EURANDOM PREPRINT SERIES

2014-013

May, 2014

Universality for first passage percolation on sparse random graphs

S. Bhamidi, R. van der Hofstad, G. Hooghiemstra
ISSN 1389-2355

UNIVERSALITY FOR FIRST PASSAGE PERCOLATION ON SPARSE RANDOM GRAPHS

SHANKAR BHAMIDI[1], REMCO VAN DER HOFSTAD[2], AND GERARD HOOGHIEMSTRA[3]

Abstract. We consider first passage percolation on the configuration model with n vertices, and general independent and identically distributed edge weights assumed to have a density. Assuming that the degree distribution satisfies a uniform $X^2 \log X$-condition, we analyze the asymptotic distribution for the minimal weight path between a pair of typical vertices, as well the number of edges on this path namely the hopcount.

The hopcount satisfies a central limit theorem (CLT). Furthermore, writing L_n for the weight of this optimal path, then we shown that $L_n - (\log n)/\alpha_n$ converges to a limiting random variable, for some sequence α_n. This sequence α_n and the norming constants for the CLT are expressible in terms of the parameters of an associated continuous-time branching process that describes the growth of neighborhoods around a uniformly chosen vertex in the random graph. The limit of $L_n - (\log n)/\alpha_n$ equals the sum of the logarithm of the product of two independent martingale limits, and a Gumbel random variable. Till date, for sparse random graph models, such results have been shown only for the special case where the edge weights have an exponential distribution, wherein the Markov property of this distribution plays a crucial role in the technical analysis of the problem.

The proofs in the paper rely on a refined coupling between shortest path trees and continuous-time branching processes, and on a Poisson point process limit for the potential closing edges of shortest-weight paths between the source and destination.

1. Introduction and results

1.1. **Motivation.** First passage percolation (FPP) is an important topic in modern probability theory, motivated by questions in a number of fields including disordered systems in statistical physics, where it arises as a building block in the analysis of complicated interacting particle systems such as the contact process, branching random walk and various epidemic models.

Let us start by describing the basic model. Let $\mathcal{G}$ be a connected graph on n vertices. Assign independent and identically distributed (i.i.d.) random edge *weights* or *lengths* to the edges of the graph. These random edge weights generate geodesics on the graph. Think of the graph as a disordered random system carrying flow between pairs of vertices in the graph via shortest paths between them. Choose two vertices in the graph uniformly at random amongst the n vertices. We will call these two vertices "typical" vertices. Two functionals of interest are the minimal weight L_n of a path between the two vertices and the number of edges H_n on the minimal path, often referred to as the *hopcount*. We assume that the common distribution of the edge weights is continuous, so that the optimal paths are a.s. unique and one can talk about objects such as the number of edges in *the* optimal path.

This model has been studied intensively, largely in the context of the integer lattice $[-N, N]^d$ (see e.g. [19, 22, 29, 38]). For the power of this model to analyze more complicated interacting

[1]Department of Statistics and Operations Research, 304 Hanes Hall, University of North Carolina, Chapel Hill, NC 27599

[2]Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

[3]DIAM, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

E-mail addresses: bhamidi@email.unc.edu, rhofstad@win.tue.nl,g.hooghiemstra@tudelft.nl.

2010 *Mathematics Subject Classification.* Primary: 60C05, 05C80, 90B15.

Key words and phrases. central limit theorem, continuous-time branching processes, extreme value theory, first passage percolation, hopcount, Malthusian rate of growth, point process convergence, Poisson point process, stable-age distribution, random graphs.

particle systems, see [32] and [16] and the references therein. In the modern context, FPP problems take on an added significance. Many real-world networks (such as the Internet at the router level or various road and rail networks) are entrusted with carrying flow between various parts of the network. These networks have both a graph theoretic structure as well as weights on edges, representing for example congestion. In the applied setting, understanding properties of both the hopcount and the minimal weight are crucial, since whilst routing is done via least weight paths, the actual time delay experienced by users scales like the hopcount (the number of "hops" a message has to perform in getting from the source to the destination). Simulation-based studies (see e.g. [13]) suggest that random edge weights have a marked effect on the geometry of the network. This has been rigorously established in various works [3, 8–10], in the specific situation of *exponential* edge weights.

In this paper, we study the behavior of the hopcount and minimal weight in the setting of random graphs with finite variance degrees and *general* continuous edge weights. Since in many applications, the distribution of edge weights is unknown, the assumption of general weights is highly relevant. From a mathematical point of view, working with general instead of exponential edge weights implies that our exploration process is *non-Markovian.* This is the first paper that studies first passage percolation on random graph models in this general setting. In a forthcoming paper [7] we will show that, due to the flexible choice of degree distribution, our results carry over to various other random graph models, including rank-1 inhomogeneous random graphs as introduced in [12].

Organization of this section. We start by introducing the configuration model in Section 1.2, where we also state our main result, Theorem 1.2. In Section 1.3, we discuss a continuous-time branching process approximation, which is necessary to identify the limiting variables in Theorem 1.2; this identification is done in Theorem 1.3. In Section 1.4, we study some examples that allow us to relate our results to existing results in the literature. We close with Section 1.5 where we present a discussion of our results and pose some open problems.

Throughout this paper, we make use of the following standard notation. We let $\xrightarrow{a.s.}$ denote convergence almost surely, $\xrightarrow{L^1}$ denote convergence in mean, $\xrightarrow{d}$ denote convergence in distribution, and $\xrightarrow{\mathbb{P}}$ convergence in probability. For a sequence of random variables $(X_n)_{n\geq 1}$, we write $X_n = O_{\mathbb{P}}(b_n)$ when $|X_n|/b_n$ is a tight sequence of random variables, and $X_n = o_{\mathbb{P}}(b_n)$ when $|X_n|/b_n \xrightarrow{\mathbb{P}} 0$, as $n \to \infty$. We write $D \sim F$ to denote that the random variable D has distribution function F. For non-negative functions $n \mapsto f(n)$, $n \mapsto g(n)$, we write $f(n) = O(g(n))$ when $f(n)/g(n)$ is uniformly bounded, and $f(n) = o(g(n))$ when $\lim_{n\to\infty} f(n)/g(n) = 0$. Furthermore, we write $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$. Finally, we say that a sequence of events $(\mathcal{E}_n)_{n\geq 1}$ occurs *with high probability* (whp) when $\mathbb{P}(\mathcal{E}_n) \to 1$.

1.2. **Configuration model and main result.** The configuration model (CM) is a random graph with vertex set $[n] := \{1, 2, \ldots, n\}$ and with prescribed degrees. Let $\boldsymbol{d} = (d_1, d_2, \ldots, d_n)$ be a given *degree sequence*, i.e., a sequence of n positive integers with total degree

$$\ell_n = \sum_{i\in[n]} d_i, \tag{1.1}$$

assumed to be even. The CM on n vertices with degree sequence $\boldsymbol{d}$ is constructed as follows: start with n vertices and d_i half-edges adjacent to vertex $i \in [n]$. Randomly choose pairs of half-edges and match the chosen pairs together to form edges. Although self-loops may occur, these become rare as $n \to \infty$ (see e.g. [11, 25]). We denote the resulting graph on $[n]$ by $\mathrm{CM}_n(\boldsymbol{d})$, with corresponding edge set $\mathcal{E}_n$.

Regularity of vertex degrees. Let us now describe our regularity assumptions on the degree sequence $\boldsymbol{d}$ as $n \to \infty$. We denote the degree of a uniformly chosen vertex V in $[n]$ by $D_n = d_V$.

The random variable D_n has distribution function F_n given by

$$F_n(x) = \frac{1}{n} \sum_{j \in [n]} \mathbb{1}_{\{d_j \le x\}}, \tag{1.2}$$

where $\mathbb{1}_A$ denotes the indicator of the event A. We write $\log(x)_+ = \log(x)$ for $x \ge 1$ and $\log(x)_+ = 0$ for $x \le 1$. Then our regularity condition is as follows:

Condition 1.1 (Regularity conditions for vertex degrees)**.**

(a) **Weak convergence of vertex degree.**
There exists a cumulative distribution function F of a discrete random variable D, taking values in $\mathbb{N}$ such that

$$\lim_{n\to\infty} F_n(x) = F(x), \tag{1.3}$$

for any continuity point x of F; i.e., $D_n \xrightarrow{d} D$.

(b) **Convergence of second moment.**

$$\lim_{n\to\infty} \mathbb{E}[D_n^2] = \lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} d_j^2 = \mathbb{E}[D^2], \tag{1.4}$$

where D_n and D have distribution functions F_n and F, respectively, and we assume that

$$\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1. \tag{1.5}$$

(c) **Uniform $X^2 \log X$-condition.** *For every $K_n \to \infty$,*

$$\lim_{n\to\infty} \mathbb{E}[D_n^2 \log{(D_n/K_n)}_+] = \lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} d_j^2 \log{(d_j/K_n)}_+ = 0. \tag{1.6}$$

By Condition 1.1(c), the random degree D_n satisfies a uniform $X^2 \log X$-condition. The degree of a vertex incident to a half-edge that is chosen uniformly at random from all half-edges has the same distribution as the random variable $D_n^\star$ given by

$$F_n^\star(x) = \mathbb{E}[D_n \mathbb{1}_{\{D_n \le x\}}]/\mathbb{E}[D_n], \quad x \in \mathbb{R}, \tag{1.7}$$

which is the size-biased version of D_n. The latter random variable satisfies a uniform $X \log X$-condition if and only if D_n satisfies a uniform $X^2 \log X$-condition. As explained in more detail in Section 1.3 below, $D_n^\star$ is closely related to a *branching-process approximation* of neighborhoods of a uniform vertex, and Condition 1.1(c) implies that this branching process satisfies a uniform $X \log X$ condition. By uniform integrability, Condition 1.1(c) follows from the assumption that $\lim_{n\to\infty} \mathbb{E}[D_n^2 \log{(D_n)}_+] = \mathbb{E}[D^2 \log{(D)}_+]$.

Note that that Conditions 1.1(a) and (c) imply that $\mathbb{E}[D_n^i] \to \mathbb{E}[D^i]$, $i = 1, 2$. When the degrees are *random* themselves, then the distribution function F_n as well as the left-hand side of (1.4) and (1.6), are *random* and we assume that the convergence in (1.3), (1.4) and (1.6) to the respective (deterministic) right-hand sides holds *in probability.* Thus, in this case, we require that, with $\mathbb{E}_n[D_n^i] = \frac{1}{n}\sum_{j\in[n]} d_j^i$ (which is now a random variable) and for every $\varepsilon > 0$ and $i \in \{1, 2\}$,

$$\lim_{n\to\infty} \mathbb{P}(|F_n(x) - F(x)| \ge \varepsilon) = 0, \quad \forall x \in \mathbb{R}, \qquad \lim_{n\to\infty} \mathbb{P}(|\mathbb{E}_n[D_n^i] - \mathbb{E}[D^i]| \ge \varepsilon) = 0. \tag{1.8}$$

A similar condition replaces (1.6).

Condition (1.5) is equivalent to the existence of a giant component in $\mathrm{CM}_n(\boldsymbol{d})$, see e.g. [27, 34, 35]. Let F be a distribution function of a random variable D, satisfying (1.5) and $\mathbb{E}[D^2 \log{(D)}_+] < \infty$. We give two canonical examples in which Condition 1.1 holds. The first is when there are precisely $n_k = \lceil nF(k)\rceil - \lceil nF(k-1)\rceil$ vertices having degree $k \ge 1$. The second is when $(d_i)_{i\in[n]}$ is an i.i.d. sequence of random variables with distribution function F (in the case that $\sum_{i\in[n]} d_i$ is odd, we increase d_n by 1, this does not affect the results).

Edge weights and shortest paths. Once the graph has been constructed, we attach an edge weight ξ_e to every edge e, where $(\xi_e)_{e\in\mathcal{E}_n}$ are i.i.d. continuous random variables with density $f_\xi\colon [0,\infty)\to[0,\infty)$ and corresponding distribution function F_ξ. Pick two vertices U_1 and U_2 at random from $[n]$ and let Γ_{12} denote the set of all paths in $\mathrm{CM}_n(\boldsymbol{d})$ between these two vertices. For any path $\pi\in\Gamma_{12}$, the weight of the path is defined as

$$\sum_{e\in\pi}\xi_e. \tag{1.9}$$

Let

$$L_n=\min_{\pi\in\Gamma_{12}}\sum_{e\in\pi}\xi_e, \tag{1.10}$$

denote the weight of the optimal (i.e., minimal weight) path between the U_1 and U_2 and let H_n denote the number of edges or the *hopcount* of this path. If the two vertices are in different components of the graph, then we let $L_n, H_n=\infty$. We are ready to state our main result. Due to the complexity of the various constants and limiting random variables arising in the theorem, we defer their complete description to the next section.

Theorem 1.2 (Joint convergence of hopcount and weight)**.** *Consider the configuration model* $\mathrm{CM}_n(\boldsymbol{d})$ *with degrees* $\boldsymbol{d}$ *satisfying Condition 1.1, and with i.i.d. edge weights distributed according to the continuous distribution* F_ξ*. Then, there exist constants* $\alpha,\gamma,\beta\in(0,\infty)$ *and* α_n,γ_n *with* $\alpha_n\to\alpha, \gamma_n\to\gamma$*, such that the hopcount* H_n *and weight* L_n *of the optimal path between two uniformly chosen vertices, conditioned on being connected, satisfy*

$$\Big(\frac{H_n-\gamma_n\log n}{\sqrt{\beta\log n}},\ L_n-\frac{1}{\alpha_n}\log n\Big)\xrightarrow{d}(Z,Q), \tag{1.11}$$

as $n\to\infty$*, where* Z *and* Q *are independent and* Z *has a standard normal distribution, while* Q *has a continuous distribution.*

This is the first time that FPP on sparse random graphs with general edge weights has been studied; for the particular case where the edge weights have an exponential distribution see e.g. [9].

In Remark 1.4 below, we will state conditions that imply that we can replace α_n and γ_n by their limits α and γ, respectively. Theorem 1.2 shows a remarkable degree of universality. For $\mathrm{CM}_n(\boldsymbol{d})$ satisfying Condition 1.1, the hopcount always satisfies a central limit theorem with mean and variance proportional to $\log n$. Also, the weight of the shortest weight path between two uniformly chosen vertices always is of order $\log n$, and the fluctuations around $\log n/\alpha_n$ converge in distribution. We will see that even the limit Q has a large degree of universality. For this, as well as to define the parameters $\alpha,\alpha_n,\beta,\gamma,\gamma_n$, we first need to describe a continuous-time branching process approximation.

1.3. **Continuous-time branching processes.** In this section, we define the limiting continuous-time branching process (CTBP) that describes the neighborhood structure of first passage percolation on $\mathrm{CM}_n(\boldsymbol{d})$. Define the size-biased distribution $F^\star$ of the random variable D with distribution function F by

$$F^\star(x)=\mathbb{E}[D\mathbb{1}_{\{D\le x\}}]/\mathbb{E}[D],\qquad x\in\mathbb{R}. \tag{1.12}$$

When Condition 1.1(a)-(b) holds, the function $F^\star$ is the weak limit as $n\to\infty$ of $F_n^\star$ in (1.7). Now let $(\mathsf{BP}^\star(t))_{t\ge0}$ denote the following CTBP: (a) At time $t=0$, we start with one individual which we refer to as the original ancestor or the root of the branching process.
(b) Each individual v in the branching process lives for a random amount of time which has distribution F_ξ, i.e., the edge weight distribution, and then dies. At the time of death the individual gives birth to $D^\star-1$ children, where $D^\star\sim F^\star$. Lifetimes and number of offspring across individuals are independent.

Note that in the above construction, by Condition 1.1(b), if we let $X_v=D^\star-1$ be the number of children of an individual, then the expected number of children satisfies

$$\mathbb{E}[X_v]=\mathbb{E}[D^\star-1]=\nu>1. \tag{1.13}$$

Further, by Condition 1.1(c), for $D^\star \sim F^\star$,

$$\mathbb{E}[D^\star \log(D^\star)+] < \infty. \tag{1.14}$$

The CTBP defined above is a *standard* Bellman-Harris process, with lifetime distribution F_ξ and offspring distributed as $D^\star - 1$ [4, 20, 23]. The Malthusian parameter α of the branching process $\mathsf{BP}^\star$ is the unique solution of the equation

$$\hat{\mu}(\alpha) = \nu \int_0^\infty \mathrm{e}^{-\alpha t} dF_\xi(t) = 1. \tag{1.15}$$

Since $\nu > 1$, we obtain that $\alpha \in (0, \infty)$. We also let α_n be the solution to (1.15) with ν replaced by

$$\nu_n = \mathbb{E}[D_n(D_n - 1)]/\mathbb{E}[D_n]. \tag{1.16}$$

Clearly, $\alpha_n \to \alpha$ when Condition 1.1 holds, and further $|\alpha_n - \alpha| = O(|\nu_n - \nu|)$.

Standard theory (see e.g., [4, 20, 23]) implies that under our assumptions on the model, namely (1.13) and (1.14), there exists a random variable $\mathcal{W}^\star$ such that

$$\mathrm{e}^{-\alpha t}|\mathsf{BP}^\star(t)| \stackrel{a.s., \mathrm{L}^1}{\longrightarrow} \mathcal{W}^\star. \tag{1.17}$$

Here the limiting random variable $\mathcal{W}^\star$ satisfies $\mathcal{W}^\star > 0$ a.s. on the event of non-extinction of the branching process and is zero otherwise. Thus α measures the true rate of exponential growth of the branching process.

By (1.15), we can define the cumulative distribution function $\bar{F}_\xi$, often referred to as the *stable-age distribution*, as

$$\bar{F}_\xi(x) = \nu \int_0^x \mathrm{e}^{-\alpha y} dF_\xi(y). \tag{1.18}$$

Let $\bar{\nu}$ be the mean and $\bar{\sigma}^2$ the variance of $\bar{F}_\xi$, i.e.,

$$\bar{\nu} = \nu \int_0^\infty x\mathrm{e}^{-\alpha x}\, dF_\xi(x), \qquad \bar{\sigma}^2 = \nu \int_0^\infty (x - \bar{\nu})^2 \mathrm{e}^{-\alpha x}\, dF_\xi(x). \tag{1.19}$$

Then $\bar{\nu}, \bar{\sigma}^2 \in (0, \infty)$, since $\alpha > 0$. We also define $\bar{F}_{n,\xi}$ to be the distribution function $\bar{F}_\xi$ in (1.18) with ν and α replaced by ν_n and α_n, and we let $\bar{\nu}_n$ and $\bar{\sigma}_n^2$ be the corresponding mean and variance.

We need a small variation of the above standard CTBP, where the root of the branching process dies immediately giving birth to a D number of children where D has distribution F, the original (i.e., non size-biased) degree distribution as in Condition 1.1(a). The details for every other individual in this branching process remain unchanged from the original description, namely each individual survives for a random amount of time with distribution F_ξ giving rise to a $D^\star - 1$ number of children where $D^\star \sim F^\star$, the size-biased distribution function $F^\star$ as in (1.12). Writing $|\mathsf{BP}(t)|$ for the number of alive individuals at time t, it is easy to see here as well that

$$\mathrm{e}^{-\alpha t}|\mathsf{BP}(t)| \stackrel{a.s., \mathrm{L}^1}{\longrightarrow} \widetilde{\mathcal{W}}. \tag{1.20}$$

Further, conditionally on $D = k$,

$$\widetilde{\mathcal{W}} \stackrel{d}{=} \widetilde{\mathcal{W}}^{\star,(1)} + \cdots + \widetilde{\mathcal{W}}^{\star,(k)},$$

where $D \sim F$, and $\widetilde{\mathcal{W}}^{\star,(i)}$ are i.i.d. with the distribution of the limiting random variable in (1.17). Let $\mathcal{W}$ denote a random variable distributed as $\widetilde{\mathcal{W}}$ conditioned to be positive, i.e., for every $x \geq 0$,

$$\mathbb{P}(\mathcal{W} \leq x) = \mathbb{P}(\widetilde{\mathcal{W}} \leq x \mid \widetilde{\mathcal{W}} > 0). \tag{1.21}$$

To simplify notation in the sequel, we will use $(\mathsf{BP}(t))_{t\geq 0}$ to denote a CTBP with the root having offspring either one (as for the standard CTBP), D or $D^\star - 1$. It will be clear from the context which setting we are in.

We are now in a position to identify the limiting random variable Q as well as the parameters $\alpha, \beta, \gamma, \alpha_n, \gamma_n$:

Theorem 1.3 (Identification of the limiting variables)**.** *The parameters $\alpha, \alpha_n, \beta, \gamma_n, \gamma$ in Theorem 1.2 satisfy that α is the Malthusian rate of growth defined in (1.15) and α_n is the solution to (1.15) with ν_n replacing ν, while*

$$\gamma = \frac{1}{\alpha\bar{\nu}}, \qquad \gamma_n = \frac{1}{\alpha_n\bar{\nu}_n}, \qquad \beta = \frac{\bar{\sigma}^2}{\bar{\nu}^3\alpha}. \tag{1.22}$$

Further, Q can be identified as

$$Q = \frac{1}{\alpha}\left(-\log \mathcal{W}^{(1)} - \log \mathcal{W}^{(2)} - \Lambda + c\right), \tag{1.23}$$

where $\mathbb{P}(\Lambda \leq x) = \mathrm{e}^{-\mathrm{e}^{-x}}$ (so that Λ is a standard Gumbel random variable), $\mathcal{W}^{(1)}, \mathcal{W}^{(2)}$ are two independent copies of the variable $\mathcal{W}$ in (1.21), also independent from Λ, and c is the constant

$$c = \log(\mathbb{E}[D](\nu-1)^2/(\nu\alpha\bar{\nu})). \tag{1.24}$$

Remark 1.4 (Asymptotic mean)**.** *In (1.11), we can replace α_n and γ_n by their limits α and γ precisely when $\gamma_n = \gamma + o(1/\sqrt{\log n})$ and $\alpha_n = \alpha + o(1/\log n)$. Since $|\alpha_n - \alpha| = O(|\nu_n - \nu|)$, $|\bar{\nu}_n - \bar{\nu}| = O(|\nu_n - \nu|)$, these conditions are equivalent to $\nu_n = \nu + o(1/\sqrt{\log n})$ and $\nu_n = \nu + o(1/\log n)$, respectively.*

Theorem 1.3 implies that also the random variable Q is remarkably universal, in the sense that it always involves two independent martingale limit variables corresponding to the branching processes, and a Gumbel distribution.

Let $L_n(i)$ denote the weight of the ith shortest path, so that $L_n = L_n(1)$, and let $H_n(i)$ denote its length. Further let $\bar{H}_n(i)$ and $\bar{L}_n(i)$ denote the re-centered and normalized quantities as in Theorem 1.2. The same proof for the optimal path easily extends to prove asymptotic results for the joint distribution of the weights and hopcount of these ranked paths. To keep the study to a manageable length, we shall skip a proof of this easy extension.

Theorem 1.5 (Multiple paths)**.** *Under the conditions of Theorem 1.2, for every $m \geq 1$,*

$$((\bar{H}_n(i), \bar{L}_n(i)))_{i\in[m]} \xrightarrow{d} ((Z_i, Q_i))_{i\in[m]}, \tag{1.25}$$

as $n \to \infty$, where for $i \in [m]$, Z_i and Q_i are independent and Z_i has a standard normal distribution, while

$$Q_i = \frac{1}{\alpha}\left(-\log \mathcal{W}^{(1)} - \log \mathcal{W}^{(2)} - \Lambda_i + c\right), \tag{1.26}$$

where $(\Lambda_i)_{i\in[m]}$ are the ordered (minimal) points of an inhomogeneous Poisson point process with intensity $\lambda(t) = \mathrm{e}^t$.

1.4. **Examples.** We treat some examples of edge-weight distributions that have appeared in the literature and have been treated via distribution-specific techniques.

(i) We start with exponential edge weights [3, 8–10]. In this case, it is immediate from (1.15) and (1.18) that
$$\alpha = \nu - 1, \qquad \bar{\nu} = \bar{\sigma} = 1/\nu,$$
hence Theorems 1.2-1.3 show that H_n converges to a normal distribution, with asymptotic mean and asymptotic variance both equal to $\frac{\nu}{\nu-1}\log n$. Furthermore, Theorem 1.2 induces the convergence of the minimal weight in [9, 10]. Observe that the random variable M, which appears in [9, (C.19)], is equal to $-\Lambda$. In [10], the special case of the Erdős-Rényi random graph with exponential edge weights was tackled. There is a small error in the expression of the limiting random variable in [10, (4.16)].

(ii) By studying weights of the form $\xi_e = 1 + E_e/k$, where (E_e) are i.i.d. exponentials with mean 1, and consecutively sending $k \to \infty$, one would expect to obtain results which are close to limiting results on the graph distance between a pair of uniformly chosen vertices in $[n]$, conditioned to be connected. Indeed, the results match up nicely with those in [17] for the Norros-Reittu model and [21] for the CM. For the sake of brevity we leave the derivation to the reader.

(iii) As a third example one can consider the CM with fixed degrees r, and where each edge is given an edge weight E^s, $s > 0$, where $E \sim \mathrm{Exp}(1)$, a variant of the weak disorder models in statistical physics [13]. One can formally consider the case $r = n-1$, although this does not satisfy the conditions of our theorem. Here the CM with fixed degrees $n-1$ resembles a complete graph on $[n]$ and the results match up nicely with those in [6], namely, a central limit theorem for H_n with asymptotic mean $s \log n$ and asymptotic variance $s^2 \log n$, while $n^s[L_n - \frac{1}{\lambda}\log n]$ converges in distribution, where $\lambda = \Gamma(1+1/s)^s$. We refer to [6] for the derivation of these parameters.

1.5. **Discussion.** In this section, we give a brief discussion of our results, possible extensions and open problems.

(a) **Universality.** Our results are universal in the sense that, as Theorems 1.2-1.3 demonstrate, the CLT for the hopcount depends only on the first two moments of the size-biased offspring distribution and on the edge-weight distribution, but not on any other property of the network model. Further, the form of the limit random variable has a universal form in terms of the martingale limits of branching processes and a Gumbel random variable.

(b) **Infinite-variance configuration model.** In [9], we have investigated the CM with exponential edge weights, but with i.i.d. degrees with $\mathbb{P}(D \geq x) \approx cx^{-(\tau-1)}$ and $\tau \in (2,3)$, so that $\mathbb{E}[D^2] = \infty$. In this case, the result for L_n is markedly different, in the sense that L_n converges in distribution *without* re-centering. Further, H_n satisfies a central limit with asymptotic mean and variance equal to a multiple of $\log n$. It would be of interest to investigate whether H_n always satisfies a central limit theorem, and, if so, whether the order of magnitude of its variance is always equal to that of its mean.

(c) **The $X \log X$-condition.** In Condition 1.1(c), we assume that the degrees satisfy a second moment condition with an additional logarithmic factor. This is equivalent to the CTBP satisfying an $X \log X$-condition (uniformly in the size n of the graph). It would be of interest to investigate what happens when this condition fails.

(d) **Flooding and diameter.** In [3], the *flooding time* and *diameter*, i.e., $\max_{j\in[n]:\, L_n(U_1,j)<\infty} L_n(U_1,j)$, respecively $\max_{i,j\in[n]:\, L_n(i,j)<\infty} L_n(i,j)$, where $L_n(i,j)$ is the minimal weight between the vertices i and j and U_1 is, as before, a randomly selected vertex, is investigated in the context of the CM with exponential edge weights. It would be of interest to investigate the flooding time for general edge weights.

(e) **Superconcentration and chaos.** Analogous to various problems in statistical physics such as random polymers or FPP on the lattice, our results suggest that the FPP optimal path problem is *chaotic.* This means that there exists $\varepsilon_n \to 0$ such that refreshing a fraction ε_n of the edge weights with new random variables with the same distribution would entirely change the actual optimal path, in the sense that the new optimal path would be "almost" disjoint of the original optimal path, see e.g. [14]. Such questions have also arisen in computer science wherein one is interested in judging the "importance" and fair price of various edges in the optimal path; if an edge being deleted causes a large change in the cost of the new optimal path, then that edge is deemed very valuable. These form the basis of various "truth and auction mechanisms" in computer science (see e.g. [5, 18, 33]).

(f) **Pandemics, gossip and other models of diffusion.** First passage percolation models as well as models using FPP as a building block have started to play an increasingly central role in the applied probability community in describing the flow of materials, ranging from viral epidemics [15], gossip algorithms [2] and more general finite Markov interchange processes [1]. Models with more general edge weight distributions have also arisen in understanding the flow of information and reconstruction of such information networks in sociology and computer science, see e.g. [30, 31] for examples in this vast field.

Organization of this paper. In Section 2, we describe the coupling between the first-passage percolation neighborhoods in $\mathrm{CM}_n(\boldsymbol{d})$ and a CTBP. In Section 3, we state our main technical result that describes a Poisson process limit for the occurrence of short paths between U_1 and U_2 which then proves our main theorem. In Section 4, we extend results for CTBPs, as proved in [20, 23, 37], to the case of infinite-variance offspring distributions, using truncation. In Section 5, we prove bounds on our coupling. In Section 6, we give a novel proof of the asymptotics for the number of alive individuals in a CTBP in an given generation and with a given residual lifetime. This proof is tailored to deal with CTBPs observed till some time t that have an offspring distribution that depends on n where $n \to \infty$ and $t = t_n \to \infty$ simultaneously. In Section 7, we prove our main technical result on the Poisson process limit.

2. Coupling

In this section, we describe a coupling between FPP on $\mathrm{CM}_n(\boldsymbol{d})$ and continuous-time branching processes. We start with an informal description.

2.1. **Informal description of shortest weight trees.** The model $\mathrm{CM}_n(\boldsymbol{d})$ with edge set $\mathcal{E}_n$ together with i.i.d. lengths (also referred to as weights) $(\xi_e)_{e\in\mathcal{E}_n}$ on the edges, was introduced in Section 1.2. Here $\xi_e \sim F_\xi$, with density f_ξ. Our ultimate goal is to calculate the limit distribution of the hopcount H_n and weight L_n of the shortest path between a uniformly chosen pair of connected vertices U_1 and U_2, when Condition 1.1 is satisfied.

To obtain a proper understanding of the shortest path between two vertices, we imagine a liquid that percolates through the edges of the CM at rate one. We start percolating the liquid simultaneously from both vertices U_1 and U_2 and we interpret the edge weight ξ_e on edge e as the distance between the two vertices incident to e. For any $t \geq 0$, the set of half-edges that are currently being wetted by the liquid, as well as the residual time to completely wet them, starting from U_i will be informally denoted as the shortest weight tree $\mathsf{SWT}^{(i)}$, $i = 1, 2$. A precise definition of these SWT's will be given in the next section. When the liquid has reached two vertices that are incident to a connecting edge between the two SWT's, then a possible shortest path has been found. Since at that moment the connecting edge has not yet been filled, we can not be sure whether the given path between U_1 and U_2 is indeed the shortest one. Hence we have to find all connecting edges between the two SWT's and take the minimum of all these path weights to determine L_n and H_n.

In the mathematical description in the next section, we build the CM simultaneously with the liquid percolating through the edges. Since we will construct the process sequentially, it is easier to index the sequence of new edge-weights added to the system as $(\xi_j)_{j\geq 1}$. The half-edges emanating from the wetted vertices are called the *alive* half-edges $\mathsf{AH}(t)$ at time t. During the building process, we form two SWT's consisting of 'alive' half-edges and vertices attached to U_i, for $i = 1, 2$. In order to perform the building process properly, we put the i.i.d. weights $(\xi_j)_{j\geq 1}$ on the half-edges instead of on the edges. Technically one has to be extremely careful in constructing the process in this fashion. Imagine a situation where the liquid reaches both a and b for some edge e formed by merging the half-edges $e = (a, b)$. Assigning independent half-edge weights ξ_a and ξ_b is then **not** the same as first passage percolation. Instead we put the weight on the half-edge that is found *first* by the liquid. We initiate the construction by putting weights $\xi_1, \ldots, \xi_{d_{U_1}}$ on the half-edges incident to U_1 and weights $\xi_{d_{U_1}+1}, \ldots, \xi_{d_{U_1}+d_{U_2}}$ on the half-edges incident to U_2. Of course, this creates a problem when these half-edges are paired to one another, which we have to take into account properly.

We construct a sequence of epoch times $(T_k)_{k\geq 0}$ that track when a decision has to be made. Start with $T_0 = 0$ and wait until the end point of the first of the $d_{U_1} + d_{U_2}$ half-edges is reached. This time is called T_1 and successive times at which further end points of 'alive' half-edges are reached are called $T_2, T_3, \ldots$. At $t = T_1$, we pair the exhausted half-edge, which we call r_1, with one of the $\ell_n - 1$ other half-edges at random; the found half-edge is called P_{r_1}. The formed edge (r_1, P_{r_1}) receives the weight of the exhausted half-edge r_1 and we connect the siblings of P_{r_1} to the newly found vertex. The sibling half-edges receive i.i.d. weights from the sequence $(\xi_j)_{j\geq 1}$, whereas the weights of the other 'alive' half-edges are updated by subtracting T_1. We repeat the whole procedure by finding the minimum of the 'alive' half-edges, and after adding this minimal weight to T_1 we find the second epoch time T_2. We continue this procedure until all half-edges are attached to one of the SWT's.

In general, the formed edge (r_1, P_{r_1}) at time $t = T_1$ receives a weight with the correct distribution. However, this only occurs when r_1 pairs with one of the $\ell_n - (d_{U_1} + d_{U_2})$ so-called 'free' half-edges, i.e., half-edges connected to vertices which are not yet wetted. When r_1 pairs with one of its $d_{U_i} - 1$ sibling half-edges (and hence a self-loop occurs) or when r_1 pairs with one of the $d_{U_{3-i}}$ half-edges incident to vertex U_{3-i} (and hence a 'collision' edge occurs), then we do not know what weight we should assign to the self-loop or collision edge, because the half-edge to which r_1 is paired is an alive half-edge and already had a weight. In order to resolve this issue, in the next section, we shall make sure that a weight is only assigned to *one* of the half-edges of an edge. This will be achieved by first investigating whether a half-edge is paired to a 'free' half-edge or to an 'alive' half-edge. In particular, we will change the order in which half-edges are paired. We are free to pair half-edges in any order we like and this property is used to remove self-loops, edges that close a cycle and collision edges beforehand. This way, all paths receive weights with the correct distribution, and after the completion of the entire construction we take the minimum of all connecting paths to find L_n and H_n.

The removal of self-loops, edges that close a cycle and collision edges is done at the epoch times $T_0, T_1, T_2, \ldots$. Conditionally on the number of 'alive' half-edges and 'free' half-edges, we know the success probabilities of Bernoulli random variables that determine whether a pairing results in attaching to a 'free' half-edge or to an 'alive' half-edge.

We will couple $\mathsf{SWT}(\cdot)$ both to an n-dependent continuous-time branching process (CTBP) denoted by $\mathsf{BP}_n(\cdot)$ and to a CTBP $\mathsf{BP}(\cdot)$ whose driving offspring distribution is the size-biased degree distribution as in (1.12) and lifetimes having distribution F_ξ. This results in a coupling $(\mathsf{SWT}(t), \mathsf{BP}_n(t), \mathsf{BP}(t))_{t\geq 0}$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathsf{SWT}(t)$ consists of the alive half-edges that are connected to U_1 and U_2 by paths of weights at most t, as well as their residual lifetimes, while $\mathsf{BP}_n(t)$ contains the same information for the n-dependent CTBP and $\mathsf{BP}(t)$ for the n-independent CTBP.

Since there are a number of ingredients in this coupling, let us start by giving the reader an intuitive mental picture of the key actors in this coupling. In the first step of the coupling, which is explained in full detail in Section 2.2, we couple the forward degrees in the $\mathsf{SWT}(t)$ to the number of offspring in the branching processes $\mathsf{BP}_n(t)$ and $\mathsf{BP}(t)$. On one and the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we introduce sequences of random variables $(B_k^{(n)})_{k\geq 1}$, $(B_k)_{k\geq 1}$, $(Y_k^{(n)})_{k\geq 1}$ and $(X_k^{(n)})_{k\geq 1}$, and a sequence of stopping times $(\tau_k)_{k\geq 1}$.

The sequences $(B_k^{(n)})_{k\geq 1}$ and $(B_k)_{k\geq 1}$ are i.i.d. and will be used as the number of offspring in a branching process, where the first branching process depends on n, while the second one is independent of n. In the coupling, there is a strong dependence between $B_k^{(n)}$ and B_k for any k. The sequence $(Y_k^{(n)})_{k\geq 1}$ will correspond to the sequence of forward degrees, i.e., the degree minus one, as the liquid in SWT percolates through the graph, while $X_k^{(n)}$ will be equal to $Y_k^{(n)}$ minus the number of pairings that result in either a self-loop, a cycle or a collision edge. Using the stopping times τ_k, the kth variable $X_k^{(n)}$ will be successfully coupled to $B_{\tau_k}^{(n)}$ precisely when $X_k^{(n)} = Y_k^{(n)}$ and $Y_k^{(n)} = B_{\tau_k}^{(n)}$.

Since this coupling is not perfect, in the second step of the coupling performed in Section 2.3, we discuss the evolution of the processes $(\mathsf{SWT}(t), \mathsf{BP}_n(t), \mathsf{BP}(t))_{t\geq 0}$, including the evolution of the children of the alive half-edges that are miscoupled. Finally, in Section 2.4, we state the main bounds on our coupling. After this high-level explanation, let us now give the details of our coupling construction.

2.2. **Coupling forward degrees of the SWT.** Before we start with the definition of the SWT's on one probability space, we introduce an abstract procedure involving two sets, and a recursive sequence of samples drawn without replacement. This procedure is used repeatedly during the building of the SWT's to remove occurring self-loops, edges that close a cycle and collision edges. The idea is as follows. Consider a partition of a set $[m]$ into A and $B = A^c$. One can achieve a uniform draw from $[m]$ in two steps, first by performing a Bernoulli experiment with success probability $p_A = |A|/m$; if the outcome of this experiment is one, then we draw an object uniformly from A, otherwise we draw a uniform object from B. In fact, we do not even have to actually draw the latter uniform element. We will perform such a construction repeatedly, with A denoting the set of alive half-edges at appropriate stopping times and where the set $[m]$ is recursively defined. When doing so, we can think of this as 'testing' whether a half-edge creates a self-loop, cycle or collision edge, or whether it connects to a 'free' half-edge. This is formalized in the following procedure. We first need to set up some notation.

The procedure takes as its input one non-empty set, the "alive" set $\mathsf{AS} = \{a_1, a_2, \ldots, a_s\}$ having s elements, and the size of the "free" set N. The elements $a_j \in \mathsf{AS}$, $j = 1, 2, \ldots, r$ for $r \leq s$ are special. We will view them as a list (namely an ordered set) $\mathsf{TS} = (a_1, a_2, \ldots, a_r)$, of r elements. Abusing notation, the procedure initializes with $\mathsf{AS}(0) := \mathsf{AS}$, $\mathsf{TS}(0) := \mathsf{TS}$ and $\mathsf{N}(0) := \mathsf{N}$ and we will sequentially update these sets and this number as follows using a sequence of conditionally independent Bernoulli random variables $(\beta_i)_{1\leq i\leq r}$ and a sequence of sets $(\mathsf{S}_i)_{i\geq 1}$:

Procedure 2.1 (Preprocessing the matchings)**.** (a) **Initialization:** *Define the success probability and set* S_1 *as*

$$p_1 = \frac{|\mathsf{AS}(0)| - 1}{|\mathsf{AS}(0)| + \mathsf{N}(0) - 1} \qquad \text{and} \qquad \mathsf{S}_1 = \mathsf{AS} \setminus \{a_1\} = \{a_2, \ldots, a_s\}. \tag{2.1}$$

Let $\beta_1 \sim \mathrm{Bernoulli}(p_1)$*.*

(i) *If* $\beta_1 = 1$*, then select element* b_1 *uniformly at random from the set* S_1 *and update the sets as* $\mathsf{AS}(1) = \mathsf{AS}(0) \setminus \{a_1, b_1\}$ *and if* $b_1 \in \mathsf{TS}(0)$ *then* $\mathsf{TS}(1) = \mathsf{TS}(0) \setminus \{a_1, b_1\}$*, else* $\mathsf{TS}(1) = \mathsf{TS}(1) \setminus \{a_1\}$*. We do not change* $\mathsf{N}(1) := \mathsf{N}(0)$*.*

(ii) *If* $\beta_1 = 0$*, then we do not select any element from* S_1 *and we update the sets as* $\mathsf{AS}(1) = \mathsf{AS}(0) \setminus \{a_1\}$ *and* $\mathsf{TS}(1) = \mathsf{TS}(0) \setminus \{a_1\}$*, while* $\mathsf{N}(1) = \mathsf{N}(0) - 1$*.*

(b) **Recursion:** *For* $k \geq 1$ *we proceed recursively as above, taking the first element at the front of the list* $\mathsf{TS}(k)$ *which, abusing notation, we still call* a_{k+1} *(note, this element may not be* a_{k+1} *in* $\mathsf{TS}(0)$*, if that element is already drawn at a previous time step) now defining*

$$p_{k+1} = \frac{|\mathsf{AS}(k)| - 1}{|\mathsf{AS}(k)| + \mathsf{N}(k) - 1} \qquad \text{and} \qquad \mathsf{S}_{k+1} = \mathsf{AS}(k) \setminus \{a_k\}, \tag{2.2}$$

generating $\beta_{k+1} \sim \mathrm{Bernoulli}(p_{k+1})$ *and proceeding as in* (i) *and* (ii) *above, with* a_1, b_1, S_1 *replaced by* $a_{k+1}, b_{k+1}, \mathsf{S}_{k+1}$*, respectively.*

(c) **Termination:** *We stop when the list equals the empty set, i.e.,* $\mathsf{TS}(k) = \varnothing$*.*

Let us give a brief and informal sketch of how we use Procedure 2.1. Consider the infomal description in Section 2.1 of the liquid percolating through a network started from two vertices U_1 and U_2 simultaneously. Assume that at some time t, the liquid from vertex U_i, where $i = 1, 2$, hits a new vertex V. The set of half-edges incident to V, called HE_V, except the one used by V to connect to $\mathsf{SWT}^{(i)}$, are now deemed active since the flow has encountered this new vertex V. Write $\mathsf{AS} = \mathsf{AH}^{(1)}(t) \cup \mathsf{AH}^{(2)}(t) \cup \mathsf{HE}_V$ for the collection of alive half-edges at this time, and $\mathsf{TS} = (\mathsf{HE}_V)$ for the set of half edges incident to V. Then the above procedure tests for each one of these newly added half-edges whether it pairs to a half-edge in $\mathsf{AH}(t)$, which corresponds to the "$\beta_k = 1$" events, or instead connects to a new half-edge not in $\mathsf{SWT}^{(1)}(t) \cup \mathsf{SWT}^{(2)}(t)$, which corresponds to the "$\beta_k = 0$" events. In the latter case, we actually do *not* connect the half-edge, but only record that the half-edge is paired to a free half-edge and thus decrease the number of free half-edges $\mathsf{N}(k)$ by 1. Further note that in each "$\beta_k = 1$" event, the new edge created could either be (a) a self-loop or cycle when the half-edge pairs to an alive half-edge in $\mathsf{SWT}^{(i)}(t)$, namely the same cluster that sees V for the first time or (b) arguably more importantly, creates a collision edge when it is paired to a half edge in $\mathsf{SWT}^{(3-i)}(t)$, the other cluster. These collision edges are the ones that potentially create the shortest path.

Let us now turn to the precise definition of the probability space for the coupling of the forward degrees of SWT and the associated branching processes. We start on one and the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the following ingredients:

(i) Two vertices U_1 and U_2 chosen at random from $[n]$;

(ii) Label the ℓ_n half-edges by $[\ell_n]$ with the half-edges of vertex 1 labelled $1, 2, \ldots d_1$, the half-edges of vertex 2 labelled $d_1 + 1, \ldots, d_1 + d_2$, etc. We will require repeated draws **with** replacement from $[\ell_n]$, which results in an i.i.d. sequence $(\sigma_i)_{i\geq 1}$. We will also require a second sequence of i.i.d. draws $(\check{\sigma}_i)_{i\geq 1}$ that is independent of the draws $(\sigma_i)_{i\geq 1}$;

(iii) To each σ_i and $\check{\sigma}_i$, we associate random variables $B_i^{(n)}$ and $\check{B}_i^{(n)}$ that correspond to the forward degree of the vertices incident to the half-edges σ_i and $\check{\sigma}_i$. To each $B_i^{(n)}$ and $\check{B}_i^{(n)}$, we associate random variables B_i and $\check{B}_i$, whose distributions only depend on $B_i^{(n)}$ and $\check{B}_i^{(n)}$ and not on any of the other randomness involved;

(iv) An i.i.d. sequence $(\xi_i)_{i\geq 1}$ of edge-weights with distribution F_ξ, and a second sequence of i.i.d. weights $(\check{\xi}_i)_{i\geq 1}$ that is independent of the edge-weights $(\xi_i)_{i\geq 1}$;

(v) Recall that vertex $j \in [n]$ had degree d_j. Recall the uniform choices $b_1, b_2, \ldots$ in Procedure 2.1 modulated by the values of the Bernoulli sequence $\beta_1, \beta_2, \ldots$. Analogously, we construct random variables $(b_j(1), \ldots, b_j(d_j))$, taking values in the set of half-edges $[\ell_n]$ modulated by a sequence of Bernoulli random variables $\beta_j(1), \ldots, \beta_j(d_j)$. The distribution of both these random variables will depend on $\big(U_1, U_2, (\sigma_i)_{i\geq 1}, (\xi_i)_{i\geq 1}\big)$. The precise laws of these ingredients will be specified as we sequentially apply Procedure 2.1 below.

Before using the above ingredients to construct SWT, let us first describe how they are used to construct the offspring of the branching processes BP_n and BP. We define, for $i \geq 1$,

$$B_i^{(n)} = \sum_{j=1}^{n} (d_j - 1)\mathbb{1}_{\{d_1+\ldots+d_{j-1}<\sigma_i\leq d_1+\ldots+d_j\}}, \tag{2.3}$$

i.e., when σ_i chooses one of the half-edges incident to vertex j, $B_i^{(n)}$ is the forward degree (i.e., degree minus one) of that vertex j. Obviously, the sequence $(B_i^{(n)})_{i\geq 1}$ is i.i.d. with common distribution given by

$$g_k^{(n)} = \mathbb{P}(B_i^{(n)} = k) = \frac{k+1}{\ell_n} \sum_{j=1}^{n} \mathbb{1}_{\{d_j=k+1\}}, \qquad k \geq 0. \tag{2.4}$$

Note that $g_k^{(n)} = \mathbb{P}(D_n^\star = k+1)$, where $D_n^\star$ has the same distribution as the size-biased version of D_n, the degree of a randomly selected vertex, see (1.7). Assuming Condition 1.1, we have that $g_k^{(n)} \to g_k$, as $n \to \infty$, where

$$g_k = \frac{(k+1)f_{k+1}}{\sum_{j=1}^{\infty} j f_j}, \qquad k \geq 0, \tag{2.5}$$

and where $f_j = F(j) - F(j-1)$.

We next construct an i.i.d. sequence $(B_i)_{i\geq 1}$ with common distribution (2.5) by using the already constructed $(B_i^{(n)})_{i\geq 1}$ sequence as follows: For each $i \geq 1$, B_i depends only on $B_i^{(n)}$ and is generated via the conditional distribution

$$\mathbb{P}(B_i = k \mid B_i^{(n)} = l) = \frac{p_{kl}^{(n)}}{\sum_{j=0}^{\infty} p_{jl}^{(n)}}, \tag{2.6}$$

where

$$p_{kl}^{(n)} = \begin{cases} \min\{g_k, g_k^{(n)}\}, & \text{for } k = l, \\ \dfrac{(g_k - \min\{g_k, g_k^{(n)}\})(g_l^{(n)} - \min\{g_l, g_l^{(n)}\})}{\frac{1}{2}\sum_{j=0}^{\infty} |g_j - g_j^{(n)}|}, & \text{for } k \neq l. \end{cases} \tag{2.7}$$

It is easy to check that $(B_i)_{i\geq 1}$ is an i.i.d. sequence of random variables having probability mass function $(g_k)_{k\geq 0}$ in (2.5). In fact, the joint distribution $(p_{kl}^{(n)})_{k,l\geq 0}$ is the one that maximizes the coupling probability between the two probability mass functions $(g_k)_{k\geq 0}$ and $(g_k^{(n)})_{k\geq 0}$ (alternatively, the coupling that minimizes the total variation distance between the two distributions [39]).

Let us now proceed to the more involved construction of the shortest weight tree SWT using the above probabilistic ingredients. The main ingredient of our construction are the continuous-time processes of 'alive' half-edges $(\mathsf{AH}(t))_{t\geq 0}$ and 'free' half-edges $(\mathsf{F}(t))_{t\geq 0}$. We also introduce two new random sequences $(Y_k^{(n)})_{k\geq 1}$ and $(X_k^{(n)})_{k\geq 0}$. We will need an additional superscript i to denote whether $Y_k^{(n)}$ and/or $X_k^{(n)}$ belongs to the SWT of $U_i, i=1,2$. The continuous-time processes $(\mathsf{AH}(t))_{t\geq 0}$ and $(\mathsf{F}(t))_{t\geq 0}$ only change at random times $T_0 = 0 < T_1 < T_2 < \ldots$ and therefore a full description of the continuous-time evolution can be given by a specification of how the random times above are constructed and how these processes "jump" at each of these times.

At time $t = T_0$, we start by testing whether any of the half-edges d_{U_i} incident to $U_i, i = 1,2$, are paired to one another. This is performed vertex by vertex, and we start with U_1. Let us define HE_j, for $j \in [n]$, as the set of d_j half-edges that belong to vertex j. We define $Y_0^{(n,1)}$ as the number of half-edges incident to U_1, i.e.,

$$Y_0^{(n,1)} = d_{U_1} = |\mathsf{HE}_{U_1}|. \tag{2.8}$$

Now put $\mathsf{AS} = \mathsf{HE}_{U_1}$, $\mathsf{TS} = (\mathsf{AS})$, where the parentheses $(\cdot)$ indicate that we consider a list instead of a set, and $\mathsf{N} = \ell_n - d_{U_1}$, and apply Procedure 2.1 to remove all half-edges from the total set of d_{U_1} half-edges that are part of a self-loop. We then define RHE_{U_1} as the set of unpaired half-edges after the self-loops incident to U_1 are removed and

$$X_0^{(n,1)} = |\mathsf{RHE}_{U_1}|, \tag{2.9}$$

as the number of unpaired half-edges of U_1 after the self-loops have been removed. We attach i.i.d. weights to each of the half-edges in RHE_{U_1} by taking the first $X_0^{(n,1)}$ weights from $(\xi_i)_{i\geq 1}$.

We continue with the d_{U_2} half-edges incident to U_2, and test whether they are paired to one of the $X_0^{(n,1)}$ remaining half-edges incident to U_1, or any of the d_{U_2} half-edges incident to U_2. We do this by applying Procedure 2.1 with $\mathsf{AS} = \mathsf{RHE}_{U_1} \cup \mathsf{HE}_{U_2}$, $\mathsf{TS} = (\mathsf{HE}_{U_2})$ and

$$\mathsf{N} = \ell_n - d_{U_1} - d_{U_2} - X_0^{(n,1)}, \tag{2.10}$$

which equals the total number of half-edges that are still available to be connected to (noting that the ones that are paired to the half-edges incident to U_1 are no longer available). A self-loop is formed when during this test a half-edge is paired to one of the d_{U_2} sibling half-edges. A so-called *collision edge* is formed when during this test a half-edge is paired to one of the $X_0^{(n,1)}$ remaining half-edges incident to vertex U_1. The weight of this collision edge is the weight of the half-edge incident to U_1, which it has already obtained in the previous step. A collision produces a path between vertices U_1 and U_2, which possibly is the minimal weight path between U_1 and U_2. We define RHE_{U_2} as the set of unpaired half-edges incident to vertex U_2 after the removal of the self-loops and collision edges. Furthermore, we define

$$Y_0^{(n,2)} = d_{U_2}, \tag{2.11}$$

and

$$X_0^{(n,2)} = |\mathsf{RHE}_{U_2}|, \tag{2.12}$$

i.e., $X_0^{(n,2)}$ denotes the number of unpaired half-edges of vertex U_2 after the test for collision edges and self-loops has been performed. We attach i.i.d. weights to each of the half-edges in RHE_{U_2} by taking the first $X_0^{(n,2)}$ available weights from the i.i.d. sequence $(\xi_i)_{i\geq 1}$ (note that the first $X_0^{(n,1)}$ weights have already been assigned to the half edges in RHE_{U_1}). By construction, the remaining $X_0^{(n,1)} + X_0^{(n,2)}$ half-edges incident to the vertices U_1, U_2 are paired to *fresh* vertices, i.e., vertices distinct from U_1 and U_2.

For the moment, we collect the possible collision edges at time T_0, together with their weights, which is equal to the weight of the half-edge incident to U_1 that forms one half of the collision edge, and continue with the description. All half-edges that are not paired to one of the other $d_{U_1} + d_{U_2} - 1$ half-edges incident to either U_1 or U_2 together form the set $\mathsf{AH}(0)$, the set of active half-edges at time 0, i.e.,

$$\mathsf{AH}(0) = \mathsf{RHE}_{U_1} \cup \mathsf{RHE}_{U_2}. \tag{2.13}$$

For $y \in \mathsf{AH}(0)$, we define the height $H(y) = 1$ and its index $I(y) = i$, if the half-edge y is connected to U_i, $i = 1, 2$, and $R_0(y)$ for the weight with distribution F_ξ that the half edge received earlier. This initiates the construction with

$$\mathsf{SWT}(0) = (y, H(y), I(y), R_0(y))_{y \in \mathsf{AH}(0)}, \tag{2.14}$$

and we let $\mathsf{SWT}^{(i)}(0) = (y, H(y), R_0(y))_{y \in \mathsf{AH}(0), I(y)=i}$, be the subset of $\mathsf{SWT}(0)$ that is connected to vertex U_i, $i = 1, 2$.

After the above initialization, let us now describe how to construct the process $\mathsf{SWT}(t) = \mathsf{SWT}^{(1)}(t) \cup \mathsf{SWT}^{(2)}(t)$ and $\mathsf{SWT}^{(i)}$ with $i = 1, 2$, for general $t > 0$. Abusing notation, we call $\mathsf{SWT}^{(i)}(t)$ the shortest path tree emanating from vertex U_i. Using the information $(\mathsf{SWT}^{(i)}(s))_{0 \le s \le t}$, we can construct the genealogical tree representing how the liquid percolates from the source U_i but this process contains much more information including edge lengths encountered by the process. As for $t = 0$, the process $\mathsf{SWT}(t)$ has a set of 'alive' half-edges $\mathsf{AH}(t)$, which we formally define below. For $y \in \mathsf{AH}(t)$, we record its index $I(y) \in \{1, 2\}$ if $y \in \mathsf{SWT}^{(I(y))}$ and we let $H(y)$ denote the graph distance of y to $U_{I(y)}$ (viewing $(\mathsf{SWT}^{(i)}(s))_{s \in [0,t]}$ as a tree). Further, for $y \in \mathsf{AH}(t)$, we let $R_t(y)$ denote the residual lifetime of y at time t. Then, we let

$$\mathsf{SWT}(t) = (y, H(y), I(y), R_t(y))_{y \in \mathsf{AH}(t)} \tag{2.15}$$

denote the set of alive half-edges together with their indices, their heights and residual lifetimes. At a later state, we will also define $\mathsf{BP}_n(t)$ and $\mathsf{BP}(t)$, the CTBP analogs of $\mathsf{SWT}(t)$.

We next recursively define the evolution of $(\mathsf{SWT}(t))_{t \ge 0}$. Define $T_1 = \min_{y \in \mathsf{AH}(0)} R_0(y)$ and denote the half-edge equal to the argument of this minimum by $y_0^\star$, hence $R_0(y_0^\star) = \min_{y \in \mathsf{AH}(0)} R_0(y)$. Since the distribution of the weights (lifetimes) admits a density f_ξ, $y_0^\star$ is a.s. *unique*. Now set

$$\mathsf{AH}(t) = \mathsf{AH}(0), \qquad 0 \le t < T_1, \tag{2.16}$$

i.e., the active set remains unchanged during the interval $[0, T_1)$. This defines the shortest weight tree in (2.15) for $0 \le t < T_1$, where $I(y)$ and $H(y)$ are defined above and $R_t(y) = R_0(y) - t$, $0 \le t < T_1$, denotes the remaining lifetime of half-edge y at time t.

At time $t = T_1$, we continue by describing the pairing of the half-edge $y_0^\star$ with $z_0 = P_{y_0^\star}$ and at this place we will introduce the coupling between $Y_1^{(n)}$ and $B_1^{(n)}$ (see (2.3)). For a half-edge y, let V_y denote the vertex incident to it. By construction, $z_0 = P_{y_0^\star}$ is chosen such that V_{z_0} is not equal to U_i, $i = 1, 2$. This is achieved by taking

$$\tau_1 = \min\{m \ge 1 \colon V_{\sigma_m} \neq U_1, V_{\sigma_m} \neq U_2\}, \tag{2.17}$$

and we define

$$Y_1^{(n)} = B_{\tau_1}^{(n)} \qquad \text{and} \qquad z_0 = \sigma_{\tau_1}. \tag{2.18}$$

When $\tau_1 = 1$, we see that $Y_1^{(n)} = B_1^{(n)}$, while when $\tau_1 > 1$, the forward degree $Y_1^{(n)}$ of the chosen vertex V_{z_0} is not successfully coupled to the random variable $B_1^{(n)}$.

At time $t = T_1$, we remove $y_0^\star$ from the set $\mathsf{AH}(t-)$. Then, for each of the $d_{V_{z_0}} - 1$ other half-edges incident to vertex V_{z_0} we test, using Procedure 2.1, with

$$\mathsf{AS} = \mathsf{AH}(t-) \cup (\mathsf{HE}_{V_{z_0}} \setminus \{z_0\}), \qquad \mathsf{TS} = (\mathsf{HE}_{V_{z_0}} \setminus \{z_0\}) \tag{2.19}$$

and

$$\mathsf{N} = \ell_n - d_{U_1} - d_{U_2} - d_{V_{z_0}} - |\mathsf{AH}(0)|, \tag{2.20}$$

which again has the interpretation of the number of available half-edges at the time of finding V_{z_0}, whether it is part of a self-loop or paired to a half-edge from the set $\mathsf{AH}(t-)$. All half-edges incident to V_{z_0} that are part of a self-loop or incident to $\mathsf{AH}(t-)$ are removed from vertex V_{z_0}; we also remove the involved half-edges from the set $\mathsf{AH}(t-)$. For all the remaining sibling half-edges x of z_0 we do the following: x is added to $\mathsf{AH}(t-)$, $I(x) = I(y_0^\star)$, $H(x) = H(y_0^\star) + 1$, while $R_{T_1}(x)$ is the next available i.i.d. lifetime from the sequence $(\xi_i)_{i \ge 1}$. We now set

$$\mathsf{AH}(t) = \mathsf{AH}(T_1-), \qquad T_1 \le t < T_2,$$

where $T_2 = T_1 + \min_{y \in \mathsf{AH}(T_1)} R_{T_1}(y)$, and where the minimizing half-edge is called $y_1^\star$.

We continue using induction, by defining $\mathsf{AH}(t)$ and $\mathsf{SWT}(t)$ during the random interval $[T_k, T_{k+1})$ for $k \geq 1$, given that the processes are defined on $[0, T_k)$. By construction, we know that $z_{k-1} = P_{y^\star_{k-1}}$ is chosen such that $V_{z_{k-1}}$ is not equal to U_i, $i = 1, 2$ or one of the previously chosen vertices $V_{z_j}, 1 \leq j \leq k-2$ (for $k = 1$, the latter is an empty condition). Therefore, we take

$$\tau_k = \min\big\{m \geq \tau_{k-1} + 1 \colon V_{\sigma_m} \notin \{U_1, U_2, V_{z_0}, \dots, V_{z_{k-2}}\}\big\}, \tag{2.21}$$

and we define

$$Y_k^{(n)} = B_{\tau_k}^{(n)} \qquad \text{and} \qquad z_{k-1} = \sigma_{\tau_k}. \tag{2.22}$$

When $\tau_k = \tau_{k-1} + 1$, we see that $Y_k^{(n)} = B^{(n)}_{\tau_{k-1}+1}$, while for $\tau_k > \tau_{k-1} + 1$, the forward degree $Y_k^{(n)}$ of the chosen vertex $V_{z_{k-1}}$ is not coupled to the random variable $B^{(n)}_{\tau_{k-1}+1}$ and we call the vertex $V_{z_{k-1}}$ *degree-miscoupled.* At time $t = T_k$, we remove $y^\star_{k-1}$ from the set $\mathsf{AH}(t-)$. Then, for each of the $d_{V_{z_{k-1}}} - 1$ other half-edges incident to vertex $V_{z_{k-1}}$, we use Procedure 2.1, with

$$\mathsf{AS} = \mathsf{AH}(t-) \cup (\mathsf{HE}_{V_{z_{k-1}}} \setminus \{z_{k-1}\}), \qquad \mathsf{TS} = (\mathsf{HE}_{V_{z_{k-1}}} \setminus \{z_{k-1}\}), \tag{2.23}$$

and

$$\mathsf{N} = \ell_n - d_{U_1} - d_{U_2} - \sum_{j=0}^{k-1} d_{V_{z_j}} - |\mathsf{AH}(T_{k-1})|, \tag{2.24}$$

to test whether it is part of a self-loop or paired to a half-edge from the set $\mathsf{AH}(t-)$. It is part of Procedure 2.1 that all half-edges incident to $V_{z_{k-1}}$ that are part of a self-loop or incident to $\mathsf{AH}(t-)$ are removed from vertex $V_{z_{k-1}}$; we also remove the involved half-edges from the set $\mathsf{AH}(t-)$. We will discuss the role of the half-edges incident to $V_{z_{k-1}}$ that are paired to half-edges in $\mathsf{AH}(t-)$ in more detail below.

We sequentially order the remaining siblings half-edges of z_{k-1} in an arbitrary order. In this order, we do the following: Let x be one such half-edge of $V_{z_{k-1}}$, then add x to $\mathsf{AH}(t-)$, and set $I(x) = I(y^\star_{k-1})$ and $H(x) = H(y^\star_{k-1}) + 1$, while $R_{T_k}(x)$ is the next in line of the i.i.d. sequence $(\xi_i)_{i\geq 1}$. We now set

$$\mathsf{AH}(t) = \mathsf{AH}(T_k-), \qquad T_k \leq t < T_{k+1}, \tag{2.25}$$

where $T_{k+1} = T_k + \min_{y \in \mathsf{AH}(T_k)} R_{T_k}(y)$, and where the minimizing half-edge is called $y^\star_k$.

For $t \in [T_k, T_{k+1})$, we define $\mathsf{SWT}(t)$ by (2.15), where $R_t(y) = R_{T_k}(y) - (t - T_k)$. Finally, we denote the number of the $d_{V_{z_{k-1}}} - 1$ other half-edges incident to vertex $V_{z_{k-1}}$ that do not form a self-loop and that are not paired to a half-edge from the set $\mathsf{AH}(t-)$ by $X_k^{(n)}$. We say that the vertex $V_{z_{k-1}}$ is successfully degree-coupled to the corresponding individual in a branching process that has offspring $B^{(n)}_{\tau_{k-1}+1}$ (this will show up again in the next section) when **both**

$$Y_k^{(n)} = B^{(n)}_{\tau_{k-1}+1} \qquad \text{and} \qquad X_k^{(n)} = Y_k^{(n)}, \tag{2.26}$$

and otherwise call is degree-miscoupled.

We finally denote $S_k^{(n)} = |\mathsf{AH}(T_k)|$, so that $S_0^{(n)} = X_0^{(n,1)} + X_0^{(n,2)}$, while S_k satisfies the recursion

$$S_k^{(n)} = S_{k-1}^{(n)} + X_k^{(n)} - 1. \tag{2.27}$$

This describes the evolution of $(\mathsf{SWT}(t))_{t\geq 0}$.

Cycle edges and collision edges. At time T_k, $k \geq 1$, we find the half-edge $y^\star_{k-1}$ that is paired to $z_{k-1} = P_{y^\star_{k-1}}$, and for each of the other half-edges x incident to $V_{z_{k-1}}$, we check, using Procedure 2.1, whether or not a self-loop has been formed or whether or not $P_x \in \mathsf{AH}(T_k-)$. The newly found half-edges that are paired to already alive half-edges in $\mathsf{AH}(T_k-)$ are special. Indeed, the edge (x, P_x) creates a cycle when $I(x) = I(P_x)$, while (x, P_x) completes a path between U_1 and U_2 when $I(x) = 3 - I(P_x)$. Precisely the latter edges can create the shortest-weight path between U_1, U_2. Let us describe these collision edges in more detail.

At time T_k and when we create a collision edge consisting of x and P_x, then we record

$$\Big(T_k, I(x), H(x), H(P_x), R_{T_k}(P_x)\Big), \tag{2.28}$$

where

$$I(x) = I(y^\star_{k-1}), \qquad H(x) = H(y^\star_{k-1}) + 1. \tag{2.29}$$

Order the times at which collision edges occur as $(T_j^{(\mathrm{col})})_{j\geq 1}$, and let (x_j, P_{x_j}) be the corresponding collision edge (so that P_{x_j} is in the other SWT as x_j). If multiple collision edges are created at the same time, then order them arbitrarily. We will see that the probability of such events in the time scale of interest converges to zero as $n\to\infty$. Write

$$\mathscr{C} := \Big(T_j^{(\mathrm{col})}, I(x_j), H(x_j), H(P_{x_j}), R_{T_j^{(\mathrm{col})}}(P_{x_j})\big)_{j\geq 1}, \tag{2.30}$$

for the collection of all collision edges collected by the process.

It is possible (albeit unlikely) that multiple half-edges incident to $V_{z_{k-1}}$ create collision edges, and if so, we collect all of them in the list in (2.30). With some abuse of notation we denote the jth collision edge by (x_j, P_{x_j}); here P_{x_j} is an alive half-edge and x_j the half-edge which pairs to P_{x_j}. Note that, at the time t of creation of the collision edge, the weight of the half-edge has already been assigned to the half-edge P_{x_j}, and the half-edge P_{x_j} has residual lifetime equal to $R_t(P_{x_j})$.

The weight of the (unique) path between U_1 and U_2 that passes through the edge (x_j, P_{x_j}) equals $2T_j^{(\mathrm{col})} + R_{T_j^{(\mathrm{col})}}(P_{x_j})$ and its hopcount is equal to $H(x_j)+H(P_{x_j})+1$, so that the shortest weight equals

$$L_n = \min_{j\geq 1}[2T_j^{(\mathrm{col})} + R_{T_j^{(\mathrm{col})}}(P_{x_j})]. \tag{2.31}$$

Let $I^\star$ denote the minimizer of $j\mapsto 2T_j^{(\mathrm{col})} + R_{T_j^{(\mathrm{col})}}(P_{x_j})$, then

$$H_n = H(x_{I^\star}) + H(P_{x_{I^\star}}) + 1. \tag{2.32}$$

Of course, (2.31) and (2.32) need a proof, which we give now:

Proof that L_n in (2.31), and H_n in (2.32) yield the minimal weight and hopcount, respectively. Observe that the weight of each path between U_1 and U_2 with weight L can be written in the form $L = 2T_i^{(\mathrm{col})} + R_{T_i^{(\mathrm{col})}}(P_{x_i})$, for some $i\geq 0$. Indeed, let $(i_0 = U_1, i_1, i_2, \ldots, i_k = U_2)$ form a path with weight L, and denote the weight on (i_{j-1}, i_j) by ξ_{e_j}, for $1\leq j\leq k$. For $k=1$, we obviously find $\xi_{e_1} = 2T_0 + \xi_{e_1}$. For general $k\geq 1$, take the maximal $j\geq 0$ such that $\xi_{e_1}+\cdots+\xi_{e_j}\leq L/2$. Then, since $L=\sum_{s=1}^k \xi_{e_s}$, we have that $\sum_{s=1}^j \xi_{e_s} \leq \sum_{s=j+1}^k \xi_{e_s}$, so that

$$L = 2\sum_{s=1}^j \xi_{e_s} + [\sum_{s=j+1}^k \xi_{e_s} - \sum_{s=1}^j \xi_{e_s}],$$

which is of the form $L = 2T_j^{(\mathrm{col})} + R_{T_j^{(\mathrm{col})}}(y)$, for some $j\geq 0$ and some half-edge y. Note that in the construction of the SWT's, instead of putting weight on the edges, we have given weights to half-edges instead. In the representation (2.31) full edge weight is given to the active half-edges and weight 0 to the ones to which they are paired. At time $T_j^{(\mathrm{col})}$ when a collision edge has been found, the path-weight of the edges belonging to the same vertex U_i as half-edge y^* add up to $T_j^{(\mathrm{col})}$, the path-weight of all completed *edges* connected to $3-U_i$ together with the residual lifetime $R_{T_j^{(\mathrm{col})}}(P_x)$ of the half-edge P_x has to be added to $T_j^{(\mathrm{col})}$ in order to yield the total weight of the path between U_1 and U_2.

The proof of (2.32) follows because the number of edges of the path between U_1 and U_2 that passes through the collision edge (x_j, P_{x_j}) is equal to the sum of the heights of the vertices incident to x_j, P_{x_j}, respectively, and we add 1 for the edge (x_j, P_{x_j}) itself. This completes the proof of the claim. ∎

2.3. **Coupling: Process level.** In the above, we have described the coupling between reduced forward degrees $(X_i^{(n)})_{i\geq 1}$ in SWT and i.i.d. random variables $((B_i^{(n)}, B_i))_{i\geq 1}$, where $(B_i^{(n)})_{i\geq 1}$ has marginal distribution (2.4) and $(B_i)_{i\geq 1}$ has marginal distribution (2.5), and they are coupled as in (2.7). We have used this coupling to describe the evolution of $(\mathsf{SWT}(t))_{t\geq 0}$, and at the end of this process, we know of each vertex that is found by the liquid, whether it is successfully degree-coupled or not. As long as no degree-miscouplings occur, this can be thought of as a coupling between SWT and two CTBPs with lifetimes having distribution F_ξ and offsprings $(B_i^{(n)})_{i\geq 1}$ and $(B_i)_{i\geq 1}$, respectively, but the evolutions will start to diverge as soon as degree-miscouplings start to appear. We now extend this coupling.

Recall that the $\mathsf{SWT}^{(i)}$ consists of half-edges and their attributes, connected to U_i, for $i = 1, 2$. We aim to couple each $\mathsf{SWT}^{(i)}$, $i = 1, 2$, to an independent CTBP $\mathsf{BP}_n^{(i)}$, so that SWT is coupled to BP_n which consists of two independent CTBPs, i.e., $\mathsf{BP}_n = (\mathsf{BP}_n^{(1)}, \mathsf{BP}_n^{(2)})$, as well as to an n-independent limiting CTBP BP that also consists of two independent CTBPs, i.e., $\mathsf{BP} = (\mathsf{BP}^{(1)}, \mathsf{BP}^{(2)})$. If $Y_k^{(n)}$ is the forward degree of vertex $V_{z_{k-1}}$, then $I(y_{k-1}^\star)$ indicates to which SWT the vertex belongs. We recall that we say that the vertex $V_{z_{k-1}}$ in the SWT is *degree-miscoupled* to the corresponding individual (which we also refer to as vertex) if

$$Y_k^{(n)} \neq B_{\tau_{k-1}+1}^{(n)} \qquad \text{or if} \qquad X_k^{(n)} \neq Y_k^{(n)}. \tag{2.33}$$

Vertices that are degree-miscoupled will appear both in the SWT as well as in the CTBP BP_n. However, after being degree-miscoupled, the evolution of vertices in the CTBP and SWT diverge, as we explain now. For the SWT, we say that an alive half-edge is *miscoupled* if the shortest-weight path to the vertex incident to that half-edge uses at least one degree-miscoupled vertex. In particular, the evolution of the SWT is such that half-edges of degree-miscoupled vertices are by definition attached to miscoupled half-edges. The same is true for the CTBP BP_n, that is, offspring of degree-miscoupled individuals are by definition miscoupled.

The weights assigned to half-edges incident to miscoupled vertices in the SWT and individuals in the CTBP are independent. For this, we have introduced a *second* sequence of i.i.d. weights $(\check{\xi}_i)_{i\geq 1}$ that is independent of the edge-weights of correctly coupled half-edges $(\xi_i)_{i\geq 1}$. Each time that a half-edge is found by the SWT, we perform Procedure 2.1 and the coupling to the CTBPs BP_n and BP. When, instead, a half-edge is found by (one of the) CTBPs only, we pair it to a uniformly chosen half-edge chosen from $[\ell_n]$ without replacement. These choices are determined by the i.i.d. sequence $(\check{\sigma}_i)_{i\geq 1}$, and, from these, the random variables $\check{B}_i^{(n)}$ and $\check{B}_i$ are obtained as explained in (2.3) and (2.7). We use the variables $\check{B}_i^{(n)}$ for the evolution of BP_n, and $\check{B}_i$ for the evolution of BP. Thus, the evolution of miscoupled individuals in the CTBPs BP_n and BP is *completely independent* of the evolution of SWT. When differences arise in BP_n and BP, also their evolutions are mutually independent.

We close this section by defining the sets of alive individuals in the coupling of $(\mathsf{SWT}(t), \mathsf{BP}_n(t), \mathsf{BP}(t))_{t\geq 0}$. Both $\mathsf{BP}_n(t)$ as well as $\mathsf{BP}(t)$ each have their sets of alive individuals that we denote by $\mathsf{AI}_n(t)$ and $\mathsf{AI}(t)$, respectively. For $\mathsf{BP}_n(t)$, we can think of these alive individuals as corresponding to repeated draws of half-edges. By our coupling, these sets of alive individuals in $\mathsf{BP}_n(t)$ and $\mathsf{BP}(t)$ are effectively coupled to the alive half-edges in $\mathsf{AH}(t)$. The successfully coupled half-edges in $\mathsf{SWT}(t)$ and $\mathsf{BP}_n(t)$ at time t form $\mathsf{AH}(t) \cap \mathsf{AI}_n(t)$, the successfully coupled individuals in $\mathsf{BP}_n(t)$ and $\mathsf{BP}(t)$ form $\mathsf{AI}_n(t) \cap \mathsf{AI}(t)$. We note that each alive half-edge, individual y in $\mathsf{AH}(t)$, $\mathsf{AI}_n(t)$ and $\mathsf{AI}(t)$ has a residual lifetime $R_y(t)$, as well as an index $I(y)$ indicating which subtree y is an element of and a height $H(y)$ denoting the generation of y. Similarly to (2.15), we then define

$$\mathsf{BP}_n(t) = (y, I(y), H(y), R_t(y))_{y\in \mathsf{AI}_n(t)}, \qquad \mathsf{BP}(t) = (y, I(y), H(y), R_t(y))_{y\in \mathsf{AI}(t)}. \tag{2.34}$$

Since the objects in (2.34) are coupled to $(\mathsf{SWT}(t))_{t\geq 0}$ in (2.15), this completes the coupling of the FPP processes $(\mathsf{SWT}(t), \mathsf{BP}_n(t), \mathsf{BP}(t))_{t\geq 0}$ and defines the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which this coupling of $(\mathsf{SWT}(t), \mathsf{BP}_n(t), \mathsf{BP}(t))$ lives. We let $(\mathcal{F}_t)_{t\geq 0}$ be the filtration generated by all the randomness used in the construction up to time t, i.e., $\mathcal{F}_t$ contains all the information needed to construct $(\mathsf{SWT}(s), \mathsf{BP}_n(s), \mathsf{BP}(s))_{s\in[0,t]}$. Under this coupling law, we can speak of convergence in probability, and we shall frequently do this in the sequel.

Summary of the coupling. For completeness and future references, we resume how differences arise in the coupling. Degree-miscouplings arise due to three effects:

(1) MISC-*miscouplings* occur between the forward degree $Y_k^{(n)}$ (which are not i.i.d. due to draws being without replacement) and the i.i.d. draws $(B_i^{(n)})_{i\geq 1}$, because $\tau_k > \tau_{k-1}+1$; and

(2) *cycle-events* occur referring to self-loops or cycles that makes $X_k^{(n)} < Y_k^{(n)}$. In these cases we remove the self-loop or the edge that gives rise to a cycle from the set of alive half-edges. This amounts to removing up to at most $Y_k^{(n)} - X_k^{(n)}$ half-edges incident to vertex $V_{z_{k-1}}$, as well as up to at most $Y_k^{(n)} - X_k^{(n)}$ half-edges to which they are paired from SWT.

(3) *collision edges* are found. In this case, precisely one of the vertices to which the collision edge is incident is degree-miscoupled. We want to emphasize here that the degree-miscoupling caused by finding the collision edge at time $T_j^{(\mathrm{col})}$ does not effect the coupling of the shortest-weight paths. When the collision edge is removed, we are left with two paths connecting U_i to one of the vertices incident to the two half-edges of which the collision edge consists. It should be checked that at any time *prior* to $T_j^{(\mathrm{col})}$ each of these paths is not miscoupled, i.e., does not contain any earlier degree-miscoupled vertices.

In all these three cases, the vertices involved are called *degree-miscoupled*, and any further offspring of degree-miscoupled vertices (in the SWT or in the CTBP) are called *miscoupled*. Thus, any miscoupling gives rise to a tress of miscoupled children half-edges in the SWT, respectively, offspring in the CTBP.

2.4. **Main coupling results.** We consider the process coupling $(\mathsf{SWT}(t), \mathsf{BP}_n(t), \mathsf{BP}(t))_{t\geq 0}$ defined in the previous section, as well as the associated filtration $(\mathcal{F}_t)_{t\geq 0}$. We recall that $\mathsf{BP}_n(t) = (\mathsf{BP}_n^{(1)}(t), \mathsf{BP}_n^{(2)}(t))$ are two independent CTBPs starting with offspring distribution D_n in the first generation and offspring law $B^{(n)} = D_n^\star - 1$ in the second and further generations, and $\mathsf{BP}(t) = (\mathsf{BP}^{(1)}(t), \mathsf{BP}^{(2)}(t))$ which are two independent CTBPs starting with offspring distribution D in the first generation and offspring distribution $B = D^\star - 1$ in the second and further generations. For this coupling $(\mathsf{SWT}(t), \mathsf{BP}_n(t), \mathsf{BP}(t))_{t\geq 0}$, we let $\mathsf{AH}(t)\triangle\mathsf{AI}_n(t)$ denote the set of miscoupled half-edges at time t. With a slight abuse of notation, we write $|\mathsf{SWT}(t)| = |\mathsf{AH}(t)|$ and $|\mathsf{SWT}(t)\triangle\mathsf{BP}_n(t)| = |\mathsf{AH}(t)\triangle\mathsf{AI}_n(t)|$. Finally, we denote the set of *all* miscoupled half-edges and individuals up to time t by

$$\bigcup_{s\in[0,t]} \mathsf{SWT}(s)\triangle\mathsf{BP}_n(s). \tag{2.35}$$

In this section, we state two key propositions concerning the coupling. Proposition 2.2(a) shows that there exists some $s_n \to \infty$ such that, whp, there are *no* miscouplings up to time s_n. In Proposition 2.2(b) and Proposition 2.3, we investigate the size of $\mathsf{SWT}(t)$ for t close to $t_n = \log n/(2\alpha_n)$.

Proposition 2.2 (Coupling the SWT to a BP)**.**

(a) *There exists $s_n \to \infty$ such that, for the coupling defined in Sections 2.2-2.3,*

$$\mathbb{P}\Big((\mathsf{SWT}(s))_{s\in[0,s_n]} = (\mathsf{BP}_n(s))_{s\in[0,s_n]} = (\mathsf{BP}(s))_{s\in[0,s_n]}\Big) = 1 - o(1). \tag{2.36}$$

Consequently, with $\mathcal{W}_{s_n}^{(i)} = \mathrm{e}^{-\alpha_n s_n}|\mathsf{SWT}^{(i)}(s_n)|$, $i = 1, 2$,

$$\lim_{\varepsilon\downarrow 0}\lim_{n\to\infty} \mathbb{P}\Big(\mathcal{W}_{s_n}^{(1)} \in [\varepsilon, 1/\varepsilon], \mathcal{W}_{s_n}^{(2)} \in [\varepsilon, 1/\varepsilon] \Big| \mathcal{W}_{s_n}^{(1)}\mathcal{W}_{s_n}^{(2)} > 0\Big) = 1. \tag{2.37}$$

(b) *Let $t_n = \log n/(2\alpha_n)$. For the coupling of $(\mathsf{SWT}(s))_{s\geq 0}$ and $(\mathsf{BP}_n(s))_{s\geq 0}$ defined in Sections 2.2-2.3, there exist sequences $\varepsilon_n \to 0$ and $B_n \to \infty$ such that, conditionally on $\mathcal{F}_{s_n}$, and for* *every* $t \leq t_n + B_n$,

$$\mathbb{P}\Big(\Big|\bigcup_{s\in[0,t]} \mathsf{SWT}(s)\triangle \mathsf{BP}_n(s)\Big| \geq \varepsilon_n\sqrt{n} \mid \mathcal{F}_{s_n}\Big) \xrightarrow{\mathbb{P}} 0. \tag{2.38}$$

The proof of Proposition 2.2 is deferred to Section 5. We warn the reader to beware for confusion between the (large) constant B_n and the i.i.d. random variables $(B_i)_{i\geq 1}$. Fix the deterministic sequence $s_n \to \infty$ from Proposition 2.2. Now let

$$t_n = \frac{1}{2\alpha_n}\log n, \qquad \bar{t}_n = \frac{1}{2\alpha_n}\log n - \frac{1}{2\alpha_n}\log\big(\mathcal{W}^{(1)}_{s_n}\mathcal{W}^{(2)}_{s_n}\big). \tag{2.39}$$

Note that $\mathrm{e}^{\alpha_n t_n} = \sqrt{n}$; it will turn out that both $|\mathsf{SWT}^{(i)}(t_n)|$, for $i = 1, 2$, are of order $\sqrt{n}$. Further, it will turn out that collision edges start to appear when these clusters grow to be of this size. Consequently, the variable t_n denotes the typical time at which collision edges start appearing, and the time $\bar{t}_n$ incorporates for stochastic fluctuations in the size of the SWT's.

For $i \in \{1, 2\}$, $k \geq 0$, and $s, t \geq 0$, we define

$$|\mathsf{SWT}^{(i)}_k[t, t+s)| = \Big|\{y \in \mathsf{AH}(t)\colon I(y) = i, H(y) = k, R_t(y) \in [0, s)\}\Big|, \tag{2.40}$$

as the number of alive half-edges at time t that (i) are in the SWT of vertex U_i, (ii) have height k, and (iii) have remaining (or residual) lifetime at most s. We further write

$$|\mathsf{SWT}^{(i)}_{\leq k}[t, t+s)| = \Big|\{y \in \mathsf{AH}(t)\colon I(y) = i, H(y) \leq k, R_t(y) \in [0, s)\}\Big|, \tag{2.41}$$

for the number of alive half-edges that have height at most k. To formulate the CLT for the height of vertices, we will choose

$$k_n(t, x) = \frac{t}{\bar{\nu}_n} + x\sqrt{t\frac{\bar{\sigma}^2}{\bar{\nu}^3}}, \tag{2.42}$$

where $\bar{\nu}_n, \bar{\nu}$ and $\bar{\sigma}^2$ are defined in (1.19).

Define the *residual life-time distribution* F_R to have density f_R given by

$$f_R(x) = \frac{\int_0^\infty \mathrm{e}^{-\alpha y} f_\xi(x+y)\,dy}{\int_0^\infty \mathrm{e}^{-\alpha y}[1 - F_\xi(y)]\,dy} = \frac{\alpha\nu}{\nu - 1}\int_0^\infty \mathrm{e}^{-\alpha y} f_\xi(x+y)\,dy. \tag{2.43}$$

Below, we let Φ denote the standard normal distribution function. Finally, for a half-edge $y \in \mathsf{AH}(t)$, we let $X^\star_y = d_{V_y} - 1$.

Proposition 2.3 (Ages and heights in SWT)**.** *Fix $j \in \{1, 2\}$, and numbers $x, y, t \in \mathbb{R}$, $s_1, s_2 > 0$, all independent of n. Then, conditionally on $\mathcal{F}_{s_n}$ and $\mathcal{W}^{(1)}_{s_n}\mathcal{W}^{(2)}_{s_n} > 0$,*

(a) *we have*

$$\begin{aligned}&\mathrm{e}^{-2\alpha_n t_n}|\mathsf{SWT}^{(j)}_{\leq k_n(t_n,x)}[\bar{t}_n + t, \bar{t}_n + t + s_1)||\mathsf{SWT}^{(3-j)}_{\leq k_n(t_n,y)}[\bar{t}_n + t, \bar{t}_n + t + s_2)| \\ &\qquad \xrightarrow{\mathbb{P}} \mathrm{e}^{2\alpha t}\Phi(x)\Phi(y)F_R(s_1)F_R(s_2),\end{aligned} \tag{2.44}$$

(b) *further*

$$\begin{aligned}&\mathrm{e}^{-2\alpha_n t_n}|\mathsf{SWT}^{(j)}_{\leq k_n(t_n,x)}[\bar{t}_n + t, \bar{t}_n + t + s_1)|\sum_v X^\star_v \mathbb{1}_{\{v\in \mathsf{SWT}^{(3-j)}_{\leq k_n(t_n,y)}[\bar{t}_n+t,\bar{t}_n+t+s_2)\}} \\ &\qquad \xrightarrow{\mathbb{P}} \nu\mathrm{e}^{2\alpha t}\Phi(x)\Phi(y)F_R(s_1)F_R(s_2).\end{aligned} \tag{2.45}$$

The proof of Proposition 2.3 is deferred to Section 6.

3. Main ingredient: Poisson Point Process limit

In this section, we state our main result that implies Theorems 1.2-1.3. To state this result we need some additional definitions.

Recall the collection of collision edges $\mathscr{C}$ from (2.30). Here the jth collision edge is given by (x_j, P_{x_j}), where P_{x_j} is an alive half-edge and x_j the half-edge which pairs to P_{x_j}. Rescaling time by $\bar{t}_n$ (see (2.39)), we define

$$\bar{T}_j^{(\mathrm{col})} = T_j^{(\mathrm{col})} - \bar{t}_n, \qquad \bar{H}_j^{(\mathrm{or})} = \frac{H(x_j) - t_n/\bar{\nu}_n}{\sqrt{\bar{\sigma}^2 t_n/\bar{\nu}^3}}, \qquad \bar{H}_j^{(\mathrm{de})} = \frac{H(P_{x_j}) - t_n/\bar{\nu}_n}{\sqrt{\bar{\sigma}^2 t_n/\bar{\nu}^3}}, \tag{3.1}$$

and write the random elements $(\Xi_j)_{j\geq 1}$ with $\Xi_j \in \mathcal{S} := \mathbb{R} \times \{1,2\} \times \mathbb{R} \times \mathbb{R} \times [0,\infty)$, by

$$\Xi_j = \big(\bar{T}_j^{(\mathrm{col})}, I(x_j), \bar{H}_j^{(\mathrm{or})}, \bar{H}_j^{(\mathrm{de})}, R_{T_j^{(\mathrm{col})}}(P_{x_j})\big). \tag{3.2}$$

Then, for sets A in the Borel σ-algebra of the space $\mathcal{S}$, we define the point process

$$\Pi_n(A) = \sum_{j\geq 1} \delta_{\Xi_j}(A), \tag{3.3}$$

where δ_x gives measure 1 to the point x. Let $\mathcal{M}(\mathcal{S})$ denote the space of all simple locally-finite point processes on $\mathcal{S}$ equipped with the vague topology (see e.g. [28]). On this space one can naturally define the notion of weak convergence of a sequence of random point processes $\Pi_n \in \mathcal{M}(\mathcal{S})$. This is the notion of convergence referred to in the following theorem.

Theorem 3.1 (PPP limit of collision edges). *Consider the distribution of the point process $\Pi_n \in \mathcal{M}(\mathcal{S})$ defined in (3.3) conditioned on $\mathcal{F}_{s_n}$ and $\mathcal{W}_{s_n}^{(1)}\mathcal{W}_{s_n}^{(2)} > 0$, for s_n as in Proposition 2.2. Then, as $n \to \infty$, Π_n converges in distribution to a Poisson Point Process (PPP) Π with intensity measure*

$$\lambda(dt \times i \times dx \times dy \times dr) = \frac{2\nu f_R(0)}{\mathbb{E}[D]} \mathrm{e}^{2\alpha t} dt \otimes \{1/2, 1/2\} \otimes \Phi(dx) \otimes \Phi(dy) \otimes F_R(dr). \tag{3.4}$$

Theorem 3.1 will be proved in Section 7.

Completion of the proof of Theorems 1.2 and 1.3. Let us now prove Theorem 1.2 subject to Theorem 3.1. First of all, by (3.1), (2.31) and (2.32),

$$\Big(\frac{H_n - \frac{1}{\alpha_n \bar{\nu}_n}\log n}{\sqrt{\frac{\bar{\sigma}^2}{\bar{\nu}^3\alpha}\log n}}, L_n - \frac{1}{\alpha_n}\log n\Big) \tag{3.5}$$

is a continuous function of the point process Π_n, and, therefore, by the continuous mapping theorem, the above random vector converges in distribution to some random limit (Z, Q).

Recall that $I^\star$ denotes the minimizer of $i \mapsto 2T_i^{(\mathrm{col})} + R_{T_i^{(\mathrm{col})}}(P_{x_i})$. By (2.31), the weight L_n as well as the value of $I^\star$, are functions of the first and the last coordinates of Π_n. The hopcount H_n is a function of the third and the fourth, instead. By the product form of the intensity in (3.4), we obtain that the limits (Z, Q) are independent. Therefore, it suffices to study their marginals.

We start with the limit distribution of the hopcount. By (3.1) and (2.32),

$$\frac{H_n - \frac{1}{\alpha_n \bar{\nu}_n}\log n}{\sqrt{\frac{\bar{\sigma}^2}{\bar{\nu}^3\alpha}\log n}} = \frac{1}{2}\sqrt{2}\bar{H}_{I^\star}^{(\mathrm{or})} + \frac{1}{2}\sqrt{2}\bar{H}_{I^\star}^{(\mathrm{de})} + o_{\mathbb{P}}(1). \tag{3.6}$$

By Theorem 3.1, the random variables $(\bar{H}_{I^\star}^{(\mathrm{or})}, \bar{H}_{I^\star}^{(\mathrm{de})})$, converge to two independent standard normals, so that also the left-hand side of (3.6) converges to a standard normal.

The limit distribution of the weight L_n is slightly more involved. By (2.39), (2.31) and (3.1),

$$\begin{aligned} L_n - \frac{1}{\alpha_n}\log n &= L_n - 2t_n = L_n - 2\bar{t}_n - \frac{1}{\alpha_n}\log(\mathcal{W}_{s_n}^{(1)}\mathcal{W}_{s_n}^{(2)}) \\ &= -\frac{1}{\alpha_n}\log(\mathcal{W}_{s_n}^{(1)}\mathcal{W}_{s_n}^{(2)}) + \min_{i\geq 1}[2\bar{T}_i^{(\mathrm{col})} + R_{T_i^{(\mathrm{col})}}(P_{x_i})]. \end{aligned} \tag{3.7}$$

By Proposition 2.2, $(\mathcal{W}^{(1)}_{s_n}, \mathcal{W}^{(2)}_{s_n}) \xrightarrow{d} (\mathcal{W}^{(1)}, \mathcal{W}^{(2)})$, which are two independent copies of the random variable in (1.21). Hence,

$$L_n - \frac{1}{\alpha_n}\log n \xrightarrow{d} -\frac{1}{\alpha}\log(\mathcal{W}^{(1)}\mathcal{W}^{(2)}) + \min_{i\geq 1}[2\pi_i + R_i], \tag{3.8}$$

where $(\pi_i)_{i\geq 1}$ is a PPP with intensity $\frac{2\nu f_R(0)}{\mathbb{E}[D]}\mathrm{e}^{2\alpha t}dt$, and $(R_i)_{i\geq 1}$ are i.i.d. random variables with distribution function F_R independently of $(\pi_i)_{i\geq 1}$.

We next identify the distribution of $M = \min_{i\geq 1}[2\pi_i + R_i]$. First, $(2\pi_i)_{i\geq 1}$ forms a Poisson process with intensity $\frac{\nu f_R(0)}{\mathbb{E}[D]}\mathrm{e}^{\alpha t}dt$. According to [36, Example 3.3 on page 137] the point process $(2\pi_i + R_i)_{i\geq 1}$ is a non-homogeneous Poisson process with mean-measure the convolution of $\mu(-\infty, x] = \int_{-\infty}^{x} \frac{\nu f_R(0)}{\mathbb{E}[D]}\mathrm{e}^{\alpha t}\,dt$ and F_R. Hence $\mathbb{P}(M \geq x)$ equals the Poisson probability of 0, where the parameter of the Poisson distribution is $(\mu * F_R)(x)$, so that

$$\mathbb{P}(M \geq x) = \exp\{-\frac{\nu f_R(0)}{\mathbb{E}[D]}\mathrm{e}^{\alpha x}\int_0^\infty F_R(z)\mathrm{e}^{-\alpha z}\,dz\}. \tag{3.9}$$

Let Λ have a Gumbel distribution, i.e., $\mathbb{P}(\Lambda \leq x) = \mathrm{e}^{-\mathrm{e}^{-x}}$, $x \in \mathbb{R}$, then, from (3.9),

$$M = \min_{i\geq 1}(2\pi_i + R_i) \stackrel{d}{=} -\alpha^{-1}\Lambda - \alpha^{-1}\log(\nu f_R(0)B/\mathbb{E}[D]), \tag{3.10}$$

with $B = \int_0^\infty F_R(z)\mathrm{e}^{-\alpha z}\,dz$. In the following lemma, we simplify these constants:

Lemma 3.2 (The constant). *The constants $B = \int_0^\infty F_R(z)\mathrm{e}^{-\alpha z}\,dz$ and $f_R(0)$ are given by*

$$B = \bar{\nu}/(\nu - 1), \qquad f_R(0) = \alpha/(\nu - 1). \tag{3.11}$$

Consequently, the constant c in the limit variable (1.23) equals

$$c = -\log(\nu f_R(0)B/\mathbb{E}[D]) = \log(\mathbb{E}[D](\nu - 1)^2/(\alpha\nu\bar{\nu})). \tag{3.12}$$

Proof. According to (2.43) and (1.15),

$$f_R(0) = \frac{\alpha\nu}{\nu - 1}\int_0^\infty \mathrm{e}^{-\alpha y} f_\xi(y)\,dy = \alpha/(\nu - 1). \tag{3.13}$$

For B, we use partial integration and substitution of (2.43). This yields

$$\begin{aligned} B &= \int_0^\infty F_R(z)\mathrm{e}^{-\alpha z}\,dz = \frac{1}{\alpha}\int_0^\infty f_R(z)\mathrm{e}^{-\alpha z}\,dz = \frac{\nu}{\nu - 1}\int_0^\infty \mathrm{e}^{-\alpha z}\int_0^\infty \mathrm{e}^{-\alpha y} f_\xi(y + z)\,dy\,dz \\ &= \frac{\nu}{\nu - 1}\int_0^\infty s f_\xi(s)\mathrm{e}^{-\alpha s}\,ds = \frac{1}{(\nu - 1)}\int_0^\infty s\bar{F}_\xi(ds) = \bar{\nu}/(\nu - 1). \end{aligned} \tag{3.14}$$

This completes the proof of Theorems 1.2 and 1.3 subject to Theorem 3.1. ∎

4. Height CLT and residual lifetime for CTBP

In this section, we set the stage for the proof of Proposition 2.3 for CTBPs. We make use of second moment methods similar to the ones in [20, 23, 24, 37], but with a suitable truncation argument to circumvent the problem of infinite-variance offspring distributions.

As in the first part of Section 1.3, we consider a (*standard*) CTBP process [20, Chapter 6], with lifetime distribution F_ξ admitting a density f_ξ, and random offspring $X = X_v$ satisfying (1.13) and the $X \log X$ in (1.14). We define

$$\eta = \nu\int_0^\infty \mathrm{e}^{-2\alpha s}\,dF_\xi(s), \qquad \text{and} \qquad m_j = K\eta^{-j},\ j \geq 1, \tag{4.1}$$

for some $K > 1$. Note that $\eta \in (0, 1)$, since α is such that $\nu\int_0^\infty \mathrm{e}^{-\alpha s}\,dF_\xi(s) = 1$. The *truncated* CTBP $\mathsf{BP}^{(\vec{m})}$ has for each individual in generation j offspring $(X \wedge m_j)$ instead of X.

We denote the number of alive individuals in the CTBP at time t by $|\mathsf{BP}(t)|$. By $|\mathsf{BP}_k(t)|$, $|\mathsf{BP}_k[t, t + s)|$, we denote the number of alive individuals in generation k at time t, number of alive individuals in generation k at time t with residual lifetime at most s, respectively. We warn the reader that $|\mathsf{BP}_k(t)|$ refers to the number of individuals in generation k, and not to the n-dependence. When dealing with n-dependent CTBPs, we will use the notation $|\mathsf{BP}_{n,k}(t)|$ instead.

Here the generation of the first individual equals 0, and the generation of its offspring is equal to 1, etc. For the truncated process $\mathsf{BP}^{(\vec{m})}(t)$, we define, analogously to the definitions above, $|\mathsf{BP}^{(\vec{m})}(t)|$, $|\mathsf{BP}_k^{(\vec{m})}(t)|$, and $|\mathsf{BP}_k^{(\vec{m})}[t,t+s)|$. Furthermore,

$$|\mathsf{BP}_{\leq k}^{(\vec{m})}[t,t+s)| = \sum_{j=0}^{k} |\mathsf{BP}_j^{(\vec{m})}[t,t+s)|, \qquad |\mathsf{BP}^{(\vec{m})}[t,t+s)| = \sum_{j=0}^{\infty} |\mathsf{BP}_j^{(\vec{m})}[t,t+s)|. \tag{4.2}$$

A key ingredient to the proof of Proposition 2.3 is Proposition 4.1 below:

Proposition 4.1 (First and second moment CLT). *Consider the branching process $\mathsf{BP}(t)$ introduced above, with i.i.d. lifetimes F_ξ admitting a density and random offspring X satisfying $\nu = \mathbb{E}[X] > 1$, and $\mathbb{E}[X \log(X)_+] < \infty$. Choose $m_j = K\eta^{-j}$ as in (4.1). Then, with $A = (\nu-1)/(\alpha\nu\bar{\nu})$,*
(a)

$$\lim_{t\to\infty} \mathrm{e}^{-\alpha t}\mathbb{E}\big[|\mathsf{BP}(t)|\big] = A, \qquad \lim_{K\to\infty} \limsup_{t\to\infty} \mathrm{e}^{-\alpha t}\mathbb{E}\Big[|\mathsf{BP}(t)| - |\mathsf{BP}^{(\vec{m})}(t)|\Big] = 0, \tag{4.3}$$

(b) there exists a constant $C > 0$, such that uniformly in $t \in [0,\infty)$,

$$\mathrm{e}^{-2\alpha t}\mathbb{E}\big[|\mathsf{BP}^{(\vec{m})}(t)|^2\big] \leq CK, \tag{4.4}$$

(c)

$$\lim_{K\to\infty} \lim_{t\to\infty} \mathrm{e}^{-\alpha t}\mathbb{E}\big[|\mathsf{BP}_{\leq k(t,x)}^{(\vec{m})}[t,t+s)|\big] = A\Phi(x)F_R(s), \tag{4.5}$$

where F_R is defined through (2.43) and $k(t,x) = t/\bar{\nu} + x\sqrt{t\bar{\sigma}^2/(\bar{\nu})^3}$.
(d) Replace in the above statements BP by BP_n, with offspring X_n, depending on n in such a way that $X_n \xrightarrow{d} X$, $\nu_n = \mathbb{E}[X_n] \to \mathbb{E}[X]$ and $\lim_n \mathbb{E}[X_n \log(X_n/K_n)_+] = 0$, for any $K_n \to \infty$. Furthermore, now define m_j by $m_j = K_n\eta_n^{-j}$, with $\eta_n = \nu_n \int_0^\infty \mathrm{e}^{-2\alpha_n s}\, dF_\xi(s)$, and replace $k(t,x)$ by $k_n(t,x)$ defined in (2.42). Then Part (a) and Part (c) hold with $\alpha = \alpha_n$ and $t = t_n$ and with the limits replaced by $\lim_{n\to\infty}$, for any sequence $t_n \to \infty$. Similarly, under these conditions and substitutions, Part (b) holds for all $n \geq 1$, with K replaced by K_n, uniformly in t.

Proof. We start by proving Proposition 4.1(a). The first claim of Part (a) is proved in [20, Theorem 17.1]. We bound the first moment of the difference between the truncated and the original branching process. Let $\nu^{(j)} = \mathbb{E}[(X \wedge m_j)]$. We compute that for $t > 0$,

$$\mathrm{e}^{-\alpha t}\mathbb{E}\Big[\sum_{k=0}^{\infty}[|\mathsf{BP}_k(t)| - |\mathsf{BP}_k^{(\vec{m})}(t)|]\Big] = \mathrm{e}^{-\alpha t}\sum_{k=0}^{\infty}[\nu^k - \prod_{j=1}^{k}\nu^{(j)}]\big[F_\xi^{\star k}(t) - F_\xi^{\star(k+1)}(t)\big], \tag{4.6}$$

where $F_\xi^{\star k}$ is the k-fold convolution of F_ξ, where, by convention, $F_\xi^{\star 0}(t) = 1$ for every $t \geq 0$. In order to bound the differences $\nu^k - \prod_{j=1}^{k}\nu^{(j)}$, we rely on the following lemma:

Lemma 4.2 (Effect of truncation on expectation CTBP). *Under the conditions of Proposition 4.1, uniformly in $k \geq 1$,*

$$[1 - \prod_{j=1}^{k}\frac{\nu^{(j)}}{\nu}] \leq (\log(1/\eta))^{-1}\mathbb{E}\big[X \log{(X/K)}_+\big] = o_K(1), \tag{4.7}$$

where $o_K(1)$ denotes a quantity that converges to zero as $K \to \infty$.

Proof. Since $\nu^{(j)} \leq \nu$, for all $j \geq 1$, it is easily shown by induction that

$$1 - \prod_{j=1}^{k}\frac{\nu^{(j)}}{\nu} \leq \sum_{j=1}^{k}\Big(1 - \frac{\nu^{(j)}}{\nu}\Big) \leq \sum_{j=1}^{\infty}\Big(1 - \frac{\nu^{(j)}}{\nu}\Big). \tag{4.8}$$

Now, using that $\nu > 1$,

$$\sum_{j=1}^{\infty}\Big(1-\frac{\nu^{(j)}}{\nu}\Big) \leq \sum_{j=1}^{\infty}\mathbb{E}[X\mathbb{1}_{\{X>m_j\}}] = \mathbb{E}\big[X\sum_{j=1}^{\infty}\mathbb{1}_{\{m_j<X\}}\big], \tag{4.9}$$

and we note that the number of j for which $m_j = K\eta^{-j} < x$ is at most $[\log(x/K)/\log(1/\eta)]\vee 0$. Therefore, the inequality in (4.7) holds. Since $\mathbb{E}\big[X\log(X/K)_+\big] = o_K(1)$, the equality in (4.7) follows. ■

By Lemma 4.2 and (4.6),

$$\mathrm{e}^{-\alpha t}\mathbb{E}\Big[\sum_{k=1}^{\infty}[|\mathsf{BP}_k(t)|-|\mathsf{BP}_k^{(\vec{m})}(t)|]\Big] = o_K(1)\mathrm{e}^{-\alpha t}\mathbb{E}\Big[\sum_{k=1}^{\infty}|\mathsf{BP}_k(t)|\Big] = o_K(1), \tag{4.10}$$

which completes the proof of Proposition 4.1(a).

We continue with the proof of the second moment estimate in Proposition 4.1(b). We follow the proof in [37], keeping attention to the truncation. We introduce the generating functions

$$h(s) = \mathbb{E}[s^X], \qquad h_j(s) = \mathbb{E}[s^{(X\wedge m_j)}], \tag{4.11}$$

where m_j is given by (4.1). Parallel to calculations in the proof of [37, Lemma 4],

$$\mathbb{E}[|\mathsf{BP}^{(\vec{m})}|^2] = h_1''(1)(\mathbb{E}[|\mathsf{BP}^{(\vec{m}_1)}|])^2 * F_\xi + h_1'(1)\mathbb{E}[|\mathsf{BP}^{(\vec{m}_1)}|^2] * F_\xi, \tag{4.12}$$

where $\vec{m}_1 = (m_2, m_3, \ldots)$ is $\vec{m}$ with the first element removed, and where for simplicity of reading the argument t has been left out. Transforming to

$$|\overline{\mathsf{BP}}^{(\vec{m})}(t)| = \mathrm{e}^{-\alpha t}|\mathsf{BP}^{(\vec{m})}(t)|, \tag{4.13}$$

we obtain, after multiplying both sides of (4.12) by $\mathrm{e}^{-2\alpha t}$,

$$\mathbb{E}[|\overline{\mathsf{BP}}^{(\vec{m})}|^2] = \frac{\eta h_1''(1)}{\nu}(\mathbb{E}[|\overline{\mathsf{BP}}^{(\vec{m}_1)}|])^2 * Q + \frac{\eta h_1'(1)}{\nu}\mathbb{E}[|\overline{\mathsf{BP}}^{(\vec{m}_1)}|^2] * Q, \tag{4.14}$$

where

$$\bar{F}_\xi(t) = \nu\int_0^t \mathrm{e}^{-\alpha y}\,dF_\xi(y), \qquad Q(t) = \eta^{-1}\int_0^t \mathrm{e}^{-\alpha y}\,d\bar{F}_\xi(y) = \eta^{-1}\nu\int_0^t \mathrm{e}^{-2\alpha y}\,dF_\xi(y), \tag{4.15}$$

and where we recall that $\eta = \int_0^\infty \mathrm{e}^{-\alpha y}\,d\bar{F}_\xi(y) < 1$ and $\nu = h'(1)$. Iteration of (4.14) yields

$$\mathbb{E}[|\overline{\mathsf{BP}}^{(\vec{m})}|^2] = \sum_{j=1}^{\infty} f_1\cdots f_{j-1}e_j(\mathbb{E}[|\overline{\mathsf{BP}}^{(\vec{m}_j)}|])^2 * Q^{\star j}, \tag{4.16}$$

where

$$e_j = \frac{\eta h_j''(1)}{\nu}, \qquad f_j = \frac{\eta h_j'(1)}{\nu}, \tag{4.17}$$

and where $\vec{m}_j = (m_{j+1}, m_{j+2}, \ldots)$. Obviously,

$$\mathbb{E}[|\overline{\mathsf{BP}}^{(\vec{m})}(t)|] \leq \mathbb{E}[|\overline{\mathsf{BP}}^{(\vec{m}_j)}(t)|] \leq \mathbb{E}[|\overline{\mathsf{BP}}(t)|] \to A, \tag{4.18}$$

by Part (a). Hence, provided that the sum $\sum_{j\geq 1} f_1\cdots f_{j-1}e_j$ converges, which we will establish in Lemma 4.3 below, we have that, uniformly in t,

$$\mathbb{E}[|\overline{\mathsf{BP}}^{(\vec{m})}(t)|^2] \leq C\sum_{j=1}^{\infty} f_1\cdots f_{j-1}e_j, \tag{4.19}$$

for some constant $C \geq A^2$.

Lemma 4.3 (Effect of truncation on variance CTBP)**.** *For $m_j = K\eta^{-j}$, and with $\nu = \mathbb{E}[X] > 1$,*

$$\sum_{j=1}^{\infty} f_1\cdots f_{j-1}e_j \leq \frac{2\nu K}{1-\eta}. \tag{4.20}$$

Proof. We bound $f_j \leq \eta$, and

$$e_j \leq \eta \mathbb{E}[(X \wedge m_j)^2] = \eta\Big(m_j^2 \mathbb{P}(X > m_j) + \mathbb{E}[X^2 \mathbb{1}_{\{X \leq m_j\}}]\Big), \tag{4.21}$$

so that

$$\sum_{j=1}^{\infty} f_1 \dots f_{j-1} e_j \leq \sum_{j=1}^{\infty} m_j^2 \mathbb{P}(X > m_j)\eta^j + \sum_{j=1}^{\infty} \mathbb{E}[X^2 \mathbb{1}_{\{X \leq m_j\}}]\eta^j. \tag{4.22}$$

We bound both terms separately. The first contribution equals

$$K^2 \sum_{j=1}^{\infty} \mathbb{P}(X > K\eta^{-j})\eta^{-j} = K^2 \mathbb{E}[\sum_{j=1}^{\infty} \eta^{-j} \mathbb{1}_{\{K\eta^{-j} < X\}}] = K^2 \mathbb{E}\big[\frac{\eta^{-a(X)} - 1}{1-\eta}\big], \tag{4.23}$$

where $a(x) = \max\{j\colon K\eta^{-j} < x\} = \lfloor \log(x/K)/\log(1/\eta) \rfloor$. Therefore, $\eta^{-a(X)} \leq X/K$, so that

$$\sum_{j=1}^{\infty} m_j^2 \mathbb{P}(X > m_j)\eta^j \leq \frac{K^2}{1-\eta}\mathbb{E}[X/K] = \frac{K\nu}{1-\eta}. \tag{4.24}$$

The second contribution is bounded in a similar way as

$$\sum_{j=1}^{\infty} \mathbb{E}[X^2 \mathbb{1}_{\{X \leq m_j\}}]\eta^j = \mathbb{E}\Big[\sum_{j=1}^{\infty} X^2 \eta^j \mathbb{1}_{\{X \leq K\eta^{-j}\}}\Big] = \mathbb{E}\Big[X^2 \sum_{j=b(X)}^{\infty} \eta^j\Big] = \mathbb{E}\Big[\frac{X^2 \eta^{b(X)}}{1-\eta}\Big], \tag{4.25}$$

where $b(x) = \min\{j\colon K\eta^{-j} \geq x\}$, so that $\eta^{b(X)} \leq K/X$. Therefore,

$$\sum_{j=1}^{\infty} \mathbb{E}[X^2 \mathbb{1}_{\{X \leq m_j\}}]\eta^j \leq \frac{K\mathbb{E}[X]}{1-\eta} = \frac{K\nu}{1-\eta}. \tag{4.26}$$

■

Proposition 4.1 (b) follows by combining (4.19) and (4.20).

For Proposition 4.1(c), we start by showing that

$$\mathrm{e}^{-\alpha t} \sum_{j=0}^{k(t,x)} \mathbb{E}[|\mathsf{BP}_j[t, t+s)|] \to A\Phi(x)F_R(s). \tag{4.27}$$

Observe that $|\mathsf{BP}[t, t+s)| = \sum_{j=0}^{\infty} |\mathsf{BP}_j[t, t+s)|$ is the total number of alive individuals at time t, with residual lifetime at most s, so that from [20, Lemma 2, Appendix Chapter VI], on the renewal equation

$$\mathbb{E}[|\mathsf{BP}[t, t+s)|] = F_\xi(t+s) - F_\xi(t) + \nu \int_0^\infty \mathbb{E}[|\mathsf{BP}(t-y, t+s-y)|]\, dF_\xi(y), \tag{4.28}$$

we readily obtain (compare the derivation of [20, Theorem 24.1]),

$$\lim_{t\to\infty} \mathrm{e}^{-\alpha t}\mathbb{E}[|\mathsf{BP}[t, t+s)|] = \lim_{t\to\infty} \mathrm{e}^{-\alpha t} \sum_{j=0}^{\infty} \mathbb{E}[|\mathsf{BP}_j[t, t+s)|] = AF_R(s),$$

with $A = (\nu - 1)/(\alpha\nu\bar{\nu})$ as given in the proposition. For fixed $s > 0$, define

$$|\overline{\mathsf{BP}}_{>m}[t, t+s)| = \sum_{j=m+1}^{\infty} |\overline{\mathsf{BP}}_j[t, t+s)| = \sum_{j=m+1}^{\infty} \mathrm{e}^{-\alpha t}|\mathsf{BP}_j[t, t+s)|. \tag{4.29}$$

Then, (4.27) follows if we show that

$$\mathbb{E}[|\overline{\mathsf{BP}}_{>k(t,x)}[t, t+s)|] \to AF_R(s) - AF_R(s)\Phi(x) = AF_R(s)\Phi(-x). \tag{4.30}$$

Conditioning on the lifetime (with cumulative distribution function equal to F_ξ) of the first individual,

$$\mathbb{E}[|\mathsf{BP}_j[t, t+s)|] = \nu \int_0^t \mathbb{E}[|\mathsf{BP}_{j-1}[t-y, t+s-y)|]\, dF_\xi(y). \tag{4.31}$$

Changing to $\overline{\mathsf{BP}}_j$ and $\bar{F}_\xi$ and iteration of (4.31) yields

$$\mathbb{E}[|\overline{\mathsf{BP}}_{>k(t,x)}[t,t+s)|] = \int_0^t \mathbb{E}[|\overline{\mathsf{BP}}[t-y,t-y+s)|]\, d\bar{F}_\xi^{\star(k(t,x)+1)}(y), \tag{4.32}$$

where $\bar{F}_\xi^{\star j}$ is the j-fold convolution of $\bar{F}_\xi$, and hence the distribution function of the independent sum of j copies of a random variable each having cumulative distribution function. $\bar{F}_\xi$. This is the point where we will use the CLT. Take an arbitrary $\varepsilon > 0$ and take t_0 so large so that for $t > t_0$,

$$|\mathbb{E}[|\overline{\mathsf{BP}}[t,t+s)|] - AF_R(s)| \leq \varepsilon. \tag{4.33}$$

Then,

$$\begin{aligned}\Big|\mathbb{E}[|\overline{\mathsf{BP}}_{>k(t,x)}[t,t+s)|] - AF_R(s)\Phi(-x)\Big| &\leq \varepsilon \bar{F}_\xi^{\star(k(t,x)+1)}(t) + AF_R(s)\big|\bar{F}_\xi^{\star(k(t,x)+1)}(t) - \Phi(-x)\big| \\ &+ \int_{t-t_0}^t \Big|\mathbb{E}[|\overline{\mathsf{BP}}[t-y,t-y+s)|] - AF_R(s)\Big|\, d\bar{F}_\xi^{\star(k(t,x)+1)}(y).\end{aligned} \tag{4.34}$$

The last term vanishes since $\mathbb{E}[|\overline{\mathsf{BP}}[t,t+s)|]$ is uniformly bounded and $\bar{F}_\xi^{\star k(t,x)}(t) - \bar{F}_\xi^{\star k(t,x)}(t-t_0) = o(1)$, as $t \to \infty$. Furthermore, with $m = k(t,x) \to \infty$,

$$k(t,x) \sim \frac{t}{\bar{\nu}} + x\sqrt{t\frac{\bar{\sigma}^2}{\bar{\nu}^3}} \qquad \Longleftrightarrow \qquad t \sim m\bar{\nu} - x\bar{\sigma}\sqrt{m}. \tag{4.35}$$

As a result, by the CLT and the fact that $\bar{\nu}$ and $\bar{\sigma}^2$ are the mean and the variance of the distribution function $\bar{F}_\xi$,

$$\lim_{t\to\infty} \bar{F}_\xi^{\star k(t,x)}(t) = \Phi(-x). \tag{4.36}$$

Together with (4.34), this proves the claim in (4.30), and hence (4.27). Finally we use the second statement of Part (a) to show that

$$\begin{aligned}&\mathrm{e}^{-\alpha t}\sum_{j=0}^{k(t,x)} \mathbb{E}[|\mathsf{BP}_j[t,t+s)|] - \mathrm{e}^{-\alpha t}\sum_{j=0}^{k(t,x)} \mathbb{E}[|\mathsf{BP}_j^{(\vec{m})}[t,t+s)|] \\ &\qquad \leq \mathrm{e}^{-\alpha t}\big(\mathbb{E}[|\mathsf{BP}(t)|] - \mathbb{E}[|\mathsf{BP}^{(\vec{m})}(t)|]\big) \to 0,\end{aligned} \tag{4.37}$$

as first $t \to \infty$ and then $K \to \infty$. This shows Proposition 4.1(c).

We continue with the proof of Proposition 4.1(a) for the n-dependent CTBP. We denote the number of alive individals at time t in the n-dependent CTBP by $|\mathsf{BP}_n(t)|$. We then have to show that for any sequence $t_n \to \infty$, as $n \to \infty$,

$$\mathrm{e}^{-\alpha_n t_n}\mathbb{E}[|\mathsf{BP}_n(t_n)|] \to A, \tag{4.38}$$

where $A = (\nu-1)/\alpha\nu\bar{\nu}$. Denote by $\varphi(s) = \int_0^\infty \mathrm{e}^{-sy} f_\xi(y)\, dy$, the Laplace transform of f_ξ, the density of the lifetime distribution F_ξ. Then

$$\int_0^\infty \mathrm{e}^{-st}\mathbb{E}[|\mathsf{BP}_n(t)|]\, dt = \frac{1-\varphi(s)}{s(1-\nu_n\varphi(s))}. \tag{4.39}$$

This equation follows directly from [20, Equation 16.1], with m replaced by ν_n and is valid when the real part of s satisfying $\mathrm{Re}(s) > \alpha_n$, where $\alpha_n > 0$ is defined as the unique solution to $\nu_n\varphi(\alpha_n) = 1$ (compare (1.15)). From the inversion formula for Laplace transforms, we obtain

$$\mathbb{E}[|\mathsf{BP}_n(t)|] = \frac{1}{2\pi i}\int_\Gamma \mathrm{e}^{st}\frac{1-\varphi(s)}{s(1-\nu_n\varphi(s))}\, ds, \tag{4.40}$$

where Γ is the path $(c_0 - i\infty, c_0 + i\infty)$, with $c_0 > \alpha_n$. Since $\alpha_n \to \alpha$ and $\nu_n \to \nu > 1$ and $\varphi(s)$ is the Laplace transform of a *probability density*, the function $s(1-\nu_n\varphi(s))$ has a simple zero

$s = \alpha_n$, but no other zeros in a small strip $|s-\alpha_n| < \varepsilon$, for some $\varepsilon > 0$. It is then easy to conclude from Cauchy's theorem, calculating the residue at $s = \alpha_n$, that

$$\begin{aligned} \mathbb{E}[|\mathsf{BP}_n(t_n)|] &= \mathrm{e}^{\alpha_n t_n} \frac{1-\varphi(\alpha_n)}{\alpha_n \cdot (-\nu_n \varphi'(\alpha_n))} \big(1 + O(\mathrm{e}^{-\varepsilon t_n})\big) \\ &= A_n \mathrm{e}^{\alpha_n t_n} \big(1 + O(\mathrm{e}^{-\varepsilon t_n})\big), \end{aligned} \tag{4.41}$$

where

$$A_n = \frac{\nu_n - 1}{\alpha_n \nu_n^2 \int_0^\infty y \mathrm{e}^{-\alpha_n y} f_\xi(y)\, dy} = \frac{\nu_n - 1}{\alpha_n \nu_n \bar{\nu}_n}. \tag{4.42}$$

Since $A_n \to A$, the claim (4.38) follows.

For the second statement in Proposition 4.1(a) for the n-dependent CTBP, we replace the inequality in (4.7) by the equivalent n-dependent statement, uniformly in $k \geq 1$,

$$1 - \prod_{j=1}^{k} \frac{\nu_n^{(j)}}{\nu_n} \leq (\log(1/\eta_n))^{-1} \mathbb{E}\big[X_n \log{(X_n/K_n)_+}\big]. \tag{4.43}$$

Since $\lim_{n\to\infty} \mathbb{E}\big[X_n \log(X_n/K_n)_+\big] = 0$, as $n \to \infty$, the statement follows as in (4.10).

For the n-dependent case of Proposition 4.1(b), we need to show that for all $n \geq 1$ and uniformly in t,

$$\mathrm{e}^{-2\alpha_n t} \mathbb{E}\big[|\mathsf{BP}_n^{(\vec{m})}(t)|^2\big] \leq C K_n, \tag{4.44}$$

for some constant C, and where K_n is the cut-off variable used in $m_j = K_n \eta_n^{-j}$. Copying the derivation which leads to (4.16), we obtain

$$\mathbb{E}[|\overline{\mathsf{BP}}_n^{(\vec{m})}|^2] = \sum_{j=1}^{\infty} f_1^{(n)} \cdots f_{j-1}^{(n)} e_j^{(n)} \big(\mathbb{E}[|\overline{\mathsf{BP}}_n{}^{(\vec{m}_j)}|]\big)^2 * Q_n^{j\star}, \tag{4.45}$$

where

$$e_j^{(n)} = \frac{\eta_n \mathbb{E}[(X_n \wedge m_j)^2]}{\nu_n}, \qquad f_j^{(n)} = \frac{\eta_n \mathbb{E}[(X_n \wedge m_j)]}{\nu_n}, \tag{4.46}$$

and

$$\bar{F}_{n,\xi}(t) = \nu_n \int_0^t \mathrm{e}^{-\alpha_n y}\, dF_\xi(y), \qquad Q_n(t) = \eta_n^{-1} \nu_n \int_0^t \mathrm{e}^{-2\alpha_n y}\, dF_\xi(y). \tag{4.47}$$

From the proof of Lemma 4.3, we readily obtain that

$$\sum_{j=1}^{\infty} f_1^{(n)} \ldots f_{j-1}^{(n)} e_j^{(n)} \leq \sum_{j=1}^{\infty} m_j^2 \mathbb{P}(X_n > m_j) \eta_n^j + \sum_{j=1}^{\infty} \mathbb{E}[X_n^2 \mathbb{1}_{\{X_n \leq m_j\}}] \eta_n^j \leq \frac{2K_n \nu_n}{1-\eta_n}. \tag{4.48}$$

Since, $\nu_n \to \nu$ and $\eta_n \to \eta$, as $n \to \infty$, we find, by combining (4.45) and (4.48), that given $\varepsilon > 0$, there is an n_0 so that for $n > n_0$, and uniformly in t,

$$\mathrm{e}^{-2\alpha_n t} \mathbb{E}\big[|\mathsf{BP}_n^{(\vec{m})}(t)|^2\big] \leq \frac{2K_n(\nu + \varepsilon)(A^2 + \varepsilon)}{(1 - \eta - \varepsilon)} \leq C K_n. \tag{4.49}$$

By enlarging the constant C we see that (4.44) holds for all $n \geq 1$ and uniformly in t.

Finally, we will give the proof of Proposition 4.1(c) for the n-dependent CTBP. We denote by $|\mathsf{BP}_{n,j}[t, t+s)|$ the number of individuals in generation j having residual lifetime at most s at time t of the CTBP with offspring given by X_n. Then, we obtain, similarly as in (4.32),

$$\mathbb{E}[|\overline{\mathsf{BP}}_{n,>k}[t, t+s)|] = \int_0^t \mathbb{E}[|\overline{\mathsf{BP}}_n[t-y, t+s-y)|]\, d\bar{F}_{n,\xi}^{\star(k+1)}(y). \tag{4.50}$$

The expectation $\mathbb{E}[\mathsf{BP}_n[t, t+s)|]$ satisfies the renewal equation

$$\mathbb{E}[|\mathsf{BP}_n[t, t+s)|] = F_\xi(t+s) - F_\xi(t) + \nu_n \int_0^t \mathbb{E}[|\mathsf{BP}_n(t-y, t+s-y)|]\, dF_\xi(y). \tag{4.51}$$

For $s>0$ fixed, we denote by

$$\tilde{K}_n(v,s)=\int_0^\infty \mathrm{e}^{-vt}\mathbb{E}[|\mathsf{BP}_n[t,t+s)|]\,dt, \qquad \tilde{f}(v,s)=\int_0^\infty \mathrm{e}^{-vt}[F_\xi(t+s)-F_\xi(t)]\,dt, \tag{4.52}$$

the Laplace transforms of $\mathbb{E}[|\mathsf{BP}_n[t,t+s)|]$ and $[F_\xi(t+s)-F_\xi(t)]$, respectively. Then (4.51) yields $\tilde{K}_n(v,s)=\tilde{f}(v,s)/(1-\nu_n\varphi(v))$. From the inversion theorem for Laplace transforms we obtain (compare (4.40)),

$$\mathbb{E}[|\mathsf{BP}_n[t,t+s)|]=\frac{1}{2\pi i}\int_\Gamma \mathrm{e}^{vt}\frac{\tilde{f}(v,s)}{(1-\nu_n\varphi(v))}\,dv, \tag{4.53}$$

where Γ is the same path as in (4.40), so that from the theory of residues, for some $\varepsilon>0$,

$$\mathbb{E}[|\mathsf{BP}_n[t_n,t_n+s)|]=\frac{\mathrm{e}^{\alpha_n t_n}\tilde{f}(\alpha_n,s)}{-\nu_n\varphi'(\alpha_n)}\big(1+O(\mathrm{e}^{-\varepsilon t_n})\big)=\mathrm{e}^{\alpha_n t_n}A_nF_{n,R}(s)\Big(1+O(\mathrm{e}^{-\varepsilon t_n})\Big), \tag{4.54}$$

with A_n defined in (4.42) and with

$$F_{n,R}(s)=\frac{\alpha_n\nu_n}{\nu_n-1}\int_0^\infty \mathrm{e}^{-\alpha_n y}[F_\xi(y+s)-F_\xi(y)]\,dy. \tag{4.55}$$

Since $A_n\to A$ and $F_{n,R}(s)\to F_R(s)$ for $n\to\infty$, we obtain that, for any sequence $t_n\to\infty$,

$$\lim_{n\to\infty}\mathbb{E}[|\overline{\mathsf{BP}}_n[t_n,t_n+s)|]=\lim_{n\to\infty}\mathrm{e}^{-\alpha_n t_n}\mathbb{E}[|\mathsf{BP}_n[t_n,t_n+s)|]=AF_R(s).$$

The n-dependent definition $k_n(t_n,x)$ yields that $m=k_n(t_n,x)\to\infty$ implies $t_n\sim m\bar{\nu}_n-x\bar{\sigma}\sqrt{m}$, so that since $\bar{\sigma}_n\to\bar{\sigma}$,

$$\bar{F}_{n,\xi}^{\star k_n(t_n,x)}(t_n)\to\Phi(-x). \tag{4.56}$$

Since $\nu_n\to\nu$, $\alpha_n\to\alpha$,we obtain, similarly as in (4.30) and for any sequence $t_n\to\infty$, that

$$\mathbb{E}[|\overline{\mathsf{BP}}_{n,>k_n(t_n,x)}[t_n,t_n+s)|]=\int_0^{t_n}\mathbb{E}[|\overline{\mathsf{BP}}_n[t_n-y,t_n-y+s)|]\,d\bar{F}_{n,\xi}^{\star(k_n(t_n,x)+1)}(y)\to AF_R(s)\Phi(-x). \tag{4.57}$$

The remaining details of the proof follow from Part (a) and an argument as in (4.37). ■

5. Bounds on the coupling: Proof of Proposition 2.2

5.1. Some simple lemmas concerning miscouplings. In (2.26), we have coupled the forward degrees in the SWT $(Y_k^{(n)})_{k\geq 1}$, as well as the (possibly) reduced forward degrees $(X_k^{(n)})_{k\geq 1}$, to an i.i.d. sequence $(B_k^{(n)})_{k\geq 1}$ with distribution equal to that of $D_n^\star-1$ given in (2.5).

We next investigate some simple consequences of this coupling. For this, it will be useful to note that when D_n, having distribution function F_n in (1.2), satisfies Condition 1.1(c), the maximal degree $\Delta_n=\max_{i\in[n]}d_i$ satisfies

$$\Delta_n=o(\sqrt{n/\log n}). \tag{5.1}$$

Indeed, suppose that $\Delta_n\geq\varepsilon\sqrt{n/\log n}$. Then, pick $K_n=n^{1/4}$ to obtain that

$$\begin{aligned}\mathbb{E}[D_n^2\log{(D_n/K_n)_+}] &= \frac{1}{n}\sum_{k=1}^n d_k^2\log(d_k/n^{1/4})_+\geq\frac{\Delta_n^2}{n}\log(\Delta_n/n^{1/4})\\ &\geq n^{-1}(\varepsilon\sqrt{n/\log n})^2\log(n^{1/4}/(\log n)^{1/2})\geq\varepsilon^2/8.\end{aligned} \tag{5.2}$$

This is in contradiction to Condition 1.1(c), so we conclude that (5.1) holds.

On $(\Omega,\mathcal{F},\mathbb{P})$ we define the sigma-algebra $\mathcal{G}_k$ by

$$\mathcal{G}_k=\sigma(d_{U_1},d_{U_2},\tau_j,X_j^{(n)},Y_j^{(n)},(B_{\tau_i}^{(n)})_{i\leq j})_{j\leq k}. \tag{5.3}$$

In the following lemma, we investigate the conditional probability of $Y_k^{(n)}\neq B_{\tau_{k-1}+1}^{(n)}$ given $\mathcal{G}_{k-1}$. In its statement, we recall the definition of $S_k^{(n)}$ in (2.27).

Lemma 5.1 (Miscoupling of forward degree)**.** *Assume that Condition* 1.1*(c) holds. For all* $k \le m_n$*, and assuming that* $m_n \le \sqrt{n \log n}$*,*

$$\mathbb{P}(Y_k^{(n)} \neq B_{\tau_{k-1}+1}^{(n)} \mid \mathcal{G}_{k-1}) \le \frac{1}{\ell_n(1-o(1))}\Big(S_0^{(n)} + \sum_{s=1}^{k-1}(Y_s^{(n)}+1)\Big) = o_{\mathbb{P}}(1). \tag{5.4}$$

Proof. We have that $Y_k^{(n)} \neq B_{\tau_{k-1}+1}^{(n)}$ precisely when we pair the half-edge $y_{k-1}^\star$ to a half-edge of a previously chosen vertex. Now let $V_{z_0}, \ldots, V_{z_{k-2}}$ be the previously chosen vertices and let $Y_s^{(n)} = B_{\tau_s}^{(n)}$, for $s \le k-1$, be the forward degree of vertex $V_{z_{s-1}}$, $s \le k-1$. Then the total number of half-edges incident to chosen vertices is at most

$$S_0^{(n)} + \sum_{s=1}^{k-1}(Y_s^{(n)}+1).$$

By (5.1), $\Delta_n = o(\sqrt{n/\log n})$, so that

$$S_0^{(n)} + \sum_{s=1}^{k-1}(Y_s^{(n)}+1) \le (k+1)\Delta_n \le (m_n+1)\Delta_n = o(n). \tag{5.5}$$

From (5.5), it is clear that we draw each time from at least $\ell_n - o(n) = \ell_n(1-o(1))$ half-edges. This shows (5.4). ∎

Lemma 5.2 (Probability of drawing at least one alive half-edge)**.** *Assume that Condition* 1.1*(c) holds. For all* $k \le m_n$*, and assuming that* $m_n \le \sqrt{n \log n}$*,*

$$\mathbb{P}(X_k^{(n)} < Y_k^{(n)} \mid \mathcal{G}_{k-1}) \le \frac{\mathbb{E}[Y_k^{(n)} \mid \mathcal{G}_{k-1}]}{\ell_n(1-o(1))}\Big(S_0^{(n)} + \sum_{s=1}^{k-1} Y_s^{(n)}\Big). \tag{5.6}$$

Proof. Recall the definition of $S_k^{(n)}$ in (2.27). We have that $X_k^{(n)} < Y_k^{(n)}$ when we pair at least one of the $Y_k^{(n)}$ half-edges to a half-edge incident to $\{U_1, U_2, V_{z_0}, \ldots, V_{z_{k-2}}\}$. Since there are precisely $Y_k^{(n)}$ half-edges that need to be paired, and the number of half-edges incident to $\{U_1, U_2, V_{z_0}, \ldots, V_{z_{k-2}}\}$, given $\mathcal{G}_{k-1}$, equals $S_{k-1}^{(n)}$, we find

$$\mathbb{P}(X_k^{(n)} < Y_k^{(n)} \mid \mathcal{G}_{k-1}, Y_k^{(n)}) \le \frac{Y_k^{(n)} \cdot S_{k-1}^{(n)}}{\ell_n - \sum_{s=1}^{k-1}(Y_s^{(n)}-1) - S_0^{(n)} - 1}. \tag{5.7}$$

Clearly, $S_{k-1}^{(n)} \le S_0^{(n)} + \sum_{s=1}^{k-1} Y_s^{(n)}$. Consequently we obtain (5.6) from the tower-property for conditional expectations. ∎

5.2. **Proof of Proposition 2.2(a).** The i.i.d. sequences $(B_i^{(n)})_{i\ge1}$ and $(B_i)_{i\ge1}$ have probability mass functions $(g_k^{(n)})_{k\ge0}$ and $(g_k)_{k\ge0}$ given in (2.4) and (2.5), respectively. Since $(g_k^{(n)})_{k\ge0}$ and $(g_k)_{k\ge0}$ are discrete distributions and since by Condition 1.1, the distribution $(g_k^{(n)})_{k\ge1}$ converges as $n\to\infty$ in distribution to $(g_k)_{k\ge1}$, it follows that

$$d_{\mathrm{TV}}(B_1^{(n)}, B_1) = \frac{1}{2}\sum_{k=0}^{\infty} |g_k - g_k^{(n)}| \to 0, \tag{5.8}$$

where d_{TV} denotes the total variation distance, see for instance [39, Theorem 6.1].

Take $s_n \to \infty$ such that

$$\mathrm{e}^{2\alpha s_n} d_{\mathrm{TV}}(B_1^{(n)}, B_1) \to 0. \tag{5.9}$$

According to (1.20), and with $i \in \{1,2\}$, we then obtain

$$\mathrm{e}^{-\alpha s_n}|\mathsf{BP}^{(i)}(s_n)| \xrightarrow{a.s.} \widetilde{\mathcal{W}}^{(i)}, \tag{5.10}$$

where $\widetilde{\mathcal{W}}^{(i)}$ are two independent copies of $\widetilde{\mathcal{W}}$. Since $\mathbb{P}(\widetilde{\mathcal{W}}^{(i)} < \infty) = 1$ and $\mathrm{e}^{\alpha s_n} \to \infty$, we conclude that $|\mathsf{BP}(s_n)| \le k_n$, whp, if we take $k_n = \lfloor \mathrm{e}^{2\alpha s_n} \rfloor$. If this k_n does not satisfy $k_n = o(\sqrt{n})$, then we lower s_n so that the corresponding value of $k_n = \lfloor \mathrm{e}^{2\alpha s_n} \rfloor$ does satisfy $k_n = o(\sqrt{n})$.

Recall the definition of $\mathcal{G}_k$ in (5.3). By Boole's inequality,

$$\mathbb{P}(X_k^{(n)} \neq B_k^{(n)} \mid \mathcal{G}_{k-1}) \leq \mathbb{P}(X_k^{(n)} < Y_k^{(n)} \mid \mathcal{G}_{k-1}) + \mathbb{P}(Y_k^{(n)} \neq B_k^{(n)} \mid \mathcal{G}_{k-1}). \tag{5.11}$$

Using Lemmas 5.1–5.2, and by taking the expectation, a lower bound for the probability of all forward-degree-couplings being successful during the first $k_n = o(\sqrt{n})$ pairings is

$$\begin{aligned} \mathbb{P}\big(\bigcap_{k=1}^{k_n}\{X_k^{(n)} = B_k^{(n)}\}\big) &= 1 - \mathbb{P}\big(\bigcup_{k=1}^{k_n}\{X_k^{(n)} \neq B_k^{(n)}\}\big) \qquad (5.12)\\ &\geq 1 - \frac{1}{\ell_n(1-o(1))}\sum_{k=1}^{k_n}\mathbb{E}[S_0^{(n)} + \sum_{s=1}^{k-1}(Y_s^{(n)}+1)]\\ &\quad - \frac{1}{\ell_n(1-o(1))}\sum_{k=1}^{k_n}\mathbb{E}[\mathbb{E}[Y_k^{(n)} \mid \mathcal{G}_{k-1}]\Big(S_0^{(n)} + \sum_{s=1}^{k-1}Y_s^{(n)}\Big)]\\ &\geq 1 - ck_n^2/n \to 1, \end{aligned}$$

where we rely on the inequality

$$\mathbb{E}[Y_k^{(n)} \mid \mathcal{G}_{k-1}] \leq \sum_{j\in[n]} \frac{d_j(d_j-1)}{\ell_n - 2k_n\Delta_n} = \nu_n(1+o(1)), \tag{5.13}$$

whenever $k_n\Delta_n = o(n)$, which follows from (5.1).

The lower bound (5.12) implies that, whp, the number of half-edges $(|\mathsf{AH}(s)|)_{s\in[0,s_n]}$ is perfectly coupled to the number of (alive) individuals of the n-dependent CTBP $(|\mathsf{AI}_n(s)|)_{s\in[0,s_n]}$, in turn, (5.9) shows that $(\mathsf{BP}_n(s))_{s\leq s_n}$ is whp perfectly coupled to $(\mathsf{BP}(s))_{s\leq s_n}$. This proves Proposition 2.2(a). ■

We close this section by investigating moments of the size-biased random variables $(Y_k^{(n)})_{k\geq 1}$, which play a crucial role in the remainder of the paper:

Lemma 5.3 (Moments of size-biased samplings)**.** *Assume that Condition 1.1 (a-c) holds. For all $k \leq m_n$, and assuming that $m_n \leq \sqrt{n\log n}$, and for any $K_n \to \infty$ such that $K_n^2 = o(n/m_n)$,*

$$\mathbb{E}[Y_k^{(n)}\mathbb{1}_{\{Y_k^{(n)}\leq K_n\}} \mid \mathcal{G}_{k-1}] = (1+o_{\mathbb{P}}(1))\nu_n, \tag{5.14}$$

$$\mathbb{E}[Y_k^{(n)}\mathbb{1}_{\{Y_k^{(n)}> K_n\}} \mid \mathcal{G}_{k-1}] = o_{\mathbb{P}}(1). \tag{5.15}$$

Proof. We use the upper bound

$$\mathbb{E}[Y_k^{(n)}\mathbb{1}_{\{Y_k^{(n)}\geq a\}} \mid \mathcal{G}_{k-1}] \leq \frac{1}{\ell_n(1-o(1))}\sum_{l\in[n]} d_l(d_l-1)\mathbb{1}_{\{d_l\geq a\}}, \tag{5.16}$$

where we again use that, since $m_n \leq \sqrt{n\log n}$,

$$\ell_n - S_0^{(n)} - \sum_{j=1}^{k-1} Y_j^{(n)} \geq \ell_n - (m_n+1)\Delta_n = \ell_n(1-o(1)). \tag{5.17}$$

This provides the necessary upper bound in (5.14) by taking $a = 0$ and from the identity $\nu_n = \sum_{l\in[n]} d_l(d_l-1)/\ell_n$. For (5.15), this also proves the necessary bound, since by Condition 1.1(c),

$$\frac{1}{\ell_n}\sum_{l\in[n]} d_l(d_l-1)\mathbb{1}_{\{d_l\geq K_n\}} = o(1). \tag{5.18}$$

For the lower bound in (5.14), we bound, instead,

$$\mathbb{E}[Y_k^{(n)}\mathbb{1}_{\{Y_k^{(n)}\leq K_n\}} \mid \mathcal{G}_{k-1}] \geq \frac{1}{\ell_n(1-o(1))}\Big[\sum_{l\in[n]} d_l(d_l-1)\mathbb{1}_{\{d_l\leq K_n\}} - \sum_{l\in[n]} d_l(d_l-1)\mathbb{1}_{\{d_l\leq K_n\}}\mathbb{1}_{\{l \text{ is chosen}\}}\Big]. \tag{5.19}$$

where the event 'l is chosen' means that vertex l belongs to the set of already chosen vertices $U_1, U_2, V_{z_0}, \dots, V_{z_{m_n-2}}$. The first term equals $\nu_n(1+o(1))$. The second term is a.s. bounded by $m_n K_n^2/\ell_n = o(1)$, since $K_n^2 = o(n/m_n)$. ∎

5.3. **Completing the coupling: Proof of Proposition 2.2(b).** In this section, we use Proposition 4.1 to prove Proposition 2.2(b). In order to bound the difference between $\mathsf{BP}(t)$ and $\mathsf{SWT}(t)$, we will introduce several events. Let $B_n, C_n, \varepsilon_n, \overline{m}_n, \underline{m}_n$ denote sequences of real numbers for which $B_n, C_n \to \infty$ and $\varepsilon_n \to 0$ arbitrarily slowly, and $\overline{m}_n \gg \sqrt{n}, \underline{m}_n \ll \sqrt{n}$. Later in this proof, we will formulate precisely how to choose these sequences.

Define the event $\mathcal{A}_n$ by

$$\mathcal{A}_n = \Big\{ \Big| \bigcup_{s\in[0,t_n+B_n]} \mathsf{SWT}(s) \triangle \mathsf{BP}_n(s) \Big| < \varepsilon_n \sqrt{n} \Big\}, \tag{5.20}$$

where we recall that $\mathsf{SWT}(s)\triangle\mathsf{BP}_n(s)$ is the set of alive half-edges at time s that are miscoupled, and where we recall further that an alive half-edge is miscoupled if the shortest-weight path from the root to the vertex incident to that half-edge uses at least one degree-miscoupled vertex. Similarly an alive individual is miscoupled if at least one of its ancestors is degree miscoupled. Note that when $\mathcal{A}_n$ holds, then $\big|\bigcup_{s\in[0,t]} \mathsf{SWT}(s)\triangle\mathsf{BP}_n(s)\big| < \varepsilon_n\sqrt{n}$ for *any* $t \le t_n + B_n$ by monotonicity in t (see Definition (2.35)).

In terms of the above notation, Proposition 2.2(b) can be reformulated as

$$\mathbb{P}(\mathcal{A}_n^c \mid \mathcal{F}_{s_n}) = o_{\mathbb{P}}(1). \tag{5.21}$$

Hence only a tiny fraction of the alive half-edges or individuals is miscoupled and the alive half-edges that are not miscoupled are connected to the root via a path containing only successfully coupled vertices.

In order to prove (5.21), we introduce the following events:

$$\begin{aligned} \mathcal{B}_n &= \{Y^{(\mathrm{BP})}(t_n+B_n) \le \overline{m}_n\} \cap \{Y^{(\mathrm{SWT})}(t_n+B_n) \le \overline{m}_n\} \\ &\qquad \cap \{Y^{(\mathrm{BP})}(t_n-B_n) \le \underline{m}_n\} \cap \{Y^{(\mathrm{SWT})}(t_n-B_n) \le \underline{m}_n\}, \end{aligned} \tag{5.22}$$

$$\mathcal{C}_n = \{|\mathsf{SWT}(t)| = |\mathsf{BP}_n(t)|,\ \forall t \le t_n - B_n\}, \tag{5.23}$$

$$\mathcal{D}_n = \{\nexists\, i \text{ such that } T_i \le t_n + B_n, X_i^{(n)} \ne B^{(n)}_{\tau_{i-1}+1}, d_{V_{z_i}} \ge C_n\}, \tag{5.24}$$

where

$$Y^{(\mathrm{BP})}(t) = |\{v\colon v \in \mathsf{BP}_n(s) \text{ for some } s \le t\}|, \tag{5.25}$$

denotes the total number of individuals ever born into the BP_n before time t and

$$Y^{(\mathrm{SWT})}(t) = |\{v\colon v \in \mathsf{AH}(s) \text{ for some } s \le t\}|, \tag{5.26}$$

denotes the number of half-edges in the SWT that have ever been alive before time t. Informally, on $\mathcal{B}_n$, the total number of half-edges in SWT and individuals in the CTBP are not too large. On $\mathcal{C}_n$, there is no early degree-miscoupled vertex, while on $\mathcal{D}_n$, there is no degree miscoupling involving a vertex that has high degree, until a late stage.

Obviously,

$$\begin{aligned} &\mathbb{P}(\mathcal{A}_n^c \mid \mathcal{F}_{s_n}) \\ &\quad \le \mathbb{P}(\mathcal{B}_n^c \mid \mathcal{F}_{s_n}) + \mathbb{P}(\mathcal{C}_n^c \cap \mathcal{B}_n \mid \mathcal{F}_{s_n}) + \mathbb{P}(\mathcal{D}_n^c \cap \mathcal{B}_n \cap \mathcal{C}_n \mid \mathcal{F}_{s_n}) + \mathbb{P}(\mathcal{A}_n^c \cap \mathcal{B}_n \cap \mathcal{C}_n \cap \mathcal{D}_n \mid \mathcal{F}_{s_n}). \end{aligned} \tag{5.27}$$

To bound conditional probabilitites of the form $\mathbb{P}(\mathcal{E}^c \mid \mathcal{F}_{s_n})$ as appearing in (5.21), we note that it suffices to prove that $\mathbb{P}(\mathcal{E}^c) = o(1)$, since then, by the Markov inequality and for every $\varepsilon > 0$,

$$\mathbb{P}\Big(\mathbb{P}(\mathcal{E}^c \mid \mathcal{F}_{s_n}) \ge \varepsilon\Big) \le \mathbb{E}[\mathbb{P}(\mathcal{E}^c \mid \mathcal{F}_{s_n})]/\varepsilon = \mathbb{P}(\mathcal{E}^c)/\varepsilon = o(1). \tag{5.28}$$

Thus, we are left to prove that

$$\mathbb{P}(\mathcal{B}_n^c) = o(1), \quad \mathbb{P}(\mathcal{C}_n^c \cap \mathcal{B}_n) = o(1), \quad \mathbb{P}(\mathcal{D}_n^c \cap \mathcal{B}_n \cap \mathcal{C}_n) = o(1), \quad \mathbb{P}(\mathcal{A}_n^c \cap \mathcal{B}_n \cap \mathcal{C}_n \cap \mathcal{D}_n) = o(1). \tag{5.29}$$

We will do so in the above order.

Lemma 5.4 (Expected number of particles born)**.** *For all $t \geq 0$,*

$$\mathbb{E}[Y^{(\mathrm{BP})}(t)] = 2\Big(1 - \frac{\mathbb{E}[D_n]F_\xi(t)}{\nu_n - 1}\Big) + \frac{\nu_n}{\nu_n - 1}\mathbb{E}[|\mathsf{BP}_n(t)|]. \tag{5.30}$$

Moreover, when $\mathrm{e}^{\alpha_n(t_n+B_n)} = o(\overline{m}_n)$ and $\mathrm{e}^{\alpha_n(t_n-B_n)} = o(\underline{m}_n)$,

$$\mathbb{P}(\mathcal{B}_n^c) = o(1). \tag{5.31}$$

Proof. Note that we grow two sets of alive half-edges and two BP's, which explains the factor 2 in (5.30). As is well known, the expected number of descendants in generation k of a BP equals ν_n^k, where ν_n denotes the mean offspring. Here, we deal with a delayed BP_n where in the first generation the mean number of offspring equals $\mathbb{E}[D_n]$; the factor $F_\xi^{\star k}(t) - F_\xi^{\star(k+1)}(t)$ represents the probability that an individual of generation k is alive at time t. Together this yields

$$\mathbb{E}[|\mathsf{BP}_n(t)|] = \sum_{k=1}^\infty 2\mathbb{E}[D_n]\nu_n^{k-1}[F_\xi^{\star k}(t) - F_\xi^{\star(k+1)}(t)], \qquad \mathbb{E}[Y^{(\mathrm{BP})}(t)] = 2 + \sum_{k=1}^\infty 2\mathbb{E}[D_n]\nu_n^{k-1}F_\xi^{\star k}(t). \tag{5.32}$$

It is not difficult to deduce (5.30) from the two identies above.

To bound $\mathbb{P}(\mathcal{B}_n^c)$, we note that we have to bound events of the form $\mathbb{P}(Y^{(\mathrm{BP})}(t) \geq m)$ and $\mathbb{P}(Y^{(\mathrm{SWT})}(t) \geq m)$ for various choices of m and t. We use the Markov inequality and (5.30) to bound

$$\mathbb{P}(Y^{(\mathrm{BP})}(t) \geq m) \leq \mathbb{E}[Y^{(\mathrm{BP})}(t)]/m \leq \frac{\nu_n}{m(\nu_n - 1)}\mathbb{E}[|\mathsf{BP}_n(t)|] + \frac{2}{m}. \tag{5.33}$$

According to (4.41), after conditioning on the offspring of the first individual, $\mathbb{E}[|\mathsf{BP}_n(t_n)|] = \mathbb{E}[D_n]A_n\mathrm{e}^{\alpha_n t_n}(1 + o(1))$, so that, since $\mathbb{E}[D_n] \to \mathbb{E}[D]$,

$$\mathbb{P}(Y^{(\mathrm{BP})}(t_n) \geq m_n) = \Theta(\mathrm{e}^{\alpha_n t_n}/m_n). \tag{5.34}$$

The conditions on t and m in Lemma 5.4 have been chosen precisely so that $\mathrm{e}^{\alpha_n(t_n-B_n)}/\underline{m}_n \to 0$, and $\mathrm{e}^{\alpha_n(t_n+B_n)}/\overline{m}_n \to 0$.

We continue with $\mathbb{P}(Y^{(\mathrm{SWT})}(t) \geq m)$. We use the same steps as above, and start by computing

$$\mathbb{E}[Y^{(\mathrm{SWT})}(t)] = 2 + 2\sum_{k=0}^\infty F_\xi^{\star k}(t)\mathbb{E}[P_k^\star], \quad k \geq 1, \tag{5.35}$$

where $P_0^\star = \ell_n/n$ and

$$P_k^\star = \sum_{|\pi|=k, \pi\subseteq \mathrm{CM}_n(\boldsymbol{d})} (d_\pi - 1)/n, \qquad k \geq 1,$$

is the sum of the number of half-edges at the ends of paths of lengths k in $\mathrm{CM}_n(\boldsymbol{d})$, from a uniformly selected starting point. Following [26, Proof of Lemma 5.1], we find that

$$\mathbb{E}[P_k^\star] = \frac{1}{n}\sum_{v_0,\dots,v_k} d_{v_0}\prod_{i=1}^k \frac{d_{v_i}(d_{v_i}-1)}{\ell_n - 2i + 1} \leq \mathbb{E}[D_n]\nu_n^k, \tag{5.36}$$

where the sum is taken over distinct vertices in $[n]$. Note that our definition of ν_n deviates from the one given in [26, (2.3)], which explains the difference between the right-hand side of (5.36) and the result in [26]. We obtain,

$$\mathbb{E}[Y^{(\mathrm{SWT})}(t)] \leq 2 + 2\sum_{k=0}^\infty F_\xi^{\star k}(t)\mathbb{E}[D_n]\nu_n^k \leq 2\mathbb{E}[D_n] + \nu_n\mathbb{E}[Y^{(\mathrm{BP})}(t)], \tag{5.37}$$

and we can repeat our arguments for $\mathbb{E}[Y^{(\mathrm{BP})}(t)]$. ∎

Lemma 5.5 (No early degree-miscoupling)**.** *When $\underline{m}_n = o(\sqrt{n})$, then*

$$\mathbb{P}(\mathcal{C}_n^c \cap \mathcal{B}_n) = o(1). \tag{5.38}$$

Proof. On $\mathcal{B}_n$, the inequality $Y^{(\mathrm{SWT})}(t_n - B_n) \leq \underline{m}_n$ holds. By (5.12), the probability that there exists a degree-miscoupling before the draw of the $\underline{m}_n$th half-edge is $o(1)$ when $\underline{m}_n = o(\sqrt{n})$. ∎

Lemma 5.6 (No late miscouplings of high degree)**.** *If $\overline{m}_n \leq \sqrt{n\log n}$, and C_n satisfies*

$$\frac{\overline{m}_n^2}{\ell_n}\sum_{i\in[n]} d_i^2 \mathbb{1}_{\{d_i\geq C_n\}} = o(n), \tag{5.39}$$

then

$$\mathbb{P}(\mathcal{D}_n^c\cap\mathcal{B}_n\cap\mathcal{C}_n) = o(1). \tag{5.40}$$

Proof. On $\mathcal{B}_n$, we have that $Y^{(\mathrm{SWT})}(t_n+B_n)\leq \overline{m}_n$, and hence for a degree-miscoupling, when $\mathcal{D}_n^c$ holds, one of the vertices $i\in[n]$ with $d_i\geq C_n$ has to be chosen twice during the first $\overline{m}_n$ pairings. By Boole's inequality and (5.39), an upper bound for this probability is

$$\overline{m}_n^2\sum_{i\in[n]}\frac{d_i^2}{(\ell_n-o(n))^2}\mathbb{1}_{\{d_i\geq C_n\}} = \frac{\overline{m}_n^2}{(\ell_n-o(n))^2}\cdot\frac{\ell_n o(n)}{\overline{m}_n^2} = o(1). \tag{5.41}$$

Proposition 5.7 (Degree-miscoupled half-edges have small offspring)**.** *If $\overline{m}_n \leq \sqrt{n\log n}$ and $\mathrm{e}^{2\alpha_n B_n}C_n\overline{m}_n^2/\ell_n = o(\varepsilon_n\sqrt{n})$, then*

$$\mathbb{P}(\mathcal{A}_n^c\cap\mathcal{B}_n\cap\mathcal{C}_n\cap\mathcal{D}_n) = o(1). \tag{5.42}$$

Proof. We split the proof into the contribution of the degree-miscoupled vertices in $|\mathsf{BP}_n(t)\setminus\mathsf{SWT}(t)|$ and those in $|\mathsf{SWT}(t)\setminus\mathsf{BP}_n(t)|$, $t\leq t_n+B_n$.

A bound on $|\mathsf{BP}_n(t)\setminus\mathsf{SWT}(t)|$. Recall that at time T_k, the vertex $V_{z_{k-1}}$ is degree-miscoupled when one of the equalities in (2.26) fails.

When $Y_k^{(n)}\neq B_{\tau_{k-1}+1}^{(n)}$, we can give an upper bound on the contribution to $\mathsf{BP}_n(\cdot)$ of the tress of miscoupled individuals by drawing from the i.i.d. sequence $(B_i^{(n)})_{i\geq 1}$. As a result, the total contribution to $|\mathsf{BP}_n(t)\setminus\mathsf{SWT}(t)|$, $t\leq t_n+B_n$ can be bounded above by

$$Y_k^{(\mathrm{BP})}(t_n+B_n-T_k), \tag{5.43}$$

where, for different $k\geq 1$, $(Y_k^{(\mathrm{BP})}(t))_{t\geq 0}$ are independent CTBPs. On $\mathcal{C}_n$, we have that $T_k\geq t_n-B_n$ so that $t_n+B_n-T_k\leq 2B_n$, while on $\mathcal{D}_n$, each degree-miscoupling starts with a vertex with degree at most C_n. Therefore, from (4.41),

$$\mathbb{E}\Big[Y_k^{(\mathrm{BP})}(t_n+B_n-T_k)\mathbb{1}_{\mathcal{C}_n\cap\mathcal{D}_n}\Big]\leq C_nA_n\mathrm{e}^{2\alpha_nB_n}(1+o(1)). \tag{5.44}$$

On $\mathcal{B}_n$, the expected number of miscoupling is at most $O(\overline{m}_n^2/\ell_n)$, hence

$$\mathbb{E}\Big[\sum_{\{k\colon T_k\leq t_n+B_n\}}Y_k^{(\mathrm{BP})}(t_n+B_n-T_k)\mathbb{1}_{\mathcal{B}_n\cap\mathcal{C}_n\cap\mathcal{D}_n}\Big]\leq O\Big(\frac{\overline{m}_n^2}{\ell_n}\Big)C_n\mathrm{e}^{2\alpha_nB_n}. \tag{5.45}$$

By assumption, the right-hand side is $o(\varepsilon_n\sqrt{n})$. Therefore, by the Markov inequality,

$$\mathbb{P}\Big(|\{\mathsf{BP}_n(s)_{s\in[0,t_n+B_n]}\setminus SWT(s)_{s\in[0,t_n+B_n]}\}|\geq\varepsilon_n\sqrt{n}\}\cap\mathcal{B}_n\cap\mathcal{C}_n\cap\mathcal{D}_n\cap\mathsf{MISC}\Big) = o(1), \tag{5.46}$$

where the intersection with MISC indicates that we only deal with degree-miscouplings of the form $Y_k^{(n)}\neq B_{\tau_{k-1}+1}^{(n)}$.

When $Y_k^{(n)}\neq X_k^{(n)}$, a self-loop or cycle-creating event occurs and the two half-edges that form the last edge in the cycle are removed from $\mathsf{SWT}(t)$, but they are kept in $\mathsf{BP}_n(t)$. In case of a removal of a collision edge, only one individual is kept in $\mathsf{BP}_n(t)$. Whether one or two individuals are kept in the $\mathsf{BP}_n(t)$ is of no consequence for the argument below.

Again, on the event $\mathcal{B}_n\cap\mathcal{C}_n$, the expected number of degree miscouplings is bounded by $O(\overline{m}_n^2/\ell_n)$. Furthermore, on the event $\mathcal{B}_n\cap\mathcal{C}_n$, the expected offspring of the half-edges involved in cycle-creating events is at most

$$\nu_n\mathbb{E}[Y^{(\mathrm{BP})}(2B_n)], \tag{5.47}$$

where $(Y^{(\mathrm{BP})}(t))_{t\geq 0}$ is the total number of individuals that have ever been alive in a CTBP where *all* individuals have i.i.d. offspring with law $(g_k^{(n)})_{k\geq 0}$. Indeed, we have no information about the remaining lifetime of the half-edge involved in an event that is caused by $Y_k^{(n)}\neq X_k^{(n)}$. As a result, rather than waiting for the residual life-time to be completed, we *instantaneously* take as offspring an i.i.d. draw from $(g_k^{(n)})_{k\geq 0}$, and start the various $\mathsf{BP}_n(t)$ from there (on the average there are ν_n of these BP_n). The total number of individuals ever alive only increases by this change. On the event $\mathcal{D}_n$, we have that $\mathbb{E}[Y^{(\mathrm{BP})}(2B_n)] \leq C_n A_n \mathrm{e}^{2\alpha_n B_n}(1+o(1))$. By assumption, $\overline{m}_n^2 C_n A_n \mathrm{e}^{2\alpha_n B_n}/\ell_n = o(\varepsilon_n\sqrt{n})$. Therefore, the total contribution to $|\mathsf{BP}_n(t)\setminus \mathsf{SWT}(t)|$, $t\leq t_n+B_n$, due to degree miscouplings of the kind $Y_k^{(n)}\neq X_k^{(n)}$ events is $o_{\mathbb{P}}(\varepsilon_n\sqrt{n})$, as required.

A bound on $|\mathsf{SWT}(t)\setminus \mathsf{BP}_n(t)|$. By construction, the number of miscoupled half-edges in $\mathsf{SWT}(s)_{s\in[0,t]}$ at any time t is bounded from above by

$$\sum_{j=1}^{\mathsf{MIS}(t)} Y_j^{(\mathrm{SWT})}(t-T_j), \tag{5.48}$$

where $\mathsf{MIS}(t)$ denotes the number of degree-miscoupled vertices and $Y_j^{(\mathrm{SWT})}(t-T_j)$ is the number of half-edges reached by the liquid during $[T_j,t)$, and which are in the tree with root $V_{z_{j-1}}$. On the event $\mathcal{C}_n$, we have that $T_1\geq t_n-B_n$. Therefore, on the event $\mathcal{C}_n$,

$$|\mathsf{SWT}(t)_{t\in[0,t_n+B_n]}\setminus \mathsf{BP}_n(t)_{t\in[0,t_n+B_n]}| \leq \sum_{j=1}^{\mathsf{MIS}(t_n+B_n)} Y_j^{(\mathrm{SWT})}(2B_n). \tag{5.49}$$

By the Markov inequality,

$$\begin{aligned}&\mathbb{P}\Big(\{|\mathsf{SWT}(t)_{t\in[0,t_n+B_n]}\setminus \mathsf{BP}_n(t)_{t\in[0,t_n+B_n]}|\geq \varepsilon_n\sqrt{n}\}\cap\mathcal{B}_n\cap\mathcal{C}_n\cap\mathcal{D}_n\Big)\\ &\quad\leq (\varepsilon_n\sqrt{n})^{-1}\mathbb{E}\Big[\mathbb{1}_{\mathcal{B}_n\cap\mathcal{D}_n}\sum_{j=1}^{\mathsf{MIS}(t_n+B_n)} Y_j^{(\mathrm{SWT})}(2B_n)\Big].\end{aligned} \tag{5.50}$$

We rewrite

$$\begin{aligned}&\mathbb{E}\Big[\mathbb{1}_{\mathcal{B}_n\cap\mathcal{D}_n}\sum_{j=1}^{\mathsf{MIS}(t_n+B_n)} Y_j^{(\mathrm{SWT})}(2B_n)\Big]\\ &\quad\leq (1+o(1))\sum_{j\in[n]}\mathbb{P}(j\ \text{ is degree-miscoupled}, \mathcal{B}_n\cap\mathcal{D}_n)\mathbb{E}[Y^{(\mathrm{SWT})}(2B_n)]\\ &\quad\leq (1+o(1))\sum_{j\in[n]}\Big(\frac{d_j\overline{m}_n}{\ell_n}\Big)^2\mathbb{1}_{\{d_j<C_n\}}\mathbb{E}[Y^{(\mathrm{SWT})}(2B_n)],\end{aligned} \tag{5.51}$$

where we use that, upon degree-miscoupling of vertex j, we redraw a vertex from the size-biased distribution, for which the number of half-edges found before time $2B_n$ is equal to $\mathbb{E}[Y^{(\mathrm{SWT})}(2B_n)](1+o(1))$ since $\overline{m}_n\leq\sqrt{n\log n}$ and $\mathcal{B}_n$ occurs. Since $\mathbb{E}[Y^{(\mathrm{SWT})}(t)]\leq 2\mathbb{E}[D_n]+\nu_n\mathbb{E}[Y^{(\mathrm{BP})}(t)]$, we obtain that

$$\mathbb{E}[Y^{(\mathrm{SWT})}(2B_n)]\leq \nu_n A_n\mathrm{e}^{2\alpha_n B_n}(1+o(1)). \tag{5.52}$$

Therefore, we arrive at

$$\mathbb{E}\Big[\mathbb{1}_{\mathcal{B}_n\cap\mathcal{D}_n}\sum_{j=1}^{\mathsf{MIS}(t_n+B_n)} Y_j^{(\mathrm{SWT})}(2B_n)\Big]\leq \nu_n A_n\mathrm{e}^{2\alpha_n B_n}(1+o(1))\sum_{j\in[n]}\Big(\frac{d_j\overline{m}_n}{\ell_n}\Big)^2\mathbb{1}_{\{d_j<C_n\}}. \tag{5.53}$$

Bounding $\sum_{j\in[n]} d_j^2\mathbb{1}_{\{d_j<C_n\}}\leq C_n\ell_n$, the right-hand side of (5.53) is bounded by $\nu_n A\mathrm{e}^{2\alpha_n B_n}(1+o(1))C_n\overline{m}_n^2/\ell_n = o(\varepsilon_n\sqrt{n})$. Combining this with (5.50) proves that $|\mathsf{SWT}(t)_{t\in[t_n+B_n]}\setminus \mathsf{BP}_n(t)_{t\in[t_n+B_n]}| = o_{\mathbb{P}}(\varepsilon_n\sqrt{n})$ on $\{\mathcal{B}_n\cap\mathcal{C}_n\cap\mathcal{D}_n\}$. ∎

Proof of Proposition 2.2(b). Take

$$\underline{m}_n=\sqrt{n}/(\log\log n)^{\alpha/2},\qquad \overline{m}_n=\sqrt{n}(\log n)^{1/4}, \tag{5.54}$$

and

$$B_n = \log\log\log n, \qquad C_n = n^{1/4}, \qquad \varepsilon_n = 1/\log n. \tag{5.55}$$

By Condition 1.1(c) applied with $K_n = \sqrt{C_n}/\mathrm{e}$, and using that $\log\sqrt{C_n} \leq \log(\mathrm{e}D_n/\sqrt{C_n})_+$ when $D_n \geq C_n$,

$$\frac{1}{n}\sum_{i\in[n]} d_i^2 \mathbb{1}_{\{d_i\geq C_n\}} = \mathbb{E}[D_n^2\mathbb{1}_{\{D_n\geq C_n\}}] \leq \mathbb{E}\Big[\frac{D_n^2\log(\mathrm{e}\cdot D_n/\sqrt{C_n})_+}{\log\sqrt{C_n}}\Big] = o((\log n)^{-1}), \tag{5.56}$$

which verifies (5.39). All other conditions in Lemmas 5.4–5.6 and Proposition 5.7 are straightforward. Therefore, (5.21) follows, which completes the proof of Proposition 2.2(b). ∎

6. Height CLT and stable age: Proof of Proposition 2.3

We first prove Proposition 2.3(a). Throughout this proof, we abbreviate $k_n = k_n(t_n, x)$ as in (2.42). The proof contains several key steps:

Reduction to a single BP. We start by showing that, in order for Proposition 2.3(a) to hold, it suffices to prove that for $j \in \{1,2\}$, $x,t\in\mathbb{R}$ and $s>0$, such that $t+s<B_n$, conditionally on $\mathcal{F}_{s_n}$ and on $\mathcal{W}^{(1)}_{s_n}\mathcal{W}^{(2)}_{s_n}>0$,

$$\mathrm{e}^{-\alpha_n t_n}|\mathsf{BP}^{(j)}_{n,\leq k_n}[\bar t_n+t,\bar t_n+t+s)| \xrightarrow{\mathbb{P}} \mathrm{e}^{\alpha t}\Phi(x)F_R(s)\sqrt{\mathcal{W}^{(j)}/\mathcal{W}^{(3-j)}}, \tag{6.1}$$

where we use (2.37) in Proposition 2.2(a) to see that $\sqrt{\mathcal{W}^{(j)}/\mathcal{W}^{(3-j)}} \in [\varepsilon, 1/\varepsilon]$ whp. Indeed, by Proposition 2.2(b) and the fact that $\mathrm{e}^{-\alpha_n\bar t_n} = \Theta_{\mathbb{P}}(n^{-1/2})$, (6.1) implies that for $t+s<B_n$,

$$\begin{aligned}\mathrm{e}^{-\alpha_n t_n}|\mathsf{SWT}^{(j)}_{\leq k_n}[\bar t_n+t,\bar t_n+t+s)| &= \mathrm{e}^{-\alpha_n t_n}|\mathsf{BP}^{(j)}_{n,\leq k_n}[\bar t_n+t,\bar t_n+t+s)| + \mathrm{e}^{-\alpha_n t_n}o_{\mathbb{P}}(\varepsilon_n\sqrt{n})\\ &\xrightarrow{\mathbb{P}} \mathrm{e}^{\alpha t}\Phi(x)F_R(s)\sqrt{\mathcal{W}^{(j)}/\mathcal{W}^{(3-j)}},\end{aligned} \tag{6.2}$$

which proves Proposition 2.3(a) by the independence of the two CTBPs involved.

Using the branching property. To prove (6.1), we note that $(\mathsf{BP}^{(j)}_n(s))_{s\geq s_n}$ is the collection of alive individuals in the different generations of a CTBP, starting from the alive particles in $(\mathsf{BP}^{(j)}_n(s_n))$. Then, conditionally on $\mathcal{F}_{s_n}$,

$$|\mathsf{BP}^{(j)}_{n,\leq k_n}[\bar t_n+t,\bar t_n+t+s)| = \sum_{i\in\mathsf{BP}^{(j)}_n(s_n)}\sum_{k=1}^{k_n-G_i^{(j)}} |\mathsf{BP}^{(i,j)}_{n,k}[\bar t_n+t-s_n-R_i,\bar t_n+t+s-s_n-R_i)|, \tag{6.3}$$

where $G_i^{(j)}$ is the generation of individual $i\in\mathsf{BP}^{(j)}_n(s_n)$, while $R_i = R_i(s_n)$ is its remaining lifetime at time s_n, and $(\mathsf{BP}^{(i,j)}(t))_{t\geq 0}$ are i.i.d. CTBPs for different i, for which the offspring for each individual has distribution $(g_k^{(n)})_{k\geq 0}$, and where the branching process starts with one individual that dies immediately and has offspring distributed as $(g_k^{(n)})_{k\geq 0}$.

Truncating the branching process. We continue by proving that we can truncate the branching process at the expense of an error term that converges to zero in probability. We let $\mathsf{BP}^{(i,j,\vec m)}_n$ denote the branching process $\mathsf{BP}^{(i,j)}_n$ obtained by truncating particles in generation l (measured from the root i) by $m_l = K_n\eta_n^{-l}$. We take $K_n\to\infty$ such that $K_n\mathrm{e}^{-\alpha_n s_n} = o(1)$. We first show that, as $t_n\to\infty$, we can replace $\mathrm{e}^{-\alpha_n t_n}|\mathsf{BP}^{(i,j)}_{n,\leq k_n}[\bar t_n,\bar t_n+s)|$ by $\mathrm{e}^{-\alpha_n t_n}|\mathsf{BP}^{(i,j,\vec m)}_{n,\leq k_n}[\bar t_n,\bar t_n+s)|$, at the expense of a $o_{\mathbb{P}}(1)$-term. Indeed, with

$$|\mathsf{BP}^{(i,j)}_n(t)| = \sum_{k=1}^\infty |\mathsf{BP}^{(i,j)}_{n,k}(t)|, \qquad |\mathsf{BP}^{(i,j,\vec m)}_n(t)| = \sum_{k=1}^\infty |\mathsf{BP}^{(i,j,\vec m)}_{n,k}(t)|, \tag{6.4}$$

by the n-dependent version of Proposition 4.1(a) in Proposition 4.1(d), which we apply to each of the individuals born at time s_n+R_j, and for each sequence $u_n\to\infty$, we have that

$$\begin{aligned}&\mathrm{e}^{-\alpha_n u_n}\mathbb{E}\Big[|\mathsf{BP}^{(i,j)}_{n,\leq k}(u_n)| - |\mathsf{BP}^{(i,j,\vec m)}_{n,\leq k}(u_n)|\Big]\\ &\quad\leq \mathrm{e}^{-\alpha_n u_n}\mathbb{E}\Big[|\mathsf{BP}^{(i,j)}_n(u_n)| - |\mathsf{BP}^{(i,j,\vec m)}_n(u_n)|\Big] = o(1).\end{aligned}$$

Therefore, using that the law of $(\mathsf{BP}^{(i,j)}_{n,\le k_n}(t))_{t\ge 0}$ only depends on $\mathcal{F}_{s_n}$ through $R_i, \bar{t}_n$,

$$\begin{aligned}
&\mathrm{e}^{-\alpha_n t_n} \sum_{i\in \mathsf{BP}^{(j)}_n(s_n)} \sum_{k=1}^{k_n-G^{(j)}_i} \mathbb{E}\Big[|\mathsf{BP}^{(i,j)}_{n,k}(\bar{t}_n+t-s_n-R_i)| - |\mathsf{BP}^{(i,j,\vec{m})}_{n,k}(\bar{t}_n+t-s_n-R_i)| \mid \mathcal{F}_{s_n}\Big]\\
&\quad\le \mathrm{e}^{-\alpha_n t_n} \sum_{i\in \mathsf{BP}^{(j)}_n(s_n)} \mathbb{E}\Big[|\mathsf{BP}^{(i,j)}_{n,\le k_n}(\bar{t}_n+t-s_n-R_i)| - |\mathsf{BP}^{(i,j,\vec{m})}_{n,\le k_n}(\bar{t}_n+t-s_n-R_i)| \mid R_i,\bar{t}_n\Big]\\
&\quad= o(1) \sum_{i\in \mathsf{BP}^{(j)}_n(s_n)} \mathrm{e}^{\alpha_n(\bar{t}_n-t_n+t-s_n-R_i)} = o_{\mathbb{P}}(1)\mathrm{e}^{-\alpha s_n} \sum_{i\in \mathsf{BP}^{(j)}_n(s_n)} \mathrm{e}^{-\alpha R_i}, \qquad (6.5)
\end{aligned}$$

since the random variable $|\bar{t}_n - t_n|$ is tight, and assuming that $s_n \to \infty$ so slowly that $s_n|\alpha_n - \alpha| = o(1)$. Since $R_i \ge 0$,

$$\mathrm{e}^{-\alpha s_n} \sum_{i\in \mathsf{BP}^{(j)}_n(s_n)} \mathrm{e}^{-\alpha R_i} \le \mathrm{e}^{-\alpha s_n}|\mathsf{BP}^{(j)}_n(s_n)| \xrightarrow{d} \widetilde{\mathcal{W}}^{(j)} = O_{\mathbb{P}}(1), \qquad (6.6)$$

by the 'perfect' coupling between BP_n and BP at time s_n stated in Proposition 2.2(a), and using that (see (1.20)),

$$\mathrm{e}^{-\alpha s_n}|\mathsf{BP}^{(j)}(s_n)| \xrightarrow{d} \widetilde{\mathcal{W}}^{(j)}. \qquad (6.7)$$

We conclude that

$$\begin{aligned}
&\mathrm{e}^{-\alpha_n t_n}|\mathsf{BP}^{(j)}_{n,\le k_n}[\bar{t}_n+t, \bar{t}_n+t+s)| \qquad (6.8)\\
&\quad= \mathrm{e}^{-\alpha_n t_n} \sum_{i\in \mathsf{BP}^{(j)}_n(s_n)} \sum_{k=1}^{k_n-G^{(j)}_i} |\mathsf{BP}^{(i,j,\vec{m})}_{n,k}[\bar{t}_n+t-s_n-R_i, \bar{t}_n+t+s-s_n-R_i)| + o_{\mathbb{P}}(1).
\end{aligned}$$

A conditional second moment method: expectation. We next use a conditional second moment estimate on the sum on the right-hand side of (6.8), conditionally on $\mathcal{F}_{s_n}$. By the n-dependent version of Proposition 4.1(c) in Proposition 4.1(d), $\bar{t}_n \to \infty$, and for each $i \in \mathsf{BP}^{(j)}_n(s_n)$,

$$\mathrm{e}^{-\alpha_n \bar{t}_n}\mathbb{E}\Big[|\mathsf{BP}^{(i,j,\vec{m})}_{n,\le k_n}[\bar{t}_n, \bar{t}_n+s)|\Big] \to A\Phi(x)F_{R}(s). \qquad (6.9)$$

Observe that also $m = k_n - \underline{k}_n \to \infty$, with $\underline{k}_n = o(\sqrt{\log n})$, implies $t_n \sim m\bar{\nu}_n - x\bar{\sigma}\sqrt{m}$, so that we can conclude from (4.56) that (6.9) also holds with k_n replaced by $k_n - \underline{k}_n$, as long as $\underline{k}_n = o(\sqrt{\log n})$. As a result, when $\bar{t}_n + t - s_n - R_i \xrightarrow{\mathbb{P}} \infty$ and $\underline{k}_n = o(\sqrt{\log n})$ and for each i,

$$\begin{aligned}
&\mathrm{e}^{-\alpha_n(\bar{t}_n+t-s_n-R_i)}\mathbb{E}\Big[|\mathsf{BP}^{(i,j,\vec{m})}_{n,\le k_n-\underline{k}_n}[\bar{t}_n+t-s_n-R_i, \bar{t}_n+t+s-s_n-R_i)| \mid \mathcal{F}_{s_n}\Big] \qquad (6.10)\\
&\quad= \mathrm{e}^{-\alpha_n(\bar{t}_n+t-s_n-R_i)}\mathbb{E}\Big[|\mathsf{BP}^{(i,j,\vec{m})}_{n,\le k_n-\underline{k}_n}[\bar{t}_n+t-s_n-R_i, \bar{t}_n+t+s-s_n-R_i)| \mid R_i,\bar{t}_n\Big]\\
&\quad= A\Phi(x)F_{R}(s)[1+o_{\mathbb{P}}(1)].
\end{aligned}$$

Further, we use the general theory of vertex characteristics in [23, Theorem 6.10.1] to conclude that

$$\sum_{i\in \mathsf{BP}^{(j)}(s_n)} \mathrm{e}^{-\alpha_n(s_n+R_i)} \xrightarrow{\mathbb{P}} \widetilde{\mathcal{W}}^{(j)}/A. \qquad (6.11)$$

Indeed, consider $Z(t) = \sum_{i\in \mathsf{BP}(t)} \mathrm{e}^{-\alpha R_i} = \sum_{i\in \mathsf{BP}(t)} \mathbb{1}_{[\tilde{\tau}_i \le t \le \tilde{\tau}_i+\xi_i]}\mathrm{e}^{-\alpha(\xi_i+\tilde{\tau}_i-t)}$, where the second sum is taken over all individuals and where $\tilde{\tau}_i$, ξ_i are the birthtime and lifetime of individual i, respectively. Then with the terminology of [23, Section 6.9], $Z(t) = Z^{\chi}(t) = \sum_{i\in \mathsf{BP}(t)} \chi_i(t-\tilde{\tau}_i)$, where the random characteristic χ_i of individual i is defined by

$$\chi_i(t) = \mathbb{1}_{[0,\xi_i]}(t)\mathrm{e}^{-\alpha(\xi_i-t)}. \qquad (6.12)$$

According to the aforementioned [23, Theorem 6.10.1],

$$\mathrm{e}^{-\alpha t}Z(t) \xrightarrow{a.s.} c_g\widetilde{\mathcal{W}}/k(\infty), \qquad (6.13)$$

where $c_g = \int_0^\infty \mathrm{e}^{-\alpha u}\mathbb{E}[\chi(u)]\,du / \int_0^\infty \mathrm{e}^{-\alpha u} u\mu(du)$, $k(\infty) = \int_0^\infty \mathrm{e}^{-\alpha u}[1-F_\xi(u)]\,du/\int_0^\infty \mathrm{e}^{-\alpha u} u\mu(du)$ and where $\mu(t) = \nu F_\xi(t)$. This yields $c_g = 1/\nu$, and $k(\infty) = (\nu-1)/(\alpha\nu^2\bar{\nu})$, so that $c_g/k(\infty) = 1/A$. The whp 'perfect' coupling between $(\mathsf{BP}_n(s))_{s\leq s_n}$ with $(\mathsf{BP}(s))_{s\leq s_n}$ stated in Proposition 2.2(a) and using that we may take $s_n \to \infty$ so slowly that $s_n(\alpha-\alpha_n) = o(1)$ implies that (6.13) implies (6.11).

This yields that, conditionally on $\mathcal{F}_{s_n}$ and $\mathcal{W}_{s_n}^{(1)}\mathcal{W}_{s_n}^{(2)} > 0$, and when $G_i^{(j)} = o_{\mathbb{P}}(\sqrt{\log n})$ (which happens whp when s_n is sufficiently small),

$$\begin{aligned}
&\mathrm{e}^{-\alpha_n t_n} \sum_{i\in\mathsf{BP}_n^{(j)}(s_n)} \mathbb{E}\Big[\sum_{k=1}^{k_n-G_i^{(j)}} |\mathsf{BP}_{n,k}^{(i,j,\vec{m})}[\bar{t}_n+t-s_n-R_i, \bar{t}_n+t+s-s_n-R_i)|\Big] \qquad (6.14)\\
&\quad= A\mathrm{e}^{\alpha t}\Phi(x)F_R(s)[1+o_{\mathbb{P}}(1)] \sum_{i\in\mathsf{BP}^{(j)}(s_n)} \mathrm{e}^{\alpha_n(\bar{t}_n-t_n-s_n-R_i)}\\
&\quad= A\mathrm{e}^{\alpha t}\Phi(x)F_R(s)[1+o_{\mathbb{P}}(1)]\mathrm{e}^{\alpha_n(\bar{t}_n-t_n)}\Big(\mathrm{e}^{-\alpha_n s_n}\sum_{i\in\mathsf{BP}^{(j)}(s_n)} \mathrm{e}^{-\alpha_n R_i}\Big)\\
&\quad\xrightarrow{\mathbb{P}} \mathrm{e}^{\alpha t}\Phi(x)F_R(s)\sqrt{\mathcal{W}^{(j)}/\mathcal{W}^{(3-j)}},
\end{aligned}$$

by (6.11), again taking $s_n \to \infty$ sufficiently slowly, and since $\mathrm{e}^{\alpha_n(t_n-\bar{t}_n)} = \sqrt{\mathcal{W}_{s_n}^{(j)}\mathcal{W}_{s_n}^{(3-j)}} \xrightarrow{\mathbb{P}} \sqrt{\mathcal{W}^{(j)}\mathcal{W}^{(3-j)}}$ (see (2.39)). Notice that in (6.14), we condition on $\mathcal{W}_{s_n}^{(j)} > 0$, so that the limit $\widetilde{\mathcal{W}}^{(j)}$ has to be replaced by $\widetilde{\mathcal{W}}^{(j)}|\widetilde{\mathcal{W}}^{(j)} > 0$ which is equal in distribution to $\mathcal{W}^{(j)}$.

A conditional second moment method: variance. We next bound, conditionally on $\mathcal{F}_{s_n}$, the variance of the sum on the right-hand side of (6.8). By conditional independence of $(\mathsf{BP}_n^{(i,j)})_{i\geq 1}$,

$$\begin{aligned}
&\mathrm{e}^{-2\alpha_n t_n}\mathrm{Var}\Big(\sum_{i\in\mathsf{BP}_n^{(j)}(s_n)}\sum_{k=1}^{k_n-G_i^{(j)}} |\mathsf{BP}_{n,k}^{(i,j,\vec{m})}[\bar{t}_n+t-s_n-R_i, \bar{t}_n+t+s-s_n-R_i)| \mid \mathcal{F}_{s_n}\Big) \qquad (6.15)\\
&\quad= \mathrm{e}^{-2\alpha_n t_n}\sum_{i\in\mathsf{BP}_n^{(j)}(s_n)}\mathrm{Var}\Big(\sum_{k=1}^{k_n-G_i^{(j)}} |\mathsf{BP}_{n,k}^{(i,j,\vec{m})}[\bar{t}_n+t-s_n-R_i, \bar{t}_n+t+s-s_n-R_i)| \mid \mathcal{F}_{s_n}\Big).
\end{aligned}$$

We bound

$$\begin{aligned}
&\mathrm{Var}\Big(\sum_{k=1}^{k_n-G_i^{(j)}} |\mathsf{BP}_{n,k}^{(i,j,\vec{m})}[\bar{t}_n+t-s_n-R_i, \bar{t}_n+t+s-s_n-R_i)| \mid \mathcal{F}_{s_n}\Big) \qquad (6.16)\\
&\quad\leq \mathbb{E}\Big[\Big(\sum_{k=1}^{k_n-G_i^{(j)}} |\mathsf{BP}_{n,k}^{(i,j,\vec{m})}(\bar{t}_n+t-s_n-R_i)|\Big)^2 \mid \mathcal{F}_{s_n}\Big]\\
&\quad\leq \mathbb{E}\Big[|\mathsf{BP}_n^{(i,j,\vec{m})}(\bar{t}_n+t-s_n-R_i)|^2 \mid \mathcal{F}_{s_n}\Big] = \mathbb{E}\Big[|\mathsf{BP}_n^{(i,j,\vec{m})}(\bar{t}_n+t-s_n-R_i)|^2 \mid R_i, \bar{t}_n\Big].
\end{aligned}$$

By the n-dependent version of Proposition 4.1(b) in Proposition 4.1(d), for each $n\geq 1$,

$$\sup_{t\geq 0}\{\mathrm{e}^{-2\alpha_n t}\mathbb{E}\big[|\mathsf{BP}_n^{(i,j,\vec{m})}(t)|^2\big]\} \leq CK_n. \qquad (6.17)$$

As a result,

$$\begin{aligned}
&\mathrm{e}^{-2\alpha_n t_n}\mathrm{Var}\Big(\sum_{i\in\mathsf{BP}_n^{(j)}(s_n)}\sum_{k=1}^{k_n-G_i^{(j)}}|\mathsf{BP}_{n,k}^{(i,j,\vec{m})}[\bar{t}_n+t-s_n-R_i,\bar{t}_n+t+s-s_n-R_i)|\mid\mathcal{F}_{s_n}\Big) \qquad (6.18)\\
&\quad\leq \mathrm{e}^{-2\alpha_n t_n}\sum_{i\in\mathsf{BP}_n^{(j)}(s_n)}\mathbb{E}\Big[|\mathsf{BP}_n^{(i,j,\vec{m})}(\bar{t}_n+t-s_n-R_i)|^2\mid R_i,\bar{t}_n\Big]\\
&\quad\leq CK_n\mathrm{e}^{-2\alpha_n t_n}\sum_{i\in\mathsf{BP}^{(j)}(s_n)}\mathrm{e}^{2\alpha_n(\bar{t}_n+t-s_n-R_i)}\\
&\quad= CK_n\mathrm{e}^{2\alpha_n(\bar{t}_n-t_n)}\mathrm{e}^{-2\alpha_n s_n+2\alpha_n t}\sum_{i\in\mathsf{BP}^{(j)}(s_n)}\mathrm{e}^{-2\alpha_n R_i}=O_{\mathbb{P}}(1)K_n\mathrm{e}^{-2\alpha_n s_n}\sum_{i\in\mathsf{BP}^{(j)}(s_n)}\mathrm{e}^{-2\alpha_n R_i},
\end{aligned}$$

since $\mathrm{e}^{2\alpha_n(\bar{t}_n-t_n)+2\alpha_n t}=O_{\mathbb{P}}(1)$. We can bound this further as in (6.6) and (6.7) by

$$K_n\mathrm{e}^{-2\alpha_n s_n}\sum_{i\in\mathsf{BP}_n^{(j)}(s_n)}\mathrm{e}^{-2\alpha_n R_i}\leq K_n\mathrm{e}^{-\alpha_n s_n}\Big(\mathrm{e}^{-\alpha_n s_n}|\mathsf{BP}_n^{(j)}(s_n)|\Big)=O_{\mathbb{P}}(1)K_n\mathrm{e}^{-\alpha_n s_n}, \qquad (6.19)$$

which is $o_{\mathbb{P}}(1)$ precisely when $K_n\mathrm{e}^{-\alpha_n s_n}=o(1)$. Since we are free to choose K_n, we can choose it such that $K_n\mathrm{e}^{-\alpha_n s_n}=o(1)$ indeed holds. By (6.18) and (6.19), the sum on the right-hand side of (6.8) is, conditionally on $\mathcal{F}_{s_n}$, concentrated around its asymptotic conditional mean given in (6.14). As a result, (6.1) follows. This completes the proof of Proposition 2.3(a). ∎

Proof of Proposition 2.3(b). In order to prove Proposition 2.3(b), we compare the statement of Proposition 2.3(b) with that of Proposition 2.3(a). Let $m=|\mathsf{SWT}^{(j)}_{\leq k(t_n,y)}[\bar{t}_n+t,\bar{t}_n+t+s_2)|$, so that $m\xrightarrow{\mathbb{P}}\infty$ on the event that $\mathcal{W}^{(1)}_{s_n}\mathcal{W}^{(2)}_{s_n}>0$, and consider the sum $\sum_{i=1}^m X_i^\star$, where $X_i^\star=d_{V_i}-1$ are forward degrees of free vertices after time $\bar{t}_n+t$, i.e., from the vertices $[n]$, we remove the set S_m of all vertices of which at least one half-edge appeared in $(\mathsf{SWT}(s))_{s\leq\bar{t}_n+t}$. We will prove that, conditionally on $\mathcal{F}_{\bar{t}_n+t}$ with $t<B_n$,

$$\frac{1}{m\nu_n}\sum_{i=1}^m X_i^\star\xrightarrow{\mathbb{P}}1, \qquad (6.20)$$

and then the proof of Proposition 2.3(b) follows from the proof of Proposition 2.3(a) and the fact that $\nu_n\to\nu$.

Without loss of generality, we may assume that $\mathcal{B}_n$ in (5.22) holds, so that $t<B_n$ implies that $|\mathsf{S}_m|\leq\overline{m}_n$, where $\overline{m}_n=\sqrt{n}(\log n)^{1/4}$ as in (5.54). As a result the sequence $(d_i)_{i\in[n]\setminus\mathsf{S}_m}$ satisfies Condition 1.1 whenever $(d_i)_{i\in[n]}$ does. Hence, Lemma 5.3 holds with B_i replaced by $X_i^\star$, so that in particular, from the Markov inequality, conditionally on $\mathcal{F}_{\bar{t}_n+t}$ and for every sequence $K_n\to\infty$ satsfying $K_n^2=o(n/\overline{m}_n)$,

$$\frac{1}{m}\sum_{i=1}^m X_i^\star\mathbb{1}_{\{X_i^\star>K_n\}}\xrightarrow{\mathbb{P}}0. \qquad (6.21)$$

We use a conditional second moment method on $\sum_{i=1}^m X_i^\star\mathbb{1}_{\{X_i^\star\leq K_n\}}$, conditionally on $\mathcal{F}_{\bar{t}_n+t}$. By (5.14) in Lemma 5.3,

$$\mathbb{E}\Big[\sum_{i=1}^m X_i^\star\mathbb{1}_{\{X_i^\star\leq K_n\}}\mid\mathcal{F}_{\bar{t}_n+t}\Big]=m\nu_n(1+o_{\mathbb{P}}(1)). \qquad (6.22)$$

This gives the asymptotics of the first conditional moment of $\sum_{i=1}^m X_i^\star\mathbb{1}_{\{X_i^\star\leq K_n\}}$. For the second moment, we start by bounding the covariances. We note that, for $1\leq i<j\leq m$,

$$\begin{aligned}
&\mathrm{Cov}\Big(X_i^\star\mathbb{1}_{\{X_i^\star\leq K_n\}},X_j^\star\mathbb{1}_{\{X_j^\star\leq K_n\}}\mid\mathcal{F}_{\bar{t}_n+t}\Big) \qquad (6.23)\\
&\quad=\mathbb{E}\Big[X_i^\star\mathbb{1}_{\{X_i^\star\leq K_n\}}\Big(\mathbb{E}[X_j^\star\mathbb{1}_{\{X_j^\star\leq K_n\}}\mid\mathcal{F}_{\bar{t}_n+t},X_1^\star,\ldots,X_i^\star]-\mathbb{E}[X_j^\star\mathbb{1}_{\{X_j^\star\leq K_n\}}\mid\mathcal{F}_{\bar{t}_n+t}]\Big)\mid\mathcal{F}_{\bar{t}_n+t}\Big].
\end{aligned}$$

By (5.14) in Lemma 5.3, as well as the fact that $i \leq \overline{m}_n = o(n)$,

$$\mathbb{E}[X_j^\star \mathbb{1}_{\{X_j^\star \leq K_n\}} \mid \mathcal{F}_{\bar{t}_n+t}, X_1^\star, \ldots, X_i^\star] - \mathbb{E}[X_j^\star \mathbb{1}_{\{X_j^\star \leq K_n\}} \mid \mathcal{F}_{\bar{t}_n+t}] = o_{\mathbb{P}}(1), \tag{6.24}$$

so that also

$$\mathrm{Cov}\Big(X_i^\star \mathbb{1}_{\{X_i^\star \leq K_n\}}, X_j^\star \mathbb{1}_{\{X_j^\star \leq K_n\}} \mid \mathcal{F}_{\bar{t}_n+t}\Big) = o_{\mathbb{P}}(1). \tag{6.25}$$

Further, a trivial bound on the second moment together with (5.14) in Lemma 5.3 yields that

$$\mathrm{Var}\Big(X_i^\star \mathbb{1}_{\{X_i^\star \leq K_n\}} \mid \mathcal{F}_{\bar{t}_n+t}\Big) \leq K_n \mathbb{E}[X_i^\star \mid \mathcal{F}_{\bar{t}_n+t}] = K_n \nu_n (1 + o_{\mathbb{P}}(1)). \tag{6.26}$$

As a result, whenever $K_n m = o(m^2)$, and $k_n^2 = o(n/\overline{m}_n)$,

$$\mathrm{Var}\Big(\sum_{i=1}^m X_i^\star \mathbb{1}_{\{X_i^\star \leq K_n\}} \mid \mathcal{F}_{\bar{t}_n+t}\Big) = o_{\mathbb{P}}(m^2), \tag{6.27}$$

which together with (6.22) proves that, conditionally on $\mathcal{F}_{\bar{t}_n+t}$,

$$\frac{1}{m\nu_n} \sum_{i=1}^m X_i^\star \mathbb{1}_{\{X_i^\star \leq K_n\}} \xrightarrow{\mathbb{P}} 1. \tag{6.28}$$

Together with (6.21), this proves (6.20), as required. ■

7. The PPP limit for collision edges: Proof of Theorem 3.1

Recall that $(\mathcal{F}_t)_{t\geq 0}$ is the filtration generated by all the randomness used in the construction up to time t, i.e.,

$$\mathcal{F}_t = \sigma\Big((\mathsf{SWT}(s), \mathsf{BP}_n(s), \mathsf{BP}(s))_{s\in[0,t]}\Big).$$

We will investigate the number of collision edges (x_i, P_{x_i}) with $I(x_i) = j \in \{1, 2\}$, $H(x_i) \leq k_n(t_n, x)$, $H(P_{x_i}) \leq k(t_n, y)$ and $R_{T_i^{(\mathrm{col})}}(P_{x_i}) \in [0, s)$ created in the time interval $[\bar{t}_n + t, \bar{t}_n + t + \varepsilon)$, where $\varepsilon > 0$ is small. We let $\mathcal{I} = [a, b) \times \{j\} \times (-\infty, x] \times (-\infty, y] \times [0, s]$ be a subset of $\mathcal{S}$, and we prove that

$$\mathbb{P}(\Pi_n(\mathcal{I}) = 0 \mid \mathcal{F}_{s_n}) \xrightarrow{\mathbb{P}} \exp\Big\{ - \int_a^b \frac{2\nu f_R(0)}{\mathbb{E}[D]} \mathrm{e}^{2\alpha t} \Phi(x)\Phi(y) F_R(s) dt\Big\}. \tag{7.1}$$

By [28, Theorem 4.7], this proves the claim.

We split

$$\mathcal{I} = \bigcup_{\ell=1}^N \mathcal{I}_\ell^{(\varepsilon)}, \tag{7.2}$$

where $\mathcal{I}_l^{(\varepsilon)} = [t_{\ell-1}^{(\varepsilon)}, t_\ell^{(\varepsilon)}) \times \{j\} \times (-\infty, x] \times (-\infty, y] \times [0, s)$, with $t_\ell^{(\varepsilon)} = a + \ell\varepsilon$ and $\varepsilon = (b-a)/N$, with $N \in \mathbb{N}$. We will let $\varepsilon \downarrow 0$ later on. For a fixed $\varepsilon > 0$, we say that a collision edge (x_i, P_{x_i}) is a *first round collision edge* when there exists $l \in [N]$ and a half-edge $y \in \mathsf{AH}(t_{l-1}^{(\varepsilon)})$ such that y is found by the liquid in the time interval $\mathcal{I}_\ell^{(\varepsilon)}$, y is paired to the half-edge P_y whose sibling half-edge x_i is paired to $P_{x_i} \in \mathsf{AH}(t_{\ell-1}^{(\varepsilon)})$ with $I(y) = j \neq I(P_{x_i}) = 3 - j$. We call all other collision edges *second round collision edges.* The second round collision edges are such that a half-edge y is found by the liquid in the time interval $\mathcal{I}_\ell^{(\varepsilon)}$ (the first round), y is paired to the half-edge P_y, one of the sibling half-edges x_i of y is then also found by the liquid in the time interval $\mathcal{I}_\ell^{(\varepsilon)}$ (the second round) and is paired to a half-edge P_{x_i}, whose sibbling half-edge z is paired to $P_z \in \mathsf{AH}(t_{\ell-1}^{(\varepsilon)})$ with $I(x_i) = j \neq I(P_z) = 3 - j$. When $\varepsilon > 0$ is quite small, the latter seems less likely, which is why we start with the first round collision edges.

Denote the point processes of first and second round collision edges by $\Pi_n^{(\mathrm{FR})}$ and $\Pi_n^{(\mathrm{SR})}$, so that $\Pi_n = \Pi_n^{(\mathrm{FR})} + \Pi_n^{(\mathrm{SR})}$. The next two lemmas investigate the point processes $\Pi_n^{(\mathrm{FR})}$ and $\Pi_n^{(\mathrm{SR})}$:

Lemma 7.1 (PPP limit for the first round collision edges)**.** *For every* $s \geq 0$, $x, y \in \mathbb{R}$, $j \in \{1,2\}$, $\varepsilon > 0$ *and* $\ell \in [N]$, *as* $n \to \infty$,

$$\mathbb{P}\big(\Pi_n^{(\mathrm{FR})}(\mathcal{I}_\ell^{(\varepsilon)}) = 0 \mid \mathcal{F}_{t_{\ell-1}^{(\varepsilon)}}\big) \xrightarrow{\mathbb{P}} \exp\big\{-\frac{2\nu}{\mathbb{E}[D]}\mathrm{e}^{2\alpha t_{\ell-1}^{(\varepsilon)}}\Phi(x)\Phi(y)F_R(s)F_R(\varepsilon)\big\}. \tag{7.3}$$

Proof. The number of half-edges $z \in \mathsf{AH}(\bar{t}_n + t_{\ell-1}^{(\varepsilon)})$ that are found by the liquid having $I(z) = j$ and $H(z) \leq k_n(t_n, x)$ is equal to

$$|\mathsf{SWT}^{(j)}_{\leq k_n(t_n,x)}[\bar{t}_n + t_{\ell-1}^{(\varepsilon)}, \bar{t}_n + t_{\ell-1}^{(\varepsilon)} + \varepsilon)|. \tag{7.4}$$

Fix such a half-edge z, and note that it is paired to P_z that has $X_z^\star = d_{V_{P_z}} - 1$ sibling half-edges. For each of these half-edges we test whether it is paired to a half-edge in $\mathsf{AH}(\bar{t}_n + t_{\ell-1}^{(\varepsilon)})$ or not. Therefore, the total number of tests performed in the time interval $[t_{\ell-1}^{(\varepsilon)}, t_\ell^{(\varepsilon)})$ is equal to

$$\sum_z X_z^\star \mathbb{1}_{\{z \in \mathsf{SWT}^{(j)}_{\leq k_n(t_n,x)}[\bar{t}_n + t_{\ell-1}^{(\varepsilon)}, \bar{t}_n + t_\ell^{(\varepsilon)})\}}. \tag{7.5}$$

By construction, we test whether these half-edges are paired to half-edges that are incident to the SWT or not. Each of these edges is paired to a half-edge $w \in \mathsf{AH}(\bar{t}_n + t_{\ell-1}^{(\varepsilon)})$ with $I(w) = 3-j$ (and thus creating a collision edge) and $H(w) \leq k(t_n, y)$ and $R_{\bar{t}_n + t_{\ell-1}^{(\varepsilon)}}(w) \in [0, s)$ with probability equal to

$$\frac{1}{\ell_n - o(n)}|\mathsf{SWT}^{(3-j)}_{\leq k(t_n,y)}[\bar{t}_n + t_{\ell-1}^{(\varepsilon)}, \bar{t}_n + t_{\ell-1}^{(\varepsilon)} + s)|. \tag{7.6}$$

Therefore, for $\varepsilon > 0$, conditionally on $\mathcal{F}_{\bar{t}_n + t_{\ell-1}^{(\varepsilon)}}$, the probability that none of the half-edges found in the time interval in between $[\bar{t}_n + t_{\ell-1}^{(\varepsilon)}, \bar{t}_n + t_\ell^{(\varepsilon)})$ creates a collision edge is asymptotically equal to

$$\prod_{v \in \mathsf{SWT}^{(j)}_{\leq k(t_n,x)}[\bar{t}_n + t_{\ell-1}^{(\varepsilon)}, \bar{t}_n + t_\ell^{(\varepsilon)})} \Big(1 - \frac{1}{\ell_n - o(n)}|\mathsf{SWT}^{(3-j)}_{\leq k(t_n,y)}[\bar{t}_n + t_{\ell-1}^{(\varepsilon)}, \bar{t}_n + t_{\ell-1}^{(\varepsilon)} + s)|\Big)^{X_v^\star} \tag{7.7}$$

$$\xrightarrow{\mathbb{P}} \exp\big\{-\frac{\nu}{\mathbb{E}[D]}\mathrm{e}^{2\alpha t_{\ell-1}^{(\varepsilon)}}\Phi(x)\Phi(y)F_R(s)F_R(\varepsilon)\big\},$$

where we use (2.45), that $\mathrm{e}^{2\alpha_n t_n} = n^{-1}$, and that $\ell_n = n\mathbb{E}[D_n]$ with $\mathbb{E}[D_n] \to \mathbb{E}[D]$. The factor 2 in (7.3) is caused by the two possibilities $j \in \{1,2\}$. ■

Lemma 7.2 (A bound on the second round collision edges)**.** *For* $x, y \in \mathbb{R}$, $j \in \{1,2\}$, $\varepsilon > 0$ *and* $\ell \in [N]$, *as* $n \to \infty$,

$$\mathbb{P}\big(\Pi_n^{(\mathrm{SR})}(\mathcal{I}_\ell^{(\varepsilon)}) \geq 1 \mid \mathcal{F}_{t_{\ell-1}^{(\varepsilon)}}\big) = O_{\mathbb{P}}(1)F_R(\varepsilon)F_\xi(\varepsilon). \tag{7.8}$$

Proof. By analogous arguments as above, the expected number of second round collision edges is of order

$$O_{\mathbb{P}}(1)\mathrm{e}^{2\alpha t}\Phi(x)\Phi(y)F_R(s)F_R(\varepsilon)F_\xi(\varepsilon), \tag{7.9}$$

since one of the half-edges z that is found by the liquid in the time interval $[\bar{t}_n + t_{\ell-1}^{(\varepsilon)}, \bar{t}_n + t_\ell^{(\varepsilon)})$ needs to satisfy that one of the $d_{V_{P_z}} - 1$ half-edges has weight at most ε, and which, upon being found, needs to create a collision edge. ■

Now we are ready to complete the proof of Theorem 3.1. We use that

$$\mathbb{P}\big(\Pi_n(\mathcal{I}) = 0 \mid \mathcal{F}_{s_n}\big) = \mathbb{E}\Big[\prod_{\ell=1}^N \mathbb{P}\big(\Pi_n(\mathcal{I}_\ell^{(\varepsilon)}) = 0 \mid \mathcal{F}_{t_{\ell-1}^{(\varepsilon)}}\big) \mid \mathcal{F}_{s_n}\Big]. \tag{7.10}$$

We start with the upper bound, for which we use that

$$\mathbb{P}\big(\Pi_n(\mathcal{I}_\ell^{(\varepsilon)}) = 0 \mid \mathcal{F}_{t_{\ell-1}^{(\varepsilon)}}\big) \leq \mathbb{P}\big(\Pi_n^{(\mathrm{FR})}(\mathcal{I}_\ell^{(\varepsilon)}) = 0 \mid \mathcal{F}_{t_{\ell-1}^{(\varepsilon)}}\big) \tag{7.11}$$

$$\xrightarrow{\mathbb{P}} \exp\big\{-\frac{2\nu}{\mathbb{E}[D]}\mathrm{e}^{2\alpha t_{\ell-1}^{(\varepsilon)}}\Phi(x)\Phi(y)F_R(s)F_R(\varepsilon)\big\},$$

by Lemma 7.1. We conclude that

$$\mathbb{P}\big(\Pi_n(\mathcal{I})=0\mid \mathcal{F}_{s_n}\big)\le \mathbb{E}\Big[\prod_{\ell=1}^{N}\exp\big\{-\frac{2\nu}{\mathbb{E}[D]}\mathrm{e}^{2\alpha t_{\ell-1}^{(\varepsilon)}}\Phi(x)\Phi(y)F_R(s)F_R(\varepsilon)\big\}\mid \mathcal{F}_{s_n}\Big] \tag{7.12}$$

$$=\exp\big\{-\sum_{\ell=1}^{N}\frac{2\nu}{\mathbb{E}[D]}\mathrm{e}^{2\alpha t_{\ell-1}^{(\varepsilon)}}\Phi(x)\Phi(y)F_R(s)F_R(\varepsilon)\big\}$$

$$\to \exp\big\{-\frac{2\nu}{\mathbb{E}[D]}f_R(0)\int_a^b \mathrm{e}^{2\alpha t}\Phi(x)\Phi(y)F_R(s)dt\big\},$$

since $\lim_{\varepsilon\downarrow 0}F_R(\varepsilon)/\varepsilon=f_R(0)$, and the Riemann approximation

$$\varepsilon\sum_{\ell=1}^{N}\mathrm{e}^{2\alpha t_{\ell-1}^{(\varepsilon)}}\to\int_a^b \mathrm{e}^{2\alpha t}dt. \tag{7.13}$$

This proves the upper bound.

For the lower bound, we instead bound

$$\mathbb{P}\big(\Pi_n(\mathcal{I})=0\mid \mathcal{F}_{s_n}\big)\ge \mathbb{E}\Big[\prod_{\ell=1}^{N}\mathbb{P}\big(\Pi_n^{(\mathrm{FR})}(\mathcal{I}_\ell^{(\varepsilon)})=0\mid \mathcal{F}_{t_{\ell-1}^{(\varepsilon)}}\big)\mid \mathcal{F}_{s_n}\Big] \tag{7.14}$$

$$-\mathbb{E}\Big[\Big(\sum_{\ell=1}^{N}\mathbb{P}\big(\Pi_n^{(\mathrm{SR})}(\mathcal{I}_\ell^{(\varepsilon)})\ge 1\mid \mathcal{F}_{t_{\ell-1}^{(\varepsilon)}}\big)\Big)\wedge 1\mid \mathcal{F}_{s_n}\Big].$$

The first term has already been dealt with, the second term is, by Lemma 7.2, bounded by

$$\mathbb{E}\Big[\Big(O_{\mathbb{P}}(1)\sum_{\ell=1}^{N}F_R(\varepsilon)F_\xi(\varepsilon)\Big)\wedge 1\mid \mathcal{F}_{s_n}\Big]=o_{\mathbb{P}}(1), \tag{7.15}$$

as $\varepsilon\downarrow 0$, by dominated convergence, since $F_R(\varepsilon)=\varepsilon f_R(0)(1+o(1))$ and $F_\xi(\varepsilon)=o(1)$. ■

Acknowledgements. The research of SB is supported by NSF-DMS grant 1105581 and SB would like to thank Eurandom and TU/e for travel support where part of this work was done. The work of RvdH was supported in part by Netherlands Organisation for Scientific Research (NWO). SB and RvdH thank Cornell University for its hospitality during the Cornell Probability Summer School in July 2012, where this work was completed. GH would like to thank the Fundación Cultural Knecht-Drenth, for his stay in Callosa d'En Sarrià (Spain), where he worked on the revision. We thank an anonymous referee for detailed comments that helped us to significantly improve the presentation.

References

[1] D. Aldous and D. Lanoue, *A lecture on the averaging process*, Probability Surveys **9** (2012), 90–102.
[2] D. J. Aldous, *When knowing early matters: Gossip, percolation and Nash equilibria*, Preprint available at `arXiv:1005.4846` (2010).
[3] H. Amini, M. Draief, and M. Lelarge, *Flooding in weighted sparse random graphs*, SIAM Journal on Discrete Mathematics **27** (2013), no. 1, 1–26.
[4] K. B. Athreya and P. E. Ney, *Branching processes*, Dover Publications Inc., Mineola, NY, (2004). Reprint of the 1972 original [Springer, New York; MR0373040]. MR2047480
[5] S. Bhamidi, *First passage percolation on locally treelike networks. I. dense random graphs*, Journal of Mathematical Physics **49** (2008), no. 12, 125218–125218.
[6] S. Bhamidi and R. v. d. Hofstad, *Weak disorder asymptotics in the stochastic mean-field model of distance*, Ann. Appl. Probab. **22** (2012), no. 1, 29–69. MR2932542
[7] S. Bhamidi, R. v. d. Hofstad, and G Hooghiemstra, *Universality for first passage percolation on sparse uniform and generalized random graphs.* Preprint (2014).
[8] S. Bhamidi, R. v. d. Hofstad, and G. Hooghiemstra, *Extreme value theory, Poisson-Dirichlet distributions, and first passage percolation on random networks*, Adv. in Appl. Probab. **42** (2010), no. 3, 706–738. MR2779556 (2012c:60226)

[9] S. Bhamidi, R. v. d. Hofstad, and G Hooghiemstra, *First passage percolation on random graphs with finite mean degrees*, Ann. Appl. Probab. **20** (2010), no. 5, 1907–1965. MR2724425 (2011m:60289)

[10] S. Bhamidi, R. v. d. Hofstad, and G Hooghiemstra, *First passage percolation on the Erdős-Rényi random graph*, Combin. Probab. Comput. **20** (2011), no. 5, 683–707. MR2825584

[11] B. Bollobás, *Random graphs*, Second edition, Cambridge Studies in Advanced Mathematics, vol. **73**, Cambridge University Press, Cambridge, (2001). MR1864966 (2002j:05132)

[12] B. Bollobás, S. Janson, and O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Structures Algorithms **31** (2007), no. 1, 3–122. MR2337396

[13] L. A. Braunstein, S. V. Buldyrev, R. Cohen, S. Havlin, and H. E. Stanley, *Optimal paths in disordered complex networks*, Physical review letters **91** (2003), no. 16, 168701.

[14] S. Chatterjee, *Chaos, concentration, and multiple valleys*, Preprint available at `arXiv:0810.4221` (2008).

[15] M. Draief and L. Massoulié, *Epidemics and rumours in complex networks*, London Mathematical Society Lecture Note Series, vol. **369**, Cambridge University Press, Cambridge, 2010. MR2582458 (2011g:05267)

[16] R. Durrett, *Lecture Notes on Particle Systems and Percolation*, Brooks/Cole Pub Co, (1988).

[17] H. v. d. Esker, R. v. d. Hofstad, and G. Hooghiemstra, *Universality for the distance in finite variance random graphs*, J. Stat. Phys. **133** (2008), no. 1, 169–202. MR2438903

[18] A. Flaxman, D. Gamarnik, and G. Sorkin, *First-passage percolation on a width-2 strip and the path cost in a vcg auction*, Internet and Network Economics (2006), 99–111.

[19] J. Hammersley, *First-passage percolation*, Journal of the Royal Statistical Society. Series B (Methodological) **28** (1966), no. 3, 491–496.

[20] T. E. Harris, *The Theory of Branching processes*, Springer-Verlag, Berlin, (1963).

[21] R. v. d. Hofstad, G. Hooghiemstra, and P. Van Mieghem, *Distances in random graphs with finite variance degrees*, Random Structures Algorithms **27** (2005), no. 1, 76–123. MR2150017 (2006d:60019)

[22] C. Howard, *Models of first-passage percolation*, Probability on discrete structures (2004), 125–173.

[23] P. Jagers, *Branching Processes with Biological Applications*, John Wiley & Sons, (1975).

[24] P. Jagers and O. Nerman, *The growth and composition of branching populations*, Advances in Applied Probability **16** (1984), no. 2, 221–259.

[25] S. Janson, *The probability that a random multigraph is simple*, Combinatorics, Probability and Computing **18** (2009), no. 1-2, 205–225.

[26] S. Janson, *Susceptibility of random graphs with given vertex degrees*, J. Comb. **1** (2010), no. 3-4, 357–387. MR2799217 (2012c:05278)

[27] S. Janson and M. Luczak, *A new approach to the giant component problem*, Random Structures Algorithms **34** (2009), no. 2, 197–216. MR2490288 (2010d:05140)

[28] O. Kallenberg, *Random measures*, Akademie-Verlag, Berlin, (1976). MR0431373 (55 #4373)

[29] H. Kesten, *Aspects of first passage percolation*, Ecole d'Eté de Probabilités de Saint Flour XIV-1984 ((1986)), 125–264.

[30] J. Leskovec, L. Backstrom, and J. Kleinberg, *Meme-tracking and the dynamics of the news cycle*, Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 497–506.

[31] J. Leskovec, M. McGlohon, F. C., and M. Hurst, *Cascading behavior in large blog graphs*, SDM Proc. 7th SIAM Int. Conf. on Data Mining, 2007, pp. 29406–29413.

[32] T. M. Liggett, *Interacting Particle Systems*, Springer Verlag, (2004).

[33] M. Mihail, C. Papadimitriou, and A. Saberi, *On certain connectivity properties of the internet topology*, Journal of Computer and System Sciences **72** (2006), no. 2, 239–251.

[34] M. Molloy and B. Reed, *A critical point for random graphs with a given degree sequence*, Random Structures Algorithms **6** (1995), no. 2-3, 161–179. MR1370952 (97a:05191)

[35] M. Molloy and B. Reed, *The size of the giant component of a random graph with a given degree sequence*, Combinatorics Probability and Computing **7** (1998), no. 3, 295–305. MR1664335 (2000c:05130)

[36] S. Resnick, *Extreme values, regular variation, and point processes*, Applied Probability. A Series of the Applied Probability Trust, vol. **4**, Springer-Verlag, New York, (1987). MR900810 (89b:60241)

[37] M. Samuels, *Distribution of the branching-process population among generations*, Journal of Applied Probability **8** (1971), no. 4, 655–667.

[38] R. Smythe and J. Wierman, *First-passage percolation on the square lattice, volume* **671** *of Lecture Notes in Mathematics*, Springer, Berlin, (1978).

[39] H. Thorisson, *Coupling, stationarity, and regeneration*, Probability and its Applications (New York), Springer-Verlag, New York, (2000). MR1741181 (2001b:60003)