Weak disorder asymptotics in the stochastic mean-field model of distance

Shankar Bhamidi [*] Remco van der Hofstad [†]

February 23, 2010

Abstract

In the recent past, there has been a concerted effort to develop mathematical models for real-world networks and analyze various dynamics on these models. One particular problem of significant importance is to understand the effect of random edge lengths or costs on the geometry and flow transporting properties of the network. Two different regimes are of great interest, the *weak* disorder regime where optimality of a path is determined by the sum of edge weights on the path and the *strong* disorder regime where optimality of a path is determined by the maximal edge weight on the path. In the context of the stochastic mean-field model of distance, we provide the first mathematically tractable model of weak disorder and show that no transition occurs at finite temperature. Indeed, we show that for every finite temperature, the number of edges on the minimal weight path (i.e., the hopcount) is always $\Theta(\log n)$ and satisfies a central limit theorem with asymptotic means and variances of order $\Theta(\log n)$, with limiting constants expressible in terms of the Malthusian rate of growth and the mean of the stable-age distribution of the associated continuous-time branching process. More precisely, we take independent and identically distributed edge weights with distribution E^s for some parameter $s > 0$, where E is an exponential random variable with mean 1. Then, the asymptotic mean and variance of the central limit theorem for the hopcount are $s \log n$ and $s^2 \log n$ respectively. We also find limiting distributional asymptotics for the value of the minimal weight path in terms of extreme value distributions and martingale limits of branching processes.

Key words: Flows, random graphs, first passage percolation, hopcount, central limit theorem, weak disorder, continuous-time branching process, stable-age distribution theory, mean-field model of distance, Cox point processes.

MSC2000 subject classification. 60C05, 05C80, 90B15.

1 Introduction

The last few years have witnessed an explosion in empirical data collected on various real-world networks, including transportation networks like road and rail networks and data transmission networks such as the Internet. This has stimulated an intense inter-disciplinary effort to formulate various mathematical network models to understand their structure as well as the evolution of such

[*]Department of Statistics and Operations Research The University of North Carolina 304 Hanes Hall Chapel Hill, NC 27510. Email: bhamidi@email.unc.edu

[†]Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: rhofstad@win.tue.nl

real-world networks. Rigorously analyzing properties of these models and deriving asymptotics as the size of the network becomes large is currently an active area of modern probability theory.

In many contexts, these models are used to model transportation networks and understanding the flow carrying properties of these models is of paramount importance. Real-world networks are described not only by their graph structure, which give us information about valid links between vertices in the network, but also by their associated *edge weights*, representing cost or time required to traverse the edge. Similar questions form the core of one of the fundamental problems in interacting particle systems, namely *first passage percolation*. In brief, one starts with a finite network model $\mathcal{K}_n$ (for example the $[-n,n]^2$ box in the integer lattice $\mathbb{Z}^2$). Each edge e is given some random edge weight l_e, usually assumed to be non-negative, independent and identically distributed (i.i.d.) across edges. We shall refer to l_e as the *length* or cost of the edge e. For any two vertices $u, v \in \mathcal{K}_n$, and a path P between the two vertices, the cost of the path $f(P)$, is some function of the edge weights on the path (see the next section where we describe two natural regimes). The optimal path $P_{\texttt{opt}}(u,v)$ between the two vertices is the path that minimizes this cost function amongst all possible such paths. Now fix two vertices in $\mathcal{K}_n$, e.g., in the case of the two-dimensional integer lattice, the origin and the point $(n,0)$. One is then interested in deriving properties of the optimal path between these two vertices, at least as the size of the network $n \to \infty$.

In the modern applied context, two particular statistics of this optimal path are of importance:
(a) $f(P_{\texttt{opt}}(u,v))$: the actual cost of the optimal path. In many situations, this gives the cost of transporting a unit of flow between the two vertices.
(b) $H(P_{\texttt{opt}}(u,v))$: the number of edges in this path. This represents the actual amount of time that a message takes in getting between the two vertices. The mental picture one should have is that the network is transporting flow between various vertices via the optimal paths and *delay*, i.e., the amount of time that a message takes in getting between vertices is the number of edges or *hops* on the optimal path. Thus this quantity is often referred to as the *hopcount*.

1.1 Weak and strong disorder

When modeling random disordered systems, two cost regimes for the cost $f(P)$ of a path P are of interest, the *strong* disorder and *weak* disorder regime. Throughout the discussion below we start with a connected network $\mathcal{K}_n$ on n vertices, with each edge assigned edge weight l_e. Fix two vertices denoted by 1 and 2 (say chosen uniformly at random amongst all vertices). We are interested in properties of the optimal path between these two vertices. Let $\mathcal{P}_{12}$ denote the set of all paths between the two vertices.

Weak disorder regime: This is the conventional setup where for any path $\mathcal{P} \in \mathcal{P}_{12}$, the cost of the path is

$$f_{\texttt{wk-dis}}(P) = \sum_{e \in P} l_e. \tag{1.1}$$

The optimal path, denoted by $P_{\texttt{wk-dis}}$, is defined by

$$P_{\texttt{wk-dis}} = \arg\min_{P \in \mathcal{P}_{12}} f_{\texttt{wk-dis}}(P). \tag{1.2}$$

In our setup, the optimal path will always be *unique*. We are then interested in the cost and hopcount of this optimal path.

Strong disorder regime: Here, for any path $P \in \mathcal{P}_{12}$, the cost of the path is given by

$$f_{\texttt{st-dis}}(P) = \max_{e \in P} l_e. \tag{1.3}$$

As before, the optimal path, denoted by $P_{\texttt{st-dis}}$, is defined by

$$P_{\texttt{st-dis}} = \arg\min_{P \in \mathcal{P}_{12}} f_{\texttt{st-dis}}(P). \tag{1.4}$$

From a statistical physics viewpoint, one is interested in parameterizing the above problem via a real-valued parameter say β, often called the "inverse temperature" of the system, such that as $\beta \to \infty$, we get the strong disorder regime, while for finite values of β, we have the weak disorder regime. One interesting way of parameterizing the above problem is to consider the original graph $\mathcal{K}_n$ with some edge random variables w_e and consider the model $\mathcal{G}_n(\beta)$ where each edge is given weight $l_e(\beta) = \exp(\beta w_e)$. The $\beta \to \infty$ regime then corresponds to the strong disorder regime with edge weights w_e, the $\beta = 0$ regime corresponds to the graph distance regime (where each edge has fixed weight 1), while finite positive values of β are supposed to model the weak disorder regime and are meant to interpolate between the graph distance regime and the strong disorder regime. What is of paramount interest is to understand if and when a transition occurs, namely given some model $\mathcal{K}_n$ of network on n vertices and edge distribution $w_e \sim F$, e.g. the uniform or exponential distribution, is there some finite value of β for which a transition occurs from the weak disorder regime to the strong disorder regime, where the graph begins to behave as in the strong disorder regime? What are the properties of the optimal paths in various regimes, and how does the hopcount scale as a function of β, at least in the $n \to \infty$ large network limit? Although a number of studies have been carried out at the simulation level (see e.g. [8] and the plethora of references therein) to understand such models of disorder in the context of various random graph models resulting in a fascination circle of conjectures, there has been no rigorous effort carried out to derive results in this context.

Our goal is to formulate a solvable model in this context and to exhibit how such questions have deep connections to the stable-age distribution theory of continuous-time branching processes as formulated by Jagers and Nerman, see e.g. [17]. Without further ado let us dive into the formulation of the model in our context.

1.2 Model formulation

Let $\mathcal{K}_n$ be the complete graph with vertex set $[n] \equiv \{1, \ldots, n\}$ and edge set $E_n = \{ij : i, j \in [n], i \neq j\}$. Each edge e is given weight $l_e = (E_e)^s$ for some fixed $s > 0$, where $(E_e)_{e \in E_n}$ are i.i.d. exponential random variables with mean 1. The optimal path between two vertices is the path that minimizes the sum of weights on that path, as in the weak disorder regime. In the context of the above discussion of strong and weak disorder, $s = 0$ corresponds to the graph distance, while $s = \infty$ corresponds to the strong disorder regime with edge weights E_e, the parameter $\beta > 0$ above is equal to s and the random variable $(w_e)_{e \in E_n}$ equals $w_e = \log(E_e)$, which has a Gumbel distribution. The advantage of this formulation is that it gives a rigorously analyzable model. The $s = 1$ regime is one of the most well-studied model in probabilistic combinatorial optimization (see e.g. [2, 3, 7, 11, 14, 19]) and often goes under the name of "stochastic mean-field model of distance". For a fixed $s \in \mathbb{R}^+$, we are interested in statistics of the optimal path, in particular, in the asymptotics for the weight and hopcount of the optimal path as $n \to \infty$.

To state the results, we shall need to set up some constructs. Let $\{Y_j\}_{j \geq 1}$ be i.i.d. mean 1 exponential random variables. Define the random variables L_i by the equation

$$L_i = \Big(\sum_{j=1}^{i} Y_j \Big)^s. \tag{1.5}$$

Let $\mathcal{P}$ be the above point process, i.e.,

$$\mathcal{P} = (L_1, L_2, \ldots). \tag{1.6}$$

While the parameter s plays an important role in our analysis, for the sake of simplicity, we shall omit it from the notation. The reader should keep in mind that all the important constructs that arise in the analysis and in description of limit results, such as the point process above, depend on this parameter. Now consider the continuous-time branching process (CTBP) where at time $t = 0$ we start with one vertex (called the root or the original ancestor), each vertex v lives forever, and has an offspring distribution $\mathcal{P}_v \sim \mathcal{P}$ as in (1.6) independent of every other vertex. Let $(\mathrm{BP}_t)_{t\geq 0}$ denote the CTBP with the above offspring distribution. The general theory of branching processes (see e.g. [17]) implies that there exists a constant $\lambda = \lambda(s)$ called the *Malthusian rate of growth* which determines the rate of explosive growth of this model. In particular, if $z_t = |\mathrm{BP}_t|$ denotes the number of individuals born by time t, then there exists a strictly positive random variable W such that

$$\mathrm{e}^{-\lambda t} z_t \xrightarrow{a.s.} W, \tag{1.7}$$

where $\xrightarrow{a.s.}$ denotes convergence almost surely. The constant λ satisfies the equation

$$\sum_{i=1}^{\infty} \mathbb{E}\left(\mathrm{e}^{-\lambda L_i}\right) = 1. \tag{1.8}$$

In this case, an explicit computation (see Lemma 3.1 below) implies that

$$\lambda = \lambda(s) = \Gamma(1 + 1/s)^s. \tag{1.9}$$

Now let $W^{(1)}, W^{(2)}$ be i.i.d. with distribution W where W is as defined above in (1.7). Define the Cox process $\mathcal{P}_{\mathsf{cox}}$ which, given $W^{(1)}$ and $W^{(2)}$ is a Poisson process on $\mathbb{R}$ with rate function given by

$$\gamma(x) = \frac{2\lambda}{s} W^{(1)} W^{(2)} \mathrm{e}^{2\lambda x}, \qquad x \in \mathbb{R}. \tag{1.10}$$

Let $\Xi^{(1)}$ denote the first point of the point process $\mathcal{P}_{\mathsf{cox}}$.

1.3 Results

We are now in a position to state our results. Recall that we started with the complete graph where each edge had distribution $l_e = E_e^s$, where $(E_e)_{e \in E_n}$ are i.i.d. exponential random variables having mean one. The first result identifies the limiting distribution for the minimal weight path while the second result below identifies the asymptotics for the number of edges on the minimal weight path.

Theorem 1.1 (The weight of the shortest-weight path) *Let $\mathcal{C} = \mathcal{C}(s)$ denote the cost of the optimal path between the two vertices 1 and 2. Then,*

$$n^s \mathcal{C} - \frac{1}{\lambda} \log n \xrightarrow{d} 2\Xi^{(1)} \tag{1.11}$$

and

$$2\Xi^{(1)} \stackrel{d}{=} \frac{1}{\lambda}\left(G - \log W^{(1)} - \log W^{(2)} - \log\left(1/s\right)\right), \tag{1.12}$$

where G is a standard Gumbel random variable independent of $W^{(1)}$ and $W^{(2)}$, and $W^{(1)}$ and $W^{(2)}$ are two independent copies of the random variable W appearing in (1.7).

Theorem 1.2 (CLT for the hopcount) *Let $H_n = H_n(s)$ denote the number of edges on the optimal path between the two vertices 1 and 2, i.e., the hopcount. Then, as $n \to \infty$,*

$$\frac{H_n - s\log n}{\sqrt{s^2 \log n}} \xrightarrow{d} Z, \tag{1.13}$$

where Z has a standard normal distribution.

Remarks: (a) Our proof in fact shows that the convergence in Theorems 1.1 and 1.2 in fact occurs *jointly* namely

$$\left(n^s \mathcal{C} - \frac{1}{\lambda}\log n, \frac{H_n - s\log n}{\sqrt{s^2\log n}}\right) \xrightarrow{d} (2\Xi^{(1)}, Z), \tag{1.14}$$

where the limiting random variables $\Xi^{(1)}, Z$ are *independent.*
(b) Not much is known about the random variable W in (1.7). Indeed, the branching property can be used in order to show that it satisfies the relation

$$W \stackrel{d}{=} \sum_{i=1}^{\infty} \mathrm{e}^{-\lambda L_i} W_i, \tag{1.15}$$

where $(W_i)_{i\geq 1}$ is an i.i.d. sequence of random variables with the same distribution as W independent of $(L_i)_{i\geq 1}$, and where L_i is defined in (1.5). Using (1.15) and properties of functionals of Poisson processes one can show that the function $\phi(u) = \mathbb{E}(\mathrm{e}^{-uW})$, defined for $u \in \mathbb{R}^+$, is the unique function satisfying the functional relationship

$$\phi(u) = \exp\left(\int_0^\infty \log\left[\phi\left(u\mathrm{e}^{-\lambda x^s}\right)\right]dx\right), \qquad \phi(0) = 1. \tag{1.16}$$

When $s = 1$, then one can see this way that W is an exponential random variable with rate 1, but for other values of s, we have no explicit form of W.
(c) The distributional equivalence given by (1.12) is proved in Lemma 2.6 below.

1.4 Discussion

In this section, we discuss the relevance of our results and how they relate to existing literature as well as various conjectures from statistical physics. The standing assumption in this discussion is that optimal paths are uniquely defined.

First vs. second order results. First order results (in our context showing for example that $H_n/s\log n \xrightarrow{\mathbb{P}} 1$, where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability) are much easier to prove than the detailed convergence in distribution proved in Theorems 1.1 and 1.2. One of the reasons for the length of this paper is that proving second order distributional convergence results in these sorts of problems proves to be much more difficult. Further, while in previous studies (e.g. [6] for various random network models) the hopcount satisfied a central limit theorem (CLT) with *matching* means and variances, Theorem 1.2 is novel in the sense that it says that, for large n, the hopcount has an approximate normal distribution with mean $s\log n$ and variance $s^2\log n$. Theorems such as Theorem 1.1 for the actual cost of the minimal weight path have been proven in a number of contexts (see e.g. [6, 14, 19]), but often prove quite tricky to handle due to the fact that we only re-center the random variables and do not divide by a normalizing factor going to ∞. Thus, one needs to be extremely careful in analyzing the contribution of various factors as $n \to \infty$. See e.g. [6] to see the various factors that could contribute to the limiting distribution in the context of exponential weights on a *random graph*.

Strong disorder regime and minimal spanning trees. Under strong disorder, it is easy to check using any of the standard greedy algorithms for constructing minimal spanning trees that the number of edges in the optimal path between any two vertices in the network has the same distribution as the number of edges between the two vertices in the *minimal spanning tree* (with edge weights l_e). More precisely, the optimal path between two vertices in the strong disorder regime is identical to the path between the two vertices in the minimal spanning tree.

In the context of our model, under strong disorder ("the $s = \infty$ regime") what is known is that for the complete graph, the hopcount of the optimal path $H(P_{\texttt{st-dis}}) \sim \Theta_{\mathbb{P}}(n^{1/3})$. Here, for two sequences of random variables $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$, we write $X_n = \Theta_{\mathbb{P}}(Y_n)$ if X_n/Y_n and Y_n/X_n are tight. This was first conjectured in [8] and recently proven in [1]. The above result in particular shows that no transition occurs for finite values of s. It might be interesting to analyze the above model when $s = s_n$ is a function of n and see when the strong disorder regime emerges ($s_n \to \infty$ regime) or the graph distance type behavior is preserved ($s_n \to 0$). In our proofs, we have kept formulas as explicit as possible in order to be able to use them later on to study the strong disorder case or the graph distance limit. Let us now heuristically discuss the strong disorder regime.

Heuristics for strong disorder. We see that the hopcount obeys a CLT with asymptotic mean and variance equal to $s \log n$ and $s^2 \log n$ respectively. It is reasonable to expect that the CLT with asymptotic mean and variance equal to $s_n \log n$ and $s_n^2 \log n$ remains valid when s_n is not too large. However, when s_n is quite large, then we should be in a phase that is close to the minimal spanning tree, for which the hopcount scales like $n^{1/3}$ and has variance of order $n^{2/3}$ (since it is not concentrated). It would be of great interest to see until what value of s_n the CLT with parameters $s_n \log n$ and $s_n^2 \log n$ remains valid. By the above, we see that for this, s_n cannot grow faster than $n^{1/3}$ for this to be true. In analogy to the scaling for the diameter of the Erdős-Rényi random graph with edge probability $p = (1+\varepsilon_n)/n$, which has size $\varepsilon_n^{-1} \log(\varepsilon_n^3 n)$ as long as $\varepsilon_n \gg n^{-1/3}$ [20], one may wonder whether the hopcount scales in leading order as $s_n \log(n/s_n^3)$, as long as $s_n \ll n^{1/3}$, and where s_n plays a similar role as $1/\varepsilon_n$.

Other edge weights. Note that in our context, the distribution of edge weights is $F(x) = 1 - \exp(-x^{1/s}) \sim x^{1/s}$ for x close to zero. One would expect that the results in the paper carry over rather easily to edge weights with distribution function F for which $F(x) = x^{1/s}(1 + o(1))$ when $x \downarrow 0$. When $F(x)$ has entirely different behavior at $x = 0$, other properties might arise. Indeed, in our current setting, we see that with high probability the shortest-weight path traverses only through edges of weights of order n^{-s}, which is the size of the minimum of n i.i.d. random variables with distribution E^s, where E is exponential with mean 1. Thus, the benefit of using edges of such small weight vastly outweigh the fact that the path thus become longer (i.e. has $\Theta_{\mathbb{P}}(\log n)$ edges). Now, when $F(x) = \mathrm{e}^{-x^{-a}}$ for some $a > 0$, then the minimum of n such random variables is $(\log n)^{-1/a}(1 + o_{\mathbb{P}}(1))$, so that the minimal weight edge in the complete graph equals $2^{-1/a}(\log n)^{-1/a}(1 + o_{\mathbb{P}}(1))$. Here, we write that $o_{\mathbb{P}}(b_n)$ to denote a random variable X_n which satisfies that $X_n/b_n \xrightarrow{\mathbb{P}} 0$. Thus, when $a > 1$, we cannot expect the optimal path to have length $\Theta_{\mathbb{P}}(\log n)$, as already the immediate path between vertices 1 and 2 has smaller weight than any path of length $\log n$.

Moreover, it is not hard to see that the minimal *two-step path* between vertices 1 and 2 has weight $2^{1+1/a}(\log n)^{-1/a}(1 + o_{\mathbb{P}}(1))$, so that the hopcount is with high probability at most $2^{1+2/a}$. Thus, this simple argument proves that the hopcount is *tight* for all $a > 0$ (as is the case for the CM with infinite mean degrees [5]). It would be of interest to investigate the limiting law of the hopcount (does the hopcount converge in probability to a constant?). In any case, it is clear that weights with distribution function $F(x) = \mathrm{e}^{-x^{-a}}$ belong to a *different universality class* as compared to edge weights E^s, where E is an exponential random variable and $s > 0$. This leads us to the following general math program:

Identify the universality classes for the weights in first passage percolation on the complete graph.

Extensions of our results to random graphs. In the context of first passage percolation on random network models, a significant amount of work both at the non-rigorous ([8, 9, 13, 22] and the references therein) as well as at the rigorous level ([4, 5, 6, 10, 16]) has been devoted to understanding such questions. What is now generally expected is that in a wide variety of network models and general edge costs, under weak disorder the hopcount scales as $\Theta(\log n)$ and satisfies central limit theorems as in Theorem 1.2. We hope that the ideas in this paper can also be applied to first passage percolation problems on various random graphs, such as the *configuration model* (CM) with any given prescribed degree distribution $\{p_k\}_{k\geq 0}$. In [6], first passage percolation with exponential weights was studied on the CM with finite mean degrees, and it is proved that similar results as on the complete graph hold in this case. Indeed, the hopcount satisfies a CLT with asymptotically *equal* mean and variance equal to $\lambda \log n$, where λ is some parameter expressible in terms of the degree distribution. We expect that when putting exponential weights raised to the power s on the edges changes this behavior, and the means and variances will become *different* constants times $\log n$. While the behavior in [6] is remarkably universal, we expect that for weights equal to powers of exponentials, when the variance of the degrees is infinite, the asymptotic ratio of mean and variance will be s as on the complete graph, while for *finite* variances degrees, the ratio may be different.

We see that the behavior of first passage percolation on the complete graph with weights E^s (as studied in this paper) gives rise to CLTs for the hopcount with means and variances of order $\log n$, while weights with distribution function $F(x) = \mathrm{e}^{-x^{-a}}$ give rise to bounded hopcounts, as is the case for the graph distance when all weights are equal to 1. Extending this to *random graphs*, it is natural to conjecture that for random graphs and weights E^s, the hopcount satisfies a CLT with asymptotic mean and variance proportional to $\log n$, while for weights with distribution function $F(x) = \mathrm{e}^{-x^{-a}}$, the hopcount behaves in a similar way as the graph distance as studied for the CM in [10, 15, 16]. This leads us to the following question:

Are the universality classes of first passage percolation on the configuration model equal to those on the complete graph?

1.5 Proof idea and overview of the paper

For the sake of notational convenience, we shall rescale each edge length by a factor $(n-1)^s$, so that each edge has distribution $(Y_e)^s$ where Y_e are distributed as exponential random variables with mean $n-1$. This does not change the optimal path while the cost of this path is scaled up by $(n-1)^s$. For the rest of the paper we shall think of the edge weights as lengths which thus induce a random metric on the complete graph and shall often refer to the optimal path between two vertices as the shortest path between them. We are interested in the optimal path between vertices 1 and 2. Consider water percolating through the network started simultaneously from two sources, vertices 1 and 2, at rate one. Then the first time of collision between the two flow processes, namely the first time the flow percolating from vertex 1 sees a vertex already visited by the the flow percolating from vertex 2 (or viceversa) gives the shortest path between the two vertices. Let $z_t^{n,(1)}$ and $z_t^{n,(2)}$ denote the number of vertices seen by the flow cluster by time t

for the flow emanating from vertex 1 and 2, respectively. For large n, the flow clusters look like independent versions of the CTBPs as formulated in Section 1.2, at least until they collide. A coupling is rigorously formulated in Sections 2.1.1 and 2.1.2. Further, they collide only when both clusters reach size $\Theta_{\mathbb{P}}(\sqrt{n})$. At a heuristic level, at any time t, the rate of collision $\gamma_n(t)$ in a small interval $[t, t+dt)$ should be

$$\gamma_n(t) \propto \left(\frac{z_t^{n,(1)} z_t^{n,(2)}}{n} \right) dt. \tag{1.17}$$

Now we use the fact that for large t, $z_t^{n,(i)} \sim W^{(i)} \mathrm{e}^{\lambda t}$, where $W^{(i)}$ is the limiting random variable for the associated CTBP defined in (1.7), to see that

$$\gamma_n(x) \propto \frac{W^{(1)} W^{(2)} \mathrm{e}^{2\lambda x}}{n}. \tag{1.18}$$

Thus collisions happen at time $(2\lambda)^{-1} \log n \pm O_{\mathbb{P}}(1)$, where $O_{\mathbb{P}}(b_n)$ denotes a random variable X_n for which $|X_n|/b_n$ is tight. If we let T_{12} denote the collision time, then the length of the optimal path equals $W_n = 2T_{12}$. The above argument gives asymptotics for the collision time and hence the length of the optimal path.

For the hopcount, we shall use general branching processes arguments to show that at large time t, if one is interested in the distribution of the generations (in our context this gives the number of individuals at various graph distances away from the root, namely the originating vertices 1 and 2), the contribution to the population comes from generations $t/\beta(s)$ and the deviations are normally distributed around this value. Here the constant $\beta(s) > 0$ denotes the mean of the stable-age distribution of the associated branching process. Intuitively, the optimal path between vertex 1 and 2 as constructed via the above simultaneous flow picture looks like the following: Suppose the connecting edge between the two clusters (v_1, v_2) arises due to the birth of a child to vertex v_1 in the flow cluster of vertex 1 and this child, v_2 has already been visited by the flow from 2. This happens at around time $(2\lambda)^{-1} \log n \pm O_{\mathbb{P}}(1)$. The hopcount H_n of the optimal path is given by the equation

$$H_n = G_1 + G_2 + 1, \tag{1.19}$$

where G_1 and G_2 are the generations of vertex v_1 and v_2 in flow cluster 1 and 2, respectively. Thus understanding the distribution amongst generations in the coupled branching processes paves the way to understanding the hopcount. The remainder of this paper involves the conversion of the above heuristic into a *rigorous argument.* The organization of rest of the paper is as follows:

- In Section 2.1, we shall couple simultaneous flow from two vertices on $\mathcal{K}_n$ with CTBPs and show that the difference is negligible;
- Section 2.2 shows that the above coupling incorporated with technical results from CTBP theory give us asymptotics for the re-centered length of the optimal path, namely Theorem 1.1.
- Section 2.3 shows how the distribution of individuals among different generations in the associated branching process proves Theorem 1.2.
- Finally, Section 3 proves all the CTBP results we need to carry out our analysis. This section is the most technical part of the paper and the point of organizing the paper in this fashion is to motivate the various results that are proved in Section 3.

2 Proofs

This section proves the main results. Proofs of the CTBP results needed are deferred to Section 3.

2.1 Dominating graph flow by continuous-time branching processes

In this section, we describe a coupling between the flows started from vertices 1 and 2 and their corresponding independent CTBPs with offspring distribution given by the point process in (1.6). We shall first start with the flow started from one vertex and then extend this to the flow simultaneously from two vertices.

2.1.1 Expansion of the flow from a single vertex

We start with some notation. Recall that $\mathcal{K}_n$ denoted the random disordered media represented by the complete graph where each undirected edge (i,j) has edge length E_{ij}^s where E_{ij} are i.i.d. exponentially distributed with mean $n-1$ (alternatively, with rate $1/(n-1)$). These edge lengths make $\mathcal{K}_n$ a metric space (with random geodesics). Let the index set of $\mathcal{K}_n$ be $[n] := \{1, 2, \ldots, n\}$ and fix vertex 1. Think of this vertex as an originator of flow of some fluid which percolates through the whole network via the geodesics at rate 1. Let $i_1 = 1, i_2, \ldots \in [n]$ be the vertices in sequential order seen by the flow. For $t \geq 0$, let $\mathrm{SWG}_t^{(1)}$ be the shortest-weight graph between vertex 1 and all the vertices that can be reached from 1 by shortest-weight paths of length at most t. More precisely, $\mathrm{SWG}_t^{(1)}$ consists of these shortest-weight paths and the weights of all of the edges used for them. Let $\{E_j^i\}_{i\geq 1, j\geq 1}$ be a doubly infinite array of mean 1 exponential random variables. Then, by the properties of the extremes of $n-1$ i.i.d. exponential random variables, each with mean $n-1$, it is easy to see that the neighbors of 1 have distances from 1 distributed as

$$\mathcal{P}_{n,1} = \left(E_1^{(1)}\right)^s, \left(E_1^{(1)} + \frac{n-1}{n-2} E_2^{(1)}\right)^s, \ldots \tag{2.1}$$

Similarly, the distribution of distances from vertex i_k (the k^{th} vertex reached by the flow from 1) to vertices other than those already seen by the flow, is distributed as

$$\mathcal{P}_{n,k} = \left(\frac{n-1}{n-k} E_1^k\right)^s, \left(\frac{n-1}{n-k} E_1^k + \frac{n-1}{n-k-1} E_2^k\right)^s, \ldots \tag{2.2}$$

Call the above the *immediate neighborhood process* of vertex k. Note that for each k, by the memoryless property of the exponential distribution, the identity of the end point of each edge in the above point process is uniformly distributed among all $[n] \setminus \{i_1, i_2, \ldots i_k\}$ vertices which have not been seen at the time when the flow hits vertex i_k. Our aim is to couple this process with a CTBP with offspring distribution given by the point process $\mathcal{P}$ defined by

$$\mathcal{P} = \{(E_1)^s, (E_1 + E_2)^s, (E_1 + E_2 + E_3)^s, \ldots\}, \tag{2.3}$$

where $\{E_i\}_{i\geq 1}$ are i.i.d. exponential rate 1 random variables. Comparing (2.3) with (2.1) and (2.2), we see that, intuitively, the $\mathrm{SWG}_t^{(1)}$ should be stochastically smaller than the corresponding CTBP driven by offspring distribution $\mathcal{P}$. The reason is that when the flow starts, then the number of edges it has to explore from vertex 1 is $n-1$, but as $\mathrm{SWG}_t^{(1)}$ increases with time, the number of edges originating from each new vertex is strictly smaller than $n-1$ due to vertices already explored by the flow. Thus, the points are being *depleted.* We shall show that asymptotically for large n, the difference is negligible. To do so, as the flow explores $\mathcal{K}_n$, we shall enlarge the graph

$\mathcal{K}_n$ with new artificial vertices to compensate for $\mathrm{SWG}_t^{(1)}$ using up vertices in $\mathcal{K}_n$ and effectively counteracting the *depletion of points* effect. For this, we shall need the following randomization ingredients:

(i) The complete graph $\mathcal{K}_n$ with random edge weights;

(ii) An infinite array of i.i.d. exponential random variables $\{E_{i,j}\}_{i\in[n],j\geq n+1}$ each with mean $n-1$;

(iii) An infinite sequence of independent branching process $\{\widetilde{\mathrm{BP}}_i(\cdot)\}_{i\geq n+1}$ each driven by the offspring distribution in (2.3).

Before diving into the construction, we shall need the following simple lemma which follows directly from the memoryless property of the exponential distribution:

Lemma 2.1 (Powers of exponential distributions) *(a) Consider the random variable E^s where E has an exponential distribution with mean $n-1$. Then, for any fixed $r>0$, the conditional distribution of $E^s \mid E^s > r$ equals that of $(\tilde{E}^s + r^{1/s})$, where $\tilde{E}$ is an independent random variable with exponential distribution with mean $n-1$.*
(b) Consider the surplus random variable $(E^s - r) \mid E^s > r$. This random variable has the same distribution as the first point of a Poisson point process with rate

$$\Lambda_r(x) = \frac{1}{s(n-1)}(r+x)^{1/s-1}, \qquad x \geq 0. \tag{2.4}$$

We shall use part (a) of Lemma 2.1 in the construction of the coupling while we shall use part (b) in the proof of the distributional result for the optimal weight. We start by proving Lemma 2.1:

Proof. Part (a) is immediate from the memoryless property of the exponential random variable. For part (b), we note that

$$\mathbb{P}(E^s - r \geq x \mid E^s > r) = \mathbb{P}(E \geq (x+r)^{1/s} \mid E > r^{1/s}) = \mathrm{e}^{-[(x+r)^{1/s} - r^{1/s}]/(n-1)}, \tag{2.5}$$

while the probability that a Poisson point process with rate (2.4) has no points before x equals

$$\mathrm{e}^{-\int_0^x \Lambda_r(y)dy} = \mathrm{e}^{-\int_0^x \frac{1}{s(n-1)}(r+y)^{1/s-1}dy} = \mathrm{e}^{-[(x+r)^{1/s} - r^{1/s}]/(n-1)}. \tag{2.6}$$

Thus, the first point of this Poisson point process has the same distribution as the conditional law $E^s - r \mid E^s > r$. ∎

Construction of the coupling: This proceeds via the following constructs:

(a) **Artificial inactive vertices:** Consider the flow traveling at rate one from vertex 1 on $\mathcal{K}_n$. Let $z_t^{n,(1)}$ denote the number of vertices in $\mathrm{SWG}_t^{(1)}$. To evoke branching process terminology, we shall often refer to this as the number of vertices *born* in the flow cluster of 1 by time t. For $1 \leq k \leq n$, we define the stopping times

$$T_k^n = \inf\{t : z_t^n = k\}, \tag{2.7}$$

so that $T_1^n = 0$. Now consider the flow from vertex 1. For $k \geq 2$, when the k^{th} vertex i_k is discovered by the flow at time T_k^n, create a new artificial vertex labeled by $n+k-1$. Let $a(i_k)$ denote the vertex in $\mathrm{SWG}_{T_k^n}^{(1)}$ to which vertex i_k is attached. Then note that for all $i_j \neq$

$a(i_k) \in \mathrm{SWG}^{(1)}_{T^n_k}$, by Lemma 2.1 and, conditionally on $\mathrm{SWG}^{(1)}_t$, the edge lengths of edge (i_j, i_k) have distribution $([t - T^n_j]^{1/s} + E)^s$ where E has an exponential distribution with mean $n-1$.

For the new artificial vertex $n - k + 1$, we attach edge lengths from each vertex $i_j \in \mathrm{SWG}^{(1)}_{T^n_k}$ of length $([t - T^n_j]^{1/s} + E_{j,n-k+1})^s$ where the $E_{j,n-k+1}$ are exponential random variables as described in the randomization needed for the coupling, and where we recall that T^n_j denotes the time of discovery of vertex i_j. We shall think of the flow having reached a distance $t - T^n_j$ on this edge. At the time of creation, we shall think of these artificial vertices as *inactive* as the flow has not yet reached this vertex. Think of these vertices as part of the network and the flow trying to get to them as well. Note that eventually the flow will reach these inactive vertices as well. Whenever the flow reaches an inactive artificial vertex, we shall think of this vertex becoming active, i.e., it is *activated*. Let $\mathcal{A}_t$ denote the set of active artificial vertices. For $k \geq 1$, let

$$T^{n,*}_k := \inf\{t : |\mathcal{A}_t| = k\} \tag{2.8}$$

be the time of activation of the k^{th} artificial vertex. Note that in this construction, edges exist only between vertices in $[n]$ and artificial vertices, no edges exist between artificial vertices.

(b) **Activation of artificial vertices:** Note that activation of inactive vertices happens at times $T^{n,*}_k$ via an edge from a vertex in $\mathrm{SWG}_{T^{n,*}_k} \subseteq [n]$ to an inactive artificial vertex $d_k \geq n+1$. Suppose at this time the set of artificial vertices (active and inactive) is $\{n+1, n+2, \ldots, n+j(T^{n,*}_k)\}$. When this happens the following constructions are performed:

(1) Remove all the edges from vertices in $[n]$ to d_k (other than the one that the flow used to get to it);

(2) Create a new inactive artificial vertex $n + j(T^{n,*}_k) + 1$. Just as before, create edges between each vertex $i \in [n]$ and vertex $n + j(T^{n,*}_k) + 1$ with edge lengths distributed as $([t - T^{n,*}_k]^{1/s} + E_{i,n+j(T^{n,*}_k)+1})^s$ and think of the flow as having already traveled $t - T^{n,*}_k$ on it;

(3) At this time, start a CTBP $\widetilde{\mathrm{BP}}_k(\cdot)$ with d_k as the ancestor. The vertices born in this branching process have no relation to the flow on $\mathcal{K}_n$ and associated inactive vertices. For time $t > T^{n,*}_k$, we shall call all the vertices in $\widetilde{\mathrm{BP}}_k(t)$, other than d_k, the *descendants* of vertex d_k at time t.

Let $\mathcal{DA}_t$ denote the set of all descendants of the associated CTBPs of active artificial vertices at time t and let

$$\mathrm{BP}^{(1)}_t = \mathrm{SWG}^{(1)}_t \cup \mathcal{A}_t \cup \mathcal{DA}_t. \tag{2.9}$$

Let $z^{(1)}_t = |\mathrm{BP}^{(1)}_t|$ denote the number of vertices reached at time t. The following proposition identifies properties of the above construction which shall all be crucial in our analysis. We shall prove this proposition in detail since we later we shall use an almost identical proposition in the context of flow from two vertices which we shall state without proof in Section 2.1.2 below.

Proposition 2.1 (Properties of the coupling) *In the above construction, the following holds: (a) The process $\{\mathrm{BP}^{(1)}_t\}_{t\geq 0}$ is a CTBP driven by the point process $\mathcal{P}$ in (2.3). The process $\{\mathrm{SWG}^{(1)}_t\}_{t\geq 0}$ is the shortest weight graph process of the flow emanating from vertex 1. As is obvious from (2.9), there is stochastic domination in the sense that for all times $t \geq 0$,*

$$\mathrm{SWG}^{(1)}_t \subseteq \mathrm{BP}^{(1)}_t. \tag{2.10}$$

In particular, $z_t^{n,(1)} = |\mathrm{SWG}_t^{(1)}| \leq z_t^{(1)} = |\mathrm{BP}_t^{(1)}|$ *for all* t.
(b) Let $\lambda = \lambda(s)$ *be the Malthusian rate of growth of* $\mathrm{BP}_t^{(1)}$ *as defined by* (1.9). *Then, given any* $\varepsilon > 0$, *there exists* $C_\varepsilon > 0$ *such that for times* $t_n = (2\lambda)^{-1} \log n - C_\varepsilon$

$$\liminf_{n\to\infty} \mathbb{P}(|\mathcal{A}_{t_n}| = 0) \geq 1 - \varepsilon. \tag{2.11}$$

(c) For any fixed $B \in \mathbb{R}$, *letting* $t_n = (2\lambda)^{-1} \log n + B$, *the sequence of random variable* $|\mathcal{A}_{t_n}| + |\mathcal{D}\mathcal{A}_{t_n}|$ *is a sequence of tight random variables. Since the processes* $(|\mathcal{A}_t| + |\mathcal{D}\mathcal{A}_t|)_{t\geq 0}$ *are monotonically increasing in* t, (2.9) *implies that* $\sup_{t\leq t_n}(z_t^{(1)} - z_t^{n,(1)})$ *is tight and, in particular, as* $n \to \infty$,

$$\sup_{t\leq t_n} \left| \frac{z_t^{n,(1)}}{z_t^{(1)}} - 1 \right| \xrightarrow{\mathbb{P}} 0, \tag{2.12}$$

Note that if $|\mathcal{A}_{t_n}| = 0$, then $\mathrm{SWG}_t^{(1)} = \mathrm{BP}_t^{(1)}$ for all $t \leq t_n$, so that part (b) yields that there is little difference between the SWG and the CTBP up to time $(2\lambda)^{-1} \log n - C_\varepsilon$.

Proof. Part (a) is obvious from construction. To prove part (b), note that by construction, if $z_t^{n,(1)} = k$, then the chance that the next vertex is an artificial inactive vertex is exactly k/n. Thus, if $z_{t_n}^{n,(1)} = k_n$ then

$$|\mathcal{A}_{t_n}| \stackrel{d}{=} \sum_{j=1}^{k_n} I_j, \tag{2.13}$$

where I_j are Bernoulli j/n random variables. Now to choose C_ε, first choose $C_\varepsilon^* > 0$ so small that $\exp(-C_\varepsilon^*/2) > 1 - \varepsilon/2$. Since $z_{t_n}^{n,(1)} \leq z_{t_n}^{(1)}$ and for the process $\{z_t^{(1)}\}_{t\geq 0}$ the asymptotics (1.7) hold, we can choose C_ε^* such that

$$\mathbb{P}(z_{t_n}^{n,(1)} > C_\varepsilon^* \sqrt{n}) < \varepsilon/2. \tag{2.14}$$

Then,

$$\begin{aligned}\mathbb{P}(|\mathcal{A}_{t_n}| > 0) &\leq \mathbb{P}(|\mathcal{A}_{t_n}| > 0, z_{t_n}^{n,(1)} < C_\varepsilon^* \sqrt{n}) + \mathbb{P}(z_{t_n}^{n,(1)} > C_\varepsilon^* \sqrt{n}) \\ &\leq (1 - \exp(-C_\varepsilon^*/2)) + \varepsilon/2 < \varepsilon,\end{aligned}$$

where the second inequality follows using a Poisson approximation in (2.13) and (2.14). This proves part (b).

Finally to prove part (c), we note the following:

- Using part (b), we choose C_ε so that with high probability no artificial vertices have been activated by time $(2\lambda)^{-1} \log n - C_\varepsilon$;
- Using (2.13) and ideas similar to the above argument one can show that the number of active artificial vertices by time $t_n = (2\lambda)^{-1} \log n + B$ can be stochastically dominated with high probability by a Poisson random variable X_B with mean $C(B)$ for some function $B \mapsto C(B)$.

These two observations together imply that with high probability

$$|\mathcal{A}_{t_n}| + |\mathcal{D}\mathcal{A}_{t_n}| \preceq_{st} \sum_{j=1}^{X_B} |\mathrm{BP}_j(B - C_\varepsilon)|, \tag{2.15}$$

where $\mathrm{BP}_j(\cdot)$ are independent CTBPs driven by $\mathcal{P}$, independent of X_B which is Poisson with mean $C(B)$ and $\preceq_{st}$ denotes stochastic domination. This proves part (c). ∎

2.1.2 Simultaneous expansion and coupling

Let us now show how the above coupling can be extended to flow originating from two vertices $1,2$ simultaneously. We shall couple the flow to two independent CTBPs $\{\mathrm{BP}_t^{(i)}\}_{i=1,2}$. All the ingredients of randomness shall be the same as in the previous section, namely, (i) the complete graph $\mathcal{K}_n$ with random edge lengths; (ii) the infinite array of exponential random variables $\{E_{i,j}\}_{(1\leq i\leq n; j\geq n+1)}$; and (iii) the infinite sequence of independent CTBPs $\{\widetilde{\mathrm{BP}}_i\}_{i\geq 1}$ driven by $\mathcal{P}$. Think of flow now emanating from the two sources $1,2$ simultaneously at rate one exploring the shortest weight structure about the two sources. We shall stop the flow when there is a *collision*, i.e., the flow from one vertex sees a vertex seen by the flow from the other vertex. As before, we let $\mathrm{SWG}_t^{(i)}$ denote the shortest weight graphs up to time t explored by the flow from each source $i=1,2$ and let

$$\mathrm{SWG}_t = \mathrm{SWG}_t^{(1)} \cup \mathrm{SWG}_t^{(2)}. \tag{2.16}$$

Let $z_t^{n,(i)} = |\mathrm{SWG}_t^{(i)}|$ and $z_t^n = z_t^{n,(1)} + z_t^{n,(2)}$. Now let T_k^n denote the stopping time

$$T_k^n = \inf\{t : z_t^n = k\}, \tag{2.17}$$

so that now $T_2^n = 0$. Let the vertex discovered at time T_k^n and attached to one of the two flow clusters be $i_k \in [n]$. We shall call this the *time of birth of the vertex* i_k. Extra care is needed as subtle issues of double counting of edges may arise.

The construction proceeds as before via two ingredients:
(a) **Artificial inactive vertices:** By convention, we shall think of the edge between 1 and 2 to belong to the flow from vertex 1. To compensate at time 0, we shall add a new artificial inactive vertex labeled by $n+1$. Compared to the other artificial vertices this shall be special in the sense that vertex 1 will not have an edge to this vertex (or the artificial vertices that replace this vertex when the flow reaches this vertex). At time 0, attach an edge $(2, n+1)$ of random length $E_{2,n+1}^s$. Now start the flow from the two sources on the vertex set $[n]\cup\{n+1\}$. The flow percolates from these two sources on the (expanded) network discovering new vertices, both actual vertices in $[n]$ as well as artificial vertices. Let SWG_t^* denote this flow process with $z_t^{n,*} = |\mathrm{SWG}_t^*|$ and let

$$\tilde{T}_k^n = \inf\{t : |z_t^{n,*}| = k\}. \tag{2.18}$$

Let i_k denote the vertex discovered by the flow at time $\tilde{T}_k^n$ (this vertex could either be an actual vertex in $[n]$ or an artificial inactive vertex). Create a new artificial vertex labeled by $n+k$. Now if i_k is in $\mathrm{SWG}_t^{(2)}$ then remove all the edges between i_k and all the vertices in $\mathrm{SWG}_{\tilde{T}_k^n}^{(2)}$ (namely *real* vertices in the actual graph $[n]$ which are part of $\mathrm{SWG}_t^{(2)}$ that have already been explored by the flow from 2). (Do the exact opposite if $i_k \in \mathrm{SWG}_t^{(1)}$.) The edges (v, i_k) for $v \in \mathrm{SWG}_{\tilde{T}_k^n}^{(1)}$ are quite special (see the beginning of Section 2.2). Call these the *potential connecting* edges as these are the edges through which collision of the two flow clusters may happen. Also perform the following constructions:

- If $i_k \neq n+1$ or any of the *replacements* of $n+1$ (this term is defined below), then attach edges between the artificial vertex $n+k$ and all $i_j \in \mathrm{SWG}_{\tilde{T}_k^n}$ with edge lengths $([t-T_{i_j}^n]^{1/s} + E_{i_j,n+k})^s$. The flow would have already flowed till distance $(t - T_{i_j})$ on this edge to this new vertex.
- If $i_k = n+1$, then replace this by a new vertex $n+k$. This vertex will be called a *replacement* of the special artificial vertex $n+1$. Also replacements of such replacements shall be called replacements. Remove all edges from $i_j \in \mathrm{SWG}_{\tilde{T}_k^n}$ to i_k and add back edges from these vertices *excluding* vertex 1 to vertex $n+1$ with edge lengths $([t-T_{i_j}^n]^{1/s} + E_{i_j,n+k})^s$. This can

be understood by noting that the flow would have already reached up to distance $(t - T_{i_j})$ on this edge to this new vertex.

Every new artificial vertex when it is born is *inactive*. Whenever the flow reaches an inactive artificial vertex we shall think of this vertex becoming active and belonging to the flow cluster from which this artificial vertex was reached. Let $\mathcal{A}_t^{(i)}$ denote the set of active artificial vertices corresponding to flow cluster $i = 1, 2$ at time t and let $\mathcal{A}_t = \mathcal{A}_t^{(1)} \cup \mathcal{A}_t^{(2)}$ be the set of artificial vertices. For $k \geq 1$, we let

$$T_k^{n,*} := \inf\{t : |\mathcal{A}_t| = k\} \tag{2.19}$$

be the time of activation of the k^{th} artificial vertex. Note that, as before, in this construction edges exist only between vertices in $[n]$ and artificial vertices, no edges exist between artificial vertices.

(b) **Activation of artificial vertices:** Note that the flow will eventually reach inactive artificial vertices. When this happens say that *activation* happens. This happens at times $T_k^{n,*}$ via an edge from a vertex in $\mathrm{SWG}_{T_k^{n,*}} \subseteq [n]$ to an inactive artificial vertex $d_k \geq n+1$ from one of the two flow clusters. When an artificial vertex gets activated, it belongs to the flow cluster that activates it and so do all its *descendants* (the notion of a descendant is defined below). Suppose that at this time, the set of artificial vertices (active and inactive) is $\{n+1, n+2, \ldots, n+j(T_k^{n,*})\}$. As described above, this inactive artificial vertex is replaced by a new inactive artificial vertex with appropriate edges and edge lengths.

Further at this time, start the CTBP $\widetilde{\mathrm{BP}}_k(\cdot)$ with d_k as the ancestor. The vertices born in this branching process have no relation to the flow on $\mathcal{K}_n$ and associated inactive vertices. For time $t > T_k^{n,*}$ we shall call all the vertices in $\widetilde{\mathrm{BP}}_k$ other than d_k the *descendants of vertex d_k*.

Let $\mathcal{DA}_t^{(i)}$ denote the set of all descendants of the associated CTBPs of active artificial vertices at time t in flow cluster $i = 1, 2$ and define the processes

$$\mathrm{BP}_t^{(i)} = \mathrm{SWG}_t^{(i)} \cup \mathcal{A}_t^{(i)} \cup \mathcal{DA}_t^{(i)}, \qquad i = 1, 2. \tag{2.20}$$

Let $z_t^{(i)} = |\mathrm{BP}_t^{(i)}|$. Finally let $\mathrm{BP}_t = \mathrm{BP}_t^{(1)} \cup \mathrm{BP}_t^{(2)}$ denote the full flow process. This completes the construction of the coupling.

The following proposition collects the properties of our construction that we shall need. It is analogous to Proposition 2.1 and we shall not give a proof. Recall that T_{12} denotes the collision time of the two flow processes.

Proposition 2.2 (Properties of the coupling) *In the above construction, the following holds:*
(a) The processes $\{\mathrm{BP}_t^{(i)}\}_{t \geq 0}$ are independent CTBPs driven by the point process $\mathcal{P}$ in (2.3). The process $\{\mathrm{SWG}_t^{(i)}\}_{0 \leq t \leq T_{12}}$ is the shortest weight graph process of the flow emanating from vertex i till the collision time. As is obvious from (2.20), there is stochastic domination in the sense that for all times $t \geq 0$,

$$\mathrm{SWG}_t^{(i)} \subseteq \mathrm{BP}_t^{(i)}. \tag{2.21}$$

In particular $z_t^{n,(i)} \leq z_t^{(i)}$ for all $t \geq 0$.
(b) Let $\lambda = \lambda(s)$ be the Malthusian rate of growth of $\mathrm{BP}_t^{(i)}$ as defined in (1.9). Then, given any $\varepsilon > 0$, there exists C_ε such that for times $t_n = (2\lambda)^{-1} \log n - C_\varepsilon$,

$$\liminf_{n \to \infty} \mathbb{P}(T_{12} > t_n, |\mathcal{A}_{t_n}^{(1)}| = 0, |\mathcal{A}_{t_n}^{(2)}| = 0) \geq 1 - \varepsilon. \tag{2.22}$$

Note that if $|\mathcal{A}_{t_n}^{(i)}| = 0$ then $\mathrm{SWG}_t^{(i)} = \mathrm{BP}_t^{(i)}$ for all $t \leq t_n$.
(c) For any fixed $B \in \mathbb{R}$, let $t_n^* = (2\lambda)^{-1} \log n + B$ and let $t_n = T_{12} \wedge t_n^*$. Then the sequence

of random variable $|\mathcal{A}^{(i)}_{t_n}| + |\mathcal{DA}^{(i)}_{t_n}|$ *is a sequence of tight random variables. Since the processes* $(|\mathcal{A}_t| + |\mathcal{DA}_t|)_{0\le t\le T_{12}}$ *are monotonically increasing in* t*,* (2.20) *implies that* $\sup_{t\le t_n}(z^{(i)}_t - z^{n,(i)}_t)$ *is tight, and, as* $n\to\infty$*,*

$$\sup_{t\le t_n}\left|\frac{z^{n,(i)}_t}{z^{(i)}_t} - 1\right| \xrightarrow{\mathbb{P}} 0. \tag{2.23}$$

2.2 Analysis of the weight of the optimal weight

Before proceeding to the main proposition in this section, we shall derive an important property of the above construction. When a vertex, say $v\in[n]$, is born into one of the flow process (to fix ideas say into the flow cluster of vertex 1) at some time t, then note that the edges it has at this time are

- edges to inactive artificial vertices.
- edges to all vertices in $[n]\setminus \mathrm{SWG}_t$.

For any vertex $v\in \mathrm{SWG}^{(1)}_t$ and, for any vertex $u\in[n]$ born into the flow cluster originating from vertex 2 at some later time $s>t$, we say that the edge connecting v to u is assigned to vertex v and **not** to u. Similarly, if vertex u is born into the flow cluster starting from 2 before vertex v which is born into flow cluster from vertex 1, then say that the edge (u,v) is assigned to vertex u. Now for any time t and any vertex $i\in \mathrm{SWG}^{(1)}_t\subseteq[n]$, let $N_t(i)$ denote the number of edges with end points in $\mathrm{SWG}^{(2)}_t$ which are assigned to it. Similarly, for a vertex $i\in\mathrm{SWG}^{(2)}_t$, $N_t(i)$ is the number of vertices in $\mathrm{SWG}^{(1)}_t$ assigned to it. Recall that our aim in sending the flow simultaneously is to analyze the collision time, namely, the first time when an edge, which we shall refer to as the *connecting edge*, forms between the two flow clusters. For any given time t and $v\in\mathrm{SWG}^{(i)}_t, i=1,2$, define the (*random*) set

$$\mathcal{N}_t(v) = \{u\in\mathrm{SWG}^{(3-i)}_t:\ \text{edge }(u,v)\text{ assigned to }v\} = \{u\in\mathrm{SWG}^{(3-i)}_t: T_u>T_v\}, \tag{2.24}$$

where, from now on, we shall use T_v to denote the time of birth of vertex v into the flow process $\{\mathrm{SWG}_t\}_{t\ge0}$ and we recall that $\mathrm{SWG}_t=\mathrm{SWG}^{(1)}_t\cup\mathrm{SWG}^{(2)}_t$.

The importance of these connecting edges is as follows: Fix some time t and vertices $i\in\mathrm{SWG}^{(1)}_t$ and $j\in\mathrm{SWG}^{(2)}_t$ with $T_j>T_i$ so that the edge between them is assigned to vertex i. Note that up till time T_j, the flow was proceeding on the edge between them at rate 1 from vertex j. Now at time T_j the flow has reached the edge from the opposite side (i.e., from vertex j) and is proceeding through the edge from *both* end points. Thus while the flow through all other non-*potential connecting* edges proceeds at rate 1, the flow through this edge proceeds at rate 2. For any time $x+T_j$, and using Lemma 2.1(b) with $r=T_j-T_i$ and the fact that the flow now proceeds at rate 2 and not 1, the intensity function for the formation of this edge at this time is

$$\lambda_{(i,j)}(x+T_j) = \frac{2}{s(n-1)}((T_j-T_i)+2x)^{1/s-1}, \qquad x\ge0. \tag{2.25}$$

In particular, for $t\ge T_j$,

$$\lambda_{(i,j)}(t) = \frac{2}{s(n-1)}((T_j-T_i)+2(t-T_j))^{1/s-1} = \frac{2}{s(n-1)}((t-T_i)+(t-T_j))^{1/s-1}. \tag{2.26}$$

This fact leads to the following proposition:

Proposition 2.3 (Collision time distribution) *If T_{12} denotes the collision time, then with respect to the filtration generated by the flow process, T_{12} has the same distribution as the first point of a Poisson point process with rate function given by*

$$\lambda_n(t) = \frac{2}{s(n-1)} \sum_{i\in \mathrm{SWG}_t^{(1)}} \sum_{j\in \mathrm{SWG}_t^{(2)}} ([t-T_j]+[t-T_i])^{1/s-1}. \tag{2.27}$$

Remark 2.4 (Extension to other graphs) *Note that a similar formula as the above remains valid for any finite graph with i.i.d. E_e^s edge weights where E_e are exponential random variables, where the sum over (i,j) is restricted to $(i,j)\in E_n$, i.e., the sum is only taken over the edges of the graph. This will be used in our analysis of more general random graph models.*

Proof. Using (2.25), Lemma 2.1 and the fact that for a finite number of independent Poisson point processes, the first point to occur in any of these processes has the same distribution as the first point in Poisson point process with rate given by the sum of rates of the corresponding point processes, we have that

$$\lambda_n(t) = 2\sum_{i\in \mathrm{SWG}_t^{(1)}} \sum_{j\in \mathcal{N}_t(i)} \frac{([t-T_i]+[t-T_j])^{1/s-1}}{s(n-1)} + 2\sum_{i\in \mathrm{SWG}_t^{(2)}} \sum_{j\in \mathcal{N}_t(i)} \frac{([t-T_i]+[t-T_j])^{1/s-1}}{s(n-1)}, \tag{2.28}$$

where we recall that $\mathcal{N}_t(i)$ denotes the set of vertices in the other flow cluster assigned to i. Now note that for every pair of vertices $(i,j), i\in \mathrm{SWG}_t^{(1)}, j\in \mathrm{SWG}_t^{(2)}$ either $i\in\mathcal{N}_t(j)$ or vice-versa and only one of these facts can happen. Rearranging the above equation gives the result. ∎

In Section 3.3 below, we shall prove the following result concerning the convergence of the two-vertex characteristic:

Theorem 2.5 (Convergence of two-vertex characteristic for BP) *Consider two independent CTBPs $\mathrm{BP}_t^{(i)}, i=1,2,$ as before. Let $W^{(i)}, i=1,2$ be the almost sure limits of $\mathrm{e}^{-\lambda t} z_t^{(i)}$. Then,*

$$\mathrm{e}^{-2\lambda t} \sum_{i\in \mathrm{BP}_t^{(1)}} \sum_{j\in \mathrm{BP}_t^{(2)}} ([t-T_j]+[t-T_i])^{1/s-1} \xrightarrow{a.s.} \lambda W^{(1)} W^{(2)}. \tag{2.29}$$

where W_i are the a.s. limits of $\mathrm{e}^{-\lambda t}|\mathrm{BP}_t^{(i)}|$ and are i.i.d. with the same distribution as W in (1.7).

Now we are ready to prove Theorem 1.1:

Completion of the proof of Theorem 1.1: First consider the rate function $\lambda_n(t)$ of the collision time given in Proposition 2.3. By Proposition 2.2, in the summation arising in this rate function, we can replace the terms $\mathrm{SWG}_t^{(i)}$ by $\mathrm{BP}_t^{(i)}$ as the effect on the rate function is asymptotically negligible, where $\mathrm{BP}_t^{(i)}$ are the independent CTBPs that have been coupled with $\mathrm{SWG}_t^{(i)}$ to understand the optimal path on $\mathcal{K}_n$. Note that while these CTBPs intrinsically depend on n since we have used the randomization in $\mathcal{K}_n$ to construct the CTBPs. By (1.7),

$$\mathrm{e}^{-\lambda t}|\mathrm{BP}_t^{(i)}| \xrightarrow{a.s.} W_n^{(i)}, \tag{2.30}$$

where $W_n^{(i)}$ are independent and identically distributed as the limit variable in (1.7).

Now, Theorem 2.5 implies that for any fixed $B>0$ and any $x\in[-B,B]$, we have,

$$\sup_{x\in B}\left|\lambda_n((2\lambda)^{-1}\log n + x) - \frac{2\lambda}{s} W_n^{(1)} W_n^{(2)} \mathrm{e}^{2\lambda x}\right| \xrightarrow{\mathbb{P}} 0. \tag{2.31}$$

Comparing the above with the definition of the Cox process in (1.10) completes the proof subject to Theorem 2.5. Theorem 2.5 is proved in Section 3.3. ■

For future reference we define the *two-vertex characteristics* $\chi_{(i,j)}(t)$ by

$$\chi_{(i,j)}(t) = ([t - T_j] + [t - T_i])^{1/s-1}. \tag{2.32}$$

We shall now quickly prove the distributional equivalence (1.12).

Lemma 2.6 (The limit of the shortest weight) *The first point $\Xi^{(1)}$ of the Cox point process with rate $\gamma(\cdot)$ as in (1.10) satisfies the distributional equivalence in (1.12).*

Proof. Since $\Xi^{(1)}$ is the first point of the Cox process with rate function γ in (1.10), we have for any fixed $y \in \mathbb{R}$, conditional on $W^{(1)}, W^{(2)}$,

$$\mathbb{P}(\Xi^{(1)} > y | W^{(1)}, W^{(2)}) = \exp\left(-\int_{-\infty}^{y} \gamma(x)dx\right) = \exp\left(-\frac{1}{s}W^{(1)}W^{(1)}\mathrm{e}^{2\lambda y}\right), \tag{2.33}$$

so that

$$\mathbb{P}\Big(\Xi^{(1)} > x - \frac{1}{2\lambda}\log\frac{W^{(1)}W^{(2)}}{s} \mid W^{(1)}, W^{(2)}\Big) = \exp(-\mathrm{e}^{2\lambda x}) = \mathbb{P}(G/(2\lambda) > y), \tag{2.34}$$

where G has the standard Gumbel distribution. This proves the result. ■

2.3 Hopcount analysis

As before, we let T_{12} be the collision time between the two flow clusters and suppose the collision happens via the formation of an edge (v_1, v_2) where $v_1 \in \mathrm{SWG}^{(1)}_{T_{12}}$ and $v_2 \in \mathrm{SWG}^{(2)}_{T_{12}}$. For $i = 1, 2$ let G_i denote the number of edges on the path from vertex i to G_i so that the hopcount is given by

$$H_n = G_1 + G_2 + 1. \tag{2.35}$$

To prove Theorem 1.2 it suffices to show that, for every fixed $r, x, y \in \mathbb{R}$ and writing $t_n = (2\lambda)^{-1}\log n$,

$$\mathbb{P}\left(T_{12} \le t_n + r, G_1 \le \lambda s t_n + xs\sqrt{\lambda t_n}, G_2 \le \lambda s t_n + ys\sqrt{\lambda t_n}\right) \to F_{12}(r)\Phi(x)\Phi(y), \tag{2.36}$$

where $F_{12}(\cdot)$ is the distribution of the random variable appearing in Theorem 1.1 and $\Phi(\cdot)$ denotes the standard normal distribution function.

For fixed time t and $v \in \mathrm{SWG}^{(i)}_t, i = 1, 2$, let $G(v)$ denote the number of edges in the optimal path between v and vertex i which started the flow. For any fixed $x \in \mathbb{R}$, let

$$\mathrm{SWG}^{(i)}_t(x) = \{v \in \mathrm{SWG}^{(i)}_t : G(v) \le \lambda s t + xs\sqrt{\lambda t}\}. \tag{2.37}$$

By Proposition 2.3 and properties of a finite number of Poisson processes, we have, for any fixed t,

$$\begin{aligned}&\mathbb{P}(T_{12} \in [t, t+dt), G_1 \le \lambda s t + xs\sqrt{\lambda t}, G_2 \le \lambda s t + ys\sqrt{\lambda t} \mid \mathrm{SWG}_t)\\ &\quad= \exp\left(-\int_0^t \lambda_n(w)dw\right)\lambda_n(t)\frac{\sum_{i\in\mathrm{SWG}^{(1)}_t(x)}\sum_{j\in\mathrm{SWG}^{(2)}_t(y)}\chi_{ij}(t)}{\sum_{i\in\mathrm{SWG}^{(1)}_t}\sum_{j\in\mathrm{SWG}^{(2)}_t}\chi_{ij}(t)}dt,\end{aligned} \tag{2.38}$$

where $\chi_{ij}(t)$ was the two-vertex characteristic defined in (2.32) and $\lambda_n(t)$ was the rate defined in (2.27). Thus, to complete the proof of (2.36), it is enough to show the following:

Theorem 2.7 (CLT from two-vertex characteristic) *The two-vertex characteristic satisfies the asymptotics, for $t \to \infty$,*

$$\frac{\sum_{i\in \mathrm{SWG}_t^{(1)}(x)} \sum_{j\in \mathrm{SWG}_t^{(2)}(y)} \chi_{ij}(t)}{\sum_{i\in \mathrm{SWG}_t^{(1)}} \sum_{j\in \mathrm{SWG}_t^{(2)}} \chi_{ij}(t)} \xrightarrow{\mathbb{P}} \Phi(x)\Phi(y). \tag{2.39}$$

Theorem 2.7 is proved in Section 3 and completes the proof subject to Theorem 2.7. In fact, together with Theorem 2.5, (2.38) proves the joint convergence of the length of the optimal path and the hopcount as remarked upon below Theorem 1.2, where the limits are *independent*. ■

3 Continuous-time branching process theory

In Sections 2.2–2.3, we have reduced the proof of our main results to the proof of Theorems 2.5 and 2.7. This section deals with a proof of Theorems 2.5 and 2.7. This section is organized as follows. In Section 3.1, we investigate properties of our CTBP. In Section 3.2, we investigate one-vertex characteristics. In Section 3.4, we analyze the two-vertex characteristic and prove Theorem 2.5. In Section 3.5, we derive a CLT for the two-vertex characteristic and prove Theorem 2.7.

3.1 Intensities and limiting parameters single CTBP

We shall first state and prove various results that we shall require regarding a *single* branching process. Let BP be a continuous-time branching process driven by the offspring point process $\mathcal{P}$ (i.e., the points given by $(L_1, L_2, \ldots)$ as in (1.5)) and let μ denote the mean intensity measure of this point process, i.e.,

$$\mu[0,t] = \mathbb{E}(\#\{i : L_i \leq t\}). \tag{3.1}$$

Now,

$$\mu[0,t] = \sum_{i=1}^{\infty} \mathbb{P}(L_i < t) = \sum_{i=1}^{\infty} \int_0^{t^{1/s}} e^{-u} \frac{u^{i-1}}{(i-1)!} dt = \int_0^{t^{1/s}} 1 du = t^{1/s}. \tag{3.2}$$

Define the Malthusian rate of growth $\lambda = \lambda(s)$ as the unique positive constant such that the measure

$$\nu(dt) = \mathrm{e}^{-\lambda t} \mu(dt). \tag{3.3}$$

is a probability measure. A simple computation shows that this is equivalent to (1.8). The following lemma collects some properties of this probability measure and the constant λ:

Lemma 3.1 (Identification of limiting parameters CTBP)
(a) The constant $\lambda = \lambda(s)$ is given by (1.9).
(b) The probability measure $\nu(dt)$ is a Gamma distribution with density

$$f(t) = \frac{\lambda^{1/s}}{\Gamma(1/s)} \mathrm{e}^{-\lambda t} t^{1/s-1}. \tag{3.4}$$

(c) Let β_1 and β_2 denote the mean and the standard deviation of ν. Then

$$\beta_1 = (s\lambda)^{-1}, \qquad \beta_2 = (\lambda\sqrt{s})^{-1}. \tag{3.5}$$

(d) Let μ^{*j} denote the j-fold convolution of the measure μ. Then

$$\mu^{*j}(du) = \frac{u^{j/s-1} \lambda^{j/s} du}{\Gamma(j/s)}. \tag{3.6}$$

Proof: To prove part (a), note that since the sum of k independent exponential random variables follows the gamma distribution, a simple computation gives that

$$\begin{aligned}1 = \sum_{i=1}^{\infty} \mathbb{E}\left(\mathrm{e}^{-\lambda L_i}\right) &= \sum_{i=1}^{\infty} \int_0^{\infty} \mathrm{e}^{-\lambda t^s} \mathrm{e}^{-t} \frac{t^{i-1}}{(i-1)!} dt = \int_0^{\infty} \mathrm{e}^{-\lambda t^s} \mathrm{e}^{-t} \sum_{i=1}^{\infty} \frac{t^{i-1}}{(i-1)!} dt \\ &= \int_0^{\infty} \mathrm{e}^{-\lambda t^s} dt = \lambda^{-1/s} \int_0^{\infty} \mathrm{e}^{-t^s} dt = \lambda^{-1/s} s^{-1} \int_0^{\infty} \mathrm{e}^{-v} v^{1/s-1} dv \\ &= \lambda^{-1/s}\Gamma(1/s)/s = \lambda^{-1/s}\Gamma(1+1/s),\end{aligned}$$

as required. Parts (b) and (c) are trivial. To prove part (d) note that, by (3.2) and [12, 4.634],

$$\begin{aligned}\mu^{*j}(du) &= du s^{-j} \int_{u_1+\cdots+u_j=u} u_1^{1/s-1} \cdots u_j^{1/s-1} du_1 \cdots du_j \\ &= \frac{u^{j/s-1} s^{-j} \Gamma(1/s)^j du}{\Gamma(j/s)} = \frac{u^{j/s-1}\Gamma(1+1/s)^j du}{\Gamma(j/s)} = \frac{u^{j/s-1}\lambda^{j/s} du}{\Gamma(j/s)}.\end{aligned} \tag{3.7}$$

■

3.2 Analysis of single-vertex characteristic

We first state a general theorem for *single* vertex characteristics of the CTBP. Consider a function $\chi\colon \mathbb{R}^+ \to \mathbb{R}^+$ which is continuous almost everywhere which (a) increases at most polynomially quickly to ∞; and (b) is integrable with respect to the Lebesgue measure near zero. Let us call such functions *regular single-vertex characteristics*. For the branching process BP_t, call

$$z_t^{\chi} = \sum_{j\in \mathrm{BP}_t} \chi(t - T_j) \tag{3.8}$$

the *branching process counted according to characteristic* χ. Branching processes counted by characteristics are some of the fundamental objects studied by Jagers and Nerman, see e.g. [18]. For example, taking $\chi(x) = 1$, we obtain $z_t^{\chi} = |\mathrm{BP}_t|$, the size of the branching process at time t. In order to investigate the hopcount, we will need to analyze not just branching processes counted according to characteristics as above but also *generation weighted* characteristics. Given a regular single vertex characteristic χ and any fixed $a \in \mathbb{R}$, define

$$z_t^{\chi}(a) = \sum_{j\in \mathrm{BP}_t} a^{G(j)} \chi(t - T_j), \tag{3.9}$$

where, as before, T_j denotes the time of birth of vertex j, while $G(j)$ denotes the *height* or *generation* of vertex j. Given any characteristic χ, when we write z_t^{χ} without the argument a, we imply the branching process counted in the usual way as in (3.8), while when we have an argument a, namely $z_t^{\chi}(a)$, we refer to the branching process counted by a generation weighted characteristic as in (3.9).

The following proposition is adapted from the general theory of CTBPs, see e.g. [17, Theorem 5.2.2] for part (a) (or see the nice treatment in [18, Theorem 3.4]). We shall give a complete proof since branching processes counted by generation weighted characteristics have not been previously analyzed. These constructs shall be crucial for us in order to prove the CLT for the hopcount.

Proposition 3.2 (Mean and co-variances of one-vertex characteristics) *For regular deterministic single-vertex characteristics* χ, *(a) the expectation $m_t^\chi(a) = \mathbb{E}[z_t^\chi(a)]$ satisfies*

$$m_t^\chi(a) = \mathbb{E}[z_t^\chi(a)] = \int_0^t \chi(t-u) \sum_{j=0}^\infty a^j \mu^{*j}(du). \tag{3.10}$$

(b) the covariances between $z_t^{\chi_1}(a_1)$ and $z_t^{\chi_2}(a_2)$ satisfy

$$\mathrm{Cov}(z_t^{\chi_1}(a_1), z_t^{\chi_2}(a_2)) = \int_0^t h_{a_1,a_2}(t-u)\tilde{\mu}_{a_1 a_2}(du), \tag{3.11}$$

where $v \mapsto h_{a_1,a_2}(v)$ is the function

$$h_{a_1,a_2}(v) = \frac{a_1 a_2}{s} \int_0^v u^{1/s-1} m_{v-u}^{\chi_1}(a_1) m_{v-u}^{\chi_2}(a_2) du, \tag{3.12}$$

and we define the generation-weighted intensity measure $\tilde{\mu}_a$ by

$$\tilde{\mu}_a(du) = \sum_{j=0}^\infty a^j \mu^{*j}(du). \tag{3.13}$$

Proof. The proof of part (a) follows the same strategy as in [18, p. 228], where the case $a = 1$ was proved. Indeed, there it is shown that the intensity measure for individuals in the k^{th} generation equals μ^{*k}. Thus,

$$m_t^\chi(a) = \mathbb{E}[z_t^\chi(a)] = \sum_{k=0}^\infty a^k \mathbb{E}\Big[\sum_{i \in N^k} \chi(t - T_i)\Big] = \int_0^t \chi(t-u) \sum_{k=0}^\infty a^k \mu^{*k}(du). \tag{3.14}$$

For part (b), we follow the identification of $\mathrm{Var}(z_t^\chi)$ in [18, Theorem 3.2 and Corollary 3.3]. We use the covariance partition

$$\mathrm{Cov}(z_t^{\chi_1}(a_1), z_t^{\chi_2}(a_2)) = \mathrm{Cov}\big(\mathbb{E}[z_t^{\chi_1}(a_1) \mid \mathcal{A}_0], \mathbb{E}[z_t^{\chi_2}(a_2) \mid \mathcal{A}_0]\big) + \mathbb{E}\Big[\mathrm{Cov}(z_t^{\chi_1}(a_1), z_t^{\chi_2}(a_2)) \mid \mathcal{A}_0\Big], \tag{3.15}$$

where $\mathcal{A}_0$ is the σ-algebra generated by the lives of the individuals in the first generation (the root is considered to be in generation zero). Then, the branching property of a CTBP gives that

$$z_t^\chi(a) = \chi(t) + a \sum_{j\colon G(j)=1} z_{t-T_j}^\chi(j; a), \tag{3.16}$$

where $\big((z_t^{\chi_1}(j; a_1), z_t^{\chi_2}(j; a_2))\big)_{j,t\geq 0}$ is, conditionally on $\mathcal{A}_0$, a sequence of i.i.d. random processes with law $((z_t^{\chi_1}(a_1), z_t^{\chi_2}(a_2)))_{t\geq 0}$. Therefore,

$$\mathrm{Cov}\big(z_t^{\chi_1}(a_1), z_t^{\chi_2}(a_2) \mid \mathcal{A}_0\big) = a_1 a_2 \sum_{j\colon G(j)=1} \mathrm{C}_{t-T_j}^{\chi_1,\chi_2}(a_1, a_2), \tag{3.17}$$

where we abbreviate

$$\mathrm{C}_t^{\chi_1,\chi_2}(a_1, a_2) = \mathrm{Cov}(z_t^{\chi_1}(a_1), z_t^{\chi_2}(a_2)). \tag{3.18}$$

Thus,

$$\mathbb{E}\Big[\mathrm{Cov}\big(z_t^{\chi_1}(a_1), z_t^{\chi_2}(a_2) \mid \mathcal{A}_0\big)\Big] = a_1 a_2 \int_0^t \mathrm{C}_{t-v}^{\chi_1,\chi_2}(a_1, a_2)\mu(du). \tag{3.19}$$

Further,

$$\mathbb{E}[z_t^\chi(a) \mid \mathcal{A}_0] = \chi(t) + a \int_0^t m_{t-u}^\chi \mathcal{P}(du), \tag{3.20}$$

where $(\mathcal{P}(t))_{t\geq 0}$ is the intensity process of the first individual. Therefore, we arrive at

$$\mathrm{C}_t^{\chi_1,\chi_2}(a_1,a_2) = h_{a_1,a_2}(t) + (a_1a_2)\int_0^t \mathrm{C}_{t-u}^{\chi_1,\chi_2}(a_1,a_2)\mu(du), \tag{3.21}$$

where

$$h_{a_1,a_2}(t) = a_1a_2\mathrm{Cov}\left(\int_0^t m_{t-u}^{\chi_1}(a_1)\mathcal{P}(du), \int_0^t m_{t-u}^{\chi_2}(a_2)\mathcal{P}(du)\right). \tag{3.22}$$

Iterating this equation yields (3.11).

As before, for $\mathcal{P}$ denoting the offspring distribution point process (given by (1.6)) and for every function $F\colon \mathbb{R}\to\mathbb{R}$, note that

$$\int_0^\infty F(x)\mathcal{P}(dx) \overset{d}{=} f(\Pi), \tag{3.23}$$

where Π is a rate 1 Poisson point process, $f(x)\equiv F(x^s)$ and where the function f applied to a point process Π is defined as $f(\Pi)\equiv\sum_{X\in\Pi} f(X)$. By properties of functionals of the Poisson point process, we have that

$$\mathrm{Cov}\left(\int_0^t F_1(u)\mathcal{P}(du), \int_0^t F_2(u)\mathcal{P}(du)\right) = \int_0^{t^{1/s}} F_1(u^s)F_2(u^s)du = s^{-1}\int_0^t u^{1/s-1}F_1(u)F_2(u)du. \tag{3.24}$$

Therefore, we obtain

$$h_{a_1,a_2}(t) = \frac{a_1a_2}{s}\int_0^t u^{1/s-1}m_{t-u}^{\chi_1}(a_1)m_{t-u}^{\chi_2}(a_2)du. \tag{3.25}$$

■

The following proposition (adapted mainly from [18, Theorem 3.5]) captures all we require to know about the asymptotics of the mean and variance of a single-vertex characteristic χ.

Proposition 3.3 (Asymptotics of mean and variance for one-vertex characteristics) *For regular single vertex characteristics χ, and all $a\geq 0$,*
(a) As $t\to\infty$,

$$\mathrm{e}^{-\lambda a^s t}\mathbb{E}(z_t^\chi(a)) \to a^s s\lambda\int_0^\infty \mathrm{e}^{-\lambda a^s y}\chi(y)dy. \tag{3.26}$$

When $a=a_t\to 1$, then the convergence holds where in the right-hand sides the value $a=1$ is substituted.
(b) As $t\to\infty$, when $a_1,a_2\geq 0$ with $a_1^s+a_2^s-a_1^sa_2^s>0$,

$$\begin{aligned}&\mathrm{e}^{-\lambda(a_1^s+a_2^s)t}\mathrm{Cov}(z_t^{\chi_1}(a_1), z_t^{\chi_2}(a_2)) \\ &\quad\to \frac{(a_1a_2)^{s+1}\lambda^2s^3}{(a_1^s+a_2^s)^{1/s}(a_1^s+a_2^s-a_1^sa_2^s)}\int_0^\infty \chi_1(x)\mathrm{e}^{-\lambda a_1^s x}dx\int_0^\infty \chi_2(x)\mathrm{e}^{-\lambda a_2^s x}dx.\end{aligned} \tag{3.27}$$

When $\vec{a}=\vec{a}_t\to(1,1)$, then the convergence holds where in the right-hand sides the value $\vec{a}=(1,1)$ is substituted.
(c) With $z_t=|\mathrm{BP}_t|$, there exists a random variable W with $W>0$ a.s. such that $\mathrm{e}^{-\lambda t}z_t$ converges almost surely and in L^2 to W and further for any single-vertex regular characteristic

$$\mathrm{e}^{-\lambda t}z_t^\chi \xrightarrow{a.s.} W\lambda\int_0^\infty \chi(x)\mathrm{e}^{-\lambda x}dx, \tag{3.28}$$

and the convergence also holds in L^2.

Proof. Part (a) for $a = 1$ is [17, Theorem 5.2.8]. For $a \neq 1$, we start from

$$m_t^\chi(a) = \int_0^t \chi(t-u) \sum_{k=0}^\infty a^k \mu^{*k}(du) = \int_0^t \chi(t-u)\tilde{\mu}_a(du). \tag{3.29}$$

Define the measure with density $p_a(u)du$ via the equation $\mathrm{e}^{-\lambda a^s u}\tilde{\mu}_a(du) = p_a(u)du + \mathrm{e}^{-\lambda a^s u}\delta_0(du)$, then we obtain

$$\mathrm{e}^{-\lambda a^s t} m_t^\chi(a) = \mathrm{e}^{-\lambda a^s t}\chi(t) + \int_0^t \chi(t-u)\mathrm{e}^{-\lambda a^s (t-u)} p_a(u)du = \mathrm{e}^{-\lambda a^s t}\chi(t) + \int_0^t \chi(v)\mathrm{e}^{-\lambda a^s v} p_a(t-v)dv. \tag{3.30}$$

By Lemma A.1(a-b), we have that $p_a(u)$ is uniformly bounded on $[1,\infty)$ and bounded by $cu^{1/s-1}$ on $[0,1]$, while and $p_a(u) \to a^s \lambda s$ when $u \to \infty$. Thus, by dominated convergence,

$$\mathrm{e}^{-\lambda a^s t} m_t^\chi(a) \to a^s \lambda s \int_0^\infty \chi(v)\mathrm{e}^{-\lambda a^s v} dv. \tag{3.31}$$

The proof when $a_t \to 1$ is identical.

See [18, Theorem 3.5] for parts (b) for $a_1 = a_2 = 1$ and for part (c). For the proof of part (b) for $(a_1, a_2) \neq (1,1)$, we start with (3.11) and (3.12). By part (a), we have that

$$\begin{aligned} \mathrm{e}^{-\lambda(a_1^s+a_2^s)t} h_{a_1,a_2}(t) &= \frac{a_1 a_2}{s}\int_0^t u^{1/s-1}\mathrm{e}^{-\lambda(a_1^s+a_2^s)u}\Big(\mathrm{e}^{-\lambda a_1^s(t-u)} m_{t-u}^{\chi_1}(a_1)\Big)\Big(\mathrm{e}^{-\lambda a_2^s(t-u)} m_{t-u}^{\chi_2}(a_2)\Big)du \\ &\sim \frac{a_1 a_2}{s} M^{\chi_1}(a_1) M^{\chi_2}(a_2)\int_0^\infty u^{1/s-1}\mathrm{e}^{-\lambda(a_1^s+a_2^s)u}du, \end{aligned} \tag{3.32}$$

where we define

$$M^\chi(a) = a^s s\lambda \int_0^\infty \mathrm{e}^{-\lambda y a^s}\chi(y)dy. \tag{3.33}$$

Further, note that, by (1.9),

$$\int_0^\infty u^{1/s-1}\mathrm{e}^{-\lambda(a_1^s+a_2^s)u}du = (a_1^s+a_2^s)^{-1/s}\lambda^{-1/s}\Gamma(1/s) = s(a_1^s+a_2^s)^{-1/s}. \tag{3.34}$$

Then we rewrite

$$\mathrm{e}^{-\lambda(a_1^s+a_2^s)t}\mathrm{Cov}(z_t^{\chi_1}(a_1), z_t^{\chi_2}(a_2)) = \int_0^t \mathrm{e}^{-\lambda(a_1^s+a_2^s)(t-u)} h_{a_1,a_2}(t-u)\mathrm{e}^{-\lambda(a_1^s+a_2^s)u}\tilde{\mu}_{a_1 a_2}(du). \tag{3.35}$$

Now, by (A.2), for u large,

$$\mathrm{e}^{-\lambda(a_1^s+a_2^s)u}\tilde{\mu}_{a_1 a_2}(du) \sim (a_1 a_2)^s \lambda s \mathrm{e}^{-\lambda(a_1^s+a_2^s-a_1^s a_2^s)u}du, \tag{3.36}$$

which is integrable, so that substitution of this asymptotics in (3.11) and using dominated convergence, proves that

$$\begin{aligned} \mathrm{e}^{-\lambda(a_1^s+a_2^s)t}\mathrm{Cov}(z_t^{\chi_1}(a_1), z_t^{\chi_2}(a_2)) &\sim s(a_1 a_2)^{1+s} M^{\chi_1}(a_1)M^{\chi_2}(a_2)(a_1^s+a_2^s)^{-1/s}\int_0^\infty \lambda \mathrm{e}^{-\lambda(a_1^s+a_2^s-a_1^s a_2^s)u}du \\ &= \frac{(a_1 a_2)^{1+s} s}{(a_1^s+a_2^s)^{1/s}(a_1^s+a_2^s-a_1^s a_2^s)} M^{\chi_1}(a_1)M^{\chi_2}(a_2). \end{aligned} \tag{3.37}$$

This proves the claim when $a_1^s + a_2^s - a_1^s a_2^s > 0$. When $\vec{a} = \vec{a}_t \to (1,1)$, then the above asymptotics holds with $\vec{a} = (1,1)$ substituted on the r.h.s. since (3.32) holds with $\vec{a} = (1,1)$ substituted on its r.h.s. ∎

3.3 Almost sure convergence of two-vertex factor: proof of Theorem 2.5

In this section, we prove Theorem 2.5. Throughout the proof, we shall abbreviate $p = 1/s - 1$. Note that, for any fixed $0 < \varepsilon < B < \infty$, we can write $z_t^{(1,2)}$ as

$$z_t^{(1,2)} = I_t^{(1)}(\varepsilon, B) + I_t^{(2)}(B) + I_t^{(3)}(\varepsilon), \tag{3.38}$$

where

$$\begin{aligned} I_t^{(1)}(\varepsilon, B) &= \sum_{j \in \mathrm{BP}_t^{(2)}: \varepsilon < t - T_j < B} \sum_{i \in \mathrm{BP}_t^{(1)}} ([t - T_i] + [t - T_j])^p, \\ I_t^{(2)}(B) &= \sum_{j \in \mathrm{BP}_t^{(2)}: t - T_j > B} \sum_{i \in \mathrm{BP}_t^{(1)}} ([t - T_i] + [t - T_j])^p, \\ I_t^{(3)}(\varepsilon) &= \sum_{j \in \mathrm{BP}_t^{(2)}: t - T_j < \varepsilon} \sum_{i \in \mathrm{BP}_t^{(1)}} ([t - T_i] + [t - T_j])^p. \end{aligned}$$

Thus to prove the result it is enough to show that for each fixed ε, B we have

$$\mathrm{e}^{-2\lambda t} I_t^{(1)}(\varepsilon, B) \xrightarrow{a.s.} W^{(1)} W^{(2)} \lambda^2 \int_\varepsilon^B \int_0^\infty (x_1 + x_2)^{1/s - 1} \mathrm{e}^{-\lambda x_1} \mathrm{e}^{-\lambda x_2} dx_1 dx_2, \tag{3.39}$$

$$\limsup_{B \to \infty} \limsup_{t \to \infty} \mathrm{e}^{-2\lambda t} I_t^{(2)}(B) = 0 \qquad \text{and} \qquad \limsup_{\varepsilon \to 0} \limsup_{t \to \infty} \mathrm{e}^{-2\lambda t} I_t^{(3)}(\varepsilon) = 0. \tag{3.40}$$

We shall start by proving (3.39). The following lemma shall be crucial in our proof:

Lemma 3.4 (Sup convergence of characteristics) *As $t \to \infty$,*

$$\sup_{x \in [\varepsilon, B]} \left| \mathrm{e}^{-\lambda t} \sum_{i \in \mathrm{BP}_t^{(1)}} (x + [t - T_i])^p - W^{(1)} \lambda \int_0^\infty (x + y)^p \mathrm{e}^{-\lambda y} dy \right| \xrightarrow{a.s.} 0, \tag{3.41}$$

where $W^{(1)}$ is the almost sure limit of $\mathrm{e}^{-\lambda t} z_t^{(1)}$.

Proof. Consider the (random) functions

$$f_t(x) = \mathrm{e}^{-\lambda t} \sum_{i \in \mathrm{BP}_t^{(1)}} (x + [t - T_i])^p \qquad x \in [\varepsilon, B]. \tag{3.42}$$

Note that for $p < 0$ these functions are monotonically decreasing, while for $p > 0$ they are increasing functions and they are all continuous when defined on the compact interval $[\varepsilon, B]$. Further, for each fixed $x \in [\varepsilon, B]$, by Proposition 3.3(c), pointwise we have on a set of measure one,

$$f_t(x) \xrightarrow{a.s.} W^{(1)} \lambda \int_0^\infty (x + y)^p \mathrm{e}^{-\lambda y} dy. \tag{3.43}$$

Thus to show the a.s. sup convergence, by the Arzela-Ascoli Theorem (see e.g. [21]), it is enough to show that the above family of functions are a.s. equicontinuous, i.e., for any $x \in [\varepsilon, B]$ and any given $\delta > 0$ there exists $\eta(x) > 0$ independent of t such that for all t:

$$|f_t(x) - f_t(y)| < \delta \qquad \text{for all } y \in [x - \eta(x), x + \eta(x)] \cap [\varepsilon, B]. \tag{3.44}$$

We separate between the cases $p < 1$ and $p \geq 1$.

Case 1: $p < 1$. In this case note that for any $l_1, l_2 > 0$ and $a > 0$, we have

$$\begin{aligned} |(l_1+a)^p - (l_2+a)^p| &= |p-1| \int_{l_1}^{l_2} (x+a)^{p-1} dx \\ &\le |p-1| \int_{l_1}^{l_2} x^{p-1} dx \qquad \text{since } p-1<0. \end{aligned} \tag{3.45}$$

By the continuity of the function $g(x) = x^p$, for any $x \in [\varepsilon, B]$ and $\delta' > 0$, we can choose $\eta'(x)$ small such that for $y \in [\varepsilon, B], |y - x| < \eta'(x)$ we have

$$|x^p - y^p| < \delta'. \tag{3.46}$$

This implies from (3.45) applied individually to the functions $g_i(x) = (x + [t - T_i])^{1/s-1}$ for $i \in \mathrm{BP}_t^{(1)}$ that

$$|f_t(x) - f_t(y)| < \delta' \mathrm{e}^{-\lambda t} z_t^{(1)},$$

where we recall that $z_t^{(1)} = |\mathrm{BP}_t^{(1)}|$. Since $\mathrm{e}^{-\lambda t} z_t^{(1)}$ converges a.s. and $\{\mathrm{e}^{-\lambda t} z_t^{(1)}\}_{t\ge 0}$ is bounded a.s., we obtain that, on a set A of measure one, for each $\omega \in A$, we can find a $\kappa(\omega)$ depending on the sample point ω but independent of t, such that

$$\sup_t \mathrm{e}^{-\lambda t} z_t^{(1)}(\omega) < \kappa(\omega).$$

Now choosing $\delta' = \delta/\kappa(\omega)$ gives us a $\eta(x) = \eta(x, \omega)$ such that (3.44) is satisfied. This proves the result for $p < 1$.

Case 2: $p \ge 1$. Here note that for any $a > 0$ and $x, y \in [\varepsilon, B]$, we have, by the mean value theorem

$$\begin{aligned} |(x+a)^p - (y+a)^p| &= p(z+a)^{p-1}|y-x| \qquad a \in [x,y] \\ &\le p(B+a)^{p-1}|y-x| \qquad \text{since } p-1 \ge 0. \end{aligned}$$

This implies that, for $x, y \in [\varepsilon, B]$,

$$|f_t(x) - f_t(y)| < H_t|x - y|, \tag{3.47}$$

where, by Proposition 3.3(c),

$$H_t = \mathrm{e}^{-\lambda t} p \sum_{i \in \mathrm{BP}_t^{(1)}} (B + [t - T_i])^{p-1} \xrightarrow{a.s.} pW^{(1)} \int_0^\infty (B+y)^{p-1} \lambda \mathrm{e}^{-\lambda y} dy. \tag{3.48}$$

This proves that (3.44) holds also when $p \ge 1$, and completes the proof of Lemma 3.4. ∎

Completion of the proof of (3.39). Write

$$\mathrm{BP}_t^{(2)}(\varepsilon, B) = \{v \in \mathrm{BP}_t^{(2)} : \varepsilon < t - T_j < B\}. \tag{3.49}$$

Then we have

$$\left| I_t^{(1)}(\varepsilon, B) - \mathrm{e}^{-\lambda t} \sum_{j \in \mathrm{BP}_t^{(2)}(\varepsilon, B)} W^{(1)} \int_0^\infty ([t - T_j] + y)^{1/s-1} \lambda \mathrm{e}^{-\lambda y} dy \right| \le Q_1(t) \mathrm{e}^{-\lambda t} z_t^{(2)}, \tag{3.50}$$

where

$$Q_1(t) = \sup_{x\in[\varepsilon,B]} \Big| \mathrm{e}^{-\lambda t} \sum_{i\in \mathrm{BP}_t^{(1)}} (x+[t-T_i])^{1/s-1} - W^{(1)} \int_0^\infty (x+y)^{1/s-1} \lambda \mathrm{e}^{-\lambda y} dy \Big|. \tag{3.51}$$

Lemma 3.4 now implies that the term on the r.h.s. of (3.50) converges to 0 a.s. Thus, to complete the proof, it is enough to show that

$$\mathrm{e}^{-\lambda t} \sum_{j\in \mathrm{BP}_t^{(2)}(\varepsilon,B)} W^{(1)} \int_0^\infty ([t-T_j]+y)^p \lambda \mathrm{e}^{-\lambda y} dy \xrightarrow{a.s.} W^{(1)}W^{(2)}\lambda^2 \int_\varepsilon^B \int_0^\infty (x_1+x_2)^p \mathrm{e}^{-\lambda(x_1+x_2)} dx_1 dx_2.$$

This follows by taking the characteristic

$$\chi_2(a) = \begin{cases} \int_0^\infty (a+x_2)^p \mathrm{e}^{-\lambda x_2} dx_2 & \text{if } \varepsilon \le a \le B, \\ 0 & \text{if } a \notin [\varepsilon, B]. \end{cases} \tag{3.52}$$

and using Proposition 3.3(c) for the branching process $\mathrm{BP}_t^{(2)}$. ■

Completion of the proof of (3.40). First consider the term $I_t^{(3)}(\varepsilon)$. Note that

$$I_t^{(3)}(\varepsilon) \le [z_t^{(2)} - z_{t-\varepsilon}^{(2)}] \sum_{j\in \mathrm{BP}_t^{(1)}} (\varepsilon + t - T_j)^p. \tag{3.53}$$

Now note that by Proposition 3.3(c)

$$\mathrm{e}^{-2\lambda t}[z_t^{(2)} - z_{t-\varepsilon}^{(2)}] \sum_{j\in \mathrm{BP}_t^{(1)}} (\varepsilon + t - T_j)^p \xrightarrow{a.s.} \Big[W^{(2)}[1-\mathrm{e}^{-\lambda\varepsilon}]\Big] \cdot \Big[W^{(1)} \int_0^\infty (\varepsilon+y)^p \mathrm{e}^{-\lambda y} dy\Big] \xrightarrow{a.s.} 0, \tag{3.54}$$

when $\varepsilon \downarrow 0$. This proves the last convergence result in (3.40).

To prove the first convergence result in (3.40), note that arguing as in the proof of (3.39), we have for all p and $x_1, x_2 > 0$,

$$(x_1 + x_2)^p \le (2^p \vee 1)(x_1^p + x_2^p), \tag{3.55}$$

where $a \vee b = \max\{a, b\}$. Thus,

$$I_t^{(2)}(B) \le (2^p \vee 1)[z_t^{(1)} z_{t-B}^{\chi_B,(2)} + z_t^{\chi,(1)} z_{t-B}^{(2)}], \tag{3.56}$$

where $\chi_B(x) = (B+x)^p, \chi(x) = x^p$. Now again, by Proposition 3.3(c),

$$\mathrm{e}^{-\lambda t}[z_t^{(1)} z_{t-B}^{\chi_B,(2)} + z_t^{\chi,(1)} z_{t-B}^{(2)}] \xrightarrow{a.s.} W^{(1)}W^{(2)}\mathrm{e}^{-\lambda B}\Big[\int_0^\infty \mathrm{e}^{-\lambda x}(B+x)^p \mathrm{e}^{-\lambda x} dx + \int_0^\infty \mathrm{e}^{-\lambda x} x^p \mathrm{e}^{-\lambda x} dx\Big], \tag{3.57}$$

which converges a.s. to 0 when $B \to \infty$. This completes the proof of (3.40). ■

3.4 Mean and variance of two-vertex characteristic

In this section, we shall analyze two-vertex characteristics. This sets the stage for the proof of the the asymptotics for the hopcount in Theorem 2.7. Define, for $\vec{a} = (a_1, a_2)$,

$$z_t^{(1,2)}(\vec{a}) = \sum_{i\in \mathrm{BP}_t^{(1)}} \sum_{j\in \mathrm{BP}_t^{(2)}} a_1^{G^{(1)}(i)} a_2^{G^{(2)}(j)} ([t-T_i] + [t-T_j])^p, \tag{3.58}$$

where we recall that $G^{(i)}(v)$ is the *generation* of vertex $v \in \mathrm{BP}_t^{(i)}$.

Lemma 3.5 (Expectation and variance of two-vertex characteristics) *Consider two independent CTBPs* $\mathrm{BP}_t^{(1)}$ *and* $\mathrm{BP}_t^{(2)}$ *driven by the offspring distribution* $\mathcal{P}$*. Then,*
(a)

$$\mathbb{E}[z_t^{(1,2)}(\vec{a})] = \int_0^t \int_0^t ([t-v]+[t-u])^p \tilde{\mu}_{a_1}(dv)\tilde{\mu}_{a_2}(du). \tag{3.59}$$

(b)

$$\mathrm{Cov}(z_t^{(1,2)}(\vec{a}), z_t^{(1,2)}(\vec{b})) = \int_0^t h^{(1)}_{\vec{a},\vec{b}}(t-u,t)\tilde{\mu}_{a_2 b_2}(du) + \int_0^t h^{(2)}_{\vec{a},\vec{b}}(t-u,t)\tilde{\mu}_{a_1 b_1}(du), \tag{3.60}$$

where

$$h^{(1)}_{\vec{a},\vec{b}}(v,t) = \frac{a_2 b_2}{s}\int_0^v \int_0^{v-u}\int_0^{v-u}\int_0^t\int_0^t u^p([t-u_2]+[v-u-u_1])^p([t-v_2]+[v-u-v_1])^p \times \tilde{\mu}_{a_1}(du_1)\tilde{\mu}_{b_1}(dv_1)\tilde{\mu}_{a_2}(du_2)\tilde{\mu}_{b_2}(dv_2)du, \tag{3.61}$$

$$h^{(2)}_{\vec{a},\vec{b}}(v,t) = \frac{a_1 b_1}{s}\int_0^v u^p \mathbb{E}[z_{v-u}^{\tilde{\chi}^{(2)}_{t,v-u,a_1}}(a_2) z_{v-u}^{\tilde{\chi}^{(2)}_{t,v-u,b_1}}(b_2)]du, \tag{3.62}$$

with $p = 1/s - 1$ *and*

$$\tilde{\chi}^{(2)}_{t,r,a_2}(x) = \int_0^r (t-u_2+x)^p \tilde{\mu}_{a_2}(du_2). \tag{3.63}$$

Proof. We shall prove part (a) by conditioning on $\mathrm{BP}_t^{(1)}$. Note we can write $z_t^{(1,2)}(\vec{a}) = z_t^{\chi^{(1)}_{t,a_1},(2)}(a_2)$, where

$$\chi^{(1)}_{t,a_1}(x) = \sum_{j\in \mathrm{BP}_t^{(1)}} a_1^{G^{(1)}(j)}(x+[t-T_j])^p. \tag{3.64}$$

Conditionally on $\mathrm{BP}_t^{(1)}$, the characteristic $\chi^{(1)}_{t,a_1}$ is deterministic. Therefore, Proposition 3.2(a) implies that

$$\mathbb{E}(z_t^{(1,2)}(\vec{a})|\mathrm{BP}_t^{(1)}) = \sum_{j\in \mathrm{BP}_t^{(1)}} a_1^{G^{(1)}(j)}\int_0^t ([t-u]+[t-T_j])^p\tilde{\mu}_{a_2}(du) = z_t^{\chi^{(2)}_{t,a_2},(1)}(a_1), \tag{3.65}$$

where $\chi^{(2)}_{t,a_2}$ is the characteristic

$$\chi^{(2)}_{t,a_2}(v) = \int_0^t ([t-u]+v)^p \tilde{\mu}_{a_2}(du). \tag{3.66}$$

We complete the proof by noting that, for all r,

$$m_r^{\chi^{(2)}_{t,a_2}}(a_1) = \mathbb{E}(z_r^{\chi^{(2)}_{t,a_2},(2)}(a_1)) = \int_0^r\int_0^t ([r-v]+[t-u])^p\tilde{\mu}_{a_1}(dv)\tilde{\mu}_{a_2}(du). \tag{3.67}$$

Taking $r=t$ proves the claim in part (a).

For part (b), we again condition on $\mathrm{BP}_t^{(1)}$, for which we use the covariance partition

$$\begin{aligned}&\mathrm{Cov}(z_t^{(1,2)}(\vec{a}), z_t^{(1,2)}(\vec{b})) \\ &= \mathrm{Cov}\big(\mathbb{E}(z_t^{(1,2)}(\vec{a})|\mathrm{BP}_t^{(1)}), z_t^{(1,2)}(\vec{b})|\mathrm{BP}_t^{(1)})\big) + \mathbb{E}\big(\mathrm{Cov}(z_t^{(1,2)}(\vec{a}), z_t^{(1,2)}(\vec{b})|\mathrm{BP}_t^{(1)})\big) = (I)_t + (II)_t.\end{aligned} \tag{3.68}$$

Let us now tackle each of these two terms separately.

Term $(I)_t$**:** For $(I)_t$, we use the explicit formula for $\mathbb{E}(z_t^{(1,2)}(\vec{a})|\mathrm{BP}_t^{(1)})$ in (3.65) and $\chi_{t,a_2}^{(2)}$ in (3.66) to obtain that

$$(I)_t = \mathrm{Cov}\big(z_t^{\chi_{t,a_2}^{(2)},(1)}(a_1), z_t^{\chi_{t,b_2}^{(2)},(1)}(b_1)\big). \tag{3.69}$$

Now using Proposition 3.2(b), we get

$$(I)_t = \int_0^t \int_0^t h_{\vec{a},\vec{b}}^{(1)}(t-u,t)\tilde{\mu}_{a_2b_2}(du). \tag{3.70}$$

where $v \mapsto h_{\vec{a},\vec{b}}^{(1)}(v,t)$ is the function

$$\begin{aligned} h_{\vec{a},\vec{b}}^{(1)}(v,t) &= \frac{a_2b_2}{s}\int_0^v u^p m_{v-u}^{\chi_{t,a_2}^{(2)}}(a_1) m_{v-u}^{\chi_{t,b_2}^{(2)}}(b_1)du \\ &= \frac{a_2b_2}{s}\int_0^v u^p \int_0^{v-u}\int_0^{v-u} \chi_{t,a_2}^{(2)}(v-u-u_1)\chi_{t,b_2}^{(2)}(v-u-v_1)\tilde{\mu}_{a_1}(du_1)\tilde{\mu}_{b_1}(dv_1)du. \\ &= \frac{a_2b_2}{s}\int_0^v \int_0^{v-u}\int_0^{v-u}\int_0^t\int_0^t u^p([t-u_2]+[v-u-u_1])^p([t-v_2]+[v-u-v_1])^p \\ &\quad \times \tilde{\mu}_{a_1}(du_1)\tilde{\mu}_{b_1}(dv_1)\tilde{\mu}_{a_2}(du_2)\tilde{\mu}_{b_2}(dv_2)du. \end{aligned} \tag{3.71}$$

Term $(II)_t$**.** We again use that, conditionally on $\mathrm{BP}_t^{(1)}$, $z_t^{(1,2)}(\vec{a}) = z_t^{\chi_{t,a_1}^{(1)},(2)}(a_2)$, where $\chi_{t,a_1}^{(1)}$ was defined in (3.64). Therefore, we can again use Proposition 3.2(b) to write

$$\mathrm{Cov}\big(z_t^{(1,2)}(\vec{a}), z_t^{(1,2)}(\vec{b})|\mathrm{BP}_t^{(1)}\big) = \int_0^t g_{\vec{a},\vec{b}}(t-u,t)\tilde{\mu}_{a_2b_2}(du), \tag{3.72}$$

where

$$g_{\vec{a},\vec{b}}(v,t) = \frac{a_2b_2}{s}\int_0^v u^p m_{v-u}^{\chi_{t,a_1}^{(1)}}(a_2) m_{v-u}^{\chi_{t,a_1}^{(1)}}(b_1)(b_2)du. \tag{3.73}$$

Therefore,

$$(II)_t = \int_0^t h_{\vec{a},\vec{b}}^{(2)}(t-u,t)\tilde{\mu}_{a_2b_2}(du), \tag{3.74}$$

where

$$h_{\vec{a},\vec{b}}^{(2)}(v,t) = \frac{a_2b_2}{s}\int_0^v u^p \mathbb{E}[m_{v-u}^{\chi_{t,a_1}^{(1)}}(a_2) m_{v-u}^{\chi_{t,b_1}^{(1)}}(b_2)]du. \tag{3.75}$$

We can now rewrite

$$\begin{aligned} m_r^{\chi_{t,a_1}^{(1)}}(a_2) &= \int_0^r \sum_{j\in\mathrm{BP}_t^{(1)}} a_1^{G^{(1)}(j)}(t-u_2+t-T_j)^p\tilde{\mu}_{a_2}(du_2) \\ &= \sum_{j\in\mathrm{BP}_t^{(1)}} a_1^{G^{(1)}(j)}\int_0^r (t-u_2+t-T_j)^p\tilde{\mu}_{a_2}(du_2) = z_t^{\tilde{\chi}_{t,r,a_2}^{(2)},(1)}(a_1), \end{aligned} \tag{3.76}$$

where

$$\tilde{\chi}_{t,r,a_2}^{(2)}(x) = \int_0^r (t-u_2+x)^p\tilde{\mu}_{a_2}(du_2). \tag{3.77}$$

This completes the proof. ■

Lemma 3.6 (Asymptotics of mean and variance of two-vertex characteristics) *Consider two independent CTBPs* $\mathrm{BP}_t^{(1)}$ *and* BP_t^2 *driven by the offspring distribution* $\mathcal{P}$*. Then,*
(a)

$$\mathrm{e}^{-\lambda(a_1^s+a_2^s)t}\mathbb{E}(z_t^{(1,2)}(\vec{a})) \to (\lambda s)^2 \int_0^\infty \int_0^\infty (x_1+x_2)^p \mathrm{e}^{-\lambda(a_1^s x_1 + a_2^s x_2)} dx_1 dx_2. \tag{3.78}$$

When $\vec{a} = \vec{a}_t \to (1,1)$*, then the convergence holds where in the right-hand sides the value* $\vec{a} = (1,1)$ *is substituted.*
(b) When $a_1^s + a_2^s - a_1^s a_2^s > 0$ *and* $b_1^s + b_2^s - b_1^s b_2^s > 0$*, there exists a constant* $A_{\mathrm{Cov}}(\vec{a}, \vec{b})$ *such that*

$$\mathrm{e}^{-\lambda[(a_1^s+a_2^s)+(b_1^s+b_2^s)]t}\mathrm{Cov}(z_t^{(1,2)}(\vec{a}), z_t^{(1,2)}(\vec{b})) \to A_{\mathrm{Cov}}(\vec{a}, \vec{b}). \tag{3.79}$$

When $\vec{a} = \vec{a}_t \to (1,1), \vec{b} = \vec{b}_t \to (1,1)$*, then the convergence holds where the r.h.s. is replaced with* $A_{\mathrm{Cov}}(\vec{1}, \vec{1})$*.*

Proof. By Lemma 3.5(a),

$$\begin{aligned}
\mathrm{e}^{-\lambda(a_1^s+a_2^s)t}\mathbb{E}(z_t^{(1,2)}(\vec{a})) &= \mathrm{e}^{-\lambda(a_1^s+a_2^s)t} \int_0^t \int_0^t ([t-u]+[t-v])^{1/s-1} \tilde{\mu}_{a_1}(du)\tilde{\mu}_{a_2}(dv) \\
&= \int_0^t \int_0^t \mathrm{e}^{-\lambda[a_1^s(t-u)+a_2^s(t-v)]}([t-u]+[t-v])^{1/s-1} p_{a_1}(u)p_{a_2}(v)dudv + o(1) \\
&= \int_0^t \int_0^t \mathrm{e}^{-\lambda[a_1^s u + a_2^s v]}(u+v)^{1/s-1} p_{a_1}(t-u)p_{a_2}(t-v)dudv + o(1),
\end{aligned} \tag{3.80}$$

where the $o(1)$ originates from part $\delta_{u,0}(du)$ in the decomposition $\mathrm{e}^{-\lambda a^s u}\tilde{\mu}_a(du) = p_a(u)du + \mathrm{e}^{-\lambda a^s u}\delta_0(du)$, and $\delta_{u,0}$ is the Dirac measure at $u=0$. Now we again use Lemma A.1(a) and (b) together with dominated convergence to obtain the claim in part (a). The proof of part (b) is similar, and we omit the details here. ■

3.5 CLT for two-vertex characteristic: Proof of Theorem 2.7

In this section, we use and extend the theory developed in the previous section to prove Theorem 2.7. The plan is as follows: We shall start by proving the result for the CTBP in the summation instead of $\mathrm{SWG}_t^{(i)}$ and then argue that the difference is negligible. The result is formulated as follows:

Theorem 3.7 (CLT for two-vertex characteristic for CTBP) *The two-vertex characteristic satisfies that, as* $t \to \infty$*,*

$$\frac{\sum_{i \in \mathrm{BP}_t^{(1)}(x)} \sum_{j \in \mathrm{BP}_t^{(2)}(y)} \chi_{ij}(t)}{\sum_{i \in \mathrm{BP}_t^{(1)}} \sum_{j \in \mathrm{BP}_t^{(2)}} \chi_{ij}(t)} \xrightarrow{\mathbb{P}} \Phi(x)\Phi(y). \tag{3.81}$$

Proof. Theorem 3.7 follows when we show that, writing $\vec{a}_t = (\mathrm{e}^{\alpha_1/\sqrt{s^2\lambda t}}, \mathrm{e}^{\alpha_2/\sqrt{s^2\lambda t}})$, for some vector $\vec{\alpha} = (\alpha_1, \alpha_2)$,

$$\frac{z_t^{(1,2)}(\vec{a}_t)\mathrm{e}^{-(\alpha_1+\alpha_2)\sqrt{\lambda t}}}{z_t^{(1,2)}} \xrightarrow{\mathbb{P}} \mathrm{e}^{\alpha_1^2/2+\alpha_2^2/2}. \tag{3.82}$$

Indeed, define the (random) measure P on pairs (X,Y) by

$$P(X \le x, Y \le y) = \frac{\sum_{i \in \mathrm{BP}_t^{(1)}(x)} \sum_{j \in \mathrm{BP}_t^{(2)}(y)} \chi_{ij}(t)}{\sum_{i \in \mathrm{BP}_t^{(1)}} \sum_{j \in \mathrm{BP}_t^{(2)}} \chi_{ij}(t)}. \tag{3.83}$$

Then, (3.81) states that the pair (X, Y) converges in distribution to a pair of independent standard normal distributions, which, in turn, follows when, for each $(\alpha_1, \alpha_2) \in \mathbb{R}$, we have that

$$E[\mathrm{e}^{\alpha_1 X + \alpha_2 Y}] = \frac{z_t^{(1,2)}(\vec{a}_t)\mathrm{e}^{-(\alpha_1+\alpha_2)\sqrt{\lambda t}}}{z_t^{(1,2)}} \xrightarrow{\mathbb{P}} \mathrm{e}^{\alpha_1^2/2+\alpha_2^2/2}, \tag{3.84}$$

where E denotes the expectation w.r.t. P.

In order to show (3.82), we show that

$$\mathrm{e}^{-2\lambda t}\Big(z_t^{(1,2)}(\vec{a}_t)\mathrm{e}^{-(\alpha_1+\alpha_2)\sqrt{\lambda t}} - \mathrm{e}^{\alpha_1^2/2+\alpha_2^2/2} z_t^{(1,2)}\Big) \xrightarrow{\mathbb{P}} 0. \tag{3.85}$$

Together with Theorem 2.5, this then implies the result, as $\mathrm{e}^{-2\lambda t} z_t^{(1,2)}$ converges a.s. to a strictly positive random variable. Thus, we are left to prove (3.85). We shall show that the convergence in (3.85) in fact holds in L^2. For this, it is immediate that it suffices to study

$$M_t(\vec{\alpha}) \equiv \mathbb{E}\Big[z_t^{(1,2)}(\vec{a}_t)\mathrm{e}^{-(\alpha_1+\alpha_2)\sqrt{\lambda t}} - \mathrm{e}^{\alpha_1^2/2+\alpha_2^2/2} z_t^{(1,2)}\Big], \tag{3.86}$$

$$Q_t(\vec{\alpha}) \equiv \mathrm{Var}(z_t^{(1,2)}(\vec{a}_t)^2), \qquad C_t(\vec{\alpha}) \equiv \mathrm{Cov}(z_t^{(1,2)}, z_t^{(1,2)}(\vec{a}_t)). \tag{3.87}$$

In terms of these quantities, we can rewrite

$$\begin{aligned}&\mathbb{E}\Big[\Big(z_t^{(1,2)}(\vec{a}_t)\mathrm{e}^{-(\alpha_1+\alpha_2)\sqrt{\lambda t}} - \mathrm{e}^{\alpha_1^2/2+\alpha_2^2/2} z_t^{(1,2)}\Big)^2\Big] \\ &\quad = M_t(\vec{\alpha})^2 + Q_t(\vec{\alpha})\mathrm{e}^{-2(\alpha_1+\alpha_2)\sqrt{\lambda t}} + \mathrm{e}^{\alpha_1^2+\alpha_2^2} Q_t(\vec{0}) - 2C_t(\vec{\alpha})\mathrm{e}^{-(\alpha_1+\alpha_2)\sqrt{\lambda t}}\mathrm{e}^{\alpha_1^2/2+\alpha_2^2/2}.\end{aligned} \tag{3.88}$$

Therefore, we shall prove that

$$\mathrm{e}^{-2\lambda t} M_t(\vec{\alpha}) = o(1), \tag{3.89}$$

and

$$\mathrm{e}^{-4\lambda t}\Big(Q_t(\vec{\alpha})\mathrm{e}^{-2(\alpha_1+\alpha_2)\sqrt{\lambda t}} + \mathrm{e}^{\alpha_1^2+\alpha_2^2} Q_t(\vec{0}) - 2C_t(\vec{\alpha})\mathrm{e}^{-(\alpha_1+\alpha_2)\sqrt{\lambda t}}\mathrm{e}^{\alpha_1^2/2+\alpha_2^2/2}\Big) = o(1). \tag{3.90}$$

For these proofs, the explicit computations of mean and covariances of $z_t^{(1,2)}(\vec{a}_t)$ and $z_t^{(1,2)} = z_t^{(1,2)}(1,1)$ are crucial.

To prove (3.89), we rewrite

$$\mathrm{e}^{-2\lambda t} M_t(\vec{\alpha}) = \mathrm{e}^{(a_1^s+a_2^s)\lambda t - 2\lambda t - (\alpha_1+\alpha_2)\sqrt{\lambda t}}\mathrm{e}^{-\lambda(a_1^s+a_2^s)t}\mathbb{E}(z_t^{(1,2)}(\vec{a})) - \mathrm{e}^{\alpha_1^2/2+\alpha_2^2/2}\mathrm{e}^{-2\lambda t}\mathbb{E}(z_t^{(1,2)}(\vec{1})). \tag{3.91}$$

Since $\vec{a}_t = (\mathrm{e}^{\alpha_1/\sqrt{s^2\lambda t}}, \mathrm{e}^{\alpha_2/\sqrt{s^2\lambda t}}) \to (1,1)$,

$$\mathrm{e}^{-2\lambda t}\mathbb{E}(z_t^{(1,2)}(\vec{1})) \to A, \qquad \mathrm{e}^{-\lambda(a_1^s+a_2^s)t}\mathbb{E}(z_t^{(1,2)}(\vec{a})) \to A, \tag{3.92}$$

where A is the limit in (3.78) in Lemma 3.6(a). By a second order Taylor expansion,

$$(a_1^s\lambda t - \lambda t - \alpha_1\sqrt{\lambda t}) = \lambda t(\mathrm{e}^{\alpha_1/\sqrt{\lambda t}} - 1 - \frac{\alpha_1}{\sqrt{\lambda t}}) = \alpha_1^2/2 + o(1). \tag{3.93}$$

Together, these two asymptotics show that (3.89) holds. The proof of (3.90) is identical, now using Lemma 3.6(b) instead, and the fact that the limit equals $A_{\mathrm{Cov}}(\vec{1}, \vec{1})$ for all contributions, since $\vec{a}_t \to (1,1)$. ∎

Completion of the proof of Theorem 2.7. We can bound

$$\Big|\frac{\sum_{i\in \mathrm{SWG}_t^{(1)}(x)}\sum_{j\in \mathrm{SWG}_t^{(2)}(y)}\chi_{ij}(t)}{\sum_{i\in \mathrm{SWG}_t^{(1)}}\sum_{j\in \mathrm{SWG}_t^{(2)}}\chi_{ij}(t)}-\frac{\sum_{i\in \mathrm{BP}_t^{(1)}(x)}\sum_{j\in \mathrm{BP}_t^{(2)}(y)}\chi_{ij}(t)}{\sum_{i\in \mathrm{BP}_t^{(1)}}\sum_{j\in \mathrm{BP}_t^{(2)}}\chi_{ij}(t)}\Big| \tag{3.94}$$
$$\leq 2\frac{\mathrm{e}^{-2\lambda t}\sum_{i\in \mathrm{BP}_t^{(1)}}\sum_{j\in \mathrm{BP}_t^{(2)}}\chi_{ij}(t)-\mathrm{e}^{-2\lambda t}\sum_{i\in \mathrm{SWG}_t^{(1)}}\sum_{j\in \mathrm{SWG}_t^{(2)}}\chi_{ij}(t)}{\mathrm{e}^{-2\lambda t}\sum_{i\in \mathrm{BP}_t^{(1)}}\sum_{j\in \mathrm{BP}_t^{(2)}}\chi_{ij}(t)}.$$

The random variable in the denominator converges in probability to $\kappa W^{(1)}W^{(2)}>0$ by Theorem 2.5, so that it suffices to prove that the numerator converges to 0 in probability.

Denote

$$z_t^{n,(1,2)}=\sum_{i\in \mathrm{SWG}_t^{(1)}}\sum_{j\in \mathrm{SWG}_t^{(2)}}\chi_{ij}(t). \tag{3.95}$$

Then, similarly to Proposition 2.2(c), and recalling that $t_n=T_{12}\wedge t_n^*$ where $t_n^*=(2\lambda)^{-1}\log n+B$ for some $B>0$, we obtain that $\sup_{t\leq t_n}(z_t^{n,(1,2)}-z_t^{(1,2)})$ is tight. From Theorem 1.1 (whose proof has been completed since it relies only on Theorem 2.5, which was proved in the previous section), we know that the collision time T_{12} is bounded by t_n^* with probability $1-o(1)$ as $B\uparrow\infty$. Therefore,

$$\mathrm{e}^{-2\lambda t}\sum_{i\in \mathrm{BP}_t^{(1)}}\sum_{j\in \mathrm{BP}_t^{(2)}}\chi_{ij}(t)-\mathrm{e}^{-2\lambda t}\sum_{i\in \mathrm{SWG}_t^{(1)}}\sum_{j\in \mathrm{SWG}_t^{(2)}}\chi_{ij}(t)=\mathrm{e}^{-2\lambda t}(z_t^{n,(1,2)}-z_t^{(1,2)})\xrightarrow{\mathbb{P}}0. \tag{3.96}$$

This completes the proof. ■

A Appendix: auxiliary results

In this section, we prove an auxiliary result on the asymptotics of the measure $\tilde{\mu}_a(du)=\sum_{j=0}^{\infty}a^j\mu^{*j}(du)$.

Lemma A.1 (Asymptotics of density of $\tilde{\mu}_a$) *(a) Let*

$$\mathrm{e}^{-\lambda a^s u}\sum_{j=1}^{\infty}a^j\mu^{*j}(du)\equiv p_a(u)du. \tag{A.1}$$

Then, for $u\in[0,1]$, there exists a constant c such that $p_a(u)\leq cu^{1/s-1}$, while, for $u\geq 1$, $p_a(u)$ is bounded and as $u\to\infty$,

$$p_1(u)\to\lambda s. \tag{A.2}$$

(b) The following scaling identity holds:

$$p_a(u)=a^s p(ua^s). \tag{A.3}$$

In particular, when $a_u\to 1$,

$$p_{a_u}(u)\to\lambda s. \tag{A.4}$$

Proof. By (3.6), we have

$$p_a(u)=\mathrm{e}^{-\lambda u}\sum_{j=1}^{\infty}a^j\frac{u^{j/s-1}\lambda^{j/s}}{\Gamma(j/s)}. \tag{A.5}$$

This form immediately proves the identity in (A.3), and therefore also (A.4) follows from (A.2). Also, this form immediately shows that $p_a(u) \leq cu^{1/s-1}$ for $u \in [0,1]$. Thus, we are left to prove (A.2).

By [12, 8.327], we have that, as $z \to \infty$,

$$z^{z-1/2}\mathrm{e}^{-z}\sqrt{2\pi} \leq \Gamma(z) \leq z^{z-1/2}\mathrm{e}^{-z}\sqrt{2\pi}(1+\frac{1}{12z}). \tag{A.6}$$

Therefore,

$$\begin{aligned} p_1(u) &= (1+o(1))\mathrm{e}^{-\lambda u}\sum_{j=0}^{\infty}\frac{1}{\sqrt{2\pi s/j}}u^{j/s-1}\lambda^{j/s}\mathrm{e}^{j/s}(j/s)^{-j/s} \qquad (A.7)\\ &= (1+o(1))\lambda\mathrm{e}^{-\lambda u}(\lambda u)^{-1}\sum_{j=0}^{\infty}\frac{\sqrt{j}}{\sqrt{2\pi s}}(\lambda use j^{-1})^{j/s}, \end{aligned}$$

where the error term converges to 0 as $z \to \infty$. Note that the r.h.s. is a function of λu, so that it suffices to prove that

$$q(v) = \mathrm{e}^{-v}v^{-1}\sqrt{s}\sum_{j=0}^{\infty}\frac{1}{\sqrt{2\pi j}}\mathrm{e}^{j/s\log(vse/j)} \to s. \tag{A.8}$$

For this, we note that $j \mapsto \mathrm{e}^{j/s\log(vse/j)}$ is maximal when $j = sv$, where it takes the value e^{v}. A second order Taylor expansion shows that when $j - sv = x$, we have

$$\mathrm{e}^{j/s\log(vse/j)} = \mathrm{e}^{v}\mathrm{e}^{-x^2/(2s^2v)}(1+o(1)). \tag{A.9}$$

Performing the approximate Gaussian sum leads to the claim in (A.8). ■

Acknowledgments. The research of SB is supported by N.S.F. Grant DMS 0704159 and by PIMS and NSERC Canada. SB would also like to thank the hospitality of EURANDOM where much of this work was done. The work of RvdH is supported in part by Netherlands Organisation for Scientific Research (NWO).

References

[1] L. Addario-Berry, N. Broutin, and B. Reed. The diameter of the minimum spanning tree of a complete graph. In *Proc. Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*, volume **2**, (2006).

[2] D. Aldous. Asymptotics in the random assignment problem. *Probability Theory and Related Fields*, **93**(4):507–534, (1992).

[3] D. Aldous and J.M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume **110** of *Encyclopaedia Math. Sci.*, pages 1–72. Springer, Berlin, (2004).

[4] S. Bhamidi. First passage percolation on locally tree like networks I: Dense random graphs. *Journal of Mathematical Physics*, **49**:125218, (2008).

[5] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. Extreme value theory, Poisson-Dirichlet distributions and first passage percolation on random networks. *To appear in Adv. Appl. Probab.*, Preprint (2009).

[6] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. First passage percolation on sparse random graphs with finite mean degrees. *To appear in Ann. Appl. Probab.*, Preprint (2009).

[7] B. Bollobás, D. Gamarnik, O. Riordan, and B. Sudakov. On the value of a random minimum weight Steiner tree. *Combinatorica*, **24**(2):187–207, (2004).

[8] L.A. Braunstein, S.V. Buldyrev, R. Cohen, S. Havlin, and H.E. Stanley. Optimal Paths in Disordered Complex Networks. *Physical Review Letters*, **91**(16):168701, (2003).

[9] Y. Chen, E. López, S. Havlin, and H.E. Stanley. Universal behavior of optimal paths in weighted networks with general disorder. *Physical Review Letters*, **96**(6):68702, (2006).

[10] H. van den Esker, R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with infinite mean degrees. *Extremes*, **8**(3):111–141, (2005).

[11] A.M. Frieze. Value of a random minimum spanning tree problem. *Discrete Appl. Math.*, **10**(1):47–56, (1985).

[12] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products.* Fourth edition prepared by Ju. V. Geronimus and M. Ju. Ceĭtlin. Translated from the Russian by Scripta Technica, Inc. Translation edited by Alan Jeffrey. Academic Press, New York, (1965).

[13] S. Havlin, L.A. Braunstein, S.V. Buldyrev, R. Cohen, T. Kalisky, S. Sreenivasan, and H. Eugene Stanley. Optimal path in random networks with disorder: A mini review. *Physica A: Statistical Mechanics and its Applications*, **346**(1-2):82–92, (2005).

[14] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. The flooding time in random graphs. *Extremes*, **5**(2):111–129, (2002).

[15] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures Algorithms*, **27**(1):76–123, (2005).

[16] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Probab.*, **12**:no. 25, 703–766 (electronic), (2007).

[17] P. Jagers. *Branching processes with biological applications.* Wiley-Interscience [John Wiley & Sons], London, (1975). Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics.

[18] P. Jagers and O. Nerman. The growth and composition of branching populations. *Adv. in Appl. Probab.*, **16**(2):221–259, (1984).

[19] S. Janson. One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combin. Probab. Comput.*, **8**(4):347–361, (1999). Random graphs and combinatorial structures (Oberwolfach, 1997).

[20] O. Riordan and N. Wormald. The diameter of sparse random graphs. Prepint (2008).

[21] W. Rudin. *Principles of mathematical analysis.* McGraw-Hill Book Co., New York, third edition, (1976). International Series in Pure and Applied Mathematics.

[22] S. Sreenivasan, T. Kalisky, L.A. Braunstein, S.V. Buldyrev, S. Havlin, and H.E. Stanley. Effect of disorder strength on optimal paths in complex networks. *Physical Review E*, **70**(4):46133, (2004).