The Annals of Applied Probability
2010, Vol. 20, No. 5, 1907–1965
DOI: 10.1214/09-AAP666

FIRST PASSAGE PERCOLATION ON RANDOM GRAPHS WITH FINITE MEAN DEGREES

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

University of North Carolina, Eindhoven University of Technology and Delft University of Technology

We study first passage percolation on the configuration model. Assuming that each edge has an independent exponentially distributed edge weight, we derive explicit distributional asymptotics for the minimum weight between two randomly chosen connected vertices in the network, as well as for the number of edges on the least weight path, the so-called *hopcount*.

We analyze the configuration model with degree power-law exponent $\tau > 2$, in which the degrees are assumed to be i.i.d. with a tail distribution which is either of power-law form with exponent $\tau - 1 > 1$, or has even thinner tails ($\tau = \infty$). In this model, the degrees have a finite first moment, while the variance is finite for $\tau > 3$, but infinite for $\tau \in (2, 3)$.

We prove a central limit theorem for the hopcount, with asymptotically equal means and variances equal to $\alpha \log n$, where $\alpha \in (0, 1)$ for $\tau \in (2, 3)$, while $\alpha > 1$ for $\tau > 3$. Here n denotes the size of the graph. For $\tau \in (2, 3)$, it is known that the graph distance between two randomly chosen connected vertices is proportional to $\log \log n$ [*Electron. J. Probab.* **12** (2007) 703–766], that is, distances are *ultra small*. Thus, the addition of edge weights causes a marked change in the geometry of the network. We further study the weight of the least weight path and prove convergence in distribution of an appropriately centered version.

This study continues the program initiated in [*J. Math. Phys.* **49** (2008) 125218] of showing that $\log n$ is the correct scaling for the hopcount under i.i.d. edge disorder, even if the graph distance between two randomly chosen vertices is of much smaller order. The case of infinite mean degrees ($\tau \in [1, 2)$) is studied in [Extreme value theory, Poisson–Dirichlet distributions and first passage percolation on random networks (2009) Preprint] where it is proved that the hopcount remains uniformly bounded and converges in distribution.

1. Introduction. The general study of *real-world* networks has seen a tremendous growth in the last few years. This growth occurred both at an empirical level of obtaining data on networks such as the Internet, transportation networks, such

Received March 2009; revised October 2009.

[1]Supported by NSF Grant DMS-07-04159.

[2]Supported in part by Netherlands Organisation for Scientific Research (NWO).

AMS 2000 subject classifications. 60C05, 05C80, 90B15.

Key words and phrases. Flows, random graph, first passage percolation, hopcount, central limit theorem, coupling to continuous-time branching processes, universality.

as rail and road networks, and biochemical networks, such as gene regulatory networks, as well as at a theoretical level in the understanding of the properties of various mathematical models for these networks.

We are interested in one specific theoretical aspect of the above vast and expanding field. The setting is as follows: Consider a transportation network whose main aim is to transport flow between various vertices in the network via the available edges. At the very basic level there are two crucial elements which affect the flow carrying capabilities and delays experienced by vertices in the network:

(a) The actual graph topology, such as the density of edges and existence of short paths between vertices in the graph distance. In this context there has been an enormous amount of interest in the concept of *small-world* networks where the typical graph distance between vertices in the network is of order $\log n$ or even smaller. Indeed, for many of the mathematical models used to model real-world transmission networks, such as the Internet, the graph distance can be of order much smaller than order $\log n$. See, for example, [13, 35], where for the configuration model with degree exponent $\tau \in (2, 3)$, the remarkable result that the graph distance between typical vertices is of order $\log\log n$ is proved. In this case, we say that the graph is *ultra small*, a phrase invented in [13]. Similar results have appeared for related models in [11, 16, 28]. The configuration model is described in more detail in Section 2. For introductions to scale-free random graphs, we refer to the monographs [12, 17], for surveys of classical random graphs focussing on the Erdős–Rényi random graph (see [8, 25]).

(b) The second factor which plays a crucial role is the edge weight or cost structure of the graph which can be thought of as representing actual economic costs or congestion costs across edges. Edge weights being identically equal to 1 gives us back the graph geometry. What can be said when the edge costs have some other behavior? The main aim of this study is to understand what happens when each edge is given an independent edge cost with mean 1. For simplicity, we have assumed that the distribution of edge costs is exponentially with mean 1 [Exp(1)], leading to first passage percolation on the graph involved. First passage percolation with exponential weights has received substantial attention (see [5, 21, 22, 30, 32, 33, 37]), in particular on the complete graph, and, more recently, also on Erdős–Rényi random graphs. However, particularly the relation to the scale-free nature of the underlying random graph and the behavior of first passage percolation on it has not yet been investigated.

In this paper, we envisage a situation where the edge weights represent actual economic costs, so that all flow is routed through minimal weight paths. The actual time delay experienced by vertices in the network is given by the number of edges on this least cost path or hopcount H_n. Thus, for two typical vertices 1 and 2 in the network, it is important to understand both the minimum weight W_n of transporting flow between two vertices as well as the hopcount H_n or the number of edges on this minimal weight path. What we shall see is the following universal behavior:

> *Even if the graph topology is of ultra-small nature, the addition of random edge weights causes a complete change in the geometry and, in particular, the number of edges on the minimal weight path between two vertices increases to* $\Theta(\log n)$.

Here we write $a_n = \Theta(b_n)$ if there exist positive constants c and C, such that, for all n, we have $cb_n \leq a_n \leq Cb_n$. For the precise mathematical results we refer to Section 3. We shall see that a remarkably universal picture emerges, in the sense that for each $\tau > 2$, the hopcount satisfies a central limit theorem (CLT) with asymptotically equal mean and variance equal to $\alpha \log n$, where $\alpha \in (0, 1)$ for $\tau \in (2, 3)$, while $\alpha > 1$ for $\tau > 3$. The parameter α is the only feature which is left from the randomness of the underlying random graph, and α is a simple function of τ for $\tau \in (2, 3)$, and of the average forward degree for $\tau > 3$. This type of universality is reminiscent of that of simple random walk, which, appropriately scaled, converges to Brownian motion, and the parameters needed for the Brownian limit are only the mean and variance of the step-size. Interestingly, for the Internet hopcount, measurements show that the hopcount is close to a normal distribution with equal mean and variance (see, e.g., [36]), and it would be of interest to investigate whether first passage percolation on a random graph can be used as a model for the Internet hopcount.

This paper is part of the program initiated in [5] to rigorously analyze the asymptotics of distances and weights of shortest-weigh paths in random graph models under the addition of edge weights. In this paper, we rigorously analyze the case of the configuration model with degree exponent $\tau > 2$, the conceptually important case in practice, since the degree exponent of a wide variety of real-world networks is conjectured to be in this interval. In [6], we investigate the case $\tau \in [1, 2)$, where the first moment of the degrees is infinite and we observe entirely different behavior of the hopcount H_n.

2. Notation and definitions. We are interested in constructing a random graph on n vertices. Given a *degree sequence*, namely a sequence of n positive integers $\mathbf{d} = (d_1, d_2, \ldots, d_n)$ with $\sum_{i=1}^{n} d_i$ assumed to be even, the configuration model (CM) on n vertices with degree sequence $\mathbf{d}$ is constructed as follows:

Start with n vertices and d_i stubs or half-edges adjacent to vertex i. The graph is constructed by randomly pairing each stub to some other stub to form edges. Let

$$l_n = \sum_{i=1}^{n} d_i \tag{2.1}$$

denote the total degree. Number the stubs from 1 to l_n in some arbitrary order. Then, at each step, two stubs which are not already paired are chosen uniformly at random among all the unpaired or *free* stubs and are paired to form a single edge in the graph. These stubs are no longer free and removed from the list of free stubs. We continue with this procedure of choosing and pairing two stubs until all the stubs are paired. Observe that the order in which we choose the stubs does not matter. Although self-loops may occur, these become rare as $n \to \infty$ (see, e.g., [8] or [23] for more precise results in this direction).

Above, we have described the construction of the CM when the degree sequence is given. Here we shall specify how we construct the actual degree sequence $\mathbf{d}$ which shall be *random*. In general, we shall let a capital letter (such as D_i) denote a random variable, while a lower case letter (such as d_i) denote a deterministic object. We shall assume that the random variables $D_1, D_2, \ldots, D_n$ are independent and identically distributed (i.i.d.) with a certain distribution function F. (When the sum of stubs $L_n = \sum_{i=1}^n D_i$ is not even then we shall use the degree sequence $D_1, D_2, \ldots, D_n$, with D_n replaced by $D_n + 1$. This does not effect our calculations.)

We shall assume that the degrees of all vertices are at least 2 and that the degree distribution F is regularly varying. More precisely, we assume

$$\mathbb{P}(D \geq 2) = 1 \quad \text{and} \quad 1 - F(x) = x^{-(\tau-1)} L(x), \tag{2.2}$$

with $\tau > 2$, and where $x \mapsto L(x)$ is a slowly varying function for $x \to \infty$. In the case $\tau > 3$, we shall replace (2.2) by the less stringent condition (3.2). Furthermore, each edge is given a random edge weight, which in this study will always be assumed to be independent and identically distributed (i.i.d.) exponential random variables with mean 1. Because in our setting the vertices are exchangeable, we let 1 and 2 be the two random vertices picked *uniformly at random* in the network.

As stated earlier, the parameter τ is assumed to satisfy $\tau > 2$, so that the degree distribution has finite mean. In some cases, we shall distinguish between $\tau > 3$ and $\tau \in (2, 3)$; in the former case, the variance of the degrees is finite, while in the latter, it is infinite. It follows from the condition $D_i \geq 2$, almost surely, that the probability that the vertices 1 and 2 are connected converges to 1.

Let $f = \{f_j\}_{j=1}^\infty$ denote the probability mass function corresponding to the distribution function F, so that $f_j = F(j) - F(j-1)$. Let $\{g_j\}_{j=1}^\infty$ denote the *size-biased* probability mass function corresponding to f, defined by

$$g_j = \frac{(j+1) f_{j+1}}{\mu}, \qquad j \geq 0, \tag{2.3}$$

where μ is the expected size of the degree, that is,

$$\mu = \mathbb{E}[D] = \sum_{j=1}^{\infty} j f_j. \tag{2.4}$$

3. Results. In this section, we state the main results for $\tau > 2$. We treat the case where $\tau > 3$ in Section 3.1 and the case where $\tau \in (2, 3)$ in Section 3.2. The case where $\tau \in [1, 2)$ is deferred to [6].

Throughout the paper, we shall denote by

$$(H_n, W_n), \tag{3.1}$$

the number of edges and total weight of the shortest-weight path between vertices 1 and 2 in the CM with i.i.d. degrees with distribution function F, where we condition the vertices 1 and 2 to be connected, and we assume that each edge in the CM has an i.i.d. exponential weight with mean 1.

3.1. *Shortest-weight paths for $\tau > 3$.* In this section, we shall assume that the distribution function F of the degrees in the CM is nondegenerate and satisfies $F(x) = 0, x < 2$, so that the random variable D is nondegenerate and satisfies $D \geq 2$, a.s., and that there exist $c > 0$ and $\tau > 3$ such that

$$1 - F(x) \leq cx^{-(\tau-1)}, \qquad x \geq 0. \tag{3.2}$$

Also, we let

$$\nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}. \tag{3.3}$$

As a consequence of the conditions we have that $\nu > 1$. The condition $\nu > 1$ is equivalent to the existence of a *giant component* in the CM, the size of which is proportional to n (see, e.g., [26, 27, 34]; for the most recent and general result, see [24]). Moreover, the proportionality constant is the survival probability of the branching process with offspring distribution $\{g_j\}_{j\geq 1}$. As a consequence of the conditions on the distribution function F, in our case, the survival probability equals 1, so that for $n \to \infty$ the graph becomes asymptotically connected in the sense that the giant component has $n(1 - o(1))$ vertices. Also, when (3.2) holds, we have that $\nu < \infty$. Throughout the paper, we shall let $\xrightarrow{d}$ denote convergence in distribution and $\xrightarrow{\mathbb{P}}$ convergence in probability.

THEOREM 3.1 (Precise asymptotics for $\tau > 3$). *Let the degree distribution F of the CM on n vertices be nondegenerate, satisfy $F(x) = 0, x < 2$ and satisfy* (3.2) *for some $\tau > 3$. Then*:

(a) *the hopcount H_n satisfies the CLT*

$$\frac{H_n - \alpha \log n}{\sqrt{\alpha \log n}} \xrightarrow{d} Z, \tag{3.4}$$

where Z has a standard normal distribution, and

$$\alpha = \frac{\nu}{\nu - 1} \in (1, \infty); \tag{3.5}$$

(b) *there exists a random variable V such that*

$$W_n - \frac{\log n}{\nu - 1} \xrightarrow{d} V. \tag{3.6}$$

In Appendix C, we shall identify the limiting random variable V as

$$V = -\frac{\log W_1}{\nu - 1} - \frac{\log W_2}{\nu - 1} + \frac{\Lambda}{\nu - 1} + \frac{\log \mu(\nu - 1)}{\nu - 1}, \tag{3.7}$$

where W_1, W_2 are two independent copies of the limiting random variable of a certain supercritical continuous-time branching process, and Λ has a Gumbel distribution.

3.2. *Analysis of shortest-weight paths for $\tau \in (2, 3)$.* In this section, we shall assume that (2.2) holds for some $\tau \in (2, 3)$ and some slowly varying function $x \mapsto L(x)$. When this is the case, the variance of the degrees is infinite, while the mean degree is finite. As a result, we have that ν in (3.3) equals $\nu = \infty$, so that the CM is always supercritical (see [24, 26, 27, 35]). In fact, for $\tau \in (2, 3)$, we shall make a stronger assumption on F than (2.2), namely, that there exists a $\tau \in (2, 3)$ and $0 < c_1 \le c_2 < \infty$ such that, for all $x \ge 0$,

$$c_1 x^{-(\tau-1)} \le 1 - F(x) \le c_2 x^{-(\tau-1)}. \tag{3.8}$$

THEOREM 3.2 [Precise asymptotics for $\tau \in (2, 3)$]. *Let the degree distribution F of the CM on n vertices be nondegenerate, satisfy $F(x) = 0, x < 2$ and satisfy* (3.8) *for some $\tau \in (2, 3)$. Then*:

(a) *the hopcount H_n satisfies the CLT*

$$\frac{H_n - \alpha \log n}{\sqrt{\alpha \log n}} \xrightarrow{d} Z, \tag{3.9}$$

where Z has a standard normal distribution and where

$$\alpha = \frac{2(\tau - 2)}{\tau - 1} \in (0, 1); \tag{3.10}$$

(b) *there exists a limiting random variable V such that*

$$W_n \xrightarrow{d} V. \tag{3.11}$$

In Section 6, we shall identify the limiting random variable V precisely as

$$V = V_1 + V_2, \tag{3.12}$$

where V_1, V_2 are two independent copies of a random variable which is the explosion time of a certain infinite-mean continuous-time branching process.

3.3. *Discussion and related literature.*

Motivation. The basic motivation of this work was to show that even though the underlying graph topology might imply that the distance between two vertices is very small, if there are edge weights representing capacities, say, then the hopcount could drastically increase. Of course, the assumption of i.i.d. edge weights is not very realistic; however, it allows us to almost completely analyze the minimum weight path. The assumption of exponentially distributed edge weights is probably not necessary [1, 22] but helps in considerably simplifying the analysis. Interestingly, hopcounts which are close to normal with asymptotically equal means and variances are observed in Internet (see, e.g., [36]). The results presented here might shed some light on the origin of this observation.

Universality for first passage percolation on the CM. Comparing Theorems 3.1 and 3.2 we see that a remarkably universal picture emerges. Indeed, the hopcount *in both cases* satisfies a CLT with equal mean and variance proportional to $\log n$, and the proportionality constant α satisfies $\alpha \in (0, 1)$ for $\tau \in (2, 3)$, while $\alpha > 1$ for $\tau > 3$. We shall see that the proofs of Theorems 3.1 and 3.2 run, to a large extent, parallel, and we shall only need to distinguish when dealing with the related branching process problem to which the neighborhoods can be coupled.

The case $\tau \in [1, 2)$ and critical cases $\tau = 2$ and $\tau = 3$. In [6], we study first passage percolation on the CM when $\tau \in [1, 2)$, that is, the degrees have infinite mean. We show that a remarkably different picture emerges, in the sense that H_n remains uniformly bounded and converges in distribution. This is due to the fact that we can think of the CM, when $\tau \in [1, 2)$, as a union of an (essentially) finite number of stars. Together with the results in Theorems 3.1–3.2, we see that only the critical cases $\tau = 2$ and $\tau = 3$ remain open. We conjecture that the CLT, with asymptotically equal means and variances, remains valid when $\tau = 3$, but that the proportionality constant α can take any value in $[1, \infty)$, depending on, for example, whether ν in (3.3) is finite or not. What happens for $\tau = 2$ is less clear to us.

Graph distances in the CM. Expanding neighborhood techniques for random graphs have been used extensively to explore shortest path structures and other properties of locally tree-like graphs. See the closely related papers [29, 31, 34, 35] where an extensive study of the CM has been carried out. Relevant to our context is [35], Corollary 1.4(i), where it has been shown that when $2 < \tau < 3$, the graph distance $\widetilde{H}_n$ between two typical vertices,which are conditioned to be connected, satisfies the asymptotics

$$\frac{\widetilde{H}_n}{\log\log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau - 2)|} \tag{3.13}$$

as $n \to \infty$, and furthermore that the fluctuations of $\tilde{H}_n - \log\log n$ remain uniformly bounded as $n \to \infty$. For $\tau > 3$, it is shown in [34], Corollary 1.3(i), and that $\tilde{H}_n - \log n$ has bounded fluctuations

$$\frac{\widetilde{H}_n}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log \nu}, \tag{3.14}$$

again with bounded fluctuations. Comparing these results with Theorems 3.1–3.2, we see the drastic effect that the addition of edge weights has on the geometry of the graph.

The degree structure. In this paper, as in [29, 31, 34, 35], we assume that the degrees are i.i.d. with a certain degree distribution function F. In the literature, also the setting where the degrees $\{d_i\}_{i=1}^n$ are deterministic, and converge in an appropriate sense to an asymptotic degree distribution is studied (see, e.g., [11, 18, 24, 26, 27]). We expect that our results can be adapted to this situation. Also, we assume that the degrees are at least 2 a.s., which ensures that two uniform vertices lie, with high probability (w.h.p.) in the giant component. We have chosen for this setting to keep the proofs as simple as possible, and we conjecture that Theorems 3.1–3.2, when instead we condition the vertices 1 and 2 to be connected, remain true verbatim in the more general case of the supercritical CM.

Annealed vs. quenched asymptotics. The problem studied in this paper, first passage percolation on a random graph, fits in the more general framework of stochastic processes in random environments, such as random walk in random environment. In such problems, there are two interesting settings, namely, when we study results when averaging out over the environment and when we freeze the environment (the so-called annealed and quenched asymptotics). In this paper, we study the *annealed* setting, and it would be of interest to extend our results to the *quenched* setting, that is, study the first-passage percolation problem *conditionally on the random graph*. We expect the results to change in this case, primarily due to the fact that we know the exact neighborhood of each point. However, when we consider the shortest-weight problem between two *uniform* vertices, we conjecture Theorems 3.1–3.2 to remain valid verbatim, due to the fact that the neighborhoods of uniform vertices converge to the same limit as in the annealed setting (see, e.g., [4, 34]).

First passage percolation on the Erdős–Rényi random graph. We recall that the Erdős–Rényi random graph $G(n, p)$ is obtained by taking the vertex set $[n] = \{1, \dots, n\}$ and letting each edge ij be present, independently of all other edges, with probability p. The study closest in spirit to our study is [5] where similar ideas were explored for dense Erdős–Rényi random graphs. The Erdős–Rényi random graph $G(n, p)$ can be viewed as a close brother of the CM, with Poisson degrees, hence with $\tau = \infty$. Consider the case where $p = \mu/n$ and $\mu > 1$. In a future paper we plan to show, parallel to the above analysis, that H_n satisfies a CLT with asymptotically equal mean and variance given by $\frac{\mu}{\mu-1}\log n$. This connects up nicely with [5] where related results were shown for $\mu = \mu_n \to \infty$, and $H_n/\log n$ was proved to converge to 1 in probability. See also [32] where related statements were proved under stronger assumptions on μ_n. Interestingly, in a recent paper, Ding et al. [15] use first passage percolation to study the diameter of the largest component of the Erdős–Rényi random graph with edge probability $p = (1+\varepsilon)/n$ for $\varepsilon = o(1)$ and $\varepsilon^3 n \to \infty$.

The weight distribution. It would be of interest to study the effect of weights even further, for example, by studying the case where the weights are i.i.d. random variables with distribution equal to E^s where E is an exponential random variable with mean 1 and $s \in [0, \infty)$. The case $s = 0$ corresponds to the graph distance $\tilde{H}_n$ as studied in [31, 34, 35] while the case $s = 1$ corresponds to the case with i.i.d. exponential weights as studied here. Even the problem on the complete graph seems to be open in this case, and we intend to return to this problem in a future paper. We conjecture that the CLT remains valid for first passage perolation on the CM when the weights are given by independent copies of E^s with asymptotic mean and variance proportional to $\log n$, but, when $s \neq 1$, we predict that the asymptotic means and variances have *different* constants.

We became interested in random graphs with edge weights from [9] where, via empirical simulations, a wide variety of behavior was predicted for the shortest-weight paths in various random graph models. The setup that we analyze is the *weak disorder* case. In [9], also a number of interesting conjectures regarding the *strong disorder case* were made, which would correspond to analyzing the minimal spanning tree of these random graph models, and which is a highly interesting problem.

Related literature on shortest-weight problems. First passage percolation, especially on the integer lattice, has been extensively studied in the last fifty years (see, e.g., [20, 30] and the more recent survey [21]). In these papers, of course, the emphasis is completely different, in the sense that geometry plays an intrinsic role and often the goal of the study is to show that there is a limiting "shape" to first passage percolation from the origin.

Janson [22] studies first passage percolation on the complete graph with exponential weights. His main results are

$$\frac{W_n^{(ij)}}{\log n/n} \xrightarrow{\mathbb{P}} 1, \qquad \frac{\max_{j\le n} W_n^{(ij)}}{\log n/n} \xrightarrow{\mathbb{P}} 2, \qquad \frac{\max_{i,j\le n} W_n^{(ij)}}{\log n/n} \xrightarrow{\mathbb{P}} 3, \tag{3.15}$$

where $W_n^{(ij)}$ denotes the weight of the shortest path between the vertices i and j. Recently the authors of [1] showed in the same set-up that $\max_{i,j\le n} H_n^{(ij)}/\log n \xrightarrow{\mathbb{P}} \alpha^\star$ where $\alpha^\star \approx 3.5911$ is the unique solution of the equation $x\log x - x = 1$. It would be of interest to investigate such questions in the CM with exponential weights.

The fundamental difference of first passage percolation on the integer lattice, or even on the complete graph, is that in our case the underlying graph is random as well, and we are lead to the delicate relation between the randomness of the graph together with that of the stochastic process, in this case first passage percolation, living on it. Finally, for a slightly different perspective to shortest weight problems, see [37] where relations between the random assignment problem and the shortest-weight problem with exponential edge weights on the complete graph are explored.

4. Overview of the proof and organization of the paper. The key idea of the proof is to first grow the shortest-weight graph (SWG) from vertex 1, until it reaches an appropriate size. After this, we grow the SWG from vertex 2 until it connects up with the SWG from vertex 1. The size to which we let the SWG from 1 grow shall be the same as the *typical size* at which the connection between the SWG from vertices 1 and 2 shall be made. However, the connection time at which the SWG from vertex 2 connects to the SWG from vertex 1 is *random*.

More precisely, we define the SWG from vertex 1, denoted by $\mathrm{SWG}^{(1)}$, recursively. The growth of the SWG from vertex 2, which is denoted by $\mathrm{SWG}^{(2)}$, is similar. We start with vertex 1 by defining $\mathrm{SWG}_0^{(1)} = \{1\}$. Then we add the edge and vertex with minimal edge weight connecting vertex 1 to one of its neighbors (or itself when the minimal edge is a self-loop). This defines $\mathrm{SWG}_1^{(1)}$. We obtain $\mathrm{SWG}_m^{(1)}$ from $\mathrm{SWG}_{m-1}^{(1)}$ by adding the edge and end vertex connected to the $\mathrm{SWG}_{m-1}^{(1)}$ with minimal edge weight. We informally let $\mathrm{SWG}_m^{(i)}$ denote the SWG from vertex $i \in \{1, 2\}$ when m edges (and vertices) have been added to it. This definition is *informal*, as we shall need to deal with self-loops and cycles in a proper way. How we do this is explained in more detail in Section 4.2. As mentioned before, we first grow $\mathrm{SWG}_m^{(1)}$ to a size a_n, which is to be chosen appropriately. After this, we grow $\mathrm{SWG}_m^{(2)}$, and we stop as soon as a vertex of $\mathrm{SWG}_{a_n}^{(1)}$ appears in $\{\mathrm{SWG}_m^{(2)}\}_{m=0}^{\infty}$, as then the shortest-weight path between vertices 1 and 2 has been found. Indeed, if on the contrary, the shortest weight path between vertex 1 and vertex 2 contains an edge not contained in the union of the two SWGs when they meet, then necessarily this edge would have been chosen in one of the two SWGs at an earlier stage, since at some earlier stage this edge must have been incident to one of the SWGs and had the minimal weight of all edges incident to that SWG. In Sections 4.2 and 4.3, we shall make these definitions precise.

Denote this first common vertex by A, and let G_i be the distance between vertex i and A, that is, the number of edges on the minimum weight path from i to A. Then we have that

$$H_n = G_1 + G_2, \tag{4.1}$$

while, denoting by T_i the weight of the shortest-weight paths from i to A, we have

$$W_n = T_1 + T_2. \tag{4.2}$$

Thus, to understand the random variables H_n and W_n, it is paramount to understand the random variables T_i and G_i, for $i = 1, 2$.

Since, for $n \to \infty$, the topologies of the neighborhoods of vertices 1 and 2 are close to being independent, it seems likely that G_1 and G_2, as well as T_1 and T_2

are close to independent. Since, further, the CM is locally tree-like, we are lead to the study of the problem on a tree.

With the above in mind, the paper is organized as follows:

In Section 4.1 we study the flow on a tree. More precisely, in Proposition 4.3, we describe the asymptotic distribution of the length and weight of the shortest-weight path between the root and the mth added vertex in a branching process with i.i.d. degrees with offspring distribution g in (2.3). Clearly, the CM *has* cycles and self-loops, and thus sometimes deviates from the tree description.

In Section 4.2, we reformulate the problem of the growth of the SWG from a fixed vertex as a problem of the SWG on a tree, where we find a way to deal with cycles by a coupling argument, so that the arguments in Section 4.1 apply quite literally. In Proposition 4.6, we describe the asymptotic distribution of the length and weight of the shortest-weight path between a fixed vertex and the mth added vertex in the SWG from the CM. However, observe that the random variables G_i described above are the generation of a vertex at the time at which the two SWGs collide, and this time is a *random* variable.

In Section 4.3, we extend the discussion to this setting and, in Section 4.4, we formulate the necessary ingredients for the collision time, that is, the time at which the connecting edge appears, in Proposition 4.4. In Section 4.5, we complete the outline.

The proofs of the key propositions are deferred to Sections 5–7.

Technical results needed in the proofs in Sections 5–7, for example on the topology of the CM, are deferred to the Appendix A.

4.1. *Description of the flow clusters in trees.* We shall now describe the construction of the SWG in the context of trees. In particular, below, we shall deal with a flow on a branching process tree, where the offspring is deterministic.

Deterministic construction: Suppose we have positive (nonrandom) integers $d_1, d_2, \ldots$. Consider the following construction of a branching process in discrete time:

CONSTRUCTION 4.1 (Flow from root of tree). The shortest-weight graph on a tree with degrees $\{d_i\}_{i=1}^{\infty}$ is obtained as follows:

1. At time 0, start with one alive vertex (the initial ancestor).
2. At each time step i, pick one of the alive vertices at random, this vertex dies giving birth to d_i children.

In the above construction, the number of offspring d_i is fixed once and for all. For a branching process tree, the variables d_i are i.i.d. *random* variables. This case shall be investigated later on, but the case of deterministic degrees is more general and shall be important for us to be able to deal with the CM.

Consider a continuous-time branching process defined as follows:

1. Start with the root which dies immediately giving rise to d_1 alive offspring.
2. Each alive offspring lives for Exp(1) amount of time, independent of all other randomness involved.
3. When the mth vertex dies it leaves behind d_m alive offspring.

The split-times (or death-times) of this branching process are denoted by $T_i, i \geq 1$. Note that the Construction 4.1 is equivalent to this continuous branching process, observed at the discrete times $T_i, i \geq 1$. The fact that the chosen alive vertex is chosen at random follows from the memoryless property of the exponential random variables that compete to become the minimal one. We quote a fundamental result from [10]. In its statement, we let

$$s_i = d_1 + \cdots + d_i - (i-1).^{3} \tag{4.3}$$

PROPOSITION 4.2 (Shortest-weight paths on a tree). *Pick an alive vertex at time $m \geq 1$ uniformly at random among all vertices alive at this time. Then*

(a) *the generation of the mth chosen vertex is equal in distribution to*

$$G_m \overset{d}{=} \sum_{i=1}^{m} I_i, \tag{4.4}$$

where $\{I_i\}_{i=1}^{\infty}$ are independent Bernoulli random variables with

$$\mathbb{P}(I_i = 1) = d_i/s_i; \tag{4.5}$$

(b) *the weight of the shortest-weight path between the root of the tree and the vertex chosen in the mth step is equal in distribution to*

$$T_m \overset{d}{=} \sum_{i=1}^{m} E_i/s_i, \tag{4.6}$$

where $\{E_i\}_{i=1}^{\infty}$ are i.i.d. exponential random variables with mean 1.

PROOF. We shall prove part (a) by induction. The statement is trivial for $m = 1$. We next assume that (4.4) holds for m where $\{I_i\}_{i=1}^{m}$ are independent Bernoulli random variables satisfying (4.5). Let G_{m+1} denote the generation of the randomly chosen vertex at time $m+1$, and consider the event $\{G_{m+1} = k\}$, $1 \leq k \leq m$. If randomly choosing one of the alive vertices at time $m+1$ results in one of the d_{m+1} newly added vertices, then, in order to obtain generation k, the previous uniform choice, that is, the choice of the vertex which was the last one to die,

[3] A new *probabilistic* proof is added, since there is some confusion between the definition s_i given here, and the definition of s_i given in [10], below equation (3.1). More precisely, in [10], s_i is defined as $s_i = d_1 + \cdots + d_i - i$, which is our $s_i - 1$.

must have been a vertex from generation $k-1$. On the other hand, if a uniform pick is conditioned on not taking one of the d_{m+1} newly added vertices, then this choice must have been a uniform vertex from generation k. Hence, we obtain, for $1 \le k \le m$,

$$\mathbb{P}(G_{m+1} = k) = \frac{d_{m+1}}{s_{m+1}} \mathbb{P}(G_m = k-1) + \left(1 - \frac{d_{m+1}}{s_{m+1}}\right) \mathbb{P}(G_m = k). \tag{4.7}$$

The proof of part (a) is now immediate from the induction hypothesis. The proof of part (b) is as follows. The minimum of s_i independent exp(1) random variables has an exponential distribution with parameter s_i, and is hence equal in distribution to E_i/s_i. We further use the memoryless property of the exponential distribution which guarantees that at each of the discrete time steps the remaining lifetimes (or weights) of the alive vertices are independent exponential variables with mean 1, independent of what happened previously. □

We note that, while Proposition 4.2 was applied in [10], Theorem 3.1, only in the case where the degrees are i.i.d., in fact, the results hold more generally for *every* tree (see, e.g., [10], equation (3.1), and the above proof). This extension shall prove to be vital in our analysis.

We next intuitively relate the above result to our setting. Start from vertex 1, and iteratively choose the edge with minimal additional weight attached to the SWG so far. As mentioned before, because of the properties of the exponential distribution, the edge with minimal additional weight can be considered to be picked uniformly at random from all edges attached to the SWG at that moment. With high probability, this edge is connected to a vertex which is not in the SWG. Let B_i denote the *forward degree* (i.e., the degree minus 1) of the vertex to which the ith edge is connected. By the results in [34, 35], $\{B_i\}_{i\ge 2}$ are close to being i.i.d. and have distribution given by (2.3). Therefore, we are lead to studying random variables of the form (4.4)–(4.5) where $\{B_i\}_{i=1}^{\infty}$ are i.i.d. random variables. Thus, this means that we study the *unconditional* law of G_m in (4.4), in the setting where the vector $\{d_i\}_{i=1}^{\infty}$ is replaced by an i.i.d. sequence of random variables $\{B_i\}_{i=1}^{\infty}$. We shall first state a CLT for G_m and a limit result for T_m in this setting.

PROPOSITION 4.3 (Asymptotics for shortest-weight paths on trees). *Let $\{B_i\}_{i=1}^{\infty}$ be an i.i.d. sequence of nondegenerate, positive integer valued, random variables satisfying*

$$\mathbb{P}(B_i > k) = k^{2-\tau} L(k), \qquad \tau > 2,$$

for some slowly varying function $k \mapsto L(k)$. Denote by $\nu = \mathbb{E}[B_1]$, for $\tau > 3$, whereas $\nu = \infty$, for $\tau \in (2,3)$. Then,

(a) *for G_m given in (4.4)–(4.5), with d_i replaced by B_i, there exists a $\beta \ge 1$ such that, as $m \to \infty$,*

$$\frac{G_m - \beta \log m}{\sqrt{\beta \log m}} \xrightarrow{d} Z, \qquad \text{where } Z \sim \mathcal{N}(0,1) \tag{4.8}$$

a standard normal variable, and where $\beta = \nu/(\nu - 1)$ *for* $\tau > 3$, *while* $\beta = 1$ *for* $\tau \in (2, 3)$;

(b) *for* T_m *given in* (4.6), *there exists a random variable* X *such that*

$$T_m - \gamma \log m \xrightarrow{d} X, \tag{4.9}$$

where $\gamma = 1/(\nu - 1)$ *when* $\tau > 3$, *while* $\gamma = 0$ *when* $\tau \in (2, 3)$.

Proposition 4.3 is proved in [10], Theorem 3.1, when $\mathrm{Var}(B_i) < \infty$, which holds when $\tau > 4$, but not when $\tau \in (2, 4)$. We shall prove Proposition 4.3 in Section 5 below. There, we shall also see that the result persists under weaker assumptions than $\{B_i\}_{i=1}^{\infty}$ being i.i.d., for example, when $\{B_i\}_{i=1}^{\infty}$ are *exchangeable* nonnegative integer valued random variables satisfying certain conditions. Such extensions shall prove to be useful when dealing with the actual (forward) degrees in the CM.

4.2. *A comparison of the flow on the CM and the flow on the tree*. Proposition 4.3 gives a CLT for the generation when considering a flow on a tree. In this section, we shall relate the problem of the flow on the CM to the flow on a tree. The key feature of this construction is that *we shall simultaneously grow the graph topology neighborhood of a vertex, as well as the shortest-weight graph from it.* This will be achieved by combining the construction of the CM as described in Section 2 with the fact that, from a given set of vertices and edges, if we grow the shortest-weight graph, each potential edge is equally likely to be the minimal one.

In the problem of finding the shortest weight path between two vertices 1 and 2, we shall grow two SWGs simultaneously from the two vertices 1 and 2, until they meet. This is the problem that we actually need to resolve in order to prove our main results in Theorems 3.1–3.2. The extension to the growth of two SWGs is treated in Section 4.3 below.

The main difference between the flow on a graph and on a tree is that on the tree there are no cycles, while on a graph there are. Thus we shall adapt the growth of the SWG for the CM in such a way that we obtain a tree (so that the results from Section 4.1 apply) while we can still retrieve all information about shortest-weight paths from the constructed graph. This will be achieved by introducing the notion of *artificial* vertices and stubs. We start by introducing some notation.

We denote by $\{\mathrm{SWG}_m\}_{m\geq 0}$ the SWG process from vertex 1. We construct this process recursively. We let SWG_0 consist only of the alive vertex 1, and we let $S_0 = 1$. We next let SWG_1 consist of the D_1 allowed stubs and of the explored vertex 1, and we let $S_1 = S_0 + D_1 - 1 = D_1$ denote the number of allowed stubs. In the sequel of the construction, the allowed stubs correspond to vertices in the shortest-weight problem on the tree in Section 4.1. This constructs SWG_1. Next, we describe how to construct SWG_m from SWG_{m-1}. For this construction, we shall have to deal with several types of stubs:

(a) The allowed stubs at time m, denoted by AS_m, are the stubs that are incident to vertices of the SWG_m and that have not yet been paired to form an edge; $S_m = |\mathrm{AS}_m|$ denotes their number;

(b) the free stubs at time m, denoted by FS_m, are those stubs of the L_n total stubs which have not yet been paired in the construction of the CM up to and including time m;

(c) the artificial stubs at time m, denoted by Art_m, are the *artificial* stubs created by breaking ties, as described in more detail below.

We note that $\mathrm{Art}_m \subset \mathrm{AS}_m$, indeed, $\mathrm{AS}_m \setminus \mathrm{FS}_m = \mathrm{Art}_m$. Then, we can construct SWG_m from SWG_{m-1} as follows. We choose one of the S_{m-1} allowed stubs uniformly at random, and then, if the stub is not artificial, pair it uniformly at random to a free stub unequal to itself. Below, we shall consistently call these two stubs the *chosen* stub and the *paired* stub, respectively. There are 3 possibilities, depending on what kind of stub we choose and what kind of stub it is paired to:

CONSTRUCTION 4.4 (The evolution of SWG for CM as SWG on a tree). (1) The chosen stub is real, that is, not artificial, and the paired stub is not one of the allowed stubs. In this case, which shall be most likely at the start of the growth procedure of the SWG, the paired stub is incident to a vertex outside SWG_{m-1}, we denote by B_m the forward degree of the vertex incident to the paired stub (i.e., its degree minus 1) and we define $S_m = S_{m-1} + B_m - 1$. Then we remove the paired and the chosen stub from AS_{m-1} and add the B_m stubs incident to the vertex incident to the paired stub to AS_{m-1} to obtain AS_m, we remove the chosen and the paired stubs from FS_{m-1} to obtain FS_m, and $\mathrm{Art}_m = \mathrm{Art}_{m-1}$.

(2) The chosen stub is real and the paired stub is an allowed stub. In this case, the paired stub is incident to a vertex in SWG_{m-1}, and we have created a cycle. In this case, we create an artificial stub replacing the paired stub and denote $B_m = 0$. Then we let $S_m = S_{m-1} - 1$, remove both the chosen and paired stubs from AS_{m-1} and add the artificial stub to obtain AS_m, and remove the chosen and paired stub from FS_{m-1} to obtain FS_m, while Art_m is Art_{m-1} together with the newly created artificial stub. In SWG_m, we also add an artificial edge to an artificial vertex in the place where the chosen stub was, the forward degree of the artificial vertex being 0. This is done because a vertex is added each time in the construction on a tree.

(3) The chosen stub is artificial. In this case, we let $B_m = 0$, $S_m = S_{m-1} - 1$ and remove the chosen stub from AS_{m-1} and Art_{m-1} to obtain AS_m and Art_m, while $\mathrm{FS}_m = \mathrm{FS}_{m-1}$.

In Construction 4.4, we always work on a tree since we replace an edge which creates a cycle, by one artificial stub, to replace the paired stub, and an artificial edge plus an artificial vertex in the SWG_m with degree 0, to replace the chosen stub. Note that the number of allowed edges at time m satisfies $S_m = S_{m-1} +$

$B_m - 1$, where $B_1 = D_1$ and, for $m \geq 2$, in cases (2) and (3), $B_m = 0$, while in case (1) (which we expect to occur in most cases), the distribution of B_m is equal to the forward degree of a vertex incident to a uniformly chosen stub. Here, the choice of stubs is without replacement.

The reason for replacing cycles as described above is that we wish to represent the SWG problem as a problem on a tree, as we now will explain informally. On a tree with degrees $\{d_i\}_{i=1}^{\infty}$, as in Section 4.1, we have that the remaining degree of vertex i at time m is precisely equal to d_i minus the number of neighbors that are among the m vertices with minimal shortest-weight paths from the root. For first passage percolation on a graph with cycles, a cycle does not only remove one of the edges of the vertex incident to it (as on the tree), but also one edge of the vertex at the other end of the cycle. Thus this is a *different* problem, and the results from Section 4.1 do not apply literally. By adding the artificial stub, edge and vertex, we artificially keep the degree of the receiving vertex the same, so that we *do* have the same situation as on a tree, and we can use the results in Section 4.1. However, we do need to investigate the relation between the problem with the artificial stubs and the original SWG problem on the CM. That is the content of the next proposition.

In its statement, we shall define the mth closest vertex to vertex 1 in the CM, with i.i.d. exponential weights, as the unique vertex of which the minimal weight path is the mth smallest among all $n-1$ vertices. Further, at each time m, we denote by *artificial vertices* those vertices which are artificially created, and we call the other vertices *real vertices*. Then we let the random time R_m be the first time j that SWG_j consists of $m+1$ real vertices, that is,

$$R_m = \min\{j \geq 0 : \mathrm{SWG}_j \text{ contains } m+1 \text{ real vertices}\}. \tag{4.10}$$

The $+1$ originates from the fact that at time $m = 0$, SWG_0 consists if 1 real vertex, namely, the vertex from which we construct the SWG. Thus, in the above set up, we have that $R_m = m$ precisely when no cycle has been created in the construction up to time m. Then our main coupling result is as follows:

PROPOSITION 4.5 (Coupling shortest-weight graphs on a tree and CM). *Jointly for all $m \geq 1$, the set of real vertices in SWG_{R_m} is equal in distribution to the set of ith closest vertices to vertex 1, for $i = 1, \dots, m$. Consequently*:

(a) *the generation of the mth closest vertex to vertex 1 has distribution G_{R_m} where G_m is defined in (4.4)–(4.5) with $d_1 = D_1$ and $d_i = B_i$, $i \geq 2$, as described in Construction 4.4*;

(b) *the weight of the shortest weight path to the mth closest vertex to vertex 1 has distribution T_{R_m}, where T_m is defined in (4.6) with $d_1 = D_1$ and $d_i = B_i$, $i \geq 2$, as described in Construction 4.4.*

We shall make use of the nice property that the sequence $\{B_{R_m}\}_{m=2}^{n}$, which consists of the forward degrees of chosen stubs that are paired to stubs which are not in the SWG, is, for the CM, an exchangeable sequence of random variables (see Lemma 6.1 below). This is due to the fact that a free stub is chosen uniformly at random, and the order of the choices does not matter. This exchangeability shall prove to be useful in order to investigate shortest-weight paths in the CM. We now prove Proposition 4.5.

PROOF OF PROPOSITION 4.5. In growing the SWG, we give exponential weights to the set $\{\mathrm{AS}_m\}_{m\geq 1}$. After pairing, we identify the exponential weight of the chosen stub to the exponential weight of the edge which it is part of. We note that by the memoryless property of the exponential random variable, each stub is chosen uniformly at random from all the allowed stubs incident to the SWG at the given time. Further, by the construction of the CM in Section 2, this stub is paired uniformly at random to one of the available free stubs. Thus the growth rules of the SWG in Construction 4.4 equal those in the above description of $\{\mathrm{SWG}_m\}_{m=0}^{\infty}$, unless a cycle is closed and an artificial stub, edge and vertex are created. In this case, the artificial stub, edge and vertex might influence the law of the SWG. However, we note that the artificial vertices are not being counted in the set of real vertices, and since artificial vertices have forward degree 0, they will not be a part of any shortest path to a real vertex. Thus the artificial vertex at the end of the artificial edge does not affect the law of the SWG. Artificial stubs that are created to replace paired stubs when a cycle is formed, and which are not yet removed at time m, will be called *dangling ends*. Now, if we only consider real vertices, then the distribution of weights and lengths of the shortest-weight paths between the starting points and those real vertices are identical. Indeed, we can decorate any graph with as many dangling ends as we like without changing the shortest-weight paths to real vertices in the graph. $\square$

Now that the flow problem on the CM has been translated into a flow problem on a related tree of which we have explicitly described its distribution, we may make use of Proposition 4.2 which shall allow us to extend Proposition 4.3 to the setting of the CM. Note that, among others, due to the fact that when we draw an artificial stub, the degrees are not i.i.d. (and not even exchangeable since the probability of drawing an artificial stub is likely to increase in time), we need to extend Proposition 4.3 to a setting where the degrees are weakly dependent. In the statement of the result, we recall that G_m is the height of the mth added vertex in the tree problem above. In the statement below, we write

$$a_n = n^{(\tau\wedge 3-2)/(\tau\wedge 3-1)} = \begin{cases} n^{(\tau-2)/(\tau-1)} & \text{for } \tau\in(2,3), \\ n^{1/2} & \text{for } \tau>3, \end{cases} \tag{4.11}$$

where, for $a,b\in\mathbb{R}$, we write $a\wedge b=\min\{a,b\}$.

Before we formulate the CLT for the hopcount of the shortest-weight graph in the CM, we repeat once more the setup of the random variables involved. Let

$S_0 = 1$, $S_1 = D_1$, and for $j \geq 2$,

$$S_j = D_1 + \sum_{i=2}^{j} (B_i - 1), \tag{4.12}$$

where, in case the chosen stub is real, that is, not artificial, and the paired stub is not one of the allowed stubs, B_i equals the forward degree of the vertex incident to the ith paired stub, whereas $B_i = 0$ otherwise. Finally, we recall that, conditionally on $D_1, B_2, B_3, \ldots, B_m$,

$$\begin{gathered} G_m = \sum_{i=1}^{m} I_i \qquad \text{where} \\ \mathbb{P}(I_1 = 1) = 1, \qquad \mathbb{P}(I_j = 1) = B_j / S_j, \qquad 2 \leq j \leq m. \end{gathered} \tag{4.13}$$

PROPOSITION 4.6 (Asymptotics for shortest-weight paths in the CM). (a) *Let the law of G_m be given in* (4.13). *Then, with $\beta \geq 1$ as in Proposition* 4.3, *and as long as $m \leq \overline{m}_n$, for any $\overline{m}_n$ such that* $\log(\overline{m}_n / a_n) = o(\sqrt{\log n})$,

$$\frac{G_m - \beta \log m}{\sqrt{\beta \log m}} \xrightarrow{d} Z \qquad \text{where } Z \sim \mathcal{N}(0, 1). \tag{4.14}$$

(b) *Let the law of T_m be given in* (4.6) *with s_i replaced by S_i given by* (4.12), *and let γ be as in Proposition* 4.3. *Then there exists a random variable X such that*

$$T_m - \gamma \log m \xrightarrow{d} X. \tag{4.15}$$

The same results apply to G_{R_m} and T_{R_m}, that is, in the statements (a) *and* (b) *the integer m can be replaced by R_m, as long as $m \leq \overline{m}_n$.*

Proposition 4.6 implies that the result of Proposition 4.3 remains true for the CM whenever m is not too large. Important for the proof of Proposition 4.6 is the coupling to a tree problem in Proposition 4.5. Proposition 4.6 shall be proved in Section 6. An important ingredient in the proof will be the comparison of the variables $\{B_m\}_{m=2}^{m_n}$, for an appropriately chosen m_n, to an i.i.d. sequence. Results in this direction have been proved in [34, 35], and we shall combine these to the following statement:

PROPOSITION 4.7 (Coupling the forward degrees to an independent sequence). *In the CM with $\tau > 2$, there exists a $\rho > 0$ such that the random vector $\{B_m\}_{m=2}^{n^\rho}$ can be coupled to an independent sequence of random variables $\{B_m^{(\mathrm{ind})}\}_{m=2}^{n^\rho}$ with probability mass function g in* (2.3) *in such a way that $\{B_m\}_{m=2}^{n^\rho} = \{B_m^{(\mathrm{ind})}\}_{m=2}^{n^\rho}$ w.h.p.*

In Proposition 4.7, in fact, we can take $\{B_m\}_{m=2}^{n^\rho}$ to be the forward degree of the vertex to which any collection of n^ρ distinct stubs has been connected.

4.3. *Flow clusters started from two vertices.* To compute the hopcount, we first grow the SWG from vertex 1 until time a_n, followed by the growth of the SWG from vertex 2 until the two SWGs meet, as we now explain in more detail. Denote by $\{\mathrm{SWG}_m^{(i)}\}_{m=0}^{\infty}$ the SWG from the vertex $i \in \{1, 2\}$, and, for $m \geq 0$, let

$$\mathrm{SWG}_m^{(1,2)} = \mathrm{SWG}_{a_n}^{(1)} \cup \mathrm{SWG}_m^{(2)}, \tag{4.16}$$

the union of the SWGs of vertex 1 and 2. We shall only consider values of m where $\mathrm{SWG}_{a_n}^{(1)}$ and $\mathrm{SWG}_m^{(2)}$ are *disjoint*, that is, they do not contain any common (real) vertices. We shall discuss the moment when they connect in Section 4.4 below.

We recall the notation in Section 4.2, and, for $i \in \{1, 2\}$, denote by $\mathrm{AS}_m^{(i)}$ and $\mathrm{Art}_m^{(i)}$ the number of allowed and artificial stubs in $\mathrm{SWG}_m^{(i)}$. We let the set of free stubs FS_m consist of those stubs which have not yet been paired in $\mathrm{SWG}_m^{(1,2)}$ in (4.16). Apart from that, the evolution of $\mathrm{SWG}_m^{(2)}$, following the evolution of $\mathrm{SWG}_{a_n}^{(1)}$, is identical as in Construction 4.4. We denote by $S_m^{(i)} = |\mathrm{AS}_m^{(i)}|$ the number of allowed stubs in $\mathrm{SWG}_m^{(i)}$ for $i \in \{1, 2\}$. We define $B_m^{(i)}$ accordingly.

The above description shows how we can grow the SWG from vertex 1 followed by the one of vertex 2. In order to state an adaptation of Proposition 4.5 to the setting where the SWGs of vertex 1 is first grown to size a_n, followed by the growth of the SWG from vertex 2 until the connecting edge appears, we let the random time $R_m^{(i)}$ be the first time l such that $\mathrm{SWG}_l^{(i)}$ consists of $m + 1$ real vertices. Then our main coupling result for two simultaneous SWGs is as follows:

PROPOSITION 4.8 (Coupling SWGs on two trees and CM from two vertices). *Jointly for $m \geq 0$, as long as the sets of real vertices in $(\mathrm{SWG}_{a_n}^{(1)}, \mathrm{SWG}_m^{(2)})$ are disjoint, these sets are equal in distribution to the sets of j_1th, respectively j_2th, closest vertices to vertex 1 and 2, respectively, for $j_1 = 1, \ldots, R_{a_n}^{(1)}$ and $j_2 = 1, \ldots, R_m^{(2)}$, respectively.*

4.4. *The connecting edge.* As described above, we grow the two SWGs until the first stub with minimal weight incident to $\mathrm{SWG}_m^{(2)}$ is paired to a stub incident to $\mathrm{SWG}_{a_n}^{(1)}$. We call the created edge linking the two SWGs the *connecting edge*. More precisely, let

$$\mathrm{C}_n = \min\{m \geq 0 : \mathrm{SWG}_{a_n}^{(1)} \cap \mathrm{SWG}_m^{(2)} \neq \varnothing\} \tag{4.17}$$

be the first time that $\mathrm{SWG}_{a_n}^{(1)}$ and $\mathrm{SWG}_m^{(2)}$ share a vertex. When $m = 0$, this means that $2 \in \mathrm{SWG}_{a_n}^{(1)}$ (which we shall show happens with small probability), while when $m \geq 1$, this means that the mth-stub of $\mathrm{SWG}^{(2)}$ which is chosen and then paired, is paired to a stub from $\mathrm{SWG}_{a_n}^{(1)}$. The path found actually is the shortest-weight path between vertices 1 and 2, since $\mathrm{SWG}_{a_n}^{(1)}$ and $\mathrm{SWG}_m^{(2)}$ precisely consists of the closest real vertices to the root i, for $i = 1, 2$, respectively.

We now study the probabilistic properties of the connecting edge. Let the edge $e = st$ be incident to $\mathrm{SWG}^{(1)}_{a_n}$, and s and t denote its two stubs. Let the vertex incident to s be i_s and the vertex incident to t be i_t. Assume that $i_s \in \mathrm{SWG}^{(1)}_{a_n}$, so that, by construction, $i_t \notin \mathrm{SWG}^{(1)}_{a_n}$. Then, conditionally on $\mathrm{SWG}^{(1)}_{a_n}$ and $\{T^{(1)}_i\}_{i=1}^{a_n}$, the weight of e is at least $T^{(1)}_{a_n} - W^{(1)}_{i_s}$, where $W^{(1)}_{i_s}$ is the weight of the shortest path from 1 to i_s. By the memoryless property of the exponential distribution, therefore, the weight on edge e equals $T^{(1)}_{a_n} - W^{(1)}_{i_s} + E_e$, where the collection (E_e), for all e incident to $\mathrm{SWG}^{(1)}_{a_n}$ are i.i.d. Exp(1) random variables. Alternatively, we can redistribute the weight by saying that the stub t has weight E_e, and the stub s has weight $T^{(1)}_{a_n} - W^{(1)}_{i_s}$. Further, in the growth of $(\mathrm{SWG}^{(2)}_m)_{m\ge 0}$, we can also think of the exponential weights of the edges incident to $\mathrm{SWG}^{(2)}_m$ being positioned on the *stubs* incident to $\mathrm{SWG}^{(2)}_m$. Hence, there is no distinction between the stubs that are part of edges connecting $\mathrm{SWG}^{(1)}_{a_n}$ and $\mathrm{SWG}^{(2)}_m$ and the stubs that are part of edges incident to $\mathrm{SWG}^{(2)}_m$, but not to $\mathrm{SWG}^{(1)}_{a_n}$. Therefore, in the growth of $(\mathrm{SWG}^{(2)}_m)_{m\ge 0}$, we can think of the minimal weight *stub* incident to $\mathrm{SWG}^{(2)}_m$ being chosen uniformly at random, and then a uniform free stub is chosen to pair it with. As a result, the distribution of the stubs chosen *at the time of connection* is equal to any of the other (real) stubs chosen along the way. This is a crucial ingredient to prove the scaling of the shortest-weight path between vertices 1 and 2.

For $i \in \{1, 2\}$, let $H^{(i)}_n$ denote the length of the shortest-weight path between vertex i and the common vertex in $\mathrm{SWG}^{(1)}_{a_n}$ and $\mathrm{SWG}^{(2)}_{\mathrm{C}_n}$, so that

$$H_n = H^{(1)}_n + H^{(2)}_n. \tag{4.18}$$

Because of the fact that at time C_n we have found the shortest-weight path, we have that

$$\big(H^{(1)}_n, H^{(2)}_n\big) \stackrel{d}{=} \big(G^{(1)}_{a_n+1} - 1, G^{(2)}_{\mathrm{C}_n}\big), \tag{4.19}$$

where $\{G^{(1)}_m\}_{m=1}^\infty$ and $\{G^{(2)}_m\}_{m=1}^\infty$ are copies of the process in (4.4), which are *conditioned on drawing a real stub*. Indeed, at the time of the connecting edge, a uniform (real) stub of $\mathrm{SWG}^{(2)}_m$ is drawn, and it is paired to a uniform (real) stub of $\mathrm{SWG}^{(1)}_{a_n}$. The number of hops in $\mathrm{SWG}^{(1)}_{a_n}$ to the end of the attached edge is therefore equal in distribution to $G^{(1)}_{a_n+1}$ conditioned on drawing a real stub. The -1 in (4.19) arises since the connecting edge is counted twice in $G^{(1)}_{a_n+1} + G^{(2)}_{\mathrm{C}_n}$. The processes $\{G^{(1)}_m\}_{m=1}^\infty$ and $\{G^{(2)}_m\}_{m=1}^\infty$ are conditionally independent given the realizations of $\{B^{(i)}_m\}_{m=2}^n$.

Further, because of the way the weight of the potential connecting edges has been distributed over the two stubs out of which the connecting edge is comprised, we have that

$$W_n = T^{(1)}_{a_n} + T^{(2)}_{\mathrm{C}_n}, \tag{4.20}$$

where $\{T_m^{(1)}\}_{m=1}^{\infty}$ and $\{T_m^{(2)}\}_{m=1}^{\infty}$ are two copies of the process $\{T_m\}_{m=1}^{\infty}$ in (4.6), again conditioned on drawing a real stub. Indeed, to see (4.20), we note that the weight of the connecting edge is equal to the sum of weights of its two stubs. Therefore, the weight of the shortest weight path is equal to the sum of the weight within $\mathrm{SWG}_{a_n}^{(1)}$, which is equal to $T_{a_n}^{(1)}$, and the weight within $\mathrm{SWG}_{\mathrm{C}_n}^{(2)}$, which is equal to $T_{\mathrm{C}_n}^{(2)}$.

In the distributions in (4.19) and (4.20) above, we always condition on drawing a real stub. Since we shall show that this occurs w.h.p., this conditioning plays a minor role.

We shall now intuitively explain why the leading order asymptotics of C_n is given by a_n where a_n is defined in (4.11). For this, we must know how many allowed stubs there are, that is, we must determine how many stubs there are incident to the union of the two SWGs at any time. Recall that $S_m^{(i)}$ denotes the number of allowed stubs in the SWG from vertex i at time m. The total number of allowed stubs incident to $\mathrm{SWG}_{a_n}^{(1)}$ is $S_{a_n}^{(1)}$, while the number incident to $\mathrm{SWG}_m^{(2)}$ is equal to $S_m^{(2)}$, and where

$$S_m^{(i)} = D_i + \sum_{l=2}^{m} (B_l^{(i)} - 1). \tag{4.21}$$

We also write $\mathrm{Art}_m = \mathrm{Art}_{a_n}^{(1)} \cup \mathrm{Art}_m^{(2)}$.

Conditionally on $\mathrm{SWG}_{a_n}^{(1)}$ and $\{(S_l^{(2)}, \mathrm{Art}_l^{(2)})\}_{l=1}^{m-1}$ and L_n, and assuming that $|\mathrm{Art}_m|$, m and S_m satisfy appropriate bounds, we obtain

$$\mathbb{P}(\mathrm{C}_n = m | \mathrm{C}_n > m-1) \approx \frac{S_{a_n}^{(1)}}{L_n}. \tag{4.22}$$

When $\tau \in (2, 3)$ and (3.8) holds, then $S_l^{(i)}/l^{1/(\tau-2)}$ can be expected to converge in distribution to a stable random variable with parameter $\tau - 2$, while, for $\tau > 3$, $S_l^{(i)}/l$ converges in probability to $\nu - 1$, where ν is defined in (3.3). We can combine these two statements by saying that $S_l^{(i)}/l^{1/(\tau\wedge 3-2)}$ converges in distribution. Note that the typical size a_n of C_n is such that, uniformly in n, $\mathbb{P}(\mathrm{C}_n \in [a_n, 2a_n])$ remains in $(\varepsilon, 1-\varepsilon)$, for some $\varepsilon \in (0, \frac{1}{2})$, which is the case when

$$\begin{aligned} \mathbb{P}(\mathrm{C}_n \in [a_n, 2a_n]) &= \sum_{m=a_n}^{2a_n} \mathbb{P}(\mathrm{C}_n = m | \mathrm{C}_n > m-1) \mathbb{P}(\mathrm{C}_n > m-1) \\ &\in (\varepsilon, 1-\varepsilon) \end{aligned} \tag{4.23}$$

uniformly as $n \to \infty$. By the above discussion, and for $a_n \le m \le 2a_n$, we have $\mathbb{P}(\mathrm{C}_n = m | \mathrm{C}_n > m-1) = \Theta(m^{1/(\tau\wedge 3-2)}/n) = \Theta(a_n^{1/(\tau\wedge 3-2)}/n)$, and $\mathbb{P}(\mathrm{C}_n > m-1) = \Theta(1)$. Then we arrive at

$$\mathbb{P}(\mathrm{C}_n \in [a_n, 2a_n]) = \Theta\big(a_n a_n^{1/(\tau\wedge 3-2)}/n\big), \tag{4.24}$$

which remains uniformly positive and bounded for a_n defined in (4.11). In turn, this suggests that

$$\mathrm{C}_n/a_n \xrightarrow{d} M \tag{4.25}$$

for some limiting random variable M.

We now discuss what happens when (2.2) holds for some $\tau \in (2,3)$, but (3.8) fails. In this case, there exists a slowly varying function $n \mapsto \ell(n)$ such that $S_l^{(i)}/(\ell(l)l^{1/(\tau-2)})$ converges in distribution. Then following the above argument shows that the right-hand side (r.h.s.) of (4.24) is replaced by $\Theta(a_n a_n^{1/(\tau-2)}\ell(a_n)/n)$ which remains uniformly positive and bounded for a_n satisfying $a_n^{(\tau-1)/(\tau-2)} \times \ell(a_n) = n$. By Bingham, Goldie and Teugels [7], Theorem 1.5.12, there exists a solution a_n to the above equation which satisfies that it is regularly varying with exponent $(\tau-2)/(\tau-1)$, so that

$$a_n = n^{(\tau-2)/(\tau-1)}\ell^*(n) \tag{4.26}$$

for some slowly varying function $n \mapsto \ell^*(n)$ which depends only on the distribution function F.

In the following proposition, we shall state the necessary result on C_n that we shall need in the remainder of the proof. In its statement, we shall use the symbol $o_{\mathbb{P}}(b_n)$ to denote a random variable X_n which satisfies that $X_n/b_n \xrightarrow{\mathbb{P}} 0$.

PROPOSITION 4.9 (The time to connection). *As $n \to \infty$, under the conditions of Theorems 3.1 and 3.2 respectively, and with a_n as in (4.11),*

$$\log \mathrm{C}_n - \log a_n = o_{\mathbb{P}}\big(\sqrt{\log n}\big). \tag{4.27}$$

Furthermore, for $i \in \{1,2\}$, and with $\beta \geq 1$ as in Proposition 4.3,

$$\left(\frac{G^{(1)}_{a_n+1} - \beta \log a_n}{\sqrt{\beta \log a_n}}, \frac{G^{(2)}_{\mathrm{C}_n} - \beta \log a_n}{\sqrt{\beta \log a_n}}\right) \xrightarrow{d} (Z_1, Z_2), \tag{4.28}$$

where Z_1, Z_2 are two independent standard normal random variables. Moreover, with γ as in Proposition 4.3, there exist random variables X_1, X_2 such that

$$\big(T^{(1)}_{a_n} - \gamma \log a_n, T^{(2)}_{\mathrm{C}_n} - \gamma \log a_n\big) \xrightarrow{d} (X_1, X_2). \tag{4.29}$$

We note that the main result in (4.28) is not a simple consequence of (4.27) and Proposition 4.6. The reason is that C_n is a *random variable*, which a priori depends on $(G^{(1)}_{a_n+1}, G^{(2)}_m)$ for $m \geq 0$. Indeed, the connecting edge is formed out of two stubs which are not artificial, and thus the choice of stubs is not completely uniform. However, since there are only few artificial stubs, we can extend the proof of Proposition 4.6 to this case. Proposition 4.9 shall be proved in Section 7.

4.5. *The completion of the proof.* By the analysis in Section 4.4, we know the distribution of the sizes of the SWGs at the time when the connecting edge appears. By Proposition 4.9, we know the number of edges and their weights used in the paths leading to the two vertices of the connecting edge together with its fluctuations. In the final step, we need to combine these results by averaging both over the *randomness* of the time when the connecting edge appears (which is a random variable), as well as over the number of edges in the shortest weight path when we know the time the connecting edge appears. Note that by (4.19) and Proposition 4.9, we have, with Z_1, Z_2 denoting independent standard normal random variables, and with $Z = (Z_1 + Z_2)/\sqrt{2}$, which is again standard normal,

$$
\begin{aligned}
H_n &\stackrel{d}{=} G^{(1)}_{a_n+1} + G^{(2)}_{\mathrm{C}_n} - 1 \\
&= 2\beta \log a_n + Z_1\sqrt{\beta \log a_n} + Z_2\sqrt{\beta \log a_n} + o_{\mathbb{P}}(\sqrt{\log n}) \qquad (4.30)\\
&= 2\beta \log a_n + Z\sqrt{2\beta \log a_n} + o_{\mathbb{P}}(\sqrt{\log n}).
\end{aligned}
$$

Finally, by (4.11), this gives (3.4) and (3.9) with

$$\alpha = \lim_{n\to\infty} \frac{2\beta \log a_n}{\log n}, \qquad (4.31)$$

which equals $\alpha = \nu/(\nu-1)$, when $\tau > 3$, since $\beta = \nu/(\nu-1)$ and $\frac{\log a_n}{\log n} = 1/2$, and $\alpha = 2(\tau-2)/(\tau-1)$, when $\tau \in (2,3)$, since $\beta = 1$ and $\frac{\log a_n}{\log n} = (\tau-2)/(\tau-1)$. This completes the proof for the hopcount.

In the description of α in (4.31), we note that when a_n contains a slowly varying function for $\tau \in (2,3)$ as in (4.26), then the result in Theorem 3.2 remains valid with $\alpha \log n$ replaced by

$$2\log a_n = \frac{2(\tau-2)}{\tau-1}\log n + 2\log \ell^*(n). \qquad (4.32)$$

For the weight of the minimal path, we make use of (4.20) and (4.29) to obtain in a similar way that

$$W_n - 2\gamma \log a_n \xrightarrow{d} X_1 + X_2. \qquad (4.33)$$

This completes the proof for the weight of the shortest path.

5. Proof of Proposition 4.3.

5.1. *Proof of Proposition* 4.3(a). We start by proving the statement for $\tau \in (2,3)$. Observe that, in this context, $d_i = B_i$, and, by (4.3), $B_1 + \cdots + B_i = S_i + i - 1$, so that the sequence $B_j/(S_i + i - 1)$, for j satisfying $1 \le j \le i$, is exchangeable for each $i \ge 1$. Therefore, we define

$$\hat{G}_m = \sum_{i=1}^{m} \hat{I}_i, \qquad \text{where } \mathbb{P}(\hat{I}_i = 1 | \{B_i\}_{i=1}^{\infty}) = \frac{B_i}{S_i + i - 1}. \qquad (5.1)$$

Thus, $\hat{I}_i$ is, conditionally on $\{B_i\}_{i=1}^{\infty}$, stochastically dominated by I_i, for each i, which, since the sequences $\{\hat{I}_i\}_{i=1}^{\infty}$ and $\{I_i\}_{i=1}^{\infty}$, conditionally on $\{B_i\}_{i=1}^{\infty}$, each have independent components, implies that $\hat{G}_m$ is stochastically dominated by G_m. We take $\hat{G}_m$ and G_m in such a way that $\hat{G}_m \le G_m$ a.s. Then, by the Markov inequality, for $\kappa_m > 0$,

$$\begin{aligned}\mathbb{P}(|G_m - \hat{G}_m| \ge \kappa_m) &\le \kappa_m^{-1}\mathbb{E}[|G_m - \hat{G}_m|] = \kappa_m^{-1}\mathbb{E}[G_m - \hat{G}_m]\\ &= \kappa_m^{-1}\sum_{i=1}^{m}\mathbb{E}\biggl[\frac{B_i(i-1)}{S_i(S_i+i-1)}\biggr] \qquad (5.2)\\ &= \kappa_m^{-1}\sum_{i=1}^{m}\frac{i-1}{i}\mathbb{E}[1/S_i],\end{aligned}$$

where, in the second equality, we used the exchangeability of $B_j/(S_i+i-1)$, $1 \le j \le i$. We will now show that

$$\sum_{i=1}^{\infty}\mathbb{E}[1/S_i] < \infty, \qquad (5.3)$$

so that for any $\kappa_m \to \infty$, we have that $\mathbb{P}(|G_m - \hat{G}_m| \le \kappa_m) \to 1$. We can then conclude that the CLT for G_m follows from the one for $\hat{G}_m$. By Deijfen et al. [14], (3.12) for $s = 1$, for $\tau \in (2,3)$ and using that $S_i = B_1 + \cdots + B_i - (i-1)$, where $\mathbb{P}(B_1 > k) = k^{2-\tau}L(k)$, there exists a slowly varying function $i \mapsto l(i)$ such that $\mathbb{E}[1/S_i] \le cl(i)i^{-1/(\tau-2)}$. When $\tau \in (2,3)$, we have that $1/(\tau-2) > 1$, so that (5.3) follows.

We now turn to the CLT for $\hat{G}_m$. Observe from the exchangeability of $B_j/(S_i + i-1)$, for $1 \le j \le i$, that for $i_1 < i_2 < \cdots < i_k$,

$$\begin{aligned}\mathbb{P}(\hat{I}_{i_1} = \cdots = \hat{I}_{i_k} = 1) &= \mathbb{E}\biggl[\prod_{l=1}^{k}\frac{B_{i_l}}{S_{i_l}+i_l-1}\biggr]\\ &= \mathbb{E}\biggl[\frac{B_{i_1}}{S_{i_1}+i_1-1}\prod_{l=2}^{k}\frac{B_{i_l}}{S_{i_l}+i_l-1}\biggr] \qquad (5.4)\\ &= \frac{1}{i_1}\mathbb{E}\biggl[\prod_{l=2}^{k}\frac{B_{i_l}}{S_{i_l}+i_l-1}\biggr] = \cdots = \prod_{l=1}^{k}\frac{1}{i_l},\end{aligned}$$

where we used that since $B_1 + \cdots + B_j = S_j + j - 1$,

$$\begin{aligned}\mathbb{E}\biggl[\frac{B_{i_1}}{S_{i_1}+i_1-1}\prod_{l=2}^{k}\frac{B_{i_l}}{S_{i_l}+i_l-1}\biggr] &= \frac{1}{i_1}\sum_{i=1}^{i_1}\mathbb{E}\biggl[\frac{B_i}{S_{i_1}+i_1-1}\prod_{l=2}^{k}\frac{B_{i_l}}{S_{i_l}+i_l-1}\biggr]\\ &= \frac{1}{i_1}\mathbb{E}\biggl[\prod_{l=2}^{k}\frac{B_{i_l}}{S_{i_l}+i_l-1}\biggr].\end{aligned}$$

Since $\hat{I}_{i_1}, \dots, \hat{I}_{i_k}$ are indicators this implies that $\hat{I}_{i_1}, \dots, \hat{I}_{i_k}$ are independent. Thus $\hat{G}_m$ has the same distribution as $\sum_{i=1}^m J_i$ where $\{J_i\}_{i=1}^\infty$ are *independent* Bernoulli random variables with $\mathbb{P}(J_i = 1) = 1/i$. It is a standard consequence of the Lindeberg–Lévy–Feller CLT that $(\sum_{i=1}^m J_i - \log m)/\sqrt{\log m}$ is asymptotically standard normally distributed.

REMARK 5.1 (Extension to exchangeable setting). Note that the CLT for G_m remains valid when (i) the random variables $\{B_i\}_{i=1}^m$ are *exchangeable*, with the same marginal distribution as in the i.i.d. case, and (ii) $\sum_{i=1}^m \mathbb{E}[1/S_i] = o(\sqrt{\log m})$.

The approach for $\tau > 3$ is different from that of $\tau \in (2,3)$. For $\tau \in (2,3)$, we coupled G_m to $\hat{G}_m$ and proved that $\hat{G}_m$ satisfies the CLT with the correct norming constants. For $\tau > 3$, the case we consider now, we first apply a *conditional* CLT, using the Lindeberg–Lévy–Feller condition, stating that, conditionally on $B_1, B_2, \dots$ satisfying

$$\lim_{m\to\infty} \sum_{j=1}^m \frac{B_j}{S_j}\left(1 - \frac{B_j}{S_j}\right) = \infty, \tag{5.5}$$

we have that

$$\frac{G_m - \sum_{j=1}^m B_j/S_j}{(\sum_{j=1}^m B_j/S_j(1 - B_j/S_j))^{1/2}} \xrightarrow{d} Z, \tag{5.6}$$

where Z is standard normal. The result (5.6) is also contained in [10].

Since $\nu = \mathbb{E}[B_j] > 1$ and $\mathbb{E}[B_j^a] < \infty$, for any $a < \tau - 2$, it is not hard to see that the random variable $\sum_{j=1}^\infty B_j^2/S_j^2$ is positive and has finite first moment, so that for $m \to \infty$,

$$\sum_{j=1}^m B_j^2/S_j^2 = O_{\mathbb{P}}(1), \tag{5.7}$$

where $O_{\mathbb{P}}(b_m)$ denotes a sequence of random variables X_m for which $|X_m|/b_m$ is tight.

We claim that

$$\sum_{j=1}^m B_j/S_j - \frac{\nu}{\nu - 1}\log m = o_{\mathbb{P}}(\sqrt{\log m}). \tag{5.8}$$

Obviously, (5.6), (5.7) and (5.8) imply Proposition 4.3(a) when $\tau > 3$.

In order to prove (5.8), we split

$$\begin{aligned} &\sum_{j=1}^m B_j/S_j - \frac{\nu}{\nu-1}\log m \\ &\quad = \left(\sum_{j=1}^m (B_j - 1)/S_j - \log m\right) + \left(\sum_{j=1}^m 1/S_j - \frac{1}{\nu - 1}\log m\right), \end{aligned} \tag{5.9}$$

and shall prove that each of these two terms on the r.h.s. of (5.9) is $o_{\mathbb{P}}(\sqrt{\log m})$. For the first term, we note from the strong law of large numbers that

$$\sum_{j=1}^{m} \log\left(\frac{S_j}{S_{j-1}}\right) = \log S_m - \log S_0 = \log m + O_{\mathbb{P}}(1). \tag{5.10}$$

Also, since $-\log(1-x) = x + O(x^2)$, we have that

$$\begin{aligned} \sum_{j=1}^{m} \log(S_j/S_{j-1}) &= -\sum_{j=1}^{m} \log\bigl(1 - (B_j - 1)/S_j\bigr) \\ &= \sum_{j=1}^{m} (B_j - 1)/S_j + O\left(\sum_{j=1}^{m} (B_j - 1)^2/S_j^2\right). \end{aligned} \tag{5.11}$$

Again, as in (5.7), for $m \to \infty$,

$$\sum_{j=1}^{m} (B_j - 1)^2/S_j^2 = O_{\mathbb{P}}(1), \tag{5.12}$$

so that

$$\sum_{j=1}^{m} (B_j - 1)/S_j - \log m = O_{\mathbb{P}}(1). \tag{5.13}$$

In order to study the second term on the right-hand side of (5.9), we shall prove a slightly stronger result than necessary, since we shall also use this later on. Indeed, we shall show that there exists a random variable Y such that

$$\sum_{j=1}^{m} 1/S_j - \frac{1}{\nu - 1} \log m \xrightarrow{\text{a.s.}} Y. \tag{5.14}$$

In fact, the proof of (5.14) is a consequence of [2], Theorem 1, since $\mathbb{E}[(B_i - 1)\log(B_i - 1)] < \infty$ for $\tau > 3$. We decided to give a separate proof of (5.14) which can be easily adapted to the exchangeable case.

To prove (5.14), we write

$$\sum_{j=1}^{m} 1/S_j - \frac{1}{\nu - 1} \log m = \sum_{j=1}^{m} \frac{(\nu - 1)j - S_j}{S_j(\nu - 1)j} + O_{\mathbb{P}}(1), \tag{5.15}$$

so that in order to prove (5.14), it suffices to prove that, uniformly in $m \geq 1$,

$$\sum_{j=1}^{m} \frac{|S_j - (\nu - 1)j|}{S_j(\nu - 1)j} < \infty \qquad \text{a.s.} \tag{5.16}$$

Thus, if we further make use of the fact that $S_j \geq \eta j$ except for at most finitely many j (see also Lemma A.4 below), then we obtain that

$$\left|\sum_{j=1}^{m} \frac{1}{S_j} - \frac{1}{\nu - 1}\log m\right| \leq \sum_{j=1}^{m} \frac{|S_j - (\nu - 1)j|}{S_j(\nu - 1)j} + O_{\mathbb{P}}(1) \leq C\sum_{j=1}^{m} \frac{|S_j^*|}{j^2}, \tag{5.17}$$

where $S_j^* = S_j - \mathbb{E}[S_j]$, since $\mathbb{E}[S_j] = (\nu - 1)j + 1$. We now take the expectation, and conclude that for any $a > 1$, Jensen's inequality for the convex function $x \mapsto x^a$, yields

$$\mathbb{E}[|S_j^*|] \leq \mathbb{E}[|S_j^*|^a]^{1/a}. \tag{5.18}$$

To bound the last expectation, we will use a consequence of the Marcinkiewicz–Zygmund inequality (see, e.g., [19], Corollary 8.2, page 152). Taking $1 < a < \tau - 2$, we have that $\mathbb{E}[|B_1|^a] < \infty$, since $\tau > 3$, so that

$$\mathbb{E}\left[\sum_{j=1}^{m} \frac{|S_j^*|}{j^2}\right] \leq \sum_{j=1}^{m} \frac{\mathbb{E}[|S_j^*|^a]^{1/a}}{j^2} \leq \sum_{j=1}^{m} \frac{c_a^{1/a}\mathbb{E}[|B_1|^a]^{1/a}}{j^{2-1/a}} < \infty. \tag{5.19}$$

This completes the proof of (5.14).

REMARK 5.2 (Discussion of exchangeable setting). When the random variables $\{B_i\}_{i=1}^m$ are *exchangeable*, with the same marginal distribution as in the i.i.d. case, and with $\tau > 3$, we note that to prove a CLT for G_m, it suffices to prove (5.7) and (5.8). The proof of (5.8) contains two steps, namely, (5.13) and (5.16). For the CLT to hold, we in fact only need that the involved quantities are $o_{\mathbb{P}}(\sqrt{\log m})$, rather than $O_{\mathbb{P}}(1)$. For this, we note that:

(a) the argument to prove (5.13) is rather flexible, and shows that if (i) $\log S_m/m = o_{\mathbb{P}}(\sqrt{\log m})$ and if (ii) the condition in (5.7) is satisfied with $O_{\mathbb{P}}(1)$ replaced by $o_{\mathbb{P}}(\sqrt{\log m})$, then (5.13) follows with $O_{\mathbb{P}}(1)$ replaced by $o_{\mathbb{P}}(\sqrt{\log m})$;

(b) for the proof of (5.16) we will make use of stochastic domination and show that each of the stochastic bounds will satisfy (5.16) with $O_{\mathbb{P}}(1)$ replaced by $o_{\mathbb{P}}(\sqrt{\log m})$ (compare Lemma A.8).

5.2. *Proof of Proposition* 4.3(b). We again start by proving the result for $\tau \in (2,3)$. It follows from (4.6) and the independence of $\{E_i\}_{i\geq 1}$ and $\{S_i\}_{i\geq 1}$ that, for the proof of (4.9), it is sufficient to show that

$$\sum_{i=1}^{\infty} \mathbb{E}[1/S_i] < \infty, \tag{5.20}$$

which holds due to (5.3).

6. Proof of Proposition 4.6. In this section, we extend the proof of Proposition 4.3 to the setting where the random vector $\{B_i\}_{i=2}^m$ is *not* i.i.d., but rather corresponds to the vector of forward degrees in the CM.

In the proofs for the CM, we shall always *condition* on the fact that the vertices under consideration are part of the giant component. As discussed below (3.3), in this case, the giant component has size $n - o(n)$, so that each vertex is in the giant component w.h.p. Further, this conditioning ensures that $S_j > 0$ for every $j = o(n)$.

We recall that the set up of the random variables involved in Proposition 4.6 is given in (4.12) and (4.13). The random variable R_m, defined in (4.10), is the first time t the SWG_t consists of $m+1$ real vertices.

LEMMA 6.1 (Exchangeability of $\{B_{R_m}\}_{m=1}^{n-1}$). *Conditionally on $\{D_i\}_{i=1}^n$, the sequence of random variables $\{B_{R_m}\}_{m=1}^{n-1}$ is exchangeable, with marginal probability distribution*

$$\mathbb{P}_n(B_{R_1} = j) = \sum_{i=2}^{n} \frac{(j+1)\mathbb{1}_{\{D_i = j+1\}}}{L_n - D_1}, \tag{6.1}$$

where $\mathbb{P}_n$ denotes the conditional probability given $\{D_i\}_{i=1}^n$.

PROOF. We note that, by definition, the random variables $\{B_{R_m}\}_{m=1}^{n-1}$ are equal to the forward degrees (where we recall that the forward degree is equal to the degree minus 1) of a vertex chosen from all vertices unequal to 1, where a vertex i is chosen with probability proportional to its degree, that is, vertex $i \in \{2, \dots, n\}$ is chosen with probability $P_i = D_i/(L_n - D_1)$. Let $K_2, \dots, K_n$ be the vertices chosen; then the sequence $K_2, \dots, K_n$ has the same distribution as draws with probabilities $\{P_i\}_{i=2}^n$ *without replacement*. Obviously, the sequence $(K_2, \dots, K_n)$ is exchangeable, so that the sequence $\{B_{R_m}\}_{m=1}^{n-1}$, which can be identified as $B_{R_m} = D_{K_{m+1}} - 1$, inherits this property. $\square$

We continue with the proof of Proposition 4.6. By Lemma 6.1, the sequence $\{B_j\}_{j=2}^m$ is exchangeable, when we condition on $|\mathrm{Art}_j| = 0$ for all $j \le m$. Also, $|\mathrm{Art}_j| = 0$ for all $j \le m$ holds precisely when $R_m = m$. In Lemma A.1 in Appendix A, the probability that $R_{m_n} = m_n$, for an appropriately chosen m_n, is investigated. We shall make crucial use of this lemma to study G_{m_n}.

PROOF OF PROPOSITION 4.6. Recall that by definition $\log(\overline{m}_n/a_n) = o(\sqrt{\log n})$. Then, we split, for some $\underline{m}_n$ such that $\log(a_n/\underline{m}_n) = o(\sqrt{\log n})$,

$$G_{\overline{m}_n} = \widetilde{G}_{\underline{m}_n} + [G_{\overline{m}_n} - \widetilde{G}_{\underline{m}_n}], \tag{6.2}$$

where $\widetilde{G}_{\underline{m}_n}$ has the same marginal distribution as $G_{\underline{m}_n}$, but also satisfies that $\widetilde{G}_{\underline{m}_n} \le G_{\overline{m}_n}$, a.s. By construction, the sequence of random variables $m \mapsto G_m$

is stochastically increasing, so that this is possible by the fact that random variable A is stochastically smaller than B if and only if we can couple A and B to $(\hat{A}, \hat{B})$ such that $\hat{A} \leq \hat{B}$, a.s.

Denote by $\mathcal{A}_m = \{R_m = m\}$ the event that the first artificial stub is chosen after time m. Then, by Lemma A.1, we have that $\mathbb{P}(\mathcal{A}^c_{\underline{m}_n}) = o(1)$. Thus, by intersecting with $\mathcal{A}_{\underline{m}_n}$ and its complement, and then using the Markov inequality, we find for any $c_n = o(\sqrt{\log n})$,

$$\begin{aligned}\mathbb{P}(|G_{\overline{m}_n} - \widetilde{G}_{\underline{m}_n}| \geq c_n) &\leq \frac{1}{c_n}\mathbb{E}[|G_{\overline{m}_n} - \widetilde{G}_{\underline{m}_n}|\mathbb{1}_{\mathcal{A}_{\underline{m}_n}}] + o(1)\\ &= \frac{1}{c_n}\mathbb{E}\big[[G_{\overline{m}_n} - \widetilde{G}_{\underline{m}_n}]\mathbb{1}_{\mathcal{A}_{\underline{m}_n}}\big] + o(1)\\ &= \frac{1}{c_n}\sum_{i=\underline{m}_n+1}^{\overline{m}_n}\mathbb{E}\bigg[\frac{B_i}{S_i}\mathbb{1}_{\mathcal{A}_{\underline{m}_n}}\bigg] + o(1).\end{aligned} \tag{6.3}$$

We claim that

$$\sum_{i=\underline{m}_n+1}^{\overline{m}_n}\mathbb{E}\bigg[\frac{B_i}{S_i}\mathbb{1}_{\mathcal{A}_{\underline{m}_n}}\bigg] = o\big(\sqrt{\log n}\big). \tag{6.4}$$

Indeed, to see (6.4), we note that $B_i = 0$, when $i \neq R_j$ for some j. Also, when $\mathcal{A}_{\underline{m}_n}$ occurs, then $R_{\underline{m}_n} = \underline{m}_n$. Thus, using also that $R_m \geq m$, so that $R_i \leq \overline{m}_n$ implies that $i \leq \overline{m}_n$,

$$\begin{aligned}\sum_{i=\underline{m}_n+1}^{\overline{m}_n}\mathbb{E}\bigg[\frac{B_i}{S_i}\mathbb{1}_{\mathcal{A}_{\underline{m}_n}}\bigg] &\leq \sum_{i=\underline{m}_n+1}^{\overline{m}_n}\mathbb{E}\bigg[\frac{B_{R_i}}{S_{R_i}}\mathbb{1}_{\{\underline{m}_n+1\leq R_i\leq\overline{m}_n\}}\bigg]\\ &\leq \sum_{i=\underline{m}_n+1}^{\overline{m}_n}\frac{1}{i-1}\mathbb{E}\bigg[\frac{S_{R_i}+R_i}{S_{R_i}}\mathbb{1}_{\{\underline{m}_n+1\leq R_i\leq\overline{m}_n\}}\bigg],\end{aligned} \tag{6.5}$$

the latter following from the exchangeability of $\{B_{R_i}\}_{i=2}^{n-1}$, because

$$S_{R_i} = D_1 + \sum_{j=2}^{R_i}(B_j - 1) = D_1 + \sum_{j=2}^{i} B_{R_j} - (R_i - 1),$$

so that

$$\sum_{j=2}^{i} B_{R_j} = S_{R_i} - D_1 + R_i - 1 \leq S_{R_i} + R_i. \tag{6.6}$$

In Lemma A.2 of the Appendix A we show that there exists a constant C such that for $i \leq \overline{m}_n$,

$$\mathbb{E}\bigg[\frac{S_{R_i}+R_i}{S_{R_i}}\mathbb{1}_{\{\underline{m}_n+1\leq R_i\leq\overline{m}_n\}}\bigg] \leq C, \tag{6.7}$$

so that, for an appropriate chosen c_n with $\log(\overline{m}_n/\underline{m}_n)/c_n \to 0$,

$$\mathbb{P}(|G_{\overline{m}_n} - \widetilde{G}_{\underline{m}_n}| \geq c_n) \leq \frac{C}{c_n} \sum_{i=\underline{m}_n+1}^{\overline{m}_n} \frac{1}{i-1} \leq \frac{C \log(\overline{m}_n/\underline{m}_n)}{c_n} = o(1), \tag{6.8}$$

since $\log(\overline{m}_n/\underline{m}_n) = o(\sqrt{\log n})$. Thus, the CLT for $G_{\overline{m}_n}$ follows from the one for $\widetilde{G}_{\underline{m}_n}$ which, since the marginal of $\widetilde{G}_{\underline{m}_n}$ is the same as the one of $G_{\underline{m}_n}$, follows from the one for $G_{\underline{m}_n}$. By Lemma A.1, we further have that with high probability, there has not been any artificial stub up to time $\underline{m}_n$, so that, again with high probability, $\{B_m\}_{m=2}^{\underline{m}_n} = \{B_{R_m}\}_{m=2}^{\underline{m}_n}$, the latter, by Lemma 6.1, being an exchangeable sequence.

We next adapt the proof of Proposition 4.3 to exchangeable sequences under certain conditions. We start with $\tau \in (2,3)$, which is relatively the more simple case. Recall the definition of G_m in (4.13). We define, for $i \geq 2$,

$$\hat{S}_i = \sum_{j=2}^{i} B_j = S_i + i - 1 - D_1. \tag{6.9}$$

Similarly to the proof of Proposition 4.3 we now introduce

$$\hat{G}_m = 1 + \sum_{i=2}^{m} \hat{I}_i, \qquad \text{where } \mathbb{P}(\hat{I}_i = 1 | \{B_i\}_{i=2}^m) = B_i/\hat{S}_i, \qquad 2 \leq i \leq m. \tag{6.10}$$

Let $\hat{Q}_i = B_i/\hat{S}_i$, $Q_i = B_i/S_i$. Then, by a standard coupling argument, we can couple $\hat{I}_i$ and I_i in such a way that $\mathbb{P}(\hat{I}_i \neq I_i | \{B_i\}_{i=2}^m) = |\hat{Q}_i - Q_i|$.

The CLT for $\hat{G}_m$ follows because, also in the exchangeable setting, $\hat{I}_2, \ldots, \hat{I}_m$ are independent and, similar to (5.2),

$$\begin{aligned}
&\mathbb{P}(|G_m - \hat{G}_m| \geq \kappa_n) \\
&\quad \leq \kappa_n^{-1} \mathbb{E}[|G_m - \hat{G}_m|] \leq \kappa_n^{-1} \mathbb{E}\Bigg[\sum_{i=1}^{m} |I_i - \hat{I}_i|\Bigg] \\
&\quad = \kappa_n^{-1} \sum_{i=2}^{m} \mathbb{E}[|\hat{Q}_i - Q_i|] \\
&\quad = \kappa_n^{-1} \sum_{i=2}^{m} \mathbb{E}\bigg[B_i \frac{|S_i - \hat{S}_i|}{S_i \hat{S}_i}\bigg] \\
&\quad \leq \kappa_n^{-1} \sum_{i=2}^{m} \mathbb{E}\bigg[B_i \frac{D_1 + (i-1)}{S_i \hat{S}_i}\bigg] \\
&\quad = \kappa_n^{-1} \sum_{i=2}^{m} \frac{1}{i-1} \mathbb{E}\bigg[\frac{D_1 + (i-1)}{S_i}\bigg]
\end{aligned} \tag{6.11}$$

$$= \kappa_n^{-1} \sum_{i=2}^{m} \left(\mathbb{E}[1/S_i] + \frac{1}{i-1} \mathbb{E}[D_1/S_i] \right)$$

$$\leq \kappa_n^{-1} \sum_{i=2}^{m} \left(\mathbb{E}[1/(S_i - D_1 + 2)] + \frac{1}{i-1} \mathbb{E}[D_1/(S_i - D_1 + 2)] \right),$$

where we used that $D_1 \geq 2$ a.s. We take $m = \underline{m}_n$, as discussed above. Since D_1 is independent of $S_i - D_1 + 2$ for $i \geq 2$ and $\mathbb{E}[D_1] < \infty$, we obtain the CLT for $G_{\underline{m}_n}$ from the one for $\hat{G}_{\underline{m}_n}$ when, for $\tau \in (2, 3)$,

$$\text{(6.12)} \quad \sum_{i=1}^{\underline{m}_n} \mathbb{E}[1/\Sigma_i] = O(1), \qquad \text{where } \Sigma_i = 1 + \sum_{j=2}^{i} (B_j - 1), \qquad i \geq 1.$$

In Lemma A.2 of the Appendix A we will prove that for $\tau \in (2, 3)$, the statement (6.12) holds. The CLT for $G_{R_{\overline{m}_n}}$ follows in an identical way.

We continue by studying the distribution of T_m and $\widetilde{T}_m$, for $\tau \in (2, 3)$. We recall that $T_m = \sum_{i=1}^{m} E_i/S_i$ [see (4.6)]. In the proof of Proposition 4.3(b) for $\tau \in (2, 3)$, we have made crucial use of (5.20), which is now replaced by (6.12). We split

$$\text{(6.13)} \qquad T_m = \sum_{i=1}^{m} E_i/S_i = \sum_{i=1}^{n^\rho} E_i/S_i + \sum_{i>n^\rho}^{m} E_i/S_i.$$

The mean of the second term converges to 0 for each $\rho > 0$ by Lemma A.2, while the first term is by Proposition 4.7 w.h.p. equal to $\sum_{i=1}^{n^\rho} E_i/S_i^{(\mathrm{ind})}$, where $S_i^{(\mathrm{ind})} = \sum_{j=1}^{i} B_j^{(\mathrm{ind})}$, and where $B_1^{(\mathrm{ind})} = D_1$, while $\{B_i^{(\mathrm{ind})}\}_{i=2}^{n^\rho}$ is an i.i.d. sequence of random variables with probability mass function g given in (2.3), which is independent from D_1. Thus, noting that also $\sum_{i>n^\rho}^{m} E_i/S_i^{(\mathrm{ind})} \xrightarrow{\mathbb{P}} 0$, and with

$$\text{(6.14)} \qquad X = \sum_{i=1}^{\infty} E_i/S_i^{(\mathrm{ind})},$$

we obtain that $T_m \xrightarrow{d} X$. The random variable X has the interpretation of the explosion time of the continuous-time branching process, where the degree of the root has distribution function F, while the degrees of the other vertices is an i.i.d. sequence of random variables with probability mass function g given in (2.3). This completes the proof of Proposition 4.6 for $\tau \in (2, 3)$, and we turn to the case $\tau > 3$.

For $\tau > 3$, we follow the steps in the proof of Proposition 4.3(a) for $\tau > 3$ as closely as possible. Again, we apply a conditional CLT as in (5.6), to obtain the CLT when (5.5) holds. From Lemma A.5 we conclude that (6.7) also holds when $\tau > 3$. Hence, as before, we may assume by Lemma A.1, that w.h.p., there has not been any artificial stub up to time $\underline{m}_n$, so that, again w.h.p., $\{B_m\}_{m=2}^{\underline{m}_n} = \{B_{R_m}\}_{m=2}^{\underline{m}_n}$, the latter, by Lemma 6.1, being an exchangeable sequence. For the exchangeable sequence $\{B_m\}_{m=2}^{m_n}$ we will then show that

$$\sum_{j=2}^{m_n} B_j^2/S_j^2 = O_{\mathbb{P}}(1). \tag{6.15}$$

The statement (6.15) is proven in Lemma A.6.

As in the proof of Proposition 4.3(a), the claim that

$$\sum_{j=2}^{\underline{m}_n} B_j/S_j - \frac{\nu}{\nu-1}\log \underline{m}_n = o_{\mathbb{P}}\big(\sqrt{\log \underline{m}_n}\big) \tag{6.16}$$

is sufficient for the CLT when $\tau > 3$. Moreover, we have shown in Remark 5.2 that (6.16) is satisfied when

$$\log\left(S_{\underline{m}_n}/\underline{m}_n\right) = o_{\mathbb{P}}\big(\sqrt{\log \underline{m}_n}\big) \tag{6.17}$$

and

$$\sum_{j=1}^{\underline{m}_n} \frac{S_j - (\nu-1)j}{S_j(\nu-1)j} = o_{\mathbb{P}}\big(\sqrt{\log \underline{m}_n}\big). \tag{6.18}$$

The proofs of (6.17) and (6.18) are given in Lemmas A.7 and A.8 of Appendix A, respectively. Again, the proof for $G_{R_{\overline{m}_n}}$ is identical. □

For the results for T_m and $\widetilde{T}_m$ for $\tau > 3$, we refer to Appendix C.

7. Proof of Proposition 4.9. In this section, we prove Proposition 4.9. We start by proving that $\log \mathrm{C}_n/a_n = o_{\mathbb{P}}(\sqrt{\log n})$, where C_n is the time at which the connecting edge appears between the SWGs of vertices 1 and 2 [recall (4.17)], as stated in (4.27). As described in Section 4.4, we shall condition vertices 1 and 2 to be in the giant component, which occurs w.h.p. and guarantees that $S_m^{(i)} > 0$ for any $m = o(n)$ and $i \in \{1, 2\}$. After this, we complete the proof of (4.28)–(4.29) in the case where $\tau \in (2, 3)$, which turns out to be relatively simplest, followed by a proof of (4.28) for $\tau > 3$. The proof of (4.29) for $\tau > 3$, which is more delicate, is deferred to Appendix C.

We start by identifying the distribution of C_n. In order for $\mathrm{C}_n = m$ to occur, apart from further requirements, the minimal stub from $\mathrm{SWG}_m^{(2)}$ must be real, that is, it may not be artificial. This occurs with probability equal to $1 - |\mathrm{Art}_m^{(2)}|/S_m^{(2)}$.

By Construction 4.4, the number of allowed stubs incident to $\mathrm{SWG}_m^{(2)}$ equals $S_m^{(2)}$, so the number of real stubs equals $S_m^{(2)} - |\mathrm{Art}_m^{(2)}|$. Similarly, the number of allowed stubs incident to $\mathrm{SWG}_{a_n}^{(1)}$ equals $S_m^{(2)}$, so the number of real stubs equals $S_{a_n}^{(1)} - |\mathrm{Art}_{a_n}^{(1)}|$. Further, the number of free stubs equals $|\mathrm{FS}_m| = L_n - a_n - m -$

$S_m + |\mathrm{Art}_m|$, where we recall that $S_m = S_{a_n}^{(1)} + S_m^{(2)}$ and $\mathrm{Art}_m = \mathrm{Art}_{a_n}^{(1)} \cup \mathrm{Art}_m^{(2)}$, and is hence bounded above by L_n and below by $L_n - a_n - m - S_m$. When the minimal-weight stub is indeed real, then it must be attached to one of the real allowed stubs incident to $\mathrm{SWG}_{a_n}^{(1)}$, which occurs with conditional probability given $\mathrm{SWG}_m^{(1,2)}$ and L_n equal to

$$\frac{S_{a_n}^{(1)} - |\mathrm{Art}_{a_n}^{(1)}|}{L_n - a_n - m - S_m + |\mathrm{Art}_m|}. \tag{7.1}$$

Thus, in order to prove Proposition 4.9, it suffices to investigate the limiting behavior of L_n, $S_m^{(i)}$ and $|\mathrm{Art}_m|$. By the law of large numbers, we known that $L_n - \mu n = o_{\mathbb{P}}(n)$ as $n \to \infty$. To study $S_m^{(i)}$ and $|\mathrm{Art}_m|$, we shall make use of results from [34, 35]. Note that we can write $S_m^{(i)} = D_i + B_2^{(i)} + \cdots + B_m^{(i)} - (m-1)$, where $\{B_m^{(i)}\}_{m=2}^{\infty}$ are close to being independent. See [34], Lemma A.2.8, for stochastic domination results on $\{B_m^{(i)}\}_{m=2}^{\infty}$ and their sums in terms of i.i.d. random variables, which can be applied in the case of $\tau > 3$. See [35], Lemma A.1.4, for bounds on tail probabilities for sums and maxima of random variables with certain tail properties.

The next step to be performed is to give criteria in terms of the processes $S_m^{(i)}$ which guarantee that the estimates in Proposition 4.9 follow. We shall start by proving that with high probability $\mathrm{C}_n \geq \underline{m}_n$, where $\underline{m}_n = \varepsilon_n a_n$, where $\varepsilon_n \downarrow 0$. This proof makes use of, and is quite similar to, the proof of Lemma A.1 given in Appendix A.

LEMMA 7.1 (Lower bound on time to connection). *Let $\underline{m}_n/a_n = o(1)$. Then*

$$\mathbb{P}(\mathrm{C}_n \leq \underline{m}_n) = o(1). \tag{7.2}$$

PROOF. Denote

$$\mathcal{E}_n = \{S_{a_n}^{(1)} \leq a_n M_n\}, \tag{7.3}$$

where $M_n = \eta_n^{-1}$ for $\tau > 3$, while $M_n = \eta_n^{-1} n^{(3-\tau)/(\tau-1)}$ for $\tau \in (2,3)$, and where $\eta_n \downarrow 0$ sufficiently slowly. Then, by (A.59) for $\tau > 3$ and (A.40) for $\tau \in (2,3)$,

$$\mathbb{P}(\mathcal{E}_n^c) = o(1), \tag{7.4}$$

since $\underline{m}_n = o(a_n)$. By the law of total probability,

$$\mathbb{P}(\mathrm{C}_n \leq \underline{m}_n) \leq \mathbb{P}(\mathcal{E}_n^c) + \sum_{m=1}^{\underline{m}_n} \mathbb{P}(\{\mathrm{C}_n = m\} \cap \mathcal{E}_n | \mathrm{C}_n > m-1) \mathbb{P}(\mathrm{C}_n > m-1). \tag{7.5}$$

Then, we make use of (7.1) and (7.4), to arrive at

$$\mathbb{P}(\mathrm{C}_n \leq \underline{m}_n) \leq \sum_{m=1}^{\underline{m}_n} \mathbb{E}\left[\mathbb{1}_{\mathcal{E}_n} \frac{S_{a_n}^{(1)}}{L_n - m - S_m}\right] + o(1). \tag{7.6}$$

As in the proof of Lemma A.1, we have that $m \leq \underline{m}_n = o(n)$ and $S_m = o(n)$, while $L_n \geq n$. Thus, (7.6) can be simplified to

$$\mathbb{P}(\mathrm{C}_n \leq \underline{m}_n) \leq \frac{1+o(1)}{n}\underline{m}_n \mathbb{E}[\mathbb{1}_{\mathcal{E}_n} S_{a_n}^{(1)}] + o(1) \leq \frac{1+o(1)}{n\eta_n} a_n \underline{m}_n. \tag{7.7}$$

When choosing $\eta_n \downarrow 0$ sufficiently slowly, for example as $\eta_n = \sqrt{\underline{m}_n/a_n}$, we obtain that $\mathbb{P}(\mathrm{C}_n \leq \underline{m}_n) = o(1)$ whenever $\underline{m}_n = o(a_n)$. $\square$

We next state an upper bound on C_n.

LEMMA 7.2 (Upper bound on time to connection). *Let* $\overline{m}_n/a_n \to \infty$, *then*,

$$\mathbb{P}(\mathrm{C}_n > \overline{m}_n) = o(1). \tag{7.8}$$

PROOF. We start by giving an explicit formula for $\mathbb{P}(\mathrm{C}_n > m)$. As before, $\mathbb{Q}_n^{(m)}$ is the conditional distribution given $\mathrm{SWG}_m^{(1,2)}$ and $\{D_i\}_{i=1}^n$. Then, by Lemma B.1,

$$\mathbb{P}(\mathrm{C}_n > m) = \mathbb{E}\left[\prod_{j=1}^{m} \mathbb{Q}_n^{(j)}(\mathrm{C}_n > j | \mathrm{C}_n > j-1)\right]. \tag{7.9}$$

Equation (7.9) is identical in spirit to [34], Lemma 4.1, where a similar identity was used for the graph distance in the CM. Now, for any sequence $\varepsilon_n \to 0$, let

$$\mathcal{B}_n = \left\{\frac{c\overline{m}_n}{n}|\mathrm{Art}_{a_n}^{(1)}| + \frac{c}{n} S_{a_n}^{(1)} \sum_{m=1}^{\overline{m}_n} \frac{|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}} \leq \varepsilon_n\right\}. \tag{7.10}$$

By Lemma B.3, the two terms appearing in the definition of $\mathcal{B}_n$ in (7.10) converge to zero in probability, so that $\mathbb{P}(\mathcal{B}_n) = 1 - o(1)$ for some $\varepsilon_n \to 0$. Then we bound

$$\mathbb{P}(\mathrm{C}_n > m) \leq \mathbb{E}\left[\mathbb{1}_{\mathcal{B}_n} \prod_{j=1}^{m} \mathbb{Q}_n^{(j)}(\mathrm{C}_n > j | \mathrm{C}_n > j-1)\right] + \mathbb{P}(\mathcal{B}_n^c). \tag{7.11}$$

We continue by noticing that according to (7.1),

$$\mathbb{Q}_n^{(m)}(\mathrm{C}_n = m | \mathrm{C}_n > m-1) = \frac{S_{a_n}^{(1)} - |\mathrm{Art}_{a_n}^{(1)}|}{|\mathrm{FS}_m|}\left(1 - \frac{|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}}\right), \tag{7.12}$$

where $|\mathrm{FS}_m|$ is the number of real free stubs which is available at time m. Combining (7.11) and (7.12) we arrive at

$$\mathbb{P}(\mathrm{C}_n > \overline{m}_n) = \mathbb{E}\left[\mathbb{1}_{\mathcal{B}_n} \prod_{m=1}^{\overline{m}_n}\left(1 - \frac{S_{a_n}^{(1)} - |\mathrm{Art}_{a_n}^{(1)}|}{|\mathrm{FS}_m|}\left(1 - \frac{|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}}\right)\right)\right] + o(1). \tag{7.13}$$

Since $|\mathrm{FS}_m| \le L_n \le n/c$, w.h.p., for some $c > 0$, and using that $1 - x \le \mathrm{e}^{-x}$, we can further bound

$$\begin{aligned}\mathbb{P}(\mathrm{C}_n > \overline{m}_n) &\le \mathbb{E}\Bigg[\mathbb{1}_{\mathcal{B}_n} \exp\Bigg\{-\frac{c}{n}\big(S_{a_n}^{(1)} - |\mathrm{Art}_{a_n}^{(1)}|\big) \sum_{m=1}^{\overline{m}_n}\Bigg(1 - \frac{|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}}\Bigg)\Bigg\}\Bigg] + o(1) \\ &\le \mathbb{E}\bigg[\mathbb{1}_{\mathcal{B}_n} \exp\bigg\{-\frac{c\overline{m}_n}{n} S_{a_n}^{(1)}\bigg\}\bigg] + e_n + o(1),\end{aligned} \tag{7.14}$$

where

$$e_n = O\Bigg(\mathbb{E}\Bigg[\mathbb{1}_{\mathcal{B}_n}\Bigg(\frac{c\overline{m}_n}{n}|\mathrm{Art}_{a_n}^{(1)}| + \frac{c}{n} S_{a_n}^{(1)} \sum_{m=1}^{\overline{m}_n} \frac{|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}}\Bigg)\Bigg]\Bigg) = O(\varepsilon_n). \tag{7.15}$$

Hence,

$$\mathbb{P}(\mathrm{C}_n > \overline{m}_n) \le \mathbb{E}\bigg[\exp\bigg\{-\frac{c\overline{m}_n}{n} S_{a_n}^{(1)}\bigg\}\bigg] + o(1). \tag{7.16}$$

When $\tau > 3$, by Lemma A.4 in the Appendix A, we have that, w.h.p., and for some $\eta > 0$,

$$S_{a_n} \ge \eta a_n, \tag{7.17}$$

so that

$$\mathbb{P}(\mathrm{C}_n > \overline{m}_n) \le \exp\bigg\{-\frac{c\eta a_n \overline{m}_n}{n}\bigg\} + o(1) = o(1), \tag{7.18}$$

as long as $\overline{m}_n a_n/n = \overline{m}_n/\sqrt{n} \to \infty$. For $\tau \in (2, 3)$, by (A.39) in Lemma A.3, and using that $n^{1/(\tau-1)}/n = 1/a_n$, we have for every $\varepsilon_n \to 0$,

$$\mathbb{P}(\mathrm{C}_n > \overline{m}_n) \le \exp\bigg\{-\frac{c\overline{m}_n \varepsilon_n}{a_n}\bigg\} + o(1) = o(1), \tag{7.19}$$

whenever $\varepsilon_n \overline{m}_n/a_n \to \infty$. By adjusting ε_n, it is hence sufficient to assume that $\overline{m}_n/a_n \to \infty$. □

Lemmas 7.1 and 7.2 complete the proof of (4.27) in Proposition 4.9. We next continue with the proof of (4.28) in Proposition 4.9. We start by showing that $\mathbb{P}(\mathrm{C}_n = 0) = o(1)$. Indeed, $\mathrm{C}_n = 0$ happens precisely when $2 \in \mathrm{SWG}_{a_n}^{(1)}$, which, by exchangeability, occurs with probability at most $a_n/n = o(1)$.

For $\mathrm{C}_n \ge 1$, we note that at time C_n, we draw a real stub. Consider the pair $(G_{a_n+1}^{(1)}, G_{\mathrm{C}_n}^{(2)})$ conditionally on $\{\mathrm{C}_n = m\}$ for a certain m. The event $\{\mathrm{C}_n = m\}$ is *equal* to the event that the last chosen stub in $\mathrm{SWG}_m^{(2)}$ is paired to a stub incident to $\mathrm{SWG}_{a_n}^{(1)}$, while this is not the case for all previously chosen stubs. For $j = 1, \dots, m$, and $i \in \{1, 2\}$, denote by $I_j^{(i)}$ the jth real vertex added to $\mathrm{SWG}^{(i)}$, and denote by $V_m^{(i)}$ the number of real vertices in $\mathrm{SWG}_m^{(i)}$. Then, for $m \ge 1$, the event

$\{\mathrm{C}_n = m\}$ is *equal* to the event that the last chosen stub in $\mathrm{SWG}_m^{(2)}$ is paired to a stub incident to $\mathrm{SWG}_{a_n}^{(1)}$, and

$$\{I_j^{(1)}\}_{j=1}^{V_{a_n}^{(1)}} \cap \{I_j^{(2)}\}_{j=1}^{V_m^{(2)}} = \varnothing. \tag{7.20}$$

As a result, conditionally on $\{\mathrm{C}_n = m\}$ and $V_{a_n}^{(1)} = k_1$, $V_m^{(2)} = k_2$, the vector consisting of both $\{B_{R_{j_1}}^{(1)}\}_{j_1=1}^{k_1}$ and $\{B_{R_{j_2}}^{(2)}\}_{j_2=1}^{k_2}$ is an exchangeable vector, with law is equal to that of $k_1 + k_2$ draws from $\{D_i - 1\}_{i=3}^{n}$ without replacement, where, for $i \in [n] \setminus \{1, 2\}$, $D_i - 1$ is drawn with probability equal to $D_i/(L_n - D_1 - D_2)$. The above explains the role of the random stopping time C_n.

We continue by discussing the limiting distributions of $(H_n^{(1)}, H_n^{(2)})$ in order to prove (4.28). For this, we note that if we condition on $\{\mathrm{C}_n = m\}$ for some m and on $\mathrm{SWG}_m^{(1,2)}$, then, by (4.19) $(H_n^{(1)}, H_n^{(2)}) \stackrel{d}{=} (G_{a_n+1}^{(1)}, G_m^{(2)})$, where the *conditional* distribution of $(G_{a_n+1}^{(1)}, G_m^{(2)})$ is as two *independent* copies of G as described in (4.4), where $\{d_j\}_{j=1}^{a_n}$ in (4.4) is given by $d_1 = D_1$ and $d_j = B_j^{(1)}$, $j \geq 2$, while, $H_n^{(2)} = G_m^{(2)}$, where $d_1 = D_2$ and $d_j = B_j^{(2)}$, $j \geq 2$. Here, we make use of the fact that $H_n^{(1)}$ is the distance from vertex 1 to the vertex to which the paired stub is connected to, which has the same distribution as the distance from vertex 1 to the vertex which has been added at time $a_n + 1$, minus 1, since the paired stub is again a *uniform* stub (conditioned to be real).

Thus, any possible dependence of $(H_n^{(1)}, H_n^{(2)})$ arises through the dependence of the vectors $\{B_j^{(1)}\}_{j=2}^{\infty}$ and $\{B_j^{(2)}\}_{j=2}^{\infty}$. However, the proof of Proposition 4.6 shows that certain weak dependency of $\{B_j^{(1)}\}_{j=2}^{\infty}$ and $\{B_j^{(2)}\}_{j=2}^{\infty}$ is allowed.

We start by completing the proof for $\tau \in (2, 3)$ which is the more simple one. Recall the split in (5.1), which was fundamental in showing the CLT for $\tau \in (2, 3)$. Indeed, let $\{\hat{I}_j^{(1)}\}_{j=1}^{\infty}$ and $\{\hat{I}_j^{(2)}\}_{j=1}^{\infty}$ be two sequences of indicators, with $\hat{I}_1^{(1)} = \hat{I}_1^{(2)} = 1$, which are, conditionally on $\{B_j^{(1)}\}_{j=2}^{\infty}$ and $\{B_j^{(2)}\}_{j=2}^{\infty}$, independent with, for $i \in \{1, 2\}$,

$$\mathbb{P}\big(\hat{I}_j^{(i)} = 1 | \{B_j^{(i)}\}_{j=2}^{\infty}\big) = B_j^{(i)}/\big(S_j^{(i)} + j - 1 - D_i\big). \tag{7.21}$$

Then, the argument in (5.4) can be straightforwardly adapted to show that the unconditional distributions of $\{\hat{I}_j^{(1)}\}_{j=2}^{\infty}$ and $\{\hat{I}_j^{(2)}\}_{j=2}^{\infty}$ are that of two independent sequences $\{J_j^{(1)}\}_{j=2}^{\infty}$ and $\{J_j^{(2)}\}_{j=2}^{\infty}$ with $\mathbb{P}(J_j^{(i)} = 1) = 1/(j-1)$. Thus, by the independence, we immediately obtain that since $\mathrm{C}_n \to \infty$ with $\log(\mathrm{C}_n/a_n) = o_{\mathbb{P}}(\sqrt{\log n})$,

$$\left(\frac{H_n^{(1)} - \beta \log a_n}{\sqrt{\beta \log a_n}}, \frac{H_n^{(2)} - \beta \log a_n}{\sqrt{\beta \log a_n}}\right) \xrightarrow{d} (Z_1, Z_2). \tag{7.22}$$

The argument to show that, since $\mathrm{C}_n \le \overline{m}_n$, $(H_n^{(1)}, H_n^{(2)})$ can be well approximated by $(G_{\underline{m}_n}^{(1)}, G_{\underline{m}_n}^{(2)})$ [recall (6.2)] only depends on the marginals of $(H_n^{(1)}, H_n^{(2)})$, and thus remains valid verbatim. We conclude that (4.28) holds.

We next prove (4.29) for $\tau \in (2,3)$. For this, we again use Proposition 4.7 to note that the forward degrees $\{B_j\}_{j=3}^{n^\rho}$ can be coupled to i.i.d. random variables $\{B_j^{(\mathrm{ind})}\}_{j=3}^{n^\rho}$, which are independent from $B_1 = D_1, B_2 = D_2$. Then we can follow the proof of Proposition 4.6(b) for $\tau \in (2,3)$ verbatim, to obtain that $(T_{a_n}^{(1)}, T_{\mathrm{C}_n}^{(2)}) \xrightarrow{d} (X_1, X_2)$, where X_1, X_2 are two independent copies of X in (6.14). This completes the proof of Proposition 4.9 when $\tau \in (2,3)$.

We proceed with the proof of Proposition 4.9 when $\tau > 3$ by studying $(H_n^{(1)}, H_n^{(2)})$. We follow the proof of Proposition 4.6(a), paying particular attention to the claimed independence of the limits (Z_1, Z_2) in (4.28). The proof of Proposition 4.6(a) is based on a *conditional* CLT, applying the Lindeberg–Lévy–Feller condition. Thus, the conditional limits (Z_1, Z_2) of

$$\left(\frac{H_n^{(1)} - \sum_{j=2}^{a_n} B_j^{(1)}/S_j^{(1)}}{(\sum_{j=2}^{a_n}(B_j^{(1)}/S_j^{(1)})(1 - B_j^{(1)}/S_j^{(1)}))^{1/2}}, \frac{H_n^{(2)} - \sum_{j=2}^{\mathrm{C}_n} B_j^{(2)}/S_j^{(2)}}{(\sum_{j=2}^{\mathrm{C}_n}(B_j^{(2)}/S_j^{(2)})(1 - B_j^{(2)}/S_j^{(2)}))^{1/2}}\right) \tag{7.23}$$

are clearly independent. The proof then continues by showing that the asymptotic mean and variance can be replaced by $\beta \log n$, which is a computation based on the marginals $\{B_j^{(1)}\}_{j=2}^{\infty}$ and $\{B_j^{(2)}\}_{j=2}^{\infty}$ only, and, thus, these results carry over verbatim, when we further make use of the fact that, w.h.p., $\mathrm{C}_n \in [\underline{m}_n, \overline{m}_n]$ for any $\underline{m}_n, \overline{m}_n$ such that $\log(\overline{m}_n/\underline{m}_n) = o(\sqrt{\log n})$. This completes the proof of (4.28) for $\tau > 3$. The proof of (4.29) for $\tau > 3$ is a bit more involved, and is deferred to Appendix C.

APPENDIX A: AUXILIARY LEMMAS FOR CLTS IN CM

In this appendix, we denote by $B_1 = D_1$, the degree of vertex 1 and $B_2, \ldots, B_m$, $m < n$, the forward degrees of the shortest weight graph SWG_m. The forward degree B_k is chosen recursively from the set FS_k, the set of free stubs at time k. Further we denote by

$$S_k = D_1 + \sum_{j=2}^{k}(B_j - 1),$$

the number of allowed stubs at time k. As before the random variable R_m denotes the first time that the shortest path graph from vertex 1 contains $m+1$ real vertices. Consequently

$$B_{R_2}, \ldots, B_{R_m},$$

$m < n$, can be seen as a sample without replacement from the degrees

$$D_2 - 1, D_3 - 1, \dots, D_n - 1.$$

A.1. The first artificial stub. We often can and will replace the sample $B_2, \dots, B_{\underline{m}_n}$, by the sample $B_{R_2}, \dots, B_{R_{\underline{m}_n}}$. The two samples have, w.h.p., the same distribution if the first artificial stub appears after time $\underline{m}_n$. This will be the content of our first lemma.

LEMMA A.1 (The first artificial stub). *Let* $\underline{m}_n/a_n \to 0$. *Then*,

$$\mathbb{P}(R_{\underline{m}_n} > \underline{m}_n) = o(1). \tag{A.1}$$

PROOF. For the event $\{R_{\underline{m}_n} > \underline{m}_n\}$ to happen it is mandatory that for some $m \le \underline{m}_n$, we have $R_m > m$, while $R_{m-1} = m - 1$. Hence

$$\mathbb{P}_n(R_{\underline{m}_n} > \underline{m}_n) = \sum_{m=2}^{\underline{m}_n} \mathbb{P}_n(R_m > m, R_{m-1} = m - 1). \tag{A.2}$$

Now, when $R_m > m$, $R_{m-1} = m - 1$, one of the S_{m-1} stubs incident to SWG_{m-1} has been drawn, so that

$$\mathbb{P}_n(R_m > m, R_{m-1} = m - 1) = \mathbb{E}_n\left[\frac{S_{m-1}}{L_n - S_{m-1} - 2m} \mathbb{1}_{\{R_{m-1}=m-1\}}\right]. \tag{A.3}$$

Since $\underline{m}_n = o(n)$, we claim that, with high probability, $S_{m-1} = o(n)$. Indeed, the maximal degree is $O_{\mathbb{P}}(n^{1/(\tau-1)})$, so that, for $m \le \underline{m}_n$,

$$S_m \le O_{\mathbb{P}}\big(mn^{1/(\tau-1)}\big) \le O_{\mathbb{P}}\big(\underline{m}_n n^{1/(\tau-1)}\big) = o_{\mathbb{P}}(n), \tag{A.4}$$

since, for $\tau > 3$, $a_n = n^{1/2}$ and $n^{1/(\tau-1)} = o(n^{1/2})$, while, for $\tau \in (2, 3)$, $a_n = n^{(\tau-2)/(\tau-1)}$, so that $\underline{m}_n n^{1/(\tau-1)} = o(n)$. Moreover, $L_n \ge n$, so that

$$\mathbb{P}_n(R_m > m, R_{m-1} = m - 1) \le \frac{C}{n} \mathbb{E}_n\big[S_{m-1} \mathbb{1}_{\{R_{m-1}=m-1\}}\big]. \tag{A.5}$$

By the remark preceding this lemma, since $R_{m-1} = m - 1$, we have that $S_{m-1} = D_1 + \sum_{j=2}^{m-1} (B_{R_j} - 1)$, so that, by Lemma 6.1,

$$\mathbb{P}_n(R_m > m, R_{m-1} = m - 1) \le \frac{C}{n} D_1 + \frac{C(m-2)}{n} \mathbb{E}_n[B_{R_2}]. \tag{A.6}$$

The first term converges to 0 in probability, while the expectation in the second term, by (6.1), equals

$$\mathbb{E}_n[B_{R_2}] = \sum_{i=2}^{n} \frac{D_i(D_i - 1)}{L_n - D_1}. \tag{A.7}$$

When $\tau > 3$, this has a bounded expectation, so that, for $a_n = \sqrt{n}$,

$$\mathbb{P}(R_{\underline{m}_n} > \underline{m}_n) \le \sum_{m=2}^{\underline{m}_n} \frac{C}{n}\mathbb{E}[D_1] + \sum_{m=2}^{\underline{m}_n} \frac{C(m-2)}{n}\mathbb{E}[B_{R_2}] \le C\frac{\underline{m}_n^2}{n} \to 0. \tag{A.8}$$

When $\tau \in (2,3)$, however, then $\mathbb{E}[D_i^2] = \infty$, and we need to be a bit more careful. In this case, we obtain from (A.6) that

$$\mathbb{P}_n(R_{\underline{m}_n} > \underline{m}_n) \le C\frac{\underline{m}_n^2}{n}\mathbb{E}_n[B_{R_2}]. \tag{A.9}$$

From (A.7), and since $L_n - D_1 \ge n-1$,

$$\mathbb{E}_n[B_{R_2}] \le \frac{C}{n-1}\sum_{i=2}^{n} D_i(D_i-1) \le \frac{C}{n-1}\sum_{i=2}^{n} D_i^2. \tag{A.10}$$

From (3.8), we obtain that $x^{(\tau-1)/2}\mathbb{P}(D_i^2 > x) \in [c_1, c_2]$ uniformly in $x \ge 0$, and since $D_1, D_2, \dots, D_n$ is i.i.d., we can conclude that $n^{-2/(\tau-1)}\sum_{i=2}^n D_i^2$ converges to a proper random variable. Hence, since $a_n/\underline{m}_n \to \infty$ we obtain, w.h.p.,

$$\mathbb{E}_n[B_{R_2}] \le \frac{a_n}{\underline{m}_n} n^{2/(\tau-1)-1} = \frac{a_n}{\underline{m}_n} n^{(3-\tau)/(\tau-1)}. \tag{A.11}$$

Combining (A.9) and (A.11), and using that $a_n = n^{(\tau-2)/(\tau-1)}$ we obtain that, w.h.p.,

$$\begin{aligned} \mathbb{P}_n(R_{\underline{m}_n} > \underline{m}_n) &\le C a_n \underline{m}_n n^{2/(\tau-1)-1} \\ &= C\frac{\underline{m}_n}{a_n} n^{2(\tau-2)/(\tau-1)+(3-\tau)/(\tau-1)-1} = C\frac{\underline{m}_n}{a_n} = o_{\mathbb{P}}(1). \end{aligned} \tag{A.12}$$

This proves the claim. □

A.2. Coupling the forward degrees to an i.i.d. sequence: Proposition 4.7. We will now prove Proposition 4.7. To this end, we denote the order statistics of the degrees by

$$D_{(1)} \le D_{(2)} \le \cdots \le D_{(n)}. \tag{A.13}$$

Let $m_n \to \infty$ and consider the i.i.d. random variables $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_{m_n}$, where $\underline{X}_i$ is taken *with replacement* from the stubs

$$D_{(1)} - 1, D_{(2)} - 1, \dots, D_{(n-m_n)} - 1, \tag{A.14}$$

that is, we sample *with replacement* from the original forward degrees $D_1 - 1, D_2 - 1, \dots, D_n - 1$, where the m_n largest degrees are discarded. Similarly, we consider the i.i.d. random variables $\overline{X}_1, \overline{X}_2, \dots, \overline{X}_{m_n}$, where $\overline{X}_i$ is taken *with re-*

placement from the stubs

$$D_{(m_n+1)} - 1, D_{(m_n+2)} - 1, \dots, D_{(n)} - 1, \tag{A.15}$$

that is, we sample *with replacement* from the original forward degrees $D_1 - 1, D_2 - 1, \dots, D_n - 1$, where the m_n smallest degrees are discarded. Then, obviously, we obtain a stochastic ordering $\underline{X}_i \le_{st} B_i \le_{st} \overline{X}_i$, compare [34], Lemma A.2.8. As a consequence, we can couple $\{B_i\}_{i=2}^{m_n}$ to m_n i.i.d. random variables $\{\underline{X}_i\}_{i=1}^{m_n-1}, \{\overline{X}_i\}_{i=1}^{m_n-1}$ such that, a.s.,

$$\underline{X}_{i-1} \le B_i \le \overline{X}_{i-1}. \tag{A.16}$$

The random variables $\{\underline{X}_i\}_{i=1}^{m_n-1}$, as well as $\{\overline{X}_i\}_{i=1}^{m_n-1}$ are i.i.d., but their distribution depends on m_n, since they are draws *with replacement* from $D_1 - 1, \dots, D_n - 1$ where the largest m_n, respectively smallest m_n, degrees have been removed [recall (A.14)]. Let the *total variation distance* between two probability mass functions p and q on $\mathbb{N}$ be given by

$$d_{\mathrm{TV}}(p,q) = \frac{1}{2}\sum_{k=0}^{\infty} |p_k - q_k|. \tag{A.17}$$

We shall show that, with $\underline{g}$ and $\overline{g}$, respectively, denoting the probability mass functions of $\underline{X}_i$ and $\overline{X}_i$, respectively, there exists $\rho' > 0$ such that w.h.p.

$$d_{\mathrm{TV}}\big(\underline{g}^{(n)}, g\big) \le n^{-\rho'}, \qquad d_{\mathrm{TV}}\big(\overline{g}^{(n)}, g\big) \le n^{-\rho'}. \tag{A.18}$$

This proves the claim for any $\rho < \rho'$, since (A.18) implies that $d_{\mathrm{TV}}(\underline{g}^{(n)}, \overline{g}^{(n)}) \le 2n^{-\rho'}$, so that we can couple $\{\underline{X}_i\}_{i=1}^{m_n-1}$ and $\{\overline{X}_i\}_{i=1}^{m_n-1}$ in such a way that $\mathbb{P}(\{\underline{X}_i\}_{i=1}^{m_n} = \{\overline{X}_i\}_{i=1}^{m_n}) \le 2m_n n^{-\rho'} = o(1)$, when $m_n = n^{\rho}$ with $\rho' < \rho$. In particular, this yields that we can couple $\{B_i\}_{i=2}^{m_n}$ to $\{\underline{X}_i\}_{i=1}^{m_n-1}$ in such a way that $\{B_i\}_{i=2}^{m_n} = \{\underline{X}_i\}_{i=1}^{m_n-1}$ w.h.p. Then, again from (A.18), we can couple $\{\underline{X}_i\}_{i=1}^{m_n-1}$ to a sequence of i.i.d. random variables $\{B_i^{(\mathrm{ind})}\}_{i=1}^{m_n-1}$ such that $\{\underline{X}_i\}_{i=1}^{m_n-1} = \{B_i^{(\mathrm{ind})}\}_{i=1}^{m_n-1}$ w.h.p. Thus, (A.18) completes the proof of Proposition 4.7.

To prove (A.18), we bound

$$d_{\mathrm{TV}}\big(\underline{g}^{(n)}, g\big) \le d_{\mathrm{TV}}\big(\underline{g}^{(n)}, g^{(n)}\big) + d_{\mathrm{TV}}\big(g^{(n)}, g\big), \tag{A.19}$$

and a similar identity holds for $d_{\mathrm{TV}}(\overline{g}^{(n)}, g)$, where

$$g_k^{(n)} = \frac{1}{L_n}\sum_{j=1}^{n} (k+1)\mathbb{1}_{\{D_j = k+1\}}. \tag{A.20}$$

In [34], (A.1.11), it is shown that there exists $\alpha_2, \beta_2 > 0$ such that

$$\mathbb{P}\big(d_{\mathrm{TV}}\big(g^{(n)}, g\big) \ge n^{-\alpha_2}\big) \le n^{-\beta_2}. \tag{A.21}$$

Thus, we are left to investigate $d_{\mathrm{TV}}(\underline{g}^{(n)}, g^{(n)})$ and $d_{\mathrm{TV}}(\overline{g}^{(n)}, g^{(n)})$. We bound

$$\begin{aligned}
d_{\mathrm{TV}}\big(\underline{g}^{(n)}, g^{(n)}\big) &= \frac{1}{2}\sum_{k=0}^{\infty} \big|\underline{g}_k^{(n)} - g_k^{(n)}\big| \\
&\le \sum_{k=0}^{\infty}(k+1)\bigg(\frac{1}{\underline{L}_n} - \frac{1}{L_n}\bigg)\sum_{j=1}^{n-m_n} \mathbb{1}_{\{D_j=k+1\}} \\
&\quad + \sum_{k=0}^{\infty}(k+1)\frac{1}{\underline{L}_n}\sum_{j=n-m_n+1}^{n} \mathbb{1}_{\{D_{(j)}=k+1\}} \qquad \text{(A.22)}\\
&\le \bigg(\frac{L_n - \underline{L}_n}{L_n\underline{L}_n}\bigg)\sum_{j=1}^{n-m_n} D_{(j)} + \frac{1}{\underline{L}_n}\sum_{j=n-m_n+1}^{n} D_{(j)} \\
&\le 2\bigg(\frac{L_n - \underline{L}_n}{\underline{L}_n}\bigg) = \frac{2}{\underline{L}_n}\sum_{j=n-m_n+1}^{n} D_{(j)},
\end{aligned}$$

where $\underline{L}_n = \sum_{j=1}^{n-m_n} D_{(j)}$. Define $b_n = \Theta(n/m_n)^{1/(\tau-1)}$. Then, from $1 - F(x) = x^{-(\tau-1)}L(x)$, and concentration results for the binomial distribution, we have, w.h.p., $D_{(n-m_n+1)} \ge b_n$, so that, w.h.p.,

$$\frac{L_n - \underline{L}_n}{\underline{L}_n} = \frac{1}{\underline{L}_n}\sum_{j=n-m_n+1}^{n} D_{(j)} \le \frac{1}{\underline{L}_n}\sum_{j=1}^{n} D_j \mathbb{1}_{\{D_j \ge b_n\}}. \qquad \text{(A.23)}$$

Now, in turn, by the Markov inequality,

$$\begin{aligned}
\mathbb{P}\bigg(\frac{1}{\underline{L}_n}\sum_{j=1}^{n} D_j \mathbb{1}_{\{D_j \ge b_n\}} \ge n^{\varepsilon} b_n^{2-\tau}\bigg) &\le n^{-\varepsilon} b_n^{\tau-2}\mathbb{E}\bigg[\frac{1}{\underline{L}_n}\sum_{j=1}^{n} D_j \mathbb{1}_{\{D_j \ge b_n\}}\bigg] \qquad \text{(A.24)}\\
&\le C n^{-\varepsilon},
\end{aligned}$$

so that

$$\mathbb{P}\big(d_{\mathrm{TV}}\big(\underline{g}^{(n)}, g^{(n)}\big) \ge n^{\varepsilon} b_n^{-(\tau-2)}\big) = o(1). \qquad \text{(A.25)}$$

Thus, w.h.p., $d_{\mathrm{TV}}(\underline{g}^{(n)}, g^{(n)}) \le n^{\varepsilon}(m_n/n)^{(\tau-2)/(\tau-1)}$, which proves (A.18) when we take $m_n = n^{\rho}$ and $\rho' = (1-\rho)(\tau-2)/(\tau-1) - \varepsilon > 0$. The upper bound for $d_{\mathrm{TV}}(\overline{g}^{(n)}, g^{(n)})$ can be treated similarly.

A.3. Auxiliary lemmas for $2 < \tau < 3$. In this section we treat some lemmas that complete the proof of Proposition 4.6(a) for $\tau \in (2, 3)$. In particular, we shall verify condition (ii) in Remark 5.1.

LEMMA A.2 (A bound on the expected value of $1/S_i$). *Fix* $\tau \in (2,3)$. *For* $\underline{m}_n, \overline{m}_n$ *such that* $\log(a_n/\underline{m}_n)$, $\log(\overline{m}_n/a_n) = o(\sqrt{\log n})$ *and for* b_n *such that* $b_n \to \infty$,

$$\text{(A.26)}\qquad \begin{aligned} &\text{(i)} && \sum_{i=1}^{\underline{m}_n} \mathbb{E}[1/\Sigma_i] = O(1),\\ &\text{(ii)} && \sum_{i=b_n}^{\underline{m}_n} \mathbb{E}[1/\Sigma_i] = o(1) \quad \textit{and}\\ &\text{(iii)} && \sup_{i \le \overline{m}_n} \mathbb{E}\big[(R_i/S_{R_i})\mathbb{1}_{\{\underline{m}_n+1 \le R_i \le \overline{m}_n\}}\big] < \infty. \end{aligned}$$

PROOF. Let $\underline{m}_n = o(a_n)$. Let

$$M_i = \max_{2 \le j \le i} (B_j - 1). \tag{A.27}$$

Then, we use that, for $1 \le i \le \underline{m}_n$,

$$\Sigma_i \equiv 1 + \sum_{j=2}^{i} (B_j - 1) \ge \max_{2 \le j \le i} (B_j - 1) - (i-2) = M_i - (i-2). \tag{A.28}$$

Fix $\delta > 0$ small, and split

$$\begin{aligned} \mathbb{E}[1/\Sigma_i] &\le \mathbb{E}\big[1/\Sigma_i \mathbb{1}_{\{\Sigma_i \le i^{1+\delta}\}}\big] + \mathbb{E}\big[1/\Sigma_i \mathbb{1}_{\{\Sigma_i > i^{1+\delta}\}}\big]\\ &\le \mathbb{P}(\Sigma_i \le i^{1+\delta}) + i^{-(1+\delta)}. \end{aligned} \tag{A.29}$$

Now, if $\Sigma_i \le i^{1+\delta}$, then $M_i \le i^{1+\delta} + i \le 2i^{1+\delta}$, and $\Sigma_j \le i^{1+\delta} + i \le 2i^{1+\delta}$ for all $j \le i$. As a result, for each $j \le i$, the conditional probability that $B_j - 1 > 2i^{1+\delta}$, given $\Sigma_{j-1} \le 2i^{1+\delta}$ and $\{D_s\}_{s=1}^n$ is at least

$$\begin{aligned} \frac{1}{L_n} \sum_{s=1}^{n} D_s \mathbb{1}_{\{D_s > 2i^{1+\delta}\}} &\ge 2i^{1+\delta} \sum_{s=1}^{n} \mathbb{1}_{\{D_s > 2i^{1+\delta}\}}/L_n\\ &= 2i^{1+\delta}\, \mathrm{BIN}\big(n, 1 - F(2i^{1+\delta})\big)/L_n. \end{aligned} \tag{A.30}$$

Further, by (3.8), for some $c > 0$, $n[1 - F(2i^{1+\delta})] \ge 2cni^{-(1+\delta)(\tau-1)}$, so that, for $i \le \underline{m}_n = o(n^{(\tau-2)/(\tau-1)})$, $ni^{-(1+\delta)(\tau-1)} \ge n^{\varepsilon}$ for some $\varepsilon > 0$. We shall use Azuma's inequality that states that for a binomial random variable $\mathrm{BIN}(N, p)$ with parameters N and p, and all $t > 0$,

$$\mathbb{P}\big(\mathrm{BIN}(N, p) \le Np - t\big) \le \exp\bigg\{-\frac{2t^2}{N}\bigg\}. \tag{A.31}$$

As a result,

$$\begin{aligned} &\mathbb{P}\Big(\mathrm{BIN}\big(n, 1 - F(2i^{1+\delta})\big) \le \mathbb{E}\big[\mathrm{BIN}\big(n, 1 - F(2i^{1+\delta})\big)\big]/2\Big)\\ &\qquad \le \mathrm{e}^{-n[1-F(2i^{1+\delta})]/2} \le \mathrm{e}^{-n^{\varepsilon}}, \end{aligned} \tag{A.32}$$

so that, with probability at least $1-\mathrm{e}^{-n^{\varepsilon}}$,

$$\frac{1}{L_n}\sum_{s=1}^{n} D_s \mathbb{1}_{\{D_s>2i^{1+\delta}\}} \geq c i^{-(1+\delta)(\tau-2)}. \tag{A.33}$$

Thus, the probability that in the first i trials, no vertex with degree at least $2i^{1+\delta}$ is chosen is bounded above by

$$\left(1-ci^{-(1+\delta)(\tau-2)}\right)^i + \mathrm{e}^{-n^{\varepsilon}} \leq \mathrm{e}^{-ci^{1-(1+\delta)(\tau-2)}} + \mathrm{e}^{-n^{\varepsilon}}, \tag{A.34}$$

where we used the inequality $1-x \leq \mathrm{e}^{-x}$, $x \geq 0$. Finally, take $\delta > 0$ so small that $1-(1+\delta)(\tau-2) > 0$; then we arrive at

$$\mathbb{E}[1/\Sigma_i] \leq i^{-(1+\delta)} + \mathrm{e}^{-ci^{1-(1+\delta)(\tau-2)}} + \mathrm{e}^{-n^{\varepsilon}}, \tag{A.35}$$

which, when summed over $i \leq \underline{m}_n$, is $O(1)$. This proves (i). For (ii), we note that, for any $b_n \to \infty$, the sum of the r.h.s. of (A.35) is $o(1)$. This proves (ii).

To prove (iii), we take $\log(a_n/\underline{m}_n), \log(\overline{m}_n/a_n) = o(\sqrt{\log n})$. We bound the expected value by

$$\overline{m}_n \mathbb{E}\left[(1/S_{R_i})\mathbb{1}_{\{\underline{m}_n+1\leq R_i\leq \overline{m}_n\}}\right].$$

For $\underline{m}_n + 1 \leq i \leq \overline{m}_n$,

$$S_i = D_1 + \sum_{j=2}^{i}(B_j-1) \geq 1 + \sum_{j=2}^{i}(B_j-1) = \Sigma_i, \tag{A.36}$$

and the above derived bound for the expectation $\mathbb{E}[1/\Sigma_i]$ remains valid for $\underline{m}_n + 1 \leq i \leq \overline{m}_n$, since also for $i \leq \overline{m}_n$, we have $ni^{-(1+\delta)(\tau+1)} \geq n^{\varepsilon}$; moreover since the r.h.s. of (A.35) is decreasing in i, we obtain

$$\mathbb{E}[1/\Sigma_i] \leq \underline{m}_n^{-(1+\delta)} + \mathrm{e}^{-c\underline{m}_n^{1-(1+\delta)(\tau-2)}} + \mathrm{e}^{-n^{\varepsilon}}. \tag{A.37}$$

Consequently,

$$\begin{aligned} &\overline{m}_n \mathbb{E}\left[(1/S_{R_i})\mathbb{1}_{\{\underline{m}_n+1\leq R_i\leq \overline{m}_n\}}\right] \\ &\quad\leq \overline{m}_n\left(\underline{m}_n^{-(1+\delta)} + \mathrm{e}^{-c\underline{m}_n^{1-(1+\delta)(\tau-2)}} + \mathrm{e}^{-n^{\varepsilon}}\right) = o(1), \end{aligned} \tag{A.38}$$

using that $\log(a_n/\underline{m}_n), \log(\overline{m}_n/a_n) = o(\sqrt{\log n})$. This proves (iii). $\square$

LEMMA A.3 (Bounds on $S_{\overline{m}_n}$). *Fix $\tau \in (2,3)$. Then, w.h.p., for every $\varepsilon_n \to 0$,*

$$S_{a_n} \geq \varepsilon_n n^{1/(\tau-1)}, \tag{A.39}$$

while, w.h.p., uniformly for all $m \leq \overline{m}_n$,

$$\mathbb{E}_n[S_m] \leq \varepsilon_n^{-1} m n^{(3-\tau)/(\tau-1)}. \tag{A.40}$$

PROOF. We prove (A.39) by noting that, by (A.33) and the fact that $\varepsilon_n \downarrow 0$,

$$\sum_{i=1}^{n} D_i \mathbb{1}_{\{D_i \geq \varepsilon_n n^{1/(\tau-1)}\}} \geq cn\big(\varepsilon_n n^{1/(\tau-1)}\big)^{-(\tau-2)} = (cn/a_n)\varepsilon_n^{-(\tau-2)}. \tag{A.41}$$

Therefore, the probability to choose none of these vertices with degree at least $\varepsilon_n n^{1/(\tau-1)}$ before time a_n is bounded by

$$\big(1 - c\varepsilon_n^{-(\tau-2)} n^{-(2-\tau)/(\tau-1)}\big)^{a_n} \leq \mathrm{e}^{-c\varepsilon_n^{-(\tau-2)}} = o(1) \tag{A.42}$$

for any $\varepsilon_n \downarrow 0$. In turn, this implies that, w.h.p., $S_{a_n} \geq \varepsilon_n n^{1/(\tau-1)} - a_n \geq \varepsilon_n n^{1/(\tau-1)}/2$, whenever ε_n is such that $\varepsilon_n n^{1/(\tau-1)} \geq 2a_n$.

To prove (A.40), we use that, w.h.p., $D_{(n)} \leq \varepsilon_n^{-1} n^{1/(\tau-1)}$ for any $\varepsilon_n \to 0$. Thus, w.h.p., using the inequality $L_n > n$,

$$\mathbb{E}_n[S_m] \leq m\mathbb{E}_n[B_2] \leq \frac{m}{n} \sum_{j=1}^{n} D_j(D_j - 1)\mathbb{1}_{\{D_j \leq \varepsilon_n^{-1} n^{1/(\tau-1)}\}}. \tag{A.43}$$

Thus, in order to prove the claimed uniform bound, it suffices to give a bound on the above sum that holds w.h.p. For this, the expected value of the sum on the r.h.s. of (A.43) equals

$$\begin{aligned} &\mathbb{E}\Bigg[\sum_{j=1}^{n} D_j(D_j - 1)\mathbb{1}_{\{D_j \leq \varepsilon_n^{-1} n^{1/(\tau-1)}\}}\Bigg] \\ &\quad \leq n \sum_{j=1}^{\varepsilon_n^{-1} n^{1/(\tau-1)}} j\mathbb{P}(D_1 > j) \\ &\quad \leq c_2 n \sum_{j=1}^{\varepsilon_n^{-1} n^{1/(\tau-1)}} j^{2-\tau} \leq \frac{c_2}{3-\tau} n\varepsilon_n^{-(3-\tau)} n^{(3-\tau)/(\tau-1)}. \end{aligned} \tag{A.44}$$

Since $\tau \in (2,3)$, $\varepsilon_n^{\tau-2} \to \infty$, so that uniformly for all $m \leq \overline{m}_n$, by the Markov inequality,

$$\begin{aligned} &\mathbb{P}\big(\mathbb{E}_n[S_m] \geq \varepsilon_n^{-1} m n^{(3-\tau)/(\tau-1)}\big) \\ &\quad \leq \varepsilon_n m^{-1} n^{-(3-\tau)/(\tau-1)} \mathbb{E}\big[\mathbb{E}_n[S_m]\mathbb{1}_{\{\max_{j=1}^{n} D_j \leq \varepsilon_n^{-1} n^{1/(\tau-1)}\}}\big] \\ &\quad \leq c_2 \varepsilon_n^{-(2-\tau)} = o(1). \end{aligned} \tag{A.45}$$

This completes the proof of (A.40). $\square$

A.4. Auxiliary lemmas for $\tau > 3$. In the lemmas below we use the coupling (A.16). We define the partial sums $\underline{S}_i$ and $\overline{S}_i$ by

$$\underline{S}_i = \sum_{j=1}^{i-1} (\underline{X}_i - 1), \qquad \overline{S}_i = \sum_{j=1}^{i-1} (\overline{X}_i - 1), \qquad i \geq 2. \tag{A.46}$$

As a consequence of (A.16), we obtain for $i \geq 2$,

$$\underline{S}_i \leq \sum_{j=2}^{i} (B_j - 1) \leq \overline{S}_i \qquad \text{a.s.} \tag{A.47}$$

LEMMA A.4 (A conditional large deviation estimate). *Fix $\tau > 2$. Then w.h.p., there exist a $c > 0$ and $\eta > 0$ sufficiently small, such that for all $i \geq 0$, and w.h.p.,*

$$\mathbb{P}_n(\underline{S}_i \leq \eta i) \leq \mathrm{e}^{-ci}. \tag{A.48}$$

The same bound applies to $\overline{S}_i$.

PROOF. We shall prove (A.48) using a conditional large deviation estimate, and an analysis of the moment generating function of $\underline{X}_1$, by adapting the proof of the upper bound in Cramér's theorem. Indeed, we rewrite and bound, for any $t \geq 0$,

$$\mathbb{P}_n(\underline{S}_i \leq \eta i) = \mathbb{P}_n(\mathrm{e}^{-t\underline{S}_i} \geq \mathrm{e}^{-t\eta i}) \leq (\mathrm{e}^{t\eta}\phi_n(t))^i, \tag{A.49}$$

where $\phi_n(t) = \mathbb{E}_n[\mathrm{e}^{-t(\underline{X}_1 - 1)}]$ is the (conditional) moment generating function of $\underline{X}_1 - 1$. Since $\underline{X}_1 - 1 \geq 0$, we have that $\mathrm{e}^{-t(\underline{X}_1 - 1)} \leq 1$, and $\underline{X}_1 \xrightarrow{d} B$, where B has the size-biased distribution in (2.3). Therefore, for every $t \geq 0$, $\phi_n(t) \xrightarrow{d} \phi(t)$, where $\phi(t) = \mathbb{E}[\mathrm{e}^{-t(B-1)}]$ is the Laplace transform of B. Since this limit is a.s. constant, we even obtain that $\phi_n(t) \xrightarrow{\mathbb{P}} \phi(t)$. Now, since $\mathbb{E}[B] = \nu > 1$, for each $0 < \eta < \mathbb{E}[B] - 1$, there exists a $t^* > 0$ and $\varepsilon > 0$ such that $\mathrm{e}^{-t^*\eta}\phi(t^*) \leq 1 - 2\varepsilon$. Then, since $\mathrm{e}^{t^*\eta}\phi_n(t^*) \xrightarrow{\mathbb{P}} \mathrm{e}^{t^*\eta}\phi(t^*)$, w.h.p. and for all n sufficiently large, $|\mathrm{e}^{t^*\eta}\phi_n(t^*) - \mathrm{e}^{t^*\eta}\phi(t^*)| \leq \varepsilon$, so that $\mathrm{e}^{-t^*\eta}\phi_n(t^*) \leq 1 - \varepsilon < 1$. The proof for $\overline{S}_i$ follows since $\overline{S}_i$ is stochastically larger than $\underline{S}_i$. This completes the proof. □

LEMMA A.5. *Fix $\tau > 3$. For $\underline{m}_n, \overline{m}_n$ such that $\log(\overline{m}_n/a_n)$, $\log(a_n/\underline{m}_n) = o(\sqrt{\log n})$,*

$$\sup_{i \leq \overline{m}_n} \mathbb{E}\big[R_i/S_{R_i} \mathbb{1}_{\{\underline{m}_n + 1 \leq R_i \leq \overline{m}_n\}}\big] < \infty. \tag{A.50}$$

PROOF. Take $\underline{m}_n + 1 \leq k \leq \overline{m}_n$ and recall the definition of $\Sigma_k < S_k$ in (6.12). For $\eta > 0$,

$$\begin{aligned}
\mathbb{E}[k/\Sigma_k] &= \mathbb{E}[k/\Sigma_k]\mathbb{1}_{\{\Sigma_k < \eta k\}} + \mathbb{E}[k/\Sigma_k]\mathbb{1}_{\{\Sigma \geq \eta k\}} \\
&\leq \mathbb{E}[k/\Sigma_k]\mathbb{1}_{\{\Sigma_k < \eta k\}} + \eta^{-1} \\
&\leq k\mathbb{P}(\Sigma_k < \eta k) + \eta^{-1} \leq k\mathbb{P}(\underline{S}_k < \eta k) + \eta^{-1},
\end{aligned}$$

since $\Sigma_k = 1 + \sum_{j=2}^{k}(B_j - 1) > \underline{S}_k$, a.s. Applying the large deviation estimate from the previous lemma, we obtain

$$\mathbb{E}[k/\Sigma_k] \le \eta^{-1} + k\mathrm{e}^{-c_2 k}$$

for each $\underline{m}_n + 1 \le k \le \overline{m}_n$. Hence,

$$\sup_{i \le \overline{m}_n} \mathbb{E}\big[R_i/S_{R_i} \mathbb{1}_{\{\underline{m}_n + 1 \le R_i \le \overline{m}_n\}}\big] \le \eta^{-1} + \overline{m}_n \mathrm{e}^{-c_2 \underline{m}_n}. \tag{A.51}$$

□

LEMMA A.6. *Fix $\tau > 3$, and let $\underline{m}_n$ be such that* $\log(a_n/\underline{m}_n) = o(\sqrt{\log n})$. *Then, for each sequence $C_n \to \infty$,*

$$\mathbb{P}_n\left(\sum_{j=2}^{m_n} B_j^2/S_j^2 > C_n\right) \xrightarrow{\mathbb{P}} 0. \tag{A.52}$$

Consequently,

$$\sum_{j=2}^{m_n} B_j^2/S_j^2 = O_{\mathbb{P}}(1). \tag{A.53}$$

PROOF. If we show that the conditional expectation of $\sum_{j=2}^{m_n} B_j^2/S_j^2$, given $\{D_i\}_{i=1}^n$, is finite, then (A.52) holds. Take $a \in (1, \min(2, \tau - 2))$; this is possible since $\tau > 3$. We bound

$$\begin{aligned}\mathbb{E}_n\left[\left(\frac{B_j}{S_j}\right)^2\right] &\le 2\left(\mathbb{E}_n\left[\left(\frac{B_j - 1}{S_j}\right)^2\right]\right) + 2\mathbb{E}_n\left[\frac{1}{(S_j)^2}\right] \\ &\le 2\left(\mathbb{E}_n\left[\left(\frac{B_j - 1}{S_j}\right)^a\right]\right) + 2\mathbb{E}_n\left[\frac{1}{(S_j)^a}\right].\end{aligned} \tag{A.54}$$

By stochastic domination and Lemma A.4, we find that, w.h.p., using $a > 1$,

$$\sum_{j=2}^{m_n} \mathbb{E}_n\left[\frac{1}{(S_j)^a}\right] < \infty.$$

We will now bound (A.52). Although, by definition

$$S_j = D_1 + \sum_{i=2}^{j}(B_i - 1)$$

for the asymptotic statements that we discuss here, we may as well replace this definition by

$$S_j = \sum_{i=2}^{j}(B_i - 1), \tag{A.55}$$

and use exchangeability, so that

$$\mathbb{E}_n\Big[\Big(\frac{B_j-1}{S_j}\Big)^a\Big]=\mathbb{E}_n\Big[\Big(\frac{B_2-1}{S_j}\Big)^a\Big],$$

since for each j, we have $\frac{B_j-1}{S_j}\stackrel{d}{=}\frac{B_1}{S_j}$. Furthermore, for $j\geq 2$,

$$\mathbb{E}_n\Big[\Big(\frac{B_2-1}{S_j}\Big)^a\Big]\leq\mathbb{E}_n\Big[\Big(\frac{B_2-1}{S_{3,j}}\Big)^a\Big],$$

where $S_{3,j}=(B_3-1)+\cdots+(B_j-1)$. Furthermore, we can replace $S_{3,j}$ by $\underline{S}_{3,j}=(\underline{X}_3-1)+\cdots+(\underline{X}_j-1)$, which are mutually independent and sampled from $D_{(1)}-1,\ldots,D_{(\underline{m}_n)}-1$, as above and which are also independent of B_2. Consequently,

$$\begin{aligned}\sum_{j=2}^{\underline{m}_n}\mathbb{E}_n\Big[\Big(\frac{B_j-1}{S_j}\Big)^2\Big]&\leq\sum_{j=2}^{\underline{m}_n}\mathbb{E}_n\Big[\Big(\frac{B_j-1}{S_j}\Big)^a\Big]\\&=\sum_{j=2}^{\underline{m}_n}\mathbb{E}_n\Big[\Big(\frac{B_2-1}{S_j}\Big)^a\Big]\\&\leq\mathbb{E}_n\Big[\Big(\frac{B_2-1}{S_2}\Big)^a\Big]+\sum_{j=3}^{\underline{m}_n}\mathbb{E}_n\Big[\Big(\frac{B_2-1}{S_{3,j}}\Big)^a\Big]\\&\leq 1+\sum_{j=3}^{\underline{m}_n}\mathbb{E}_n\Big[\Big(\frac{B_2-1}{\underline{S}_{3,j}}\Big)^a\Big]\\&=1+\mathbb{E}_n[(B_2-1)^a]\sum_{j=3}^{\underline{m}_n}\mathbb{E}_n\Big[\Big(\frac{1}{\underline{S}_{3,j}}\Big)^a\Big].\end{aligned}\tag{A.56}$$

Finally, the expression $\sum_{j=3}^{\underline{m}_n}\mathbb{E}_n[1/\underline{S}^a_{2,j}]$ can be shown to be finite as above. □

LEMMA A.7 (Logarithmic asymptotics of $S_{\overline{m}_n}$). *Fix $\tau>3$, and let $\underline{m}_n$ be such that* $\log(a_n/\underline{m}_n)=o(\sqrt{\log n})$. *Then,*

$$\log S_{\underline{m}_n}-\log\underline{m}_n=o_{\mathbb{P}}\big(\sqrt{\log\underline{m}_n}\big).\tag{A.57}$$

PROOF. As in the previous lemma we define w.l.o.g. S_j by (A.55). Then,

$$S_j\leq_{st}\overline{S}_j,$$

where $\overline{S}_j$ is a sum of i.i.d. random variables $\overline{X}_i - 1$, where the $\overline{X}_i$ are sampled from $D_1, \dots, D_n$ with replacement, where $\underline{m}_n$ of the vertices with the smallest degree(s) have been removed. Using the Markov inequality,

$$\begin{aligned}\mathbb{P}_n\bigl(\log(S_{\underline{m}_n}/\underline{m}_n) > c_n\bigr) &= \mathbb{P}_n(S_{\underline{m}_n}/\underline{m}_n > \mathrm{e}^{c_n}) \\ &\le \mathrm{e}^{-c_n}\mathbb{E}_n[\overline{S}_{\underline{m}_n}/\underline{m}_n] = \mathrm{e}^{-c_n}\mathbb{E}_n[\overline{X}_i - 1].\end{aligned} \tag{A.58}$$

We shall prove below that, for $\tau > 3$, $\mathbb{E}_n[\overline{X}_1] \xrightarrow{\mathbb{P}} \nu$ so that

$$\mathbb{E}_n[S_m] \le \nu m\bigl(1 + o_{\mathbb{P}}(1)\bigr). \tag{A.59}$$

Indeed, from [34], Proposition A.1.1, we know that there are $\alpha, \beta > 0$, such that

$$\mathbb{P}(|\nu_n - \nu| > n^{-\alpha}) \le n^{-\beta}, \tag{A.60}$$

where

$$\nu_n = \sum_{j=1}^{\infty} j g_j^{(n)} = \sum_{j=1}^{\infty} j(j+1)\frac{1}{L_n}\sum_{i=1}^{n} \mathbb{1}_{\{D_i = j+1\}} = \frac{1}{L_n}\sum_{i=1}^{n} D_i(D_i - 1). \tag{A.61}$$

Define $\overline{\nu}_n = \mathbb{E}_n[\overline{X}_1]$. Then we claim that there exists $\alpha, \beta > 0$ such that

$$\mathbb{P}(|\overline{\nu}_n - \nu_n| > n^{-\alpha}) \le n^{-\beta}. \tag{A.62}$$

To see (A.62), by definition of $\overline{\nu}_n = \mathbb{E}_n[\overline{X}_1]$,

$$\begin{aligned}|\overline{\nu}_n - \nu_n| &= \left|\frac{1}{\overline{L}_n}\sum_{i=\underline{m}_n+1}^{n} D_{(i)}\bigl(D_{(i)} - 1\bigr) - \frac{1}{L_n}\sum_{i=1}^{n} D_i(D_i - 1)\right| \\ &\le \left|\frac{1}{\overline{L}_n}\sum_{i=\underline{m}_n+1}^{n} D_{(i)}\bigl(D_{(i)} - 1\bigr) - \frac{1}{L_n}\sum_{i=\underline{m}_n+1}^{n} D_{(i)}\bigl(D_{(i)} - 1\bigr)\right| \\ &\quad + \left|\frac{1}{L_n}\sum_{i=\underline{m}_n+1}^{n} D_{(i)}\bigl(D_{(i)} - 1\bigr) - \frac{1}{L_n}\sum_{i=1}^{n} D_{(i)}\bigl(D_{(i)} - 1\bigr)\right|.\end{aligned} \tag{A.63}$$

The first term on the r.h.s. of (A.64) is with probability at least $1 - n^{-\beta}$ bounded above by $n^{-\alpha}$, w.h.p., since it is bounded by

$$\left(\frac{L_n - \overline{L}_n}{\overline{L}_n}\right)\frac{1}{L_n}\sum_{i=1}^{n} D_i(D_i - 1),$$

and since, using (A.23) and (A.24), $(L_n - \overline{L}_n)/\overline{L}_n = o_{\mathbb{P}}(n^{-\alpha})$ for some $\alpha > 0$. The second term on the r.h.s. of (A.64) is bounded by

$$\frac{1}{L_n}\sum_{j=1}^{\underline{m}_n} D_{(j)}^2 \le \frac{1}{L_n}\sum_{j=1}^{\underline{m}_n} D_j^2 = o_{\mathbb{P}}(n^{-\alpha}), \tag{A.64}$$

since $\tau > 3$. This completes the proof of (A.62). Combining (A.58) with $c_n = o(\sqrt{\log \underline{m}_n})$ and the fact that $\mathbb{E}_n[\overline{X}_1] \xrightarrow{\mathbb{P}} \nu$, we obtain an upper bound for the left-hand side of (A.57).

For the lower bound, we simply make use of the fact that, by Lemma A.4 and w.h.p., $S_{\underline{m}_n} \ge \eta \underline{m}_n$, so that $\log S_{\underline{m}_n} - \log \underline{m}_n \ge \log \eta = o_{\mathbb{P}}(\sqrt{\log \underline{m}_n})$. $\square$

LEMMA A.8. *Fix* $\tau > 3$, *and let* $\underline{m}_n$ *be such that* $\log(a_n/\underline{m}_n) = o(\sqrt{\log n})$. *Then*,

$$\sum_{j=1}^{\underline{m}_n} \frac{S_j - (\nu-1)j}{S_j(\nu-1)j} = \sum_{j=1}^{\underline{m}_n}\left[\frac{1}{(\nu-1)j} - \frac{1}{S_j}\right] = o_{\mathbb{P}}(\sqrt{\log \underline{m}_n}). \tag{A.65}$$

PROOF. We can stochastically bound the sum (A.65) by

$$\sum_{j=1}^{\underline{m}_n}\left[\frac{1}{(\nu-1)j} - \frac{1}{\underline{S}_j}\right] \le \sum_{j=1}^{\underline{m}_n}\left[\frac{1}{(\nu-1)j} - \frac{1}{S_j}\right] \le \sum_{j=1}^{\underline{m}_n}\left[\frac{1}{(\nu-1)j} - \frac{1}{\overline{S}_j}\right]. \tag{A.66}$$

We now proceed by proving (A.65) both with S_j replaced by $\overline{S}_j$, and with S_j replaced by $\underline{S}_j$. In the proof of Lemma A.7 we have shown that $\mathbb{E}_n[\overline{X}_1]$ converges, w.h.p., to ν. Consequently, we can copy the proof of Proposition 4.3(a) to show that, w.h.p.,

$$\sum_{j=1}^{\underline{m}_n} \frac{\overline{S}_j - (\nu-1)j}{\overline{S}_j(\nu-1)j} = o_{\mathbb{P}}(\sqrt{\log \underline{m}_n}). \tag{A.67}$$

Indeed, assuming that $\overline{S}_j > \varepsilon j$ for all $j > j_0$, independent of n (recall Lemma A.4), we can use the bound

$$\begin{aligned}\sum_{j=j_0}^{\underline{m}_n} \frac{|\overline{S}_j - (\nu-1)j|}{\overline{S}_j(\nu-1)j} &\le C\sum_{j=j_0}^{\underline{m}_n} \frac{|\overline{S}_j - (\nu-1)j|}{j^2} \\ &\le C\sum_{j=j_0}^{\underline{m}_n} \frac{|\overline{S}_j^*|}{j^2} + O_{\mathbb{P}}(|\nu - \overline{\nu}_n|\log \overline{m}_n),\end{aligned} \tag{A.68}$$

where $\overline{S}_j^* = \overline{S}_j - (\overline{\nu}_n - 1)j$, is for fixed n the sum of i.i.d. random variables with mean 0. Combining (A.60) and (A.62), we obtain that $O_{\mathbb{P}}(|\nu - \nu_n|\log \overline{m}_n) = o_{\mathbb{P}}(1)$, so we are left to bound the first contribution in (A.68).

According to the Marcinkiewicz–Zygmund inequality [recall (5.19)], for $a \in (1,2)$,

$$\begin{aligned}\mathbb{E}_n[|\overline{S}_j^*|^a] &\le B_a^* \mathbb{E}\Bigg[\sum_{k=1}^{j} \big(\overline{X}_k - (\overline{\nu}_n - 1)\big)^2\Bigg]^{a/2} \\ &\le B_a^* \sum_{k=1}^{j} \mathbb{E}_n[|\overline{X}_k - (\overline{\nu}_n - 1)]|^a] = jC_a \mathbb{E}_n[|\overline{X}_1 - (\overline{\nu}_n - 1)|^a],\end{aligned}$$

When we take $1 < a < \tau - 2$, where $\tau - 2 > 1$, then uniformly in n, we have that $\mathbb{E}_n[|\overline{X}_1 - \overline{\nu}_n|^a] < c_a$ because

$$\begin{aligned}\mathbb{E}_n[|\overline{X}_1|^a] &= \sum_{s=1}^{\infty} s^a g_s^{(n)} = \frac{1}{L_n} \sum_{i=1}^{n} D_i^a (D_i - 1) \\ &\le \frac{1}{L_n} \sum_{i=1}^{n} D_i^{a+1} \xrightarrow{\text{a.s.}} \frac{\mathbb{E}[D_1^{a+1}]}{\mu} < \infty,\end{aligned}$$

since $a < \tau - 2$, so that

$$\begin{aligned}\mathbb{E}_n\Bigg[\sum_{j=1}^{\overline{m}_n} \frac{|\overline{S}_j^*|}{j^2}\Bigg] &\le \sum_{j=1}^{\overline{m}_n} \frac{\mathbb{E}_n[|\overline{S}_j^*|^a]^{1/a}}{j^2} \\ &= \sum_{j=1}^{\overline{m}_n} \frac{(c_a)^{1/a} \mathbb{E}_n[|\overline{X}_1 - (\nu_n - 1)|^a]^{1/a}}{j^{2-1/a}} < \infty,\end{aligned} \tag{A.69}$$

since $a > 1$, and the last bound being true a.s. and uniform in n. The proof for $\underline{S}_j$ is identical, where now, instead of (A.64), we use that there exists $\alpha > 0$ such that, w.h.p.,

$$\frac{1}{L_n} \sum_{j=n-\overline{m}_n+1}^{n} D_{(j)}^2 = o_{\mathbb{P}}(n^{-\alpha}), \tag{A.70}$$

using the argument in (A.23)–(A.24). □

APPENDIX B: ON THE DEVIATION FROM A TREE

In this section, we do the necessary preliminaries needed for the proof of Proposition 4.9 in Section 7. One of the ingredients is writing $\mathbb{P}(\mathrm{C}_n > m)$ as the expectation of the product of conditional probabilities [see (7.9) and Lemma B.1]. A second issue of Section 7 is to estimate the two error terms in (7.15). We will deal with these two error terms in Lemma B.3. Lemma B.2 is a preparation for Lemma B.3 and gives an upper bound for the expected number of artificial stubs, which in turn is bounded by the expected number of closed cycles.

In the statement of the following lemma, we recall that $\mathbb{Q}_n^{(j)}$ denotes the conditional distribution given $\mathrm{SWG}_j^{(1,2)}$ and $\{D_i\}_{i=1}^n$.

LEMMA B.1 (Conditional product form tail probabilities C_n).

$$\mathbb{P}(\mathrm{C}_n > m) = \mathbb{E}\Bigg[\prod_{j=1}^{m} \mathbb{Q}_n^{(j)}(\mathrm{C}_n > j \mid \mathrm{C}_n > j-1)\Bigg]. \tag{B.1}$$

PROOF. By the tower property of conditional expectations, we can write

$$\begin{aligned}\mathbb{P}(\mathrm{C}_n > m) &= \mathbb{E}\big[\mathbb{Q}_n^{(1)}(\mathrm{C}_n > m)\big]\\ &= \mathbb{E}\big[\mathbb{Q}_n^{(1)}(\mathrm{C}_n > 1)\mathbb{Q}_n^{(1)}(\mathrm{C}_n > m \mid \mathrm{C}_n > 1)\big].\end{aligned} \tag{B.2}$$

Continuing this further, for all $1 \le k \le m$,

$$\begin{aligned}&\mathbb{Q}_n^{(k)}(\mathrm{C}_n > m \mid \mathrm{C}_n > k)\\ &\quad = \mathbb{E}_n^{(k)}\big[\mathbb{Q}_n^{(k+1)}(\mathrm{C}_n > m \mid \mathrm{C}_n > k)\big]\\ &\quad = \mathbb{E}_n^{(k)}\big[\mathbb{Q}_n^{(k+1)}(\mathrm{C}_n > k+1 \mid \mathrm{C}_n > k)\mathbb{Q}_n^{(k+1)}(\mathrm{C}_n > m \mid \mathrm{C}_n > k+1)\big],\end{aligned} \tag{B.3}$$

where $\mathbb{E}_n^{(k)}$ denotes the expectation w.r.t. $\mathbb{Q}_n^{(k)}$. In particular,

$$\begin{aligned}\mathbb{P}(\mathrm{C}_n > m) &= \mathbb{E}\big[\mathbb{Q}_n^{(1)}(\mathrm{C}_n > m)\big]\\ &= \mathbb{E}\big[\mathbb{Q}_n^{(1)}(\mathrm{C}_n > 1)\mathbb{E}_n^{(1)}\big[\mathbb{Q}_n^{(2)}(\mathrm{C}_n > 2 \mid \mathrm{C}_n > 1)\\ &\qquad \times \mathbb{Q}_n^{(2)}(\mathrm{C}_n > m \mid \mathrm{C}_n > 2)\big]\big]\\ &= \mathbb{E}\big[\mathbb{Q}_n^{(1)}(\mathrm{C}_n > 1)\mathbb{Q}_n^{(2)}(\mathrm{C}_n > 2 \mid \mathrm{C}_n > 1)\\ &\qquad \times \mathbb{Q}_n^{(2)}(\mathrm{C}_n > m \mid \mathrm{C}_n > 2)\big],\end{aligned} \tag{B.4}$$

where the last equality follows since $\mathbb{Q}_n^{(1)}(\mathrm{C}_n > 1)$ is measurable w.r.t. $\mathbb{Q}_n^{(2)}$ and the tower property. Continuing this indefinitely, we arrive at (B.1). □

LEMMA B.2 (The number of cycles closed). (a) *Fix* $\tau \in (2,3)$. *Then*, *w.h.p.*, *there exist* $\overline{m}_n$ *with* $\overline{m}_n/a_n \to \infty$ *and* $C > 0$ *such that for all* $m \le \overline{m}_n$ *and all* $\varepsilon_n \downarrow 0$,

$$\mathbb{E}_n\big[R_m^{(i)} - m\big] \le \varepsilon_n^{-1}\bigg(\frac{m}{a_n}\bigg)^2, \qquad i = 1, 2. \tag{B.5}$$

(b) *Fix* $\tau > 3$. *Then*, *there exist* $\overline{m}_n$ *with* $\overline{m}_n/a_n \to \infty$ *and* $C > 0$ *such that for all* $m \le \overline{m}_n$,

$$\mathbb{E}\big[R_m^{(i)} - m\big] \le Cm^2/n, \qquad i = 1, 2. \tag{B.6}$$

PROOF. Observe that

$$R_m^{(i)} - m \le \sum_{j=1}^{m} U_j, \tag{B.7}$$

where U_j is the indicator that a cycle is closed at time j. Since closing a cycle means choosing an allowed stub, which occurs with conditional probability at most $S_{j-1}^{(i)}/(L_n-2j-1)$, we find that

$$\mathbb{E}[U_j|S_{j-1}^{(i)},L_n]=S_{j-1}^{(i)}/(L_n-2j-1), \tag{B.8}$$

so that

$$\mathbb{E}_n[R_m^{(i)}-m]\le\sum_{j=1}^m\mathbb{E}_n[U_j]=\sum_{j=1}^m\mathbb{E}_n[S_{j-1}^{(i)}/(L_n-2j-1)]. \tag{B.9}$$

When $\tau>3$, and using that, since $j\le\overline{m}_n=o(n)$, we have $L_n-2j-1\ge 2n-2j-1\ge n$ a.s., we arrive at

$$\mathbb{E}[R_m^{(i)}-m]\le\frac{1}{n}\sum_{j=1}^m\mathbb{E}[S_{j-1}^{(i)}]\le\frac{\mu}{n}+\frac{1}{n}\sum_{j=2}^m C(j-1)\le Cm^2/n. \tag{B.10}$$

When $\tau\in(2,3)$, we have to be a bit more careful. In this case, we apply (A.40) to the r.h.s. of (B.9), so that, w.h.p., and uniformly in m,

$$\mathbb{E}_n[R_m^{(i)}-m]\le\frac{m^2}{n}\varepsilon_n^{-1}n^{(3-\tau)/(\tau-1)}=\varepsilon_n^{-1}\left(\frac{m}{a_n}\right)^2. \tag{B.11}$$

This proves (B.5). $\square$

LEMMA B.3 (Treatment of error terms). *As $n\to\infty$, there exists $\overline{m}_n$ with $\overline{m}_n/a_n\to\infty$, such that*

$$\frac{\overline{m}_n}{n}|\mathrm{Art}_{a_n}^{(1)}|=o_{\mathbb{P}}(1),\qquad \frac{S_{a_n}^{(1)}}{n}\sum_{m=1}^{\overline{m}_n}\frac{|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}}=o_{\mathbb{P}}(1). \tag{B.12}$$

PROOF. We start with the first term. By Lemma B.2, for $\tau>3$,

$$\mathbb{E}[|\mathrm{Art}_m^{(i)}|]\le\mathbb{E}[R_m^{(i)}-m]\le Cm^2/n,\qquad m\le\overline{m}_n. \tag{B.13}$$

As a result, we have that

$$\frac{\overline{m}_n}{n}\mathbb{E}[|\mathrm{Art}_{a_n}^{(1)}|]\le C\overline{m}_n^3/n^2=o(1). \tag{B.14}$$

Again by Lemma B.2, but now for $\tau\in(2,3)$, w.h.p. and uniformly in $m\le\overline{m}_n$, where $\overline{m}_n$ is determined in Lemma B.2,

$$\frac{\overline{m}_n}{n}\mathbb{E}_n[|\mathrm{Art}_{a_n}^{(1)}|]\le\frac{\overline{m}_n}{n}\varepsilon_n^{-1}=o(1), \tag{B.15}$$

whenever $\varepsilon_n^{-1}\to\infty$ sufficiently slowly.

Using (B.7) and $|\mathrm{Art}_m| \leq R_m - m$, and using also that, w.h.p. and for all $j \leq m$, $S_m^{(2)} \geq S_{j-1}^{(2)}$, we obtain that

$$\begin{aligned}\mathbb{E}_n\left[\frac{|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}}\right] &\leq \mathbb{E}_n\left[\frac{\sum_{j=1}^m U_j}{S_m^{(2)}}\right] = \sum_{j=1}^m \mathbb{E}_n\left[\frac{U_j}{S_m^{(2)}}\right] \\ &\leq \sum_{j=1}^m \mathbb{E}_n\left[\frac{U_j}{S_{j-1}^{(2)}}\right] \leq \sum_{j=1}^m \mathbb{E}_n[1/(L_n - 2j - 1)] \\ &\leq m/n,\end{aligned} \tag{B.16}$$

where we used (B.8) in the one-but-last inequality.

When $\tau > 3$, we thus further obtain

$$\frac{1}{n}\sum_{m=1}^{\overline{m}_n} \mathbb{E}\left[\frac{|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}}\right] \leq \frac{1}{n}\sum_{m=1}^{\overline{m}_n} m/n = O(\overline{m}_n^2/n^2), \tag{B.17}$$

so that, also using the bound on $S_{a_n}^{(1)}$ that holds w.h.p. as proved in (7.4),

$$\frac{1}{n}\sum_{m=1}^{\overline{m}_n} \frac{S_{a_n}^{(1)}|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}} = o_{\mathbb{P}}(1). \tag{B.18}$$

When $\tau \in (2, 3)$, by (B.16),

$$\sum_{m=1}^{\overline{m}_n} \mathbb{E}_n\left[\frac{|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}}\right] \leq \sum_{m=1}^{\overline{m}_n} m/n \leq \overline{m}_n^2/n, \tag{B.19}$$

so that, again using the bound on $S_{a_n}^{(1)}$ that holds w.h.p. as proved in (7.4),

$$\begin{aligned}\frac{S_{a_n}^{(1)}}{n}\sum_{m=1}^{\overline{m}_n} \frac{|\mathrm{Art}_m^{(2)}|}{S_m^{(2)}} &= O_{\mathbb{P}}\bigl(\eta_n^{-1} n^{-2+(3-\tau)/(\tau-1)} a_n \overline{m}_n^2\bigr) \\ &= O_{\mathbb{P}}\bigl(\eta_n^{-1}(\overline{m}_n/a_n)^2 n^{-1/(\tau-1)}\bigr) = o_{\mathbb{P}}(1),\end{aligned} \tag{B.20}$$

since $a_n = n^{(\tau-2)/(\tau-1)}$ and whenever $\overline{m}_n/a_n, \eta_n^{-1} \to \infty$ sufficiently slowly such that $n^{-1/(\tau-1)}\eta_n^{-1}(\overline{m}_n/a_n)^2 = o(1)$. □

APPENDIX C: WEAK CONVERGENCE OF THE WEIGHT FOR $\tau > 3$

In this section we prove Propositions 4.3(b) and 4.6(b), for $\tau > 3$. Moreover, we show weak convergence of C_n/a_n and prove (4.29) for $\tau > 3$. We start with Proposition 4.3(b).

For this, we rewrite T_m [compare (4.6), with s_i replaced by S_i],

$$T_m - \frac{1}{\nu - 1}\log m = \sum_{i=1}^m \frac{E_i - 1}{S_i} + \left[\sum_{i=1}^m \frac{1}{S_i} - \frac{1}{\nu - 1}\log m.\right] \tag{C.1}$$

The second term on the r.h.s. of (C.1) converges a.s. to some Y by (5.14); thus, it suffices to prove that $\sum_{i=1}^{m}(E_i - 1)/S_i$ converges a.s. For this, we use that the second moment equals, due to the independence of $\{E_i\}_{i=1}^{\infty}$ and $\{S_i\}_{i=1}^{\infty}$ and the fact that $\mathbb{E}[E_i] = \mathrm{Var}(E_i) = 1$,

$$\mathbb{E}\left[\left(\sum_{i=1}^{m}\frac{E_i - 1}{S_i}\right)^2\right] = \mathbb{E}\left[\sum_{i=1}^{m} 1/S_i^2\right], \tag{C.2}$$

which converges uniformly in m. This shows that

$$T_m - \frac{1}{\nu - 1}\log m \xrightarrow{d} \sum_{i=1}^{\infty}\frac{E_i - 1}{S_i} + Y, \tag{C.3}$$

which completes the proof for T_m for $\tau > 3$.

We continue the proof of Proposition 4.9 by showing that, for $\tau > 3$, (4.25) holds.

LEMMA C.1 (Weak convergence of connection time). *Fix* $\tau > 3$, *then*,

$$\mathrm{C}_n/a_n \xrightarrow{d} M, \tag{C.4}$$

where M *has an exponential distribution with mean* $\mu/(\nu - 1)$, *that is*,

$$\mathbb{P}(M > x) = \exp\left\{-\frac{\nu - 1}{\mu}x\right\}. \tag{C.5}$$

PROOF. The proof is somewhat sketchy; we leave the details to the reader. We again make use of the product structure in Lemma B.1 [recall (B.1)], and simplify (7.12), by taking complementary probabilities, to

$$\mathbb{Q}_n^{(m)}(\mathrm{C}_n > m + 1|\mathrm{C}_n > m) \approx 1 - S_{a_n}^{(1)}/L_n. \tag{C.6}$$

For $m \le \overline{m}_n$, error terms that are left out can easily be seen to be small by Lemma B.3. We next simplify by substitution of $L_n = \mu n$, and using that $\mathrm{e}^{-x} \approx 1 - x$, for x small, to obtain that

$$\mathbb{Q}_n^{(m)}(\mathrm{C}_n > m + 1|\mathrm{C}_n > m) \approx \exp\{-S_{a_n}^{(1)}/(\mu n)\}. \tag{C.7}$$

Substituting the above approximation into (B.1) for $m = a_n x$ yields

$$\mathbb{P}(\mathrm{C}_n > a_n x) \approx \mathbb{E}\left[\exp\left\{-a_n x\frac{S_{a_n}^{(1)}}{\mu n}\right\}\right] = \exp\left\{-\frac{(\nu - 1)}{\mu n}a_n^2 x\right\}, \tag{C.8}$$

where we approximate $S_m^{(1)} \approx (\nu - 1)m$. Since $a_n = \sqrt{n}$, we arrive at (C.4)–(C.5). $\square$

We now complete the proof of (4.29) for $\tau > 3$. It is not hard to prove from (C.1) that

$$\big(T^{(1)}_{a_n} - \gamma \log a_n, T^{(2)}_{\mathrm{C}_n} - \gamma \log \mathrm{C}_n\big) \xrightarrow{d} (X_1, X_2), \tag{C.9}$$

where (X_1, X_2) are two independent random variables with distribution given by

$$\begin{aligned} X_1 &= \sum_{i=1}^{\infty} \frac{E_i - 1}{S_i^{(\mathrm{ind})}} + \lim_{m\to\infty}\Bigg[\Bigg(\sum_{i=1}^{m} 1/S_i^{(\mathrm{ind})}\Bigg) - \log m\Bigg] \\ &= \sum_{i=1}^{\infty} \frac{E_i - 1}{S_i^{(\mathrm{ind})}} + \sum_{i=1}^{\infty}\bigg(\frac{1}{S_i^{(\mathrm{ind})}} - \frac{1}{(\nu-1)i}\bigg) + \gamma^{(\mathrm{e})}, \end{aligned} \tag{C.10}$$

where $\gamma^{(\mathrm{e})}$ is the Euler–Mascheroni constant. By Lemma C.1,

$$\big(T^{(1)}_{a_n} - \gamma \log a_n, T^{(2)}_{\mathrm{C}_n} - \gamma \log a_n\big) \xrightarrow{d} (X_1, X_2 + \gamma \log M), \tag{C.11}$$

where M is the weak limit of C_n/a_n defined in (4.25). We conclude that

$$W_n - \gamma \log n \xrightarrow{d} V = X_1 + X_2 + \gamma \log M. \tag{C.12}$$

Since $(\nu - 1)M/\mu$ is an exponential variable with mean 1, $\Lambda = \log((\nu-1)M/\mu)$ has a Gumbel distribution.

Finally let us derive the distribution of X_i. The random variables X_i are related to a random variable W, which appears as a limit in a supercritical continuous-time branching process as described in Section 4.1. Indeed, denoting by $Z(t)$ the number of alive individuals in a continuous-time branching process where the root has degree D having distribution function F, while all other vertices in the tree have degree $\{B_i^{(\mathrm{ind})}\}_{i=2}^{\infty}$, which are i.i.d. random variables with probability mass function g in (2.3). Then, W arises as

$$Z(t)\mathrm{e}^{-(\nu-1)t} \xrightarrow{\text{a.s.}} W. \tag{C.13}$$

We note the following general results about the limiting distributional asymptotics of continuous-time branching processes.

PROPOSITION C.2 (The limiting random variables). (a) *The limiting random variable W has the following explicit construction*:

$$W = \sum_{j=1}^{D} \widetilde{W}_j \mathrm{e}^{-(\nu-1)\xi_j}. \tag{C.14}$$

Here D has distribution F, ξ_i are i.i.d. exponential random variables with mean one independent of $\widetilde{W}_i$, which are independent and identically distributed with Laplace transform $\phi(t) = \mathbb{E}(\mathrm{e}^{-t\widetilde{W}})$ given by the formula

$$\phi^{-1}(x) = (1-x)\exp\bigg\{\int_1^x \bigg(\frac{\nu-1}{h(s)-s} + \frac{1}{1-s}\bigg)\,ds\bigg\}, \qquad 0 < x \le 1, \tag{C.15}$$

and $h(\cdot)$ is the probability generating function of the size-biased probability mass function g [*see* (2.3)].

(b) *Let T_m be the random variables defined as*

$$T_m = \sum_{i=1}^{m} E_i / S_i^{(\mathrm{ind})}, \tag{C.16}$$

where E_i are i.i.d. exponential random variables with mean one, and recall that $S_i^{(\mathrm{ind})}$ is a random walk where the first step has distribution D where $D \sim F$ and the remaining increments have distribution $B - 1$ where B has the size biased distribution. Then

$$T_m - \frac{\log m}{\nu - 1} \xrightarrow{\text{a.s.}} -\frac{\log(W/(\nu - 1))}{\nu - 1}, \tag{C.17}$$

where W is the martingale limit in (C.13) *in part* (a).

(c) *The random variables X_i, $i = 1, 2$, are i.i.d. with $X_i \stackrel{d}{=} -\frac{\log(W/(\nu-1))}{\nu-1}$.*

PROOF. These results follow from results about continuous-time branching processes (everything relevant to this result is taken from [3]). Part (b) is proved in [3], Theorem 2, page 120. To prove part (a) recall the continuous-time version of the construction described in Section 4.1, where we shall let $D \sim F$ denote the number of offspring of the initial root and, for $i \geq 2$, $B_i \sim g$, the size-biased biased probability mass function (2.3). Then note that for any t sufficiently large we can decompose $Z(t)$, the number of alive nodes at time t as

$$Z(t)\mathrm{e}^{-(\nu-1)t} = \sum_{i=1}^{D} \widetilde{Z}_i(t - \xi_i)\mathrm{e}^{-(\nu-1)t}. \tag{C.18}$$

Here D, ξ_i and the processes $\widetilde{Z}_i(\cdot)$ are all independent of each other, $D \sim F$ denotes the number of offspring of the root, ξ_i are lifetimes of these offspring and are distributed as i.i.d. exponential random variables with mean 1 and $\widetilde{Z}_j(\cdot)$, corresponding to the subtrees attached below offspring j of the root, are independent continuous-time branching processes where each individual lives for an exponential mean 1 amount of time and then dies, giving birth to a random number of offspring where the number of offspring has distribution $B \sim g$ as in (2.3).

Now known results (see [3], Theorem 1, page 111 and Theorem 3, page 116) imply that

$$\widetilde{Z}_i(t)\mathrm{e}^{-(\nu-1)t} \xrightarrow{\text{a.s.}} \widetilde{W}_i,$$

where $\widetilde{W}_i$ have Laplace transform given by (C.15). Part (a) now follows by comparing (C.14) with (C.18).

Part (c) follows from part (b) and observing that

$$T_m - \frac{1}{(\nu - 1)} \log m = \sum_{i=1}^{m} \frac{E_i - 1}{S_i^{(\mathrm{ind})}} + \sum_{i=1}^{m} \frac{1}{S_i^{(\mathrm{ind})}} - \frac{1}{(\nu - 1)} \log m,$$

and a comparison with (C.10). This completes the proof. □

Thus, with Λ a Gumbel distribution, the explicit distribution of the re-centered minimal weight paths is given by

$$V = -\frac{\log(W_1/(\nu-1))}{\nu-1} - \frac{\log(W_2/(\nu-1))}{\nu-1} + \gamma\Lambda - \gamma\log(\nu-1)/\mu, \tag{C.19}$$

since $\log M = \Lambda - \log((\nu-1)/\mu)$. Rearranging terms establishes the claims on the limit V below Theorem 3.1, and completes the proof of (4.29) in Proposition 4.9(b) for $\tau > 3$.

Acknowledgments. The authors thank Allan Sly for help with a preliminary version of the paper. G. H. thanks the Collegium Budapest for the opportunity to work on the revision of this paper during his visit.

REFERENCES

[1] ADDARIO-BERRY, L., BROUTIN, N. and LUGOSI, G. (2010). The longest minimum-weight path in a complete graph. Preprint. *Combin. Probab. Comput.* **19** 1–19.

[2] ATHREYA, K. B. and KARLIN, S. (1967). Limit theorems for the split times of branching processes. *J. Math. Mech.* **17** 257–277. MR0216592

[3] ATHREYA, K. B. and NEY, P. E. (2004). *Branching Processes*. Dover, Mineola, NY. MR2047480

[4] BERTOIN, J. and SIDORAVICIUS, V. (2009). The structure of typical clusters in large sparse random configurations. *J. Stat. Phys.* **135** 87–105. MR2505727

[5] BHAMIDI, S. (2008). First passage percolation on locally treelike networks. I. Dense random graphs. *J. Math. Phys.* **49** 125218. MR2484349

[6] BHAMIDI, S., VAN DER HOFSTAD, R. and HOOGHIEMSTRA, G. (2009). Extreme value theory, Poisson–Dirichlet distributions and first passage percolation on random networks. Preprint.

[7] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1989). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. MR1015093

[8] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966

[9] BRAUNSTEIN, L. A., BULDYREV, S. V., COHEN, R., HAVLIN, S. and STANLEY, H. E. (2003). Optimal paths in disordered complex networks. *Phys. Rev. Lett.* **91** 168701.

[10] BÜHLER, W. J. (1971). Generations and degree of relationship in supercritical Markov branching processes. *Z. Wahrsch. Verw. Gebiete* **18** 141–152. MR0295444

[11] CHUNG, F. and LU, L. (2003). The average distance in a random graph with given expected degrees. *Internet Math.* **1** 91–113. MR2076728

[12] CHUNG, F. and LU, L. (2006). *Complex Graphs and Networks. CBMS Regional Conference Series in Mathematics* **107**. Amer. Math. Soc., Providence, RI. MR2248695

[13] COHEN, R. and HAVLIN, S. (2003). Scale-free networks are ultrasmall. *Phys. Rev. Lett.* **90** 058701.

[14] DEIJFEN, M., VAN DEN ESKER, H., VAN DER HOFSTAD, R. and HOOGHIEMSTRA, G. (2009). A preferential attachment model with random initial degrees. *Ark. Mat.* **47** 41–72. MR2480915

[15] DING, J., KIM, J. H., LUBETZKY, E. and PERES, Y. (2009). Diameters in supercritical random graphs via first passage percolation. Preprint. Available at http://arxiv.org/abs/0906.1840.

[16] DOMMERS, S., VAN DER HOFSTAD, R. and HOOGHIEMSTRA, G. (2010). Diameters in preferential attachment graphs. *J. Stat. Phys.* **139** 72–107.

[17] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge Univ. Press, Cambridge. MR2271734

[18] FERNHOLZ, D. and RAMACHANDRAN, V. (2007). The diameter of sparse random graphs. *Random Structures Algorithms* **31** 482–516. MR2362640

[19] GUT, A. (2005). *Probability: A Graduate Course*. Springer, New York. MR2125120

[20] HAMMERSLEY, J. M. and WELSH, D. J. A. (1965). First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In *Bernouilli–Bayes–Laplace: Anniversary Volume* 61–110. Springer, New York. MR0198576

[21] HOWARD, C. D. (2004). Models of first-passage percolation. In *Probability on Discrete Structures. Encyclopaedia Math. Sci.* **110** 125–173. Springer, Berlin. MR2023652

[22] JANSON, S. (1999). One, two and three times $\log n/n$ for paths in a complete graph with random weights. *Combin. Probab. Comput.* **8** 347–361. MR1723648

[23] JANSON, S. (2009). The probability that a random multigraph is simple. *Combin. Probab. Comput.* **18** 205–225. MR2497380

[24] JANSON, S. and LUCZAK, M. J. (2009). A new approach to the giant component problem. *Random Structures Algorithms* **34** 197–216. MR2490288

[25] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. Wiley, New York. MR1782847

[26] MOLLOY, M. and REED, B. (1995). A critical point for random graphs with a given degree sequence. In *Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs'93" (Poznań, 1993)* **6** 161–179. Wiley, New York. MR1370952

[27] MOLLOY, M. and REED, B. (1998). The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.* **7** 295–305. MR1664335

[28] NORROS, I. and REITTU, H. (2006). On a conditionally Poissonian graph process. *Adv. in Appl. Probab.* **38** 59–75. MR2213964

[29] REITTU, H. and NORROS, I. (2004). On the power law random graph model of massive data networks. *Performance Evaluation* **55** 3–23.

[30] SMYTHE, R. T. and WIERMAN, J. C. (1978). *First-Passage Percolation on the Square Lattice. Lecture Notes in Math.* **671**. Springer, Berlin.

[31] VAN DEN ESKER, H., VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. and ZNAMENSKI, D. (2005). Distances in random graphs with infinite mean degrees. *Extremes* **8** 111–141. MR2275914

[32] VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. and VAN MIEGHEM, P. (2001). First-passage percolation on the random graph. *Probab. Engrg. Inform. Sci.* **15** 225–237. MR1828576

[33] VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. and VAN MIEGHEM, P. (2002). The flooding time in random graphs. *Extremes* **5** 111–129. MR1965974

[34] VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. and VAN MIEGHEM, P. (2005). Distances in random graphs with finite variance degrees. *Random Structures Algorithms* **27** 76–123. MR2150017

[35] VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. and ZNAMENSKI, D. (2007). Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Probab.* **12** 703–766 (electronic). MR2318408

[36] VAN MIEGHEM, P., HOOGHIEMSTRA, G. and VAN DER HOFSTAD, R. (2000). A scaling law for the hopcount. Technical report 2000125, Delft Univ. Technology. Available at http://www.nas.ewi.tudelft.nl/people/Piet/.

[37] WÄSTLUND, J. (2006). Random assignment and shortest path problems. In *Proceedings of the Fourth Colloquium on Mathematics and Computer Science, Institut. Élie Cartan, Nancy France. DMTCS Proc.* **AG** 31–38.

S. BHAMIDI
DEPARTMENT OF STATISTICS
AND OPERATIONS RESEARCH
304 HANES HALL
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599
USA
E-MAIL: bhamidi@email.unc.edu

R. VAN DER HOFSTAD
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

G. HOOGHIEMSTRA
EEMCS
DELFT UNIVERSITY OF TECHNOLOGY
MEKELWEG 4
2628 CD DELFT
THE NETHERLANDS
E-MAIL: g.hooghiemstra@tudelft.nl