal{B}_{\text{fr}}$ . An elementary calculation [applying (6.3) twice] shows that, for a pair  $\mathcal{P}^{(1)}, \mathcal{P}^{(2)}$  of Poisson point processes with intensities  $\nu_1, \nu_2$ ,

$$\begin{aligned} & \mathbb{E} \left( \iint f(x_1, x_2) d\mathcal{P}^{(1)}(x_1) d\mathcal{P}^{(2)}(x_2) \iint g(y_1, y_2) d\mathcal{P}^{(1)}(y_1) d\mathcal{P}^{(2)}(y_2) \right) \\ &= \iint f(x_1, x_2) g(x_1, x_2) d\nu_1(x_1) d\nu_2(x_2) + \iiint f(x_1, x_2) g(x_1, y_2) d\nu_1(x_1) d\nu_2(x_2) d\nu_2(y_2) \\ &+ \iiint f(x_1, x_2) g(y_1, x_2) d\nu_1(x_1) d\nu_1(y_1) d\nu_2(x_2) \\ &+ \iint f(x_1, x_2) d\nu_1(x_1) d\nu_2(x_2) \iint g(y_1, y_2) d\nu_1(y_1) d\nu_2(y_2). \end{aligned} \quad (9.24)$$

We apply (9.24) to (9.20). In the notation of (9.24), we have  $x_j = (t_j, bp^{(j)})$ , and  $f(x_1, x_2) = g(x_1, x_2) = e^{-\lambda_n(\tilde{a}_1)t_1} e^{-\lambda_n(\tilde{a}_2)t_2} \bar{z}_{t-T_{\text{unfr}}-t_1, t-T_{\text{unfr}}-t_2}^{\chi_n^{(K)}, bp^{(1)}, bp^{(2)}}(\tilde{a})$ . Integration against the intensity measure  $d\nu_j = d\mu_{n,\text{fr}}^{(j)} \otimes d\mathbb{P}(BP^{(j)} \in \cdot)$  is equivalent to a branching process expectation together with an integration over  $t_j$ , and we will therefore obtain first or second moments of one- or two-vertex characteristics from the various terms in (9.24). Namely, for  $\vec{\xi}, \vec{\zeta} \in \mathbb{R}^2$  and writing  $b_j = \exp(\zeta_j/\sqrt{\log(n/s_n^3)})$ , we obtain

$$\begin{aligned} & s_n^6 e^{-2t^* - \frac{1}{4}\|\vec{\xi}\|^2 - 2u^* - \frac{1}{4}\|\vec{\zeta}\|^2 - o(1)} \mathbb{E} \left( \hat{Z}'_{n,t^*}(\vec{\xi}) \hat{Z}'_{n,u^*}(\vec{\zeta}) \mid \mathcal{B}_{\text{fr}} \right) \\ &= \iint e^{-(\lambda_n(\tilde{a}_1) + \lambda_n(\tilde{b}_1))t_1} e^{-(\lambda_n(\tilde{a}_2) + \lambda_n(\tilde{b}_2))t_2} \bar{M}_{t-T_{\text{unfr}}-\tilde{t}, u-T_{\text{unfr}}-\tilde{t}}^{\chi_n^{(K)}, \chi_n^{(K)}}(\tilde{a}, \tilde{b}) d\mu_{n,\text{fr}}^{(1)}(t_1) d\mu_{n,\text{fr}}^{(2)}(t_2) \\ &+ \iiint e^{-(\lambda_n(\tilde{a}_1) + \lambda_n(\tilde{b}_1))t_1} e^{-\lambda_n(\tilde{a}_2)t_2} e^{-\lambda_n(\tilde{b}_2)u_2} \bar{M}_{t-T_{\text{unfr}}-t_1, u-T_{\text{unfr}}-t_1}^{\rho_{t-T_{\text{unfr}}-t_1, \tilde{a}_2}, \rho_{u-T_{\text{unfr}}-u_1, \tilde{b}_2}}(\tilde{a}_1, \tilde{b}_1) \\ &\quad \times d\mu_{n,\text{fr}}^{(1)}(t_1) d\mu_{n,\text{fr}}^{(2)}(t_2) d\mu_{n,\text{fr}}^{(2)}(u_2) \\ &+ \iiint e^{-(\lambda_n(\tilde{a}_2) + \lambda_n(\tilde{b}_2))t_2} e^{-\lambda_n(\tilde{a}_1)t_1} e^{-\lambda_n(\tilde{b}_1)u_1} \bar{M}_{t-T_{\text{unfr}}-t_2, u-T_{\text{unfr}}-t_2}^{\rho_{t-T_{\text{unfr}}-t_2, \tilde{a}_1}, \rho_{u-T_{\text{unfr}}-u_1, \tilde{b}_1}}(\tilde{a}_2, \tilde{b}_2) \\ &\quad \times d\mu_{n,\text{fr}}^{(1)}(t_2) d\mu_{n,\text{fr}}^{(2)}(t_1) d\mu_{n,\text{fr}}^{(2)}(u_1) \\ &+ \iint e^{-\lambda_n(\tilde{a}_1)t_1} e^{-\lambda_n(\tilde{a}_2)t_2} \bar{m}_{t-T_{\text{unfr}}-\tilde{t}}^{\chi_n^{(K)}}(\tilde{a}) d\mu_{n,\text{fr}}^{(1)}(t_1) d\mu_{n,\text{fr}}^{(2)}(t_2) \\ &\quad \times \iint e^{-\lambda_n(\tilde{b}_1)u_1} e^{-\lambda_n(\tilde{b}_2)u_2} \bar{m}_{u-T_{\text{unfr}}-\tilde{u}}^{\chi_n^{(K)}}(\tilde{b}) d\mu_{n,\text{fr}}^{(1)}(u_1) d\mu_{n,\text{fr}}^{(2)}(u_2), \end{aligned} \quad (9.25)$$

where  $\rho_{t_2, \tilde{a}_2}(t'_1) = \bar{m}_{t'_2}^{\chi_n^{(K)}}(t'_1, \cdot)(\tilde{a}_2)$  is the characteristic from (7.5). Abbreviate

$$D = e^{-2t^* - \frac{1}{4}\|\vec{\xi}\|^2} \hat{Z}'_{n,t^*}(\vec{\xi}) - e^{-2u^*} \hat{Z}'_{n,u^*}(\vec{\zeta}) = e^{-2t^* - \frac{1}{4}\|\vec{\xi}\|^2} \hat{Z}'_{n,t^*}(\vec{\xi}) - e^{-2u^*} \hat{Z}'_{n,u^*}(\vec{0}). \quad (9.26)$$

Then

$$\begin{aligned} \mathbb{E} \left( D^2 \mid \mathcal{B}_{\text{fr}} \right) &= e^{-4t^* - \frac{2}{4}\|\vec{\xi}\|^2} \mathbb{E} \left( \hat{Z}'_{n,t^*}(\vec{\xi}) \hat{Z}'_{n,t^*}(\vec{\xi}) \mid \mathcal{B}_{\text{fr}} \right) - 2e^{-2t^* - \frac{1}{4}\|\vec{\xi}\|^2 - 2u^*} \mathbb{E} \left( \hat{Z}'_{n,t^*}(\vec{\xi}) \hat{Z}'_{n,u^*}(\vec{0}) \mid \mathcal{B}_{\text{fr}} \right) \\ &+ e^{-4u^*} \mathbb{E} \left( \hat{Z}'_{n,u^*}(\vec{0}) \hat{Z}'_{n,u^*}(\vec{0}) \mid \mathcal{B}_{\text{fr}} \right). \end{aligned} \quad (9.27)$$

Applying (9.25) to each of the 3 terms in (9.27) gives 12 summands. From the first term in the right-hand side of (9.25), we obtain

$$\begin{aligned} & \frac{e^{o(1)}}{s_n^2} \iint s_n^{-4} \left[ e^{-2\lambda_n(\tilde{a}_1)t_1} e^{-2\lambda_n(\tilde{a}_2)t_2} \bar{M}_{t-T_{\text{unfr}}-\tilde{t}, t-T_{\text{unfr}}-\tilde{t}}^{\chi_n^{(K)}, \chi_n^{(K)}} \left( \tilde{\tilde{a}}, \tilde{\tilde{a}} \right) \right. \\ & \quad - 2e^{-(\lambda_n(\tilde{a}_1)+\lambda_n(1))t_1} e^{-(\lambda_n(\tilde{a}_2)+\lambda_n(1))t_2} \bar{M}_{t-T_{\text{unfr}}-\tilde{t}, u-T_{\text{unfr}}-\tilde{t}}^{\chi_n^{(K)}, \chi_n^{(K)}} \left( \tilde{\tilde{a}}, \tilde{1} \right) \\ & \quad \left. + e^{-2\lambda_n(1)t_1} e^{-2\lambda_n(1)t_2} \bar{M}_{u-T_{\text{unfr}}-\tilde{t}, u-T_{\text{unfr}}-\tilde{t}}^{\chi_n^{(K)}, \chi_n^{(K)}} \left( \tilde{1}, \tilde{1} \right) \right] d\mu_{n,\text{fr}}^{(1)}(t_1) d\mu_{n,\text{fr}}^{(2)}(t_2). \end{aligned} \quad (9.28)$$

As in (9.22), Theorem 3.15 and Corollary 6.5 imply that the integrand in (9.28) is at most  $(4\varepsilon + o(1))e^{-2\lambda_n(1)t_1 + o(1)}e^{-2\lambda_n(1)t_2 + o(1)}$  in absolute value, uniformly for  $\lambda_n(1)[t_1 \vee t_2] \leq K'$ , and is otherwise bounded by  $4K''e^{-2\lambda_n(1\wedge\tilde{a}_1)t_1}e^{-2\lambda_n(1\wedge\tilde{a}_2)t_2}$ . Using Lemmas 9.1 and 9.2, it easily follows that the quantity in (9.28) is at most  $5\varepsilon$  in absolute value, whp. From the second term in the right-hand side of (9.25), we obtain similarly

$$\begin{aligned} & \frac{e^{o(1)}}{s_n^3} \iiint s_n^{-3} \left[ e^{-2\lambda_n(\tilde{a}_1)t_1} e^{-2\lambda_n(\tilde{a}_2)(t_2+u_2)} \bar{M}_{t-T_{\text{unfr}}-t_2, \tilde{a}_2, \rho_{t-T_{\text{unfr}}-u_2, \tilde{a}_2}}^{\rho_{t-T_{\text{unfr}}-t_2, \tilde{a}_2}, \rho_{t-T_{\text{unfr}}-u_2, \tilde{a}_2}} (\tilde{a}_1, \tilde{a}_1) \right. \\ & \quad - 2e^{-(\lambda_n(\tilde{a}_1)+\lambda_n(1))t_1} e^{-\lambda_n(\tilde{a}_2)t_2} e^{-\lambda_n(1)u_2} \bar{M}_{t-T_{\text{unfr}}-t_1, t-T_{\text{unfr}}-t_1}^{\rho_{t-T_{\text{unfr}}-t_2, \tilde{a}_2}, \rho_{t-T_{\text{unfr}}-u_2, \tilde{a}_2}} (\tilde{a}_1, 1) \\ & \quad \left. + e^{-2\lambda_n(1)t_1} e^{-2\lambda_n(1)(t_2+u_2)} \bar{M}_{u-T_{\text{unfr}}-t_1, u-T_{\text{unfr}}-t_1}^{\rho_{u-T_{\text{unfr}}-t_2, 1}, \rho_{u-T_{\text{unfr}}-u_2, 1}} (1, 1) \right] d\mu_{n,\text{fr}}^{(1)}(t_1) d\mu_{n,\text{fr}}^{(2)}(t_2) d\mu_{n,\text{fr}}^{(2)}(u_2). \end{aligned} \quad (9.29)$$

Arguing from Proposition 7.1 instead of Theorem 3.15, the quantity in (9.29) is again at most  $5\varepsilon$  in absolute value, whp. The third and fourth terms from (9.25) are analogous. This verifies the second moment condition (3.56).

**Verification of (3.58):** Finally we verify condition (b) of Theorem 3.29. For  $R \in (0, \infty)$ , let  $v_{j,R}$  be the first vertex in  $\mathcal{B}^{(j)}$  born after time  $T_{\text{unfr}}$  with  $\left| BP_{f_n(1)}^{(v)} \right| \geq Rs_n^2$ , so that  $v$  is  $Rs_n^2$ -lucky in the sense of [Part I, Definition 2.14]. (We will use a similar notion in Definition 9.3.) It follows from [Part I, Proposition 2.17 and Lemma 2.18] that

$$T_{v_{j,R}}^{\mathcal{B}} = T_{\text{unfr}} + O_{\mathbb{P}}(f_n(1)). \quad (9.30)$$

To see this, note that [Part I, Lemma 2.18] gives  $T_{v_{j,r}} - T_{v_{j,r}} = O_{\mathbb{P}}(f_n(1))$  for each  $r > 0$ , so it suffices to show that  $T_{v_{j,r}}^{\mathcal{B}} = T_{\text{unfr}} + O_{\mathbb{P}}(f_n(1))$  for some  $r > 0$  sufficiently small. Let  $\varepsilon > 0$  and use Lemmas 9.1 and 9.2 to find  $K < \infty$  such that  $\mu_{n,\text{fr}}^{(j)}([0, Kf_n(1)]) \geq \frac{1}{2}s_n$  with probability at least  $1 - \varepsilon$  for  $n$  sufficiently large. Assuming this event occurs, the number of vertices born in the time interval  $(T_{\text{unfr}}, T_{\text{unfr}} + Kf_n(1))$  is conditionally Poisson with mean at least  $\frac{1}{2}s_n$ . By [Part I, Proposition 2.17] we can choose  $r > 0$  such that, for  $n$  sufficiently large, each such vertex has probability at least  $1/\varepsilon s_n$  of being  $rs_n^2$ -lucky. Hence at least one vertex born by time  $T_{\text{unfr}} + Kf_n(1)$  will be  $rs_n^2$ -lucky, with probability at least  $1 - \varepsilon - e^{-1/2\varepsilon} - o(1)$ . This verifies  $T_{v_{j,r}}^{\mathcal{B}} = T_{\text{unfr}} + O_{\mathbb{P}}(f_n(1))$  and hence proves (9.30).

Now let  $\varepsilon > 0$  and choose  $n_0 \in \mathbb{N}$  and  $C' < \infty$  such that  $\lambda_n(1)(f_n(1) + f_n(1+1/s_n)) \leq C'$  for all  $n \geq n_0$ . Choose  $R \in (0, \infty)$  such that  $\frac{16}{(\log 2)^4} e^{4C'} (1 + 2R)/R^2 \leq \varepsilon$ . After possibly increasing  $n_0$ , (9.30) and (6.15) yield a constant  $C'' \in (0, \infty)$  such that with probability at least  $1 - \varepsilon$ ,  $\lambda_n(1)(T_{v_{j,R}}^{\mathcal{B}} - T_{\text{unfr}}) \leq C''$ . Denote this event by  $\mathcal{A}$  and assume for the remainder of the proof that  $\mathcal{A}$  holds.

Set  $\mathcal{L}_R^{(j)}$  to be the collection of descendants  $w$  of  $v_{j,R}$  such that  $T_w^{\mathcal{B}} - T_{v_{j,R}}^{\mathcal{B}} \leq f_n(1)$  (thus  $|\mathcal{L}_R^{(j)}| \geq Rs_n^2$  by definition). We will repeat the previous arguments used to analyze  $Z_{n,t^*}$ ,



with  $\mathcal{L}_R^{(j)}$  playing the role of  $\mathcal{B}_{\text{fr}}^{(j)}$ . Instead of  $\partial\mathcal{L}_R^{(j)}$ , we consider the subset

$$\mathcal{U}_R^{(j)} = \left\{ v \in \partial\mathcal{L}_R^{(j)} : T_{p(v)}^{\mathcal{B}} + f_n(1) < T_v^{\mathcal{B}} < T_{p(v)}^{\mathcal{B}} + f_n(1 + 1/s_n) \right\} \quad (9.31)$$

[the set of immediate children of  $\mathcal{L}_R^{(j)}$  born to a parent of age between times  $f_n(1)$  and  $f_n(1 + 1/s_n)$ ]. Since  $v_{j,R}$  is born after  $T_{\text{unfr}}$ , it follows immediately that  $\mathcal{B}_{\text{fr}}^{(j)} \cap \mathcal{U}_R^{(j)} = \emptyset$ . Moreover the arrival time  $T_v^{\mathcal{B}}$  of any  $v \in \mathcal{U}_R^{(j)}$  satisfies

$$T_{v_{j,R}}^{\mathcal{B}} + f_n(1) < T_{p(v)}^{\mathcal{B}} + f_n(1) < T_v^{\mathcal{B}} < T_{p(v)}^{\mathcal{B}} + f_n(1 + 1/s_n) \leq T_{v_{j,R}}^{\mathcal{B}} + f_n(1) + f_n(1 + 1/s_n). \quad (9.32)$$

Let  $\tilde{\mathcal{P}}_{n,R}^{(j)} = \sum_{v \in \mathcal{U}_R^{(j)}} \delta_{(T_v^{\mathcal{B}} - T_{\text{unfr}}, BP(v))}$  [cf. (9.8)]. The definitions of  $v_{j,R}$  and  $\mathcal{L}_R^{(j)}$  depend only on descendants born to parents of age at most  $f_n(1)$ , whereas  $\mathcal{U}_R^{(j)}$  consists of descendants born to parents of age greater than  $f_n(1)$ . It follows that  $\tilde{\mathcal{P}}_{n,R}^{(j)}$  is a Cox process conditionally on  $\mathcal{L}_R^{(j)}$  with intensity measure  $d\tilde{\mu}_{n,R}^{(j)} \otimes d\mathbb{P}(BP^{(1)} \in \cdot)$ , where  $\tilde{\mu}_{n,R}$  is the measure such that, for all measurable, nonnegative functions  $h$  on  $\mathbb{R}$ ,

$$\int h(y) d\tilde{\mu}_{n,R}^{(j)}(y) = \sum_{w \in \mathcal{L}_R^{(j)}} \int_{f_n(1)}^{f_n(1+1/s_n)} h(y - (T_{\text{fr}}^{(j)} - T_w)) d\mu_n(y). \quad (9.33)$$

[These are the analogues of  $\mathcal{P}_{n,\text{unfr}}^{(j)}$  and  $\mu_{n,\text{fr}}^{(j)}$ , see (9.2 and (9.8).] In particular, the total mass of  $\tilde{\mu}_{n,R}^{(j)}$  is at least  $Rs_n$ . Using (9.32) we find that its support is in  $[T_{v_{j,R}}^{\mathcal{B}} - T_{\text{unfr}} + f_n(1), T_{v_{j,R}}^{\mathcal{B}} - T_{\text{unfr}} + f_n(1) + f_n(1 + 1/s_n)]$  and, uniformly over  $t_j$  in that set and  $t^* \geq 0$ ,

$$\lambda_n(1)(T_{v_{j,R}}^{\mathcal{B}} - T_{\text{unfr}}) \leq \lambda_n(1)t_j \leq \lambda_n(1)(T_{v_{j,R}}^{\mathcal{B}} - T_{\text{unfr}}) + C', \quad (9.34)$$

$$\lambda_n(1)(t - T_{\text{unfr}} - t_j) \geq \frac{1}{2} \log(n/s_n^3) - C'' - C'. \quad (9.35)$$

In particular, the fact that the right-hand side of (9.35) tends to infinity implies that, possibly after increasing  $n_0$ ,  $\mathcal{U}_R^{(j)} \subset \mathcal{B}_t^{(j)}$  whenever  $n \geq n_0$  and  $t^* \geq 0$ . Furthermore, Theorem 3.15 and Proposition 7.1 yield a constant  $K < \infty$  such that

$$\bar{m}_{t-T_{\text{unfr}}-t_1, t-T_{\text{unfr}}-t_2}^{\chi_n^{(K)}}(\vec{1}) \geq s_n/2, \quad (9.36)$$

$$\bar{M}_{t-T_{\text{unfr}}-\vec{t}, u-T_{\text{unfr}}-\vec{t}}^{\chi_n^{(K)}, \chi_n^{(K)}}(\vec{1}, \vec{1}) \leq \frac{2}{(\log 2)^2} s_n^4, \quad (9.37)$$

$$\bar{M}_{t-T_{\text{unfr}}-t_{j'}, u-T_{\text{unfr}}-t_i}^{\rho_t-T_{\text{unfr}}-t_j, 1 \cdot \rho_u-T_{\text{unfr}}-u_{j',1}}(1, 1) \leq \frac{2}{\log 2} s_n^3, \quad (9.38)$$

for  $\{j, j'\} = \{1, 2\}$ , all  $t_1, t_2$  in the support of  $\tilde{\mu}_{n,R}^{(j)}$  and  $n \geq n_0$ . Using this  $K$  to truncate, let  $\tilde{Z}_{n,t^*,R}^{(K)}$  denote the restriction of  $Z_{n,t^*}^{(K)}$  to pairs  $(v_1, v_2)$  for which each  $v_j$  is a descendant of some vertex of  $\mathcal{U}_R^{(j)}$ . As in the argument leading to (9.20), we conclude that

$$\begin{aligned} |Z_{n,t^*}^*| &\geq |\tilde{Z}_{n,t^*,R}^{(K)}| = \frac{e^{2t^*}}{s_n^3} \int d\tilde{\mathcal{P}}_{n,R}^{(1)}(t_1, bp^{(1)}) \\ &\quad \times \int d\tilde{\mathcal{P}}_{n,R}^{(2)}(t_2, bp^{(2)}) e^{-\lambda_n(1)(t_1+t_2)} z_{t-T_{\text{unfr}}-t_1, t-T_{\text{unfr}}-t_2}^{\chi_n^{(K)}, bp^{(1)}, bp^{(2)}}(\vec{1}). \end{aligned} \quad (9.39)$$

Hence, on  $\mathcal{A}$ , we may use (9.34) and (9.36) to obtain

$$\mathbb{E} \left( \left| \tilde{Z}_{n,t^*,R}^{(K)} \right| \middle| \mathcal{L}_R^{(1)}, \mathcal{L}_R^{(2)} \right) \quad (9.40)$$

$$\begin{aligned} &\geq \frac{e^{2t^*}}{s_n^3} \int d\tilde{\mu}_{n,R}^{(1)}(t_1) \int d\tilde{\mu}_{n,R}^{(2)}(t_2) \exp \left( -\lambda_n(1) \left( T_{v_{1,R}}^{\mathcal{B}} - T_{\text{unfr}} + T_{v_{2,R}}^{\mathcal{B}} - T_{\text{unfr}} \right) - 2C' \right) \frac{s_n}{2} \\ &= \frac{1}{2} \exp \left( 2t^* - \lambda_n(1) \left( T_{v_{1,R}}^{\mathcal{B}} - T_{\text{unfr}} + T_{v_{2,R}}^{\mathcal{B}} - T_{\text{unfr}} \right) - 2C' \right) \frac{\left| \tilde{\mu}_{n,R}^{(1)} \right|}{s_n} \frac{\left| \tilde{\mu}_{n,R}^{(2)} \right|}{s_n} \end{aligned} \quad (9.41)$$

$$\geq \frac{R^2}{2} \exp \left( 2t^* - 2(C'' + C') \right). \quad (9.42)$$

Similarly, using (9.24) to compute the variance of  $\left| \tilde{Z}_{n,t^*,R}^{(K)} \right|$  as in (9.25), and employing (9.34) and (9.37)–(9.38) for the estimation, we find that, on  $\mathcal{A}$ ,

$$\begin{aligned} \text{Var} \left( \left| \tilde{Z}_{n,t^*,R}^{(K)} \right| \middle| \mathcal{L}_R^{(1)}, \mathcal{L}_R^{(2)} \right) &\leq \frac{2}{(\log 2)^2} \exp \left( 4t^* - 2\lambda_n(1) \left( T_{v_{1,R}}^{\mathcal{B}} - T_{\text{unfr}} + T_{v_{2,R}}^{\mathcal{B}} - T_{\text{unfr}} \right) \right) \\ &\quad \times \left( \frac{\left| \tilde{\mu}_{n,R}^{(1)} \right| \left| \tilde{\mu}_{n,R}^{(2)} \right| s_n^4}{s_n^6} + \frac{\left| \tilde{\mu}_{n,R}^{(1)} \right| \left| \tilde{\mu}_{n,R}^{(2)} \right|^2 s_n^3 + \left| \tilde{\mu}_{n,R}^{(1)} \right|^2 \left| \tilde{\mu}_{n,R}^{(2)} \right| s_n^3}{s_n^6} \right). \end{aligned} \quad (9.43)$$

Abbreviate the conditional mean from (9.40) as  $m$ . Chebyshev's inequality, (9.41) and (9.43) give

$$\begin{aligned} \mathbb{P} \left( \left| \tilde{Z}_{n,t^*,R}^{(K)} \right| \geq \frac{1}{2}m \middle| \mathcal{L}_R^{(1)}, \mathcal{L}_R^{(2)} \right) &\geq 1 - \frac{16}{(\log 2)^4} e^{4C'} \left( \frac{s_n^2}{\left| \tilde{\mu}_{n,R}^{(1)} \right| \left| \tilde{\mu}_{n,R}^{(2)} \right|} + \frac{s_n}{\left| \tilde{\mu}_{n,R}^{(1)} \right|} + \frac{s_n}{\left| \tilde{\mu}_{n,R}^{(2)} \right|} \right) \\ &\geq 1 - \frac{16}{(\log 2)^4} e^{4C'} \left( \frac{1}{R^2} + 2\frac{1}{R} \right) \geq 1 - \varepsilon, \end{aligned} \quad (9.44)$$

on  $\mathcal{A}$ . For given  $C < \infty$ , we use (9.42) to choose  $t^*$  sufficiently large that  $\frac{1}{2}m \geq C$ . The claim now follows from  $\left| Z_{n,t^*}^* \right| \geq \left| \tilde{Z}_{n,t^*,R} \right|$  and (9.44).  $\square$

## 9.4 No Collisions from the Frozen Cluster: Proof of Lemma 3.32 (b)

In this section we prove Lemma 3.32 (b), which will show that whp the collision edge neither starts nor ends in the frozen cluster.

**Definition 9.3** Let  $\varepsilon_1$  denote the constant from Lemma 4.5 for  $K = 1$ . Call a vertex  $v \in \mathcal{T}^{(j)}$  *lucky* if  $\left| BP_{f_n(1)}^{(v)} \right| \geq s_n^2/\varepsilon_1$ , and set

$$T_{\text{lucky}}^{(j)} = \inf \left\{ T_v : v \in \mathcal{T}^{(j)} \setminus \{\emptyset_j\} \text{ is lucky and } T_v > T_{p(v)} + f_n(1) \right\}, \quad (9.45)$$

the first time that a lucky vertex is born to a parent of age greater than  $f_n(1)$ .

The notion that  $v$  is lucky is the same as the notion of  $v$  being  $s_n^2/\varepsilon_1$ -lucky in [Part I, Definition 2.14 and (2.24)]. We will use this fact later on, when we apply results from Part I.

In view of Definition 3.17 and Lemma 4.5, we have

$$v \in \mathcal{T}^{(j)} \text{ is lucky} \implies T_{\text{fr}}^{(j)} \leq T_v + f_n(1). \quad (9.46)$$

In other words, a lucky vertex has enough descendants in time  $f_n(1)$  that the integral in the definition (3.31) of the freezing time must be at least  $s_n$ .

It has been proved in [Part I, Lemma 6.4 and Proposition 2.17] that the distribution of

$$\sum_{v \in B P_{T_{\text{lucky}}^{(j)}}^{(j)}} \left( f_n^{-1} \left( T_{\text{lucky}}^{(j)} - T_v \right) - 1 \right)^+ \quad \text{is exponential with rate } \mathbb{P}(v \text{ is lucky}), \quad (9.47)$$

and that there is a constant  $\delta > 0$  so that for sufficiently large  $n$ ,

$$\mathbb{P}(v \text{ is lucky}) \geq \delta/s_n. \quad (9.48)$$

Now we are in the position to prove Lemma 3.32 (b):

**Proof of Lemma 3.32 (b)** It suffices to show that the Cox intensity  $Z_{n,\bar{t}}$  satisfies

$$\sum_{v_1 \in \mathcal{B}_{\bar{t}}^{(1)}} \sum_{v_2 \in \mathcal{B}_{\bar{t}}^{(2)}} \mathbb{1}_{\{v_1, v_2\} \cap \mathcal{B}_{\text{fr}} \neq \emptyset} Z_{n,\bar{t}}(\{v_1\} \times \{v_2\}) = o_{\mathbb{P}}(1), \quad (9.49)$$

where we recall that  $Z_{n,t}(\{v_1\} \times \{v_2\}) = \frac{1}{n} \mu_n(\Delta R_{v_1, v_2}, R_1(t) - R_1(T_{v_1}^{\mathcal{B}}) + R_2(t) - R_2(T_{v_2}^{\mathcal{B}}))$ .

We begin with the contribution to  $Z_{n,\bar{t}}$  arising from the restriction of  $\mu_n$  to  $(f_n(1), \infty)$ . Note that, by construction,  $R_j(\bar{t}) - R_j(T_{\text{fr}}^{(j)}) = \lambda_n(1)^{-1}(\frac{1}{2} \log(n/s_n^3) + K)$  and  $(R_j(T_{\text{fr}}^{(j)}) - R_j(T_{v_j}^{\mathcal{B}}))^+ = (T_{\text{fr}}^{(j)} - T_{v_j}^{\mathcal{B}})^+$ . Hence

$$\begin{aligned} & \frac{1}{n} \mu_n|_{(f_n(1), \infty)} (\Delta R_{v_1, v_2}, R_1(\bar{t}) - R_1(T_{v_1}^{\mathcal{B}}) + R_2(\bar{t}) - R_2(T_{v_2}^{\mathcal{B}})) \\ & \leq \frac{1}{n} \mu_n \left( f_n(1), \frac{\log(n/s_n^3) + 2K}{\lambda_n(1)} + R_1(T_{\text{fr}}^{(1)}) - R_1(T_{v_1}^{\mathcal{B}}) + R_2(T_{\text{fr}}^{(2)}) - R_2(T_{v_2}^{\mathcal{B}}) \right) \\ & \leq \frac{1}{n} \mu_n \left( f_n(1), 3 \frac{\log(n/s_n^3) + 2K}{\lambda_n(1)} \right) + \sum_{j=1}^2 \frac{1}{n} \mu_n \left( f_n(1), f_n(1) \vee 3 (T_{\text{fr}}^{(j)} - T_{v_j}^{\mathcal{B}})^+ \right). \end{aligned} \quad (9.50)$$

Lemma 4.3 and (6.15) imply that the first term in (9.50) is  $O(1)(\log(n/s_n^3) + K)/(ns_n)$ . In the second term, the  $j$ th summand is zero if  $v_j \notin \mathcal{B}_{\text{fr}}^{(j)}$ . For  $v_j \in \mathcal{B}_{\text{fr}}^{(j)}$ , we consider separately the intervals

$$\begin{aligned} I_1 &= \left( f_n(1), f_n(1) \vee (T_{\text{lucky}}^{(j)} - T_{v_j}^{\mathcal{B}})^+ \right), \\ I_2 &= \left( f_n(1) \vee (T_{\text{lucky}}^{(j)} - T_{v_j}^{\mathcal{B}})^+, f_n(1) \vee (T_{\text{fr}}^{(j)} - T_{v_j}^{\mathcal{B}})^+ \right), \\ I_3 &= \left( f_n(1) \vee (T_{\text{fr}}^{(j)} - T_{v_j}^{\mathcal{B}})^+, f_n(1) \vee 3 (T_{\text{fr}}^{(j)} - T_{v_j}^{\mathcal{B}})^+ \right), \end{aligned} \quad (9.51)$$

where  $T_{\text{lucky}}^{(j)}$  was defined in Definition 9.3. The definition of  $\mu_n$  gives  $\mu_n(I_1) = (f_n^{-1}(T_{\text{lucky}}^{(j)} - T_{v_j}^{\mathcal{B}}) - 1)^+$ . For  $I_2$ , note from (9.46) that  $I_2$  is a subinterval of  $(f_n(1), \infty)$  of length at most  $f_n(1)$ , so  $\mu_n(I_2) \leq O(1/s_n)$  by Lemma 4.3. For  $I_3$ , note from Lemma 4.1 that  $f_n(3^{1/\varepsilon_0 s_n} m) \geq 3f_n(m)$  for any  $m \geq 1$ . It follows that  $f_n^{-1}(3y) \leq 3^{1/\varepsilon_0 s_n} f_n^{-1}(y) = (1 + O(1/s_n)) f_n^{-1}(y)$  uniformly over  $y \geq f_n(1)$ , so that  $\mu_n(I_3) \leq O(1/s_n) f_n^{-1}(T_{\text{fr}}^{(j)})$ . We conclude that

$$\begin{aligned} & \frac{1}{n} \mu_n|_{(f_n(1), \infty)} \left( \Delta R_{v_1, v_2}, R_1(\bar{t}) - R_1(T_{v_1}^{\mathcal{B}}) + R_2(\bar{t}) - R_2(T_{v_2}^{\mathcal{B}}) \right) \\ & \leq O(1) \left( \frac{\log(n/s_n^3)}{ns_n} + \sum_{j=1}^2 \mathbb{1}_{\{v_j \in \mathcal{B}_{\text{fr}}^{(j)}\}} \left( \frac{1}{n} \left( f_n^{-1}(T_{\text{lucky}}^{(j)} - T_{v_j}^{\mathcal{B}}) - 1 \right)^+ + \frac{1 + f_n^{-1}(T_{\text{fr}}^{(j)})}{ns_n} \right) \right). \end{aligned} \quad (9.52)$$

By Theorem 3.18 (a),  $f_n^{-1}(T_{\text{fr}}^{(j)}) = O_{\mathbb{P}}(1)$ . Sum (9.52) over  $v_1 \in \mathcal{B}_{\bar{t}}^{(1)}, v_2 \in \mathcal{B}_{\bar{t}}^{(2)}$  with  $\{v_1, v_2\} \cap \mathcal{B}_{\text{fr}} \neq \emptyset$  and use  $T_{v_j}^{\mathcal{B}} = T_{v_j}$  for  $v_j \in \mathcal{B}_{\text{fr}}^{(j)}$  to obtain

$$O(1) \sum_{\{j, j'\}=\{1,2\}} \left| \mathcal{B}_{\bar{t}}^{(j')} \right| \left( \frac{\log(n/s_n^3) + O_{\mathbb{P}}(1)}{ns_n} \left| \mathcal{B}_{\text{fr}}^{(j)} \right| + \frac{1}{n} \sum_{v_j \in \mathcal{B}_{\text{fr}}^{(j)}} \left( f_n^{-1}(T_{\text{lucky}}^{(j)} - T_{v_j}) - 1 \right)^+ \right). \quad (9.53)$$

By Lemma 3.32 (a),  $\left| \mathcal{B}_{\bar{t}}^{(j')} \right| = O_{\mathbb{P}}(\sqrt{ns_n})$ , and by Theorem 3.18 (b),  $\left| \mathcal{B}_{\text{fr}}^{(j)} \right| = O_{\mathbb{P}}(s_n^2)$ . In the sum over  $v_j \in \mathcal{B}_{\text{fr}}^{(j)}$ , only terms with  $v_j \in \mathcal{B}_{T_{\text{lucky}}^{(j)}}^{(j)}$  can contribute, so (9.46)–(9.48) imply that the inner sum in (9.53) is  $O_{\mathbb{P}}(s_n)$ . Hence (9.53) is  $O_{\mathbb{P}}(\log(n/s_n^3)/\sqrt{n/s_n^3})$ , which is  $o_{\mathbb{P}}(1)$  since  $n/s_n^3 \rightarrow \infty$ .

We now turn to the contribution to  $Z_{n, \bar{t}}$  arising from the restriction of  $\mu_n$  to  $[0, f_n(1)]$  and split the sum into three groups of vertex pairs. Let  $(J, J')$  denote the random ordering of  $\{1, 2\}$  for which  $T_{\text{fr}}^{(J)} < T_{\text{fr}}^{(J')}$ . The first group of vertex pairs are those with  $v_J \in \mathcal{B}_{\bar{t}}^{(J)}$  and  $v_{J'} \in \mathcal{B}_{T_{\text{fr}}^{(J)}}^{(J')}$ . That is, the vertex in the slower-growing cluster is born before the faster cluster freezes. We show that the number  $\left| \mathcal{B}_{T_{\text{fr}}^{(J)}}^{(J')} \right|$  of such choices for  $v_{J'}$  is  $O_{\mathbb{P}}(1)$ . By Theorem 3.18 (a),  $f_n^{-1}(T_{\text{fr}}^{(j)}) \xrightarrow{\mathbb{P}} M^{(j)}$  for  $j \in \{1, 2\}$  and  $M^{(1)} \neq M^{(2)}$  a.s. Hence  $T_{\text{fr}}^{(J)} < f_n(M^{(J')})$  whp. A vertex  $v' \in \mathcal{B}_{\text{fr}}^{(J')}$  with  $T_{v'}^{\mathcal{B}} < f_n(M^{(J')})$  must be connected to the root  $\emptyset_{J'}$  by edges that all have PWIT edge weight less than  $M^{(J')}$ . The number of such vertices is finite and independent of  $n$ , and is in particular  $O_{\mathbb{P}}(1)$ . Since the measure  $\mu_n|_{[0, f_n(1)]}$  has total mass 1 by construction, the total contribution to  $Z_{n, \bar{t}}$  of this group of vertex pairs is at most  $\frac{1}{n} \left| \mathcal{B}_{\bar{t}}^{(J)} \right| O_{\mathbb{P}}(1)$ , which is  $o_{\mathbb{P}}(1)$  by Lemma 3.32 (a).

The second group are pairs  $(v_1, v_2)$  with  $T_{v_{J'}}^{\mathcal{B}} \geq T_{\text{fr}}^{(J)} \geq T_{v_J}^{\mathcal{B}}$ . For these pairs, by (3.49),

$$\Delta R_{v_1, v_2} = R_J(T_{v_{J'}}^{\mathcal{B}}) - R_J(T_{v_J}^{\mathcal{B}}) \geq R_J(T_{\text{fr}}^{(J)}) - R_J(T_{v_J}^{\mathcal{B}}) = T_{\text{fr}}^{(J)} - T_{v_J}^{\mathcal{B}}. \quad (9.54)$$

We can therefore bound the contribution to  $Z_{n, \bar{t}}$  in terms of the contribution to  $\mu_{n, \text{fr}}^{(J)}$ :

$$\begin{aligned} & \frac{1}{n} \sum_{v_1, v_2: T_{v_{J'}}^{\mathcal{B}} \geq T_{\text{fr}}^{(J)} \geq T_{v_J}^{\mathcal{B}}} \mu_n|_{[0, f_n(1)]} \left( \Delta R_{v_1, v_2}, R_1(\bar{t}) - R_1(T_{v_1}^{\mathcal{B}}) + R_2(\bar{t}) - R_2(T_{v_2}^{\mathcal{B}}) \right) \\ & \leq \frac{1}{n} \left| \mathcal{B}_{\bar{t}}^{(J')} \right| \sum_{v_J \in \mathcal{B}_{\text{fr}}^{(J)}} \mu_n \left( f_n(1) \wedge (T_{\text{fr}}^{(J)} - T_{v_J}^{\mathcal{B}}), f_n(1) \right) \\ & \leq \frac{1}{n} O_{\mathbb{P}}(\sqrt{ns_n}) \sum_{v_J \in \mathcal{B}_{\text{fr}}^{(J)}} e^{\lambda_n(1)f_n(1)} \int_0^\infty e^{-\lambda_n(1)y} \mu_n(T_{\text{fr}}^{(J)} - T_{v_J}^{\mathcal{B}} + dy) \leq O_{\mathbb{P}}(\sqrt{s_n^3/n}) \end{aligned} \quad (9.55)$$

by Lemma 3.32 (a), Lemma 9.1 and (6.15). The last group of vertex pairs are those with  $T_{v_j}^{\mathcal{B}} \geq T_{\text{fr}}^{(j')} \geq T_{v_{j'}}^{\mathcal{B}}$ . These pairs satisfy  $\Delta R_{v_1, v_2} \geq T_{\text{fr}}^{(j')} - T_{v_{j'}}^{\mathcal{B}}$ , instead of (9.54), and their contribution can therefore be handled as in (9.55) with  $J$  and  $J'$  interchanged.  $\square$

## 9.5 Properties of the Freezing Time and Frozen Cluster: Proof of Theorem 3.18

**Proof of Theorem 3.18** The proof uses the results of [Part I, Theorem 2.15], which applies because the assumptions in [Part I, Conditions 2.2–2.3] are weaker than those of Conditions 2.1–2.3. We adopt the terminology of [Part I, Theorem 2.15], adding superscripts in the obvious way to indicate the two copies of the PWIT.

Let  $\varepsilon_1$  denote the constant from Lemma 4.5 with  $K = 1$ . Then Lemma 4.5 asserts that any vertex  $v \in BP_t^{(j)}$  of age at most 1, i.e.  $t - T_v \leq f_n(1)$ , contributes at least  $\varepsilon_1/s_n$  to the sum in the formula (3.31) defining  $T_{\text{fr}}^{(j)}$ , for  $n$  sufficiently large. In particular, if there are at least  $s_n^2/\varepsilon_1$  such vertices at time  $t$ , then the sum in (3.31) will be at least  $s_n$ . We conclude that then  $v$  is lucky, as in Definition 9.3. Thus,

$$T_{\text{fr}}^{(j)} \leq T_{\text{lucky}}^{(j)} \quad (9.56)$$

for  $n$  sufficiently large.

On the other hand, by Lemma 4.4 [with  $\bar{\varepsilon} = 1$  and  $\varepsilon$  chosen sufficiently small that  $\varepsilon/f_n(1) \leq \lambda_n(1)$ ; this is possible, for  $n$  sufficiently large, by Lemma 3.12], a single vertex can contribute at most

$$\begin{aligned} \int_t^\infty e^{-\lambda_n(1)(y-t)} d\mu_n(y) &\leq \mu_n(0, Kf_n(1)) + \int e^{-\varepsilon y'/f_n(1)} \mathbb{1}_{\{y' \geq Kf_n(1)\}} \mu_n(t + dy') \\ &\leq 1 + O(1/s_n) + 1/s_n \end{aligned} \quad (9.57)$$

to the sum in (3.31). In particular, this contribution is  $O(1)$ . It follows that  $\mathcal{B}_{\text{fr}}$  must contain at least  $\delta s_n$  vertices, for some  $\delta > 0$  chosen sufficiently small, and thus that

$$T_{\text{fr}}^{(j)} \geq T_{\text{size } \delta s_n}^{(j)}. \quad (9.58)$$

The proof now follows from (9.56), (9.58) and [Part I, Theorem 2.15] (with  $\sigma_n = s_n/\sqrt{\varepsilon_1}$  and  $\sigma_n = \sqrt{\delta s_n}$ , respectively).  $\square$

### A short guide to notation:

- $SWT_t^{(j)}$  is SWT from vertex  $j \in \{1, 2\}$ .
- $S_t^{(j)}$  is the SWT from vertex  $j \in \{1, 2\}$  such that  $\mathcal{S}^{(1)}$  and  $\mathcal{S}^{(2)}$  cannot merge and with an appropriate freezing procedure.
- $\mathcal{S}_t = \mathcal{S}_t^{(1)} \cup \mathcal{S}_t^{(2)}$ .
- $BP^{(j)}$  is branching process copy number  $j$  where  $j \in \{1, 2\}$ , without freezing.
- $\mathcal{B}^{(j)}$  is branching process copy number  $j$  where  $j \in \{1, 2\}$ , with freezing.
- $\mathcal{B}_t$  is the union of 2 CTBPs with the appropriate freezing of one of them.
- $\tilde{\mathcal{B}}_t$  is the union of 2 CTBPs with the appropriate freezing of one of them, and the resulting thinning. Thus,  $\tilde{\mathcal{B}}_t$  has the same law as the frozen  $\mathcal{S}_t$ .
- $f_n$  is the function with  $Y_e^{(K_n)} \stackrel{d}{=} f_n(nE)$ , where  $E$  is exponential with mean 1.
- $\mu_n$  is the image of the Lebesgue measure on  $(0, \infty)$  under  $f_n$ .
- $\lambda_n(a)$  is the exponential growth rate of the CTBP, cf. (3.16).
- $z_t^X(a)$  and  $z_t^X(\vec{a})$  are the generation-weighted vertex characteristics from one and two vertices, respectively, cf. (3.14) and (3.22).

- $\bar{m}_i^X(a)$  and  $\bar{m}_i^X(\bar{a})$  are the expected, rescaled vertex characteristics, cf. (3.19) and (3.23), respectively.

**Acknowledgements** A substantial part of this work has been done at Eurandom and Eindhoven University of Technology. ME and JG are grateful to both institutions for their hospitality. The work of JG was carried out in part while at Leiden University (supported in part by the European Research Council Grant VARIS 267356) and the Technion, and JG thanks his hosts at those institutions for their hospitality. The work of JG at the University of Auckland is supported by the Marsden Fund, administered by the Royal Society of New Zealand. The work of RvdH is supported by the Netherlands Organisation for Scientific Research (NWO) through VICI Grant 639.033.806 and the Gravitation NETWORKS Grant 024.002.003.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Addario-Berry, L.: The local weak limit of the minimum spanning tree of the complete graph. Preprint. [arXiv:1301.1667v2](https://arxiv.org/abs/1301.1667v2) [math.PR], (2013)
2. Addario-Berry, L., Griffiths, S., Kang, R.J.: Invasion percolation on the Poisson-weighted infinite tree. *Ann. Appl. Probab.* **22**(3), 931–970 (2012)
3. Addario-Berry, L., Broutin, N., Goldschmidt, C., Miermont, G.: The scaling limit of the minimum spanning tree of the complete graph. *Ann. Probab.* **45**(5), 3075–3144 (2017)
4. Adriaans, E., Komjáthy, J.: Weighted distances in scale-free configuration models. *J. Stat. Phys.* **173**(3–4), 1082–1109 (2018)
5. Amini, H., Peres, Y.: Shortest-weight paths in random regular graphs. *SIAM J. Discrete Math.* **28**(2), 656–672 (2014)
6. Amini, H., Draief, M., Lelarge, M.: Flooding in weighted random graphs. In: *ANALCO11—Workshop on Analytic Algorithmics and Combinatorics*, pp. 1–15. SIAM, Philadelphia (2011)
7. Antunović, T., Dekel, Y., Mossel, E., Peres, Y.: Competing first passage percolation on random regular graphs. *Random Struct. Algorithms* **50**(4), 534–583 (2017)
8. Barabási, A., Albert, R.: Emergence of scaling in random networks. *Science* **286**(5439), 509–512 (1999)
9. Baroni, E., van der Hofstad, R., Komjáthy, J.: Nonuniversality of weighted random graphs with infinite variance degree. *J. Appl. Probab.* **54**(1), 146–164 (2017)
10. Baroni, E., van der Hofstad, R., Komjáthy, J.: Tight fluctuations of weight-distances in random graphs with infinite-variance degrees. *J. Stat. Phys.* **174**(4), 906–934 (2019)
11. Bhamidi, S., van der Hofstad, R.: Weak disorder asymptotics in the stochastic mean-field model of distance. *Ann. Appl. Probab.* **22**(1), 29–69 (2012)
12. Bhamidi, S., van der Hofstad, R.: Diameter of the stochastic mean-field model of distance. *Comb. Probab. Comput.* **26**(6), 797–825 (2017)
13. Bhamidi, S., van der Hofstad, R., Hooghiemstra, G.: Extreme value theory, Poisson–Dirichlet distributions, and first passage percolation on random networks. *Adv. Appl. Probab.* **42**(3), 706–738 (2010)
14. Bhamidi, S., van der Hofstad, R., Hooghiemstra, G.: First passage percolation on random graphs with finite mean degrees. *Ann. Appl. Probab.* **20**(5), 1907–1965 (2010)
15. Bhamidi, S., van der Hofstad, R., Hooghiemstra, G.: First passage percolation on the Erdős–Rényi random graph. *Comb. Probab. Comput.* **20**(5), 683–707 (2011)
16. Bhamidi, S., van der Hofstad, R., Hooghiemstra, G.: Weak disorder in the stochastic mean-field model of distance II. *Bernoulli* **19**(2), 363–386 (2013)
17. Bhamidi, S., van der Hofstad, R., Hooghiemstra, G.: Universality for first passage percolation on sparse random graphs. *Ann. Probab.* **45**(4), 2568–2630 (2017)
18. Braunstein, L.A., Buldyrev, S.V., Cohen, R., Havlin, S., Stanley, H.E.: Optimal paths in disordered complex networks. *Phys. Rev. Lett.* **91**(16), 168701 (2003)

19. Durrett, R.: Random Graph Dynamics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (2007)
20. Eckhoff, M., Goodman, J., van der Hofstad, R., Nardi, F.R.: Short paths for first passage percolation on the complete graph. *J. Stat. Phys.* **151**, 1056–1088 (2013)
21. Eckhoff, M., Goodman, J., van der Hofstad, R., Nardi, F.R.: Long paths in first passage percolation on the complete graph I. Local PWIT dynamics. *Electron. J. Probab.* **25**, 1–45 (2020). <https://doi.org/10.1214/20-EJP484>
22. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events for Insurance and Finance. Applications of Mathematics-Statistical Modelling and Applied Probability, vol. 33. Springer, Berlin (1997)
23. Havlin, S., Braunstein, L.A., Buldyrev, S.V., Cohen, R., Kalisky, T., Sreenivasan, S., Stanley, E.H.: Optimal path in random networks with disorder: a mini review. *Physica A* **346**(1–2), 82–92 (2005)
24. van der Hofstad, R.: Random Graphs and Complex Networks. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 1. Cambridge University Press, Cambridge (2017)
25. van der Hofstad, R., Komjáthy, J.: Explosion and distances in scale-free percolation. Preprint. [arXiv:1706.02597](https://arxiv.org/abs/1706.02597) [math.PR], (2017)
26. van der Hofstad, R., Hooghiemstra, G., Znamenski, D.: Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Probab.* **12**(25), 703–766 (2007)
27. Janson, S.: One, two and three times  $\log n/n$  for paths in a complete graph with random weights. *Comb. Probab. Comput.* **8**(4), 347–361 (1999)
28. Kolossváry, I., Komjáthy, J.: First passage percolation on inhomogeneous random graphs. *Adv. Appl. Probab.* **47**(2), 589–610 (2015)
29. Komjáthy, J., Lodewijks, B.: Explosion in weighted hyperbolic random graphs and geometric inhomogeneous random graphs. *Stoch. Process. Appl.* **130**(3), 1309–1367 (2020)
30. Komjáthy, J., Vadon, V.: First passage percolation on the Newman–Watts small world model. *J. Stat. Phys.* **162**(4), 959–993 (2016)
31. Newman, M.E.J.: The structure and function of complex networks. *SIAM Rev.* **45**(2), 167–256 (2003). (electronic)
32. Newman, M.E.J.: Networks: An Introduction. Oxford University Press, Oxford (2010)
33. Sreenivasan, S., Kalisky, T., Braunstein, L.A., Buldyrev, S.V., Havlin, S., Stanley, H.E.: Effect of disorder strength on optimal paths in complex networks. *Phys. Rev. E* **70**(4), 46133 (2004)
34. Thorisson, H.: Coupling, Stationarity, and Regeneration. Probability and Its Applications. Springer, New York (2000)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Maren Eckhoff<sup>1</sup> · Jesse Goodman<sup>2</sup>  · Remco van der Hofstad<sup>3</sup>  · Francesca R. Nardi<sup>3</sup>

✉ Remco van der Hofstad  
rhofstad@win.tue.nl

Maren Eckhoff  
eckhoff.maren@gmail.com

Jesse Goodman  
jesse.goodman@auckland.ac.nz

Francesca R. Nardi  
f.r.nardi@tue.nl

<sup>1</sup> Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

<sup>2</sup> Department of Statistics, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

<sup>3</sup> Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands