*The Annals of Applied Probability*
2018, Vol. 28, No. 3, 1739–1792
https://doi.org/10.1214/17-AAP1342


# TYPICAL DISTANCES IN THE DIRECTED CONFIGURATION MODEL

BY PIM VAN DER HOORN[1] AND MARIANA OLVERA-CRAVIOTO

*Northeastern University, Boston and University of California, Berkeley*

We analyze the distribution of the distance between two nodes, sampled uniformly at random, in digraphs generated via the directed configuration model, in the supercritical regime. Under the assumption that the covariance between the in-degree and out-degree is finite, we show that the distance grows logarithmically in the size of the graph. In contrast with the undirected case, this can happen even when the variance of the degrees is infinite. The main tool in the analysis is a new coupling between a breadth-first graph exploration process and a suitable branching process based on the Kantorovich–Rubinstein metric. This coupling holds uniformly for a much larger number of steps in the exploration process than existing ones, and is therefore of independent interest.

## CONTENTS



Received November 2015; revised August 2017.

[1]Supported by ARO Grant number W911NF1610391.

*MSC2010 subject classifications.* Primary 05C80; secondary 60B10.

*Key words and phrases.* Random digraphs, directed configuration model, typical distances, branching processes, couplings, Kantorovich–Rubinstein distance.

**1. Introduction.** When proposing a mathematical model for studying the typical characteristics of complex networks, one of the first things to try to mimic is the degree distribution, that is, the proportion of nodes having a certain number of neighbors. Perhaps the easiest way to do this, is by sampling a random graph from a prescribed degree sequence through the configuration or pairing model, originally introduced and analyzed in Bollobás (1980), Wormald (1978). In the undirected case, the construction of the graph begins by assigning to each node a number of stubs or half-edges according to the given degree sequence, and determines the edges by randomly pairing the stubs, each time by choosing uniformly among all the unpaired stubs. Conditionally on the resulting graph having no multiple edges or self-loops, it is well known that it has the distribution of a uniformly chosen graph among all those having the corresponding degree sequence [see, e.g., Bollobás (2001), Van Der Hofstad (2016)]. In the directed setting, each node is given a number of inbound and outbound stubs according to its in-degree and out-degree, and the pairing is done by matching an inbound half-edge with an outbound one. Again, conditionally on having no self-loops or multiple edges in the same direction, the resulting graph is uniformly chosen among those having the prescribed degrees.

The versatility of the configuration model and its ability to match any prescribed degree distribution makes it useful for analyzing the structural properties of networks as well as of processes on them Goh et al. (2003), Miller (2009), Newman (2002), Chen, Litvak and Olvera-Cravioto (2017). One such property is the typical distance between nodes. In particular, for the undirected configuration model constructed from an i.i.d. degree sequence, it is known that the hopcount between two randomly chosen nodes in a graph with $n$ nodes, conditioned on them being in the same component, grows logarithmically in $n$ when the degree distribution has finite variance [van der Hofstad, Hooghiemstra and Van Mieghem (2005), van den Esker, van der Hofstad and Hooghiemstra (2008)], as $\log\log n$ when it has infinite variance but finite mean [van der Hofstad, Hooghiemstra and Znamenski (2007)], and is bounded if the mean is infinite [van den Esker et al. (2005)]. These results reflect what has been observed in many real networks, that is, the typical distance between connected nodes is very small compared to the size of the network, and that this distance gets shorter the more variable the degrees are.

In this paper, we provide an analysis of the distance between two randomly chosen nodes in the supercritical directed configuration model,[2] conditioned on the existence of a directed path from one to the other, under the assumption that the covariance between in- and out-degree is finite. We focus on the supercritical regime, since the existence of a directed path between two randomly selected nodes is a rare event in the critical and subcritical regimes. The directed nature of the graphs introduces some subtle differences compared to the undirected case,

[2]The supercritical regime ensures the existence of a giant strongly connected component.

starting with the problem of constructing degree sequences having a prescribed joint distribution. More precisely, in the undirected configuration model one can obtain a degree sequence having distribution $F$ by simply sampling i.i.d. observations from $F$ and adding one to the last node in case the sum is odd [Arratia and Liggett (2005)]. For the directed case, on the other hand, one needs to guarantee that the sum of the in-degrees is equal to that of the out-degrees, an event that can have asymptotically zero probability (e.g., when the in-degree and out-degree are allowed to be different, and nodes are independent).

A more important difference between the undirected and directed cases is that the dependence between the in- and out-degree in the latter plays an important role in the behavior of the distance between nodes. More precisely, the main contribution of this paper is a theorem stating that the hopcount, that is, the length of the shortest directed path between two nodes, grows logarithmically in the number of nodes, which unlike in the undirected case, can occur even when the variance of the degrees is infinite. Intuitively, the length of the shortest directed path between any two nodes will always be larger than the shortest undirected path. However, what is surprising, is that this distance does not necessarily get shorter as the variability of the degrees grows larger, and whether it gets shorter or not depends on the level of dependence between the in- and out-degree. Together with prior results on the existence and the size of a giant strongly connected component in random directed graphs [Cooper and Frieze (2004), Penrose (2016)], our results provide valuable insights into the differences and similarities between the directed and undirected cases.

The second contribution of the paper is a novel coupling between a breadth-first graph exploration process and a Galton–Watson tree. This coupling is based on the Kantorovich–Rubinstein distance between two probability measures [see, e.g., Villani (2008)], and has the advantage of being uniformly accurate for a considerably longer time than existing constructions. Specifically, the coupling holds for a number of steps in the graph exploration process equivalent to discovering $n^{1-\epsilon}$ nodes, for arbitrarily small $\epsilon > 0$, compared to a constant number of nodes in Penrose (2016), $n^{1/2-\epsilon}$ nodes in Norros, Reittu et al. (2006) and Durrett (2010) (Theorem 2.2.2), or $n^{1/2+\epsilon_0}$ nodes, for a very small $\epsilon_0 > 0$, in van der Hofstad, Hooghiemstra and Van Mieghem (2005), van den Esker, van der Hofstad and Hooghiemstra (2008). Moreover, the coupled branching process has a deterministic offspring distribution that does not depend on $n$ or the degree sequences, avoiding the need to consider intermediate tree constructions. The generality of our main coupling result, and the wide range of applications where a so-called branching process argument is used, makes it of independent interest.

The paper is organized as follows: Section 2 contains an overview of our results for the typical distance between two randomly chosen nodes, with the main theorem presented in Section 2.1. The corresponding assumptions are given in terms of the realized degree sequences, that is, the fixed degree sequences from which the graphs are constructed according to the pairing model. In Section 3, we provide an algorithm that can be used to generate degree sequences satisfying our main assumptions for any prescribed joint distribution. We also include in that section numerical examples validating the accuracy of our theoretical approximations for the hopcount. Our coupling results are given in Section 4, and in Section 5 we give a more detailed derivation of the main theorem. All the proofs are postponed until Section 6.

**2. Notation and main results.** Throughout the paper, we consider a directed random graph generated via the directed configuration model (DCM), that is, given two sequences $\{d_1^-, d_2^-, \dots, d_n^-\}$ and $\{d_1^+, d_2^+, \dots, d_n^+\}$ of nonnegative integers satisfying

$$l_n = \sum_{i=1}^{n} d_i^- = \sum_{i=1}^{n} d_i^+,$$

we construct the graph by assigning to each node $i \in \{1, 2, \dots, n\}$ a number of inbound and outbound half-edges according to $(d_i^-, d_i^+)$, respectively. To determine the edges in the graph, we pair each inbound stub with an outbound stub chosen uniformly at random among all unpaired stubs. This pairing process is equivalent to matching the inbound half-edges with a permutation, uniformly chosen at random, of the outbound half-edges. We refer to the sequence $(\mathbf{d}^-, \mathbf{d}^+) = (\{d_1^-, \dots, d_n^-\}, \{d_1^+, \dots, d_n^+\})$ as the bi-degree sequence of the graph.

Our analysis of the typical distances in the DCM will be done in the large graph limiting regime, that is, when the number of nodes $n \to \infty$. This means that we are considering a sequence of graphs, indexed by $n$, each having its own bi-degree sequence, say $(\mathbf{d}_n^-, \mathbf{d}_n^+) = (\{d_{n,1}^-, \dots, d_{n,n}^-\}, \{d_{n,1}^+, \dots, d_{n,n}^+\})$.

As mentioned in the Introduction, sampling a bi-degree sequence $(\mathbf{d}_n^-, \mathbf{d}_n^+)$ having a prescribed joint distribution is not as straightforward as in the undirected case, so we allow the bi-degree sequence itself to be generated through a random process, as long as the realized bi-degree sequence satisfies our regularity conditions with high probability. To emphasize the possibility that the bi-degree sequence may itself be random, we will use the notation $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ to refer to the bi-degree sequence of a graph on $n$ nodes. In particular, we use $D_i^-$ and $D_i^+$ to denote the in-degree and out-degree, respectively, of node $i$, and use $L_n = \sum_{i=1}^n D_i^- = \sum_{i=1}^n D_i^+$ to denote the total number of edges in the graph. To show that bi-degree sequences satisfying our main assumptions are easy to construct, we provide in Section 3.1 an algorithm based on i.i.d. samples from the prescribed degree distribution.

In view of our previous remarks, we need to be able to distinguish between the unconditional probability space and the conditional probability space given the bi-degree sequence $(\mathbf{D}_n^-, \mathbf{D}_n^+)$. To this end, let $\mathscr{F}_n$ denote the sigma-algebra generated by the bi-degree sequence $(\mathbf{D}_n^-, \mathbf{D}_n^+)$, and define $\mathbb{P}_n$ and $\mathbb{E}_n$ to be the corresponding conditional probability and expectation, respectively, given $\mathscr{F}_n$, that is, $\mathbb{P}_n(\cdot) =$

$E[1(\cdot)|\mathscr{F}_n]$ and $\mathbb{E}_n[\cdot] = E[\cdot|\mathscr{F}_n]$. We point out that under the probability $\mathbb{P}_n$, the bi-degree sequence is fixed, as in the classical configuration model.

Before we can state the assumptions imposed in our main theorems, we need to define the following (random) probability mass functions:

$$g_n^+(t) = \frac{1}{n}\sum_{r=1}^{n} 1(D_r^+ = t), \qquad g_n^-(t) = \frac{1}{n}\sum_{r=1}^{n} 1(D_r^- = t),$$

$$f_n^+(t) = \frac{1}{L_n}\sum_{r=1}^{n} 1(D_r^+ = t)D_r^-, \qquad f_n^-(t) = \frac{1}{L_n}\sum_{r=1}^{n} 1(D_r^- = t)D_r^+,$$

for $t = 0, 1, 2, \dots$, and let $G_n^+, G_n^-, F_n^+, F_n^-$ denote their corresponding cumulative distribution functions.

We point out that the probability mass functions $g_n^+$ and $g_n^-$ correspond to the marginal distributions of the out-degree and in-degree, respectively, of a uniformly chosen node in the graph, while $f_n^+$ (resp. $f_n^-$) is the distribution of the out-degree (resp., in-degree) of a uniformly chosen inbound (resp., outbound) neighbor of that node, also known as the size-biased out-degree (resp., in-degree) distribution.

*Notation*: Throughout the manuscript, we use the superscript $\pm$ to mean that the property/result holds for the distributions or random variables with the $\pm$ symbol substituted consistently with either the $+$ or $-$ symbol.

The main assumption needed throughout the paper is given below.

ASSUMPTION 2.1. The bi-degree sequence $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ satisfies:

(a) There exist probability mass functions $g^+, g^-, f^+$ and $f^-$ on the nonnegative integers, such that, for some $\varepsilon > 0$,

$$\sum_{k=0}^{\infty}\left|\sum_{i=0}^{k}\bigl(g_n^{\pm}(i) - g^{\pm}(i)\bigr)\right| \le n^{-\varepsilon} \quad \text{and} \quad \sum_{k=0}^{\infty}\left|\sum_{i=0}^{k}\bigl(f_n^{\pm}(i) - f^{\pm}(i)\bigr)\right| \le n^{-\varepsilon},$$

with $\nu \triangleq \sum_{j=0}^{\infty} jg^+(j) = \sum_{j=0}^{\infty} jg^-(j) < \infty$ and $\mu \triangleq \sum_{j=0}^{\infty} jf^+(j) = \sum_{j=0}^{\infty} jf^-(j) \in (1, \infty)$.

(b) For some $0 < \kappa \le 1$ and some constant $K_\kappa < \infty$,

$$\sum_{r=1}^{n}\bigl((D_r^-)^\kappa + (D_r^+)^\kappa\bigr)D_r^+ D_r^- \le K_\kappa n.$$

REMARK 2.2. Note that by requiring that $\mu > 1$, we are assuming that the graph is in the supercritical regime, where with high probability there exists a unique strongly connected component of linear size; see Cooper and Frieze (2004), Penrose (2016). In this regime, the probability that there exists a directed path between the two randomly chosen nodes is asymptotically positive, while it is a rare event in both the critical ($\mu = 1$) and subcritical ($\mu < 1$) cases.

To provide some insights into these assumptions and relate them to the construction of the coupling in Section 4, it is useful to define first the Kantorovich–Rubinstein distance (also known as Wasserstein metric of order one), which is a metric on the space of probability measures. In particular, convergence in this sense is equivalent to weak convergence plus convergence of the first absolute moments.

DEFINITION 2.3. Let $M(\mu, \nu)$ denote the set of joint probability measures on $\mathbb{R} \times \mathbb{R}$ with marginals $\mu$ and $\nu$. Then the Kantorovich–Rubinstein distance between $\mu$ and $\nu$ is given by

$$d_1(\mu, \nu) = \inf_{\pi \in M(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} |x - y| \, d\pi(x, y).$$

We point out that $d_1$ is only strictly speaking a distance when both $\mu$ and $\nu$ have finite first absolute moments. Moreover, it is well known that

$$d_1(\mu, \nu) = \int_0^1 \left|F^{-1}(u) - G^{-1}(u)\right| du = \int_{-\infty}^{\infty} \left|F(x) - G(x)\right| dx,$$

where $F$ and $G$ are the cumulative distribution functions of $\mu$ and $\nu$, respectively, and $f^{-1}(t) = \inf\{x \in \mathbb{R} : f(x) \geq t\}$ denotes the pseudo-inverse of $f$. It follows that the optimal coupling of two real random variables $X$ and $Y$ is given by $(X, Y) = (F^{-1}(U), G^{-1}(U))$, where $U$ is uniformly distributed in $[0, 1]$.

With some abuse of notation, for two distribution functions $F$ and $G$ we use $d_1(F, G)$ to denote the Kantorovich–Rubinstein distance between their corresponding probability measures. We refer the interested reader to Villani (2008) for more details.

REMARK 2.4. (i) In terms of the previous definition, the first condition in Assumption 2.1 can also be written as

$$d_1(G_n^{\pm}, G^{\pm}) \leq n^{-\varepsilon} \quad \text{and} \quad d_1(F_n^{\pm}, F^{\pm}) \leq n^{-\varepsilon}.$$

Furthermore, since

$$\nu_n = \frac{L_n}{n} \quad \text{and} \quad \mu_n = \frac{1}{L_n} \sum_{r=1}^{n} D_r^- D_r^+$$

are the common means of $g_n^+$, $g_n^-$, and $f_n^+$, $f_n^-$, respectively, it follows from Definition 2.3 that

$$|\nu_n - \nu| \leq n^{-\varepsilon} \quad \text{and} \quad |\mu_n - \mu| \leq n^{-\varepsilon}.$$

Hence, the first set of assumptions simply state that the empirical degree distributions and the empirical size-biased degree distributions converge weakly, along with their means.

(ii) The second condition in Assumption 2.1 implies that

$$\sum_{i=0}^{\infty} i^{1+\kappa} f^{\pm}(i) \leq \liminf_{n\to\infty} \sum_{i=0}^{\infty} i^{1+\kappa} f_n^{\pm}(i) \leq K_\kappa/\nu < \infty,$$

that is, $f^{\pm}$ has finite moments of order $1+\kappa$.

(iii) We point out that any bi-degree sequence satisfying Assumption 2.1 will also be "proper" in the sense of Cooper and Frieze (2004), provided that $g^+$ and $g^-$ have finite variance and that the maximum degree is smaller or equal than $n^{1/2}/\log n$. Hence, under these additional conditions, the results in Cooper and Frieze (2004) regarding the bow-tie structure of the supercritical directed configuration model hold.

Since, as mentioned earlier, the bi-degree sequence $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ may itself be generated through a random process, we only require that Assumption 2.1 holds with high probability. More precisely, if we let

$$\Omega_n = \{\max\{d_1(G_n^+, G^+), d_1(G_n^-, G^-), d_1(F_n^+, F^+), d_1(F_n^-, F^-)\} \leq n^{-\varepsilon}\}$$
$$\cap \left\{\sum_{r=1}^{n} ((D_r^-)^\kappa + (D_r^+)^\kappa) D_r^+ D_r^- \leq K_\kappa n\right\},$$

then our condition will be that $P(\Omega_n) \to 1$ as $n \to \infty$. In Section 3.1, we show that the *i.i.d. algorithm* presented there satisfies this condition.

2.1. *Main result.* Our main result, Theorem 2.5 below, establishes that the distance between two randomly chosen nodes grows logarithmically in the size of the graph, and characterizes the spread around the logarithmic term.

In the statement of our results, we use $H_n$ to denote the hopcount, or distance, between two randomly chosen nodes in a graph of size $n$. Since the graph is directed, we say that the hopcount between node $i$ and node $j$ is $k$ if there exists a directed path of length $k$ from $i$ to $j$; if there is no directed path from $i$ to $j$ we say that the hopcount is infinite. Since the two nodes are chosen at random, we can assume without loss of generality that $H_n$ is the hopcount from the first node to the second one.

The last thing we need to do before stating Theorem 2.5 is to introduce the limiting random variables appearing in the characterization of the hopcount. To this end, let $g^{\pm}$ and $f^{\pm}$ be the probability mass functions from Assumption 2.1. Throughout the paper, we will use $\{\hat{Z}_k^{\pm} : k \geq 0\}$, $\hat{Z}_0^{\pm} = 1$, to denote a delayed Galton–Watson process where nodes in the tree have offspring according to distribution $f^{\pm}$, with the exception of the root node which has a number of offspring distributed according to $g^{\pm}$. Note that $W_k^{\pm} = \hat{Z}_k^{\pm}/(\nu\mu^{k-1})$ is a mean one martingale with respect to the filtration generated by the process $\{\hat{Z}_k^{\pm} : k \geq 1\}$. Hence, by the martingale convergence theorem,

$$W^{\pm} = \lim_{k\to\infty} \hat{Z}_k^{\pm}/(\nu\mu^{k-1}) \qquad \text{a.s.}$$

exists and satisfies $E[W^\pm] \leq 1$.

To see that under Assumption 2.1 $W^+$ and $W^-$ are nontrivial, it is useful to define first $\{\mathcal{Z}_k^\pm : k \geq 0\}$ to be a (nondelayed) Galton–Watson process having offspring distribution $f^\pm$ and let

$$\mathcal{W}^\pm = \lim_{k\to\infty} \mathcal{Z}_k^\pm/\mu^k \qquad \text{a.s.}$$

be its corresponding martingale limit. Now recall from Remark 2.4 that $f^+$ and $f^-$ have finite moments of order $1+\kappa > 1$, which implies that $\sum_{j=1}^\infty j(\log j) f^\pm(j) < \infty$, a necessary and sufficient condition for $\mathcal{W}^\pm$ to be nontrivial and satisfy $E[\mathcal{W}^\pm] = 1$ [see, e.g., Athreya and Ney (2004)]. Moreover, if $q^\pm = P(\mathcal{Z}_m^\pm = 0 \text{ for some } m)$ denotes the probability of extinction of $\{\mathcal{Z}_k^\pm : k \geq 0\}$, then $q^\pm < 1$ and $P(\mathcal{W}^\pm = 0) = q^\pm$ [Lemma 6.1 contains an expression for $P(W^\pm = 0)$]. Furthermore, provided $f^\pm$ is not degenerate, $\mathcal{W}^\pm$ possesses a density on $(0,\infty)$ [see Athreya and Ney (2004) p. 52], which implies that $W^\pm$ does as well. Interestingly, the degenerate case appears when studying $d$-regular graphs (i.e., where all nodes have in-degree $d$ and out-degree $d$), in which case $W^\pm = \mathcal{W}^\pm \equiv 1$ a.s. Hence, the randomness of $W^\pm$ is due to the variability of the degrees.

We are now ready to state the main result of the paper; $\lfloor x \rfloor$ denotes the largest integer smaller or equal to $x$.

THEOREM 2.5. *Let $\{\mathcal{G}_n : n \geq 1\}$ be a sequence of graphs generated through the DCM from a sequence of bi-degree sequences $\{(\mathbf{D}_n^-, \mathbf{D}_n^+) : n \geq 1\}$ satisfying $P(\Omega_n) \to 1$ as $n \to \infty$. Let $H_n$ denote the hopcount between two randomly chosen nodes in $\mathcal{G}_n$. Then there exist random variables $\{\mathcal{H}_n\}_{n\in\mathbb{N}}$ such that for each* (*fixed*) *$t \in \mathbb{Z}$,*

$$\lim_{n\to\infty} \left| P\left(H_n - \lfloor \log_\mu n \rfloor = t \,|\, H_n < \infty\right) - P(\mathcal{H}_n = t)\right| = 0, \tag{2.1}$$

*where $\mathcal{H}_n$ has distribution*

$$P(\mathcal{H}_n \leq x) = 1 - E\left[\exp\left\{-\frac{\nu}{\mu - 1} \cdot \frac{\mu^{\lfloor \log_\mu n \rfloor + \lfloor x \rfloor}}{n} W^+W^-\right\} \Big| W^+W^- > 0\right], \quad x \in \mathbb{R}. \tag{2.2}$$

Theorem 2.5 shows that the hopcount between two randomly chosen nodes, conditionally on it being finite, is $\lfloor \log_\mu n \rfloor$ plus a random fluctuation having the same distribution as $\mathcal{H}_n$. This variation in the hopcount length comes from the specific locations of the randomly chosen nodes within the graph, and the distribution of $\mathcal{H}_n$ is determined by the randomness of $W^+$ and $W^-$, which in turn is determined by that of $f^+$ and $f^-$. As pointed out earlier, $W^+$ and $W^-$ become deterministic when analyzing $d$-regular graphs, in which case (2.2) can be explicitly computed.

As a straightforward corollary, we obtain the asymptotic equivalence of $H_n$ and $\log_\mu n$ in probability.

COROLLARY 2.6. *Under the same assumptions as Theorem 2.5, and for any* $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{H_n}{\log_\mu n} - 1\right| > \epsilon \middle| H_n < \infty\right) = 0.$$

Theorem 2.5 shows that the directed distance between two randomly chosen nodes in the DCM scales logarithmically in the size of the graph, which is consistent with existing results for the undirected configuration model (CM) under the assumption that the degree distribution has finite variance [van der Hofstad, Hooghiemstra and Van Mieghem (2005)]. We remark that no assumption is made concerning the simplicity of the graph, since the hopcount is unaffected by the existence of multiple edges and self-loops. For instance, removing all self-loops and merging all duplicate edges into a single edge, as is done in the erased configuration model, will not change the hopcount.

To understand the result of Theorem 2.5, including the appearance of the martingale limits $W^{\pm}$, note that directed graphs with high connectivity consist of a strongly connected component (SCC), a set of nodes with directed paths going into the SCC (the inbound wing), a set of nodes with directed paths exiting the SCC (the outbound wing), and some additional secondary structures. This, so-called, bow-tie structure has been observed experimentally in the web graph Broder et al. (2000), and has been established for the supercritical directed configuration model in Cooper and Frieze (2004); see also Timár et al. (2017) for a more detailed analysis of the secondary structures. More precisely, the work in Cooper and Frieze (2004) shows that the inbound wing consists of nodes whose out-component is of linear size but whose in-component is small [i.e., of order $o(n)$], the outbound wing consists of nodes whose in-component is of linear size but whose out-component is small, and the SCC is the set of nodes having both linear size in-component and linear size out-component. The branching processes $\hat{Z}_k^+$ and $\hat{Z}_k^-$ describe the breadth-first exploration process of the out-component of the first randomly chosen node and the in-component of the second one, respectively, whose sizes are approximately $W^+\nu\mu^{k-1}$ and $W^-\nu\mu^{k-1}$. We refer the reader to van der Hofstad, Hooghiemstra and Znamenski (2007) (pp. 712–714) for a more detailed explanation relating the hopcount with the branching processes appearing in the limit.

The interesting difference between the directed and undirected cases lies in the observation that Assumption 2.1 can hold with high probability for degree sequences having infinite variance (as shown in Section 3.1), hence showing that the distance remains logarithmic even when in its undirected counterpart becomes of order $\log\log n$ [van der Hofstad, Hooghiemstra and Znamenski (2007)]. To explain this, note that distances in the CM get smaller as the degree distribution gets heavier (i.e., more variable) presumably because of the appearance of nodes with extremely large degrees that should create shortcuts between nodes in their connected component.[3] In contrast, when the graph is directed, increasing the variability of the in- and out-degree distributions does not necessarily imply the appearance of more shortcuts, since even if there are more nodes with very large in-degrees or very large out-degrees, they may not be the same nodes, for example, when the in-degree is independent of the out-degree, it is unlikely that a node has both large in-degree and large out-degree. Our results are consistent with the intuition that if the nodes with very large in-degrees are the same as those with very large out-degrees (i.e., positively correlated in- and out-degrees), then more shortcuts should be created and the distances will get smaller.

To complement the main theorem, we also compute the asymptotic probability that the hopcount is finite, which can be expressed in terms of the survival properties of the delayed branching processes $\{\hat{Z}_k^+ : k \geq 1\}$ and $\{\hat{Z}_k^- : k \geq 1\}$.

PROPOSITION 2.7. *Let $\{\mathcal{G}_n : n \geq 1\}$ be a sequence of graphs generated through the DCM from a sequence of bi-degree sequences $\{(\mathbf{D}_n^-, \mathbf{D}_n^+) : n \geq 1\}$ satisfying $P(\Omega_n) \to 1$ as $n \to \infty$ let $H_n$ denote the hopcount between two randomly chosen nodes in $\mathcal{G}_n$. Then*

$$\lim_{n\to\infty} P(H_n < \infty) = s^+ s^-, \tag{2.3}$$

*where $s^\pm = P(W^\pm > 0)$.*

To provide some insights into this probability, we refer again to the bow-tie structure of the supercritical directed configuration model, where $S$ is the SCC, $K^-$ is the inbound wing, and $K^+$ is the outbound wing. As the work in Cooper and Frieze (2004) shows, if we let $L^-$ and $L^+$ denote the set of nodes with in-component, respectively out-component, of linear size, then $S = L^- \cap L^+$, $K^- = L^- \cap (L^+)^c$ and $K^+ = L^+ \cap (L^-)^c$. Moreover, the proof of our main coupling result (Theorem 4.1) shows that $s^+$ ($s^-$) is the asymptotic probability that a randomly chosen node in the graph belongs to $L^+$ ($L^-$), which is consistent with Theorem 1.2[4] in Cooper and Frieze (2004), suggesting that the bow-tie structure proved there should hold even under the weaker assumptions of this paper.

With respect to the martingale limits $W^+$ and $W^-$ appearing in (2.2), we point out that although it is in general difficult to compute them analytically, it can easily be done numerically, for example, by using the Population Dynamics algorithm described in Chen, Litvak and Olvera-Cravioto (2017). We use this algorithm in Section 3 below for validating our theoretical results.

[3]The role that high degree nodes play in the creation of shortcuts is best understood through the notion of betweenness centrality, which computes the fraction of shortest paths that go through a given node. For the undirected configuration model, it was shown numerically in Goh et al. (2003) that the betweenness centrality is positively correlated with the degree, which is consistent with our intuitive explanation of why distances get smaller the more spread out the degree distribution becomes.

[4]See Remark 2.4(iii).

**3. Construction of a bi-degree sequence and numerical examples.** To illustrate the accuracy of the approximation for the hopcount between two randomly chosen nodes provided by Theorem 2.5, we give in this section several numerical examples for different choices of the bi-degree sequence. This requires us to construct a sequence of bi-degree sequences $\{(\mathbf{D}_n^-, \mathbf{D}_n^+) : n \geq 1\}$ satisfying Assumption 2.1 with high probability for some prescribed joint distribution for the in- and out-degrees. As pointed out earlier, there are many ways of constructing such sequences, but for the sake of completeness, we include here an algorithm based on i.i.d. samples from the prescribed degree distribution.

3.1. *The i.i.d. algorithm.* Let $G(x, y)$ be a joint distribution function on $\mathbb{N}^2$ such that if $(\mathscr{D}^-, \mathscr{D}^+)$ is distributed according to $G$, then $E[\mathscr{D}^-] = E[\mathscr{D}^+]$, $E[|\mathscr{D}^- - \mathscr{D}^+|^{1+\kappa}] < \infty$ and $E[(\mathscr{D}^-\mathscr{D}^+)^{1+\kappa}] < \infty$ for some $0 < \kappa \leq 1$. Set $\delta = c\kappa/(1+\kappa)$, for some $0 < c < 1$ if $\kappa < 1$ or choose any $0 < \delta < 1/2$ if $\kappa = 1$.

STEP 1: Sample $\{(\mathscr{D}_i^-, \mathscr{D}_i^+)\}_{i=1}^n$ as i.i.d. vectors distributed according to $G(x, y)$.

STEP 2: Define $\Delta_n = \sum_{i=1}^n (\mathscr{D}_i^- - \mathscr{D}_i^+)$. If $|\Delta_n| \leq n^{1-\delta}$, proceed to STEP 3; else, repeat STEP 1.

STEP 3: Select $|\Delta_n|$ indices from $\{1, 2, \ldots, n\}$ uniformly at random (without replacement) and set

$$D_i^- = \mathscr{D}_i^- + \tau_i \quad \text{and} \quad D_i^+ = \mathscr{D}_i^+ + \chi_i, \qquad i = 1, 2, \ldots, n,$$

where

$$\tau_i = 1(\Delta_n \leq 0 \text{ and } i \text{ was selected}) \quad \text{and} \quad \chi_i = 1(\Delta_n > 0 \text{ and } i \text{ was selected}).$$

This algorithm was first introduced in Chen and Olvera-Cravioto (2013) for the special case where $\mathscr{D}^-$ and $\mathscr{D}^+$ are independent. There, it was shown that the degree sequences generated by the algorithm are graphical w.h.p., that is, they can be used to construct simple graphs. Moreover, the empirical joint distribution of the degrees in a simple graph generated through either the *repeated* DCM or the *erased* DCM, converges in probability to $G(x, y)$;[5] see Theorems 2.3 and 2.4 in Chen and Olvera-Cravioto (2013).[6]

Note that the $\{\mathscr{D}_i^- - \mathscr{D}_i^+\}$ are zero-mean random variables with $E[|\mathscr{D}^- - \mathscr{D}^+|^{1+\kappa}] < \infty$. Hence, using Burkholder's inequality (see Lemma 6.2), one obtains that

$$P(|\Delta_n| > n^{1-\delta}) = O(n^{1-(1+\kappa)(1-\delta)}) = O(n^{-(1-c)\kappa}),$$

[5]Note that the joint degree distribution of the nodes in the resulting graph is not $G(x, y)$, since this distribution is changed by STEP 1 and STEP 2 of the algorithm, as well as by the pairing process itself.

[6]These theorems are stated for the case when $\mathscr{D}^+$ and $\mathscr{D}^-$ are independent, but a close look at the proofs shows that they remain valid when they are dependent.

for $\kappa < 1$, while it is $O(n^{2\delta-1})$ when $\kappa = 1$. Hence the probability of success in STEP 2 is $1 - O(n^{-a})$ for some $a > 0$.

We also point out that the i.i.d. algorithm only requires the moment conditions $E[|\mathscr{D}^- - \mathscr{D}^+|^{1+\kappa}] < \infty$ and $E[(\mathscr{D}^-)^{1+\kappa}(\mathscr{D}^+)^{1+\kappa}] < \infty$ and, therefore, can be used to generate any light-tailed degree sequence as well as the vast majority of scale-free (heavy-tailed) degree distributions. It also includes as a special case the $d$-regular bi-degree sequence.

The following result shows that the degree sequences generated by this algorithm satisfy Assumption 2.1 with high probability.

THEOREM 3.1. *Let $G^-$ denote the marginal distribution of $\mathscr{D}^-$ and $G^+$ denote that of $\mathscr{D}^+$; define $F^+(x) = E[1(\mathscr{D}^+ \le x)\mathscr{D}^-]/\nu$ and $F^-(x) = E[1(\mathscr{D}^- \le x)\mathscr{D}^+]/\nu$, where $\nu = E[\mathscr{D}^+] = E[\mathscr{D}^-]$. Then, for any $0 < \varepsilon < \delta$ and $E[(\mathscr{D}^-)^{1+\kappa}\mathscr{D}^+ + \mathscr{D}^-(\mathscr{D}^+)^{1+\kappa}] < K_\kappa < \infty$, we have*

$$\lim_{n\to\infty} P(\Omega_n) = 1.$$

3.2. *The hopcount distribution.* In order to compute the hopcount distribution, we constructed 20 graphs of size $n = 10^6$, using the DCM for different choices of bi-degree sequence. For each of these graphs, we computed the neighborhood function, which gives for each $t > 0$ the number of pairs of nodes at distance at least $t$. For the computation of the neighborhood function, we used the HyperBall algorithm Boldi and Vigna (2013), which is part of the Webgraph Framework Boldi and Vigna (2004). We used HyperBall since it implements the HyperANF algorithm Boldi, Rosa and Vigna (2011), which is designed to give a tight approximation of the neighborhood function of large graphs. From the neighborhood function, we determined, for all finite $t$, the number of shortest paths of length $t$. In this way, we compute the distance between all pairs of nodes, with finite distance, in 20 independently generated graphs. We then took the empirical distribution of these values as a approximation of the hopcount distribution.

We point out that since $H_n$ was defined as the hopcount between two randomly selected nodes, the natural unbiased estimator for the distribution of $H_n$ is the one obtained from randomly selecting pairs of nodes in independent graphs and using the corresponding empirical distribution function. However, this approach is computationally too intensive considering the amount of effort needed to generate one graph. Our approach is considerably more efficient, and although the empirical distribution function it generates does not consist of i.i.d. samples (samples from the same graph are positively correlated), it produces results that are in close agreement with the theoretical approximation in Theorem 2.5. Additional experiments not included in this paper showed that the two approaches produce similar results, with the method used in this paper exhibiting smaller variance.

The three examples below illustrate the accuracy of the approximation provided by Theorem 2.5 for different choices of bi-degree sequences. All three examples are special cases of the i.i.d. algorithm, and thus satisfy Assumption 2.1.

3.2.1. *$d$-regular bi-degree sequence.* A $d$-regular bi-degree sequence satisfies $D_i^+ = d = D_i^-$ for all $1 \le i \le n$. It readily follows that the probability densities $g^\pm$ and $f^\pm$ have just one atom at $d$. Moreover, we have $\hat{Z}_k^\pm = d^k = \mu^k$ for all $k \ge 1$, hence $W^\pm = 1$ and

$$P(\mathcal{H}_n \le x) = 1 - \exp\left\{-\frac{d^{\lfloor \log_d n \rfloor + \lfloor x \rfloor}}{(d-1)n}\right\}, \qquad x \in \mathbb{R}. \tag{3.1}$$

In Figure 1(a), we plotted the probability mass functions of both the hopcount distribution and that of its theoretical limit (3.1). The plots are indistinguishable in the figure, with a Kolmogorov–Smirnov distance of $1.3 \times 10^{-4}$. This shows that for nonrandom sequences, the approximation provided by Theorem 2.5 is almost exact.

3.2.2. *I.I.D. bi-degree sequence with independent in- and out-degrees.* Following the result from Theorem 3.1, we computed the hopcount distribution for bi-degree sequences, generated by the i.i.d. algorithm, using as the in- and out-degree distributions Poisson mixed with Pareto rates, and keeping the in-degree and out-degree independent of each other. More precisely, we chose $\Lambda_1$ and $\Lambda_2$ to be independent Pareto random variables, both with scale parameter 1 and shape parameter $3/2$, and then set $\mathscr{D}^-$ and $\mathscr{D}^+$ to be i.i.d. with conditional distributions

$$P(\mathscr{D}^- = k | \Lambda_1 = \lambda) = P(\mathscr{D}^+ = k | \Lambda_2 = \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad k = 0, 1, 2, \dots.$$

It can be verified [see Proposition 8.4 in Grandell (1997)] that

$$P(\mathscr{D}^- \ge k) \sim c_1 k^{-3/2} \quad \text{and} \quad P(\mathscr{D}^+ \ge k) \sim c_2 k^{-3/2},$$

as $k \to \infty$, for some constants $c_1, c_2 > 0$.

Note that the independence between $\mathscr{D}^-$ and $\mathscr{D}^+$ implies that the size-biased distributions $f^+$ and $f^-$ are equal to the unbiased ones, that is, $f^\pm = g^\pm$. Hence, $\mu = \nu = 3$ and the branching processes $\{\hat{Z}_k^+ : k \ge 1\}$ and $\{\hat{Z}_k^- : k \ge 1\}$ are not delayed.

In order to compute our theoretical approximation for the hopcount, we also need to compute $P(\mathcal{H}_n > k)$, which is written in terms of $W^+$ and $W^-$. Since $W^+$ and $W^-$ are not known in general, we estimate them numerically using the approach from Chen, Litvak and Olvera-Cravioto (2017), which describes a bootstrap algorithm for simulating the endogenous solutions of branching linear recursions. For this, we first observe that $W^+$ and $W^-$ satisfy the following stochastic fixed-point equations:

$$W^- \stackrel{d}{=} \sum_{i=1}^{\mathscr{D}^-} \frac{W_i^-}{\mu} \quad \text{and} \quad W^+ \stackrel{d}{=} \sum_{i=1}^{\mathscr{D}^+} \frac{W_i^+}{\mu},$$

where $W_i^\pm$ are i.i.d. copies of $W^\pm$, independent of $\mathscr{D}^-$ and $\mathscr{D}^+$. Using the algorithm in Chen, Litvak and Olvera-Cravioto (2017) for 30 generations of the trees

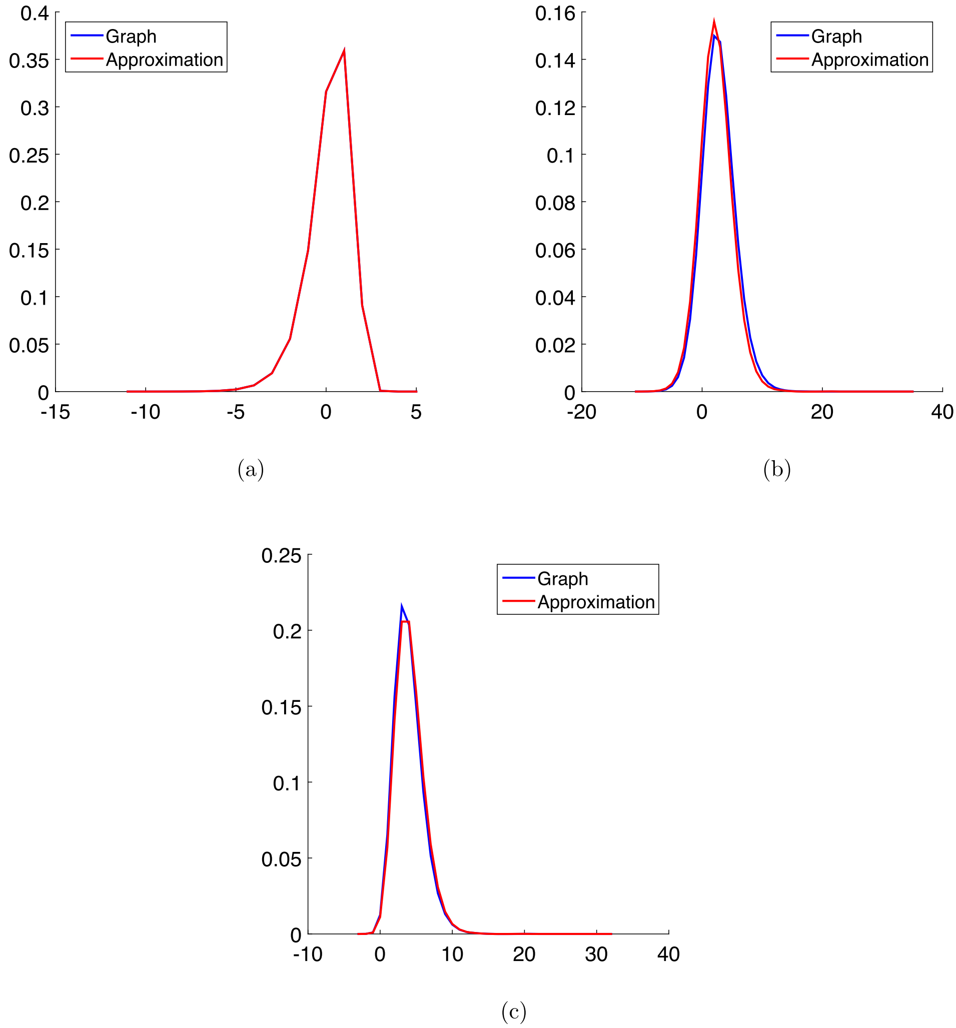


FIG. 1. *Hopcount probability mass function compared to the approximation provided by Theorem* 2.5 *for*: (a) *a* 3*-regular bi-degree sequence*; (b) *a bi-degree sequence generated by the i.i.d. algorithm with independent in- and out-degrees*; *and* (c) *a bi-degree sequence generated by the i.i.d. algorithm with dependent in- and out-degrees*. *The Kolmogorov–Smirnov distance in each case is*: (a) $1.3 \times 10^{-4}$, (b) 0.0583, *and* (c) 0.0353. *In all cases the graphs had* $n = 10^6$ *nodes*.

with a sample pool of size $10^6$, we obtained $10^6$ observations for each of $W^+$ and $W^-$, with the sample for $W^+$ independent of that for $W^-$. We then used these samples to estimate

$$E\left[\exp\left\{-\frac{\nu}{\mu-1}\cdot\frac{\mu^{\lfloor\log_\mu n\rfloor+k}}{n}W^+W^-\right\}\middle|W^+W^->0\right] \qquad \text{for } k=0,1,\ldots.$$

The results for the hopcount distribution are shown in Figure 1(b). The Kolmogorov–Smirnov distance in this case is 0.0583.

3.2.3. *I.I.D. bi-degree sequence with dependent in- and out-degrees*. Our third and last example is for a bi-degree sequence obtained using the i.i.d. algorithm but for the case where $\mathscr{D}^-$ and $\mathscr{D}^+$ are dependent. We take the extreme case where $D_i^- = D_i^+$ for all $1 \le i \le n$. To obtain such a sequence, we generate the $D_i^-$ by sampling from a Zipf distribution with corpus size $10^3$ and exponent $7/2$ and set $D_i^+ = D_i^-$, that is,

$$P(\mathscr{D}^+ = t) = t^{-7/2}/\zeta(7/2) \qquad \text{for all } t = 1, 2, \dots,$$

where $\zeta(s)$ is the Riemann zeta function. Observe that since the exponent is larger than 3, the distribution has finite $2 + \varepsilon$ moment, for $0 < \varepsilon < 1/2$. Therefore, it follows from Theorem 3.1 that this bi-degree sequence satisfies Assumption 2.1 with high probability. We used a Zipf distribution here since then the sized-bias distribution will again be Zipf with exponent $5/2$.

The $W^+$ and $W^-$ were again simulated using the algorithm in Chen, Litvak and Olvera-Cravioto (2017) with the same number of generations and the same pool size as for the independent case above, but with the appropriate sized-biased distribution and the corresponding delay for the first generation of the tree.

The results for the hopcount are shown in Figure 1(c), and the Kolmogorov–Smirnov distance is 0.0353.

**4. Coupling with a branching process.** Given a directed graph $\mathcal{G}_n$ of size $n$ the shortest directed path from node $v_1$ to node $v_2$ can be computed by starting two breadth-first exploration processes, one to uncover the out-component of $v_1$, call this $B^+(v_1)$, and another one to uncover the in-component of $v_2$, call it $B^-(v_2)$. If $B^+(v_1) \cap B^-(v_2) \neq \varnothing$, then there exists a finite $(v_1, v_2)$-path, whereas if this intersection is empty, there is none. We point out that since shortest paths do not contain cycles, the exploration of the components, either inbound or outbound, requires only that we keep track of edges with nodes not previously uncovered.

The first step in proving Theorem 2.5 is to couple the breadth-first exploration processes described above, starting from uniformly chosen nodes in $\mathcal{G}_n$, with two independent branching processes. This is a well-known approach for analyzing the properties of random graphs, also referred to as a branching process argument.

The main result of this section is Theorem 4.1, along with its more immediately useful corollary (Corollary 4.2), which is the key ingredient in the proof of Theorem 2.5.

4.1. *Exploration of new stubs*. Similar to the construction in van der Hofstad, Hooghiemstra and Van Mieghem (2005), we start by designating all the $n$ nodes as *inactive*, meaning they have not been uncovered yet, and setting $Z_0^{\pm} = 1$ [note that in van der Hofstad, Hooghiemstra and Van Mieghem (2005) it is the stubs themselves that are labeled, not the nodes]. Let $\varnothing$ denote the fictional first stub, and set $A_0^{\pm} = \{\varnothing\}$; call this initialization step 0. The process $\{Z_k^{\pm} : k \geq 0\}$ will keep track of the number of outbound (inbound) stubs discovered during the $k$th step of the exploration process, as we will now describe. The superscript $\pm$ refers to whether the exploration follows the outbound stubs (for which we use the superscript $+$), or the inbound stubs (for which we use the superscript $-$).

In step 1, we randomly select a node and set $Z_1^{\pm} = j$ if it has $j$ outbound (inbound) stubs; we set its state to *active*, meaning it has already been uncovered. To identify each of the outbound (inbound) stubs, we index them 1 through $j$ and let $A_1^{\pm} = \{1, \dots, j\}$ be the set of the indices of the newly discovered stubs. For the second step of the exploration process, we will need to traverse all $Z_1^{\pm}$ outbound (inbound) stubs, which we do sequentially and in lexicographic order with respect to their indexes. Here, we say that we have traversed an outbound (inbound) stub if we have identified the node it leads to and discovered how many outbound (inbound) stubs this new node has. If the stub is pointing to an *inactive* node, we label the node as *active*, index all its outbound (inbound) stubs with a name of the form $(i, j)$, $j \geq 1$, and then proceed to explore the next outbound (inbound) stub. If the stub is pointing to an *active* node, no new outbound (inbound) stubs are discovered. Once we are done exploring all $Z_1^{\pm}$ outbound (inbound) stubs, we set $Z_2^{\pm}$ to be the number of newly discovered outbound (inbound) stubs and let $A_2^{\pm}$ denote the set of their indices.

In general, in step $k$ we will traverse all $Z_{k-1}^{\pm}$ outbound (inbound) stubs, in lexicographic order, discovering new nodes, and hence new outbound (inbound) stubs. If outbound (inbound) stub $\mathbf{i} = (i_1, \dots, i_{k-1})$ is paired with an inbound (outbound) stub belonging to an *inactive* node, then the outbound (inbound) stubs of the newly discovered node receive an index of the form $(i_1, \dots, i_{k-1}, i_k)$, $i_k \geq 1$; if outbound (inbound) stub $\mathbf{i}$ is paired with an inbound (outbound) stub belonging to an *active* node, then no new outbound (inbound) stubs are discovered. Once we have traversed all $Z_{k-1}^{\pm}$ outbound (inbound) stubs we set $Z_k^{\pm}$ to be the number of new outbound (inbound) stubs discovered in step $k$. The process continues until all $L_n$ outbound (inbound) stubs have been traversed.

Note that the process $\{Z_k^{\pm} : k \geq 0\}$ defines a labeled tree, where the "individuals" are the outbound (inbound) stubs discovered in step $k$ ($Z_0^{\pm} = 1$), not the nodes of the graph themselves. In addition to keeping track of $Z_k^{\pm}$, we will also keep track of "time" in the exploration process, where time $t$ means we have traversed $t$ outbound (inbound) stubs.

4.2. *Construction of the coupling.* To study the distance between two randomly chosen nodes, we will couple the exploration of the graph described above with a branching process. To do this, we first note that the exploration process is equivalent to assigning to outbound stub $\mathbf{i} \neq \varnothing$ a number of offspring $\chi_{\mathbf{i}}^+$ chosen according to the (random) probability mass function

$$h_{\mathbf{i}}^+(t) = \begin{cases} \dfrac{1}{L_n - T_{\mathbf{i}}^+} \displaystyle\sum_{r=1}^n 1(D_r^+ = t) D_r^- \mathcal{I}_r(T_{\mathbf{i}}^+), \\ \quad t = 1, 2, \ldots, \\ \dfrac{1}{L_n - T_{\mathbf{i}}^+} \left\{ \displaystyle\sum_{r=1}^n 1(D_r^+ = 0) D_r^- \mathcal{I}_r(T_{\mathbf{i}}^+) + V_{\mathbf{i}}^- \right\}, \\ \quad t = 0, \end{cases} \tag{4.1}$$

where $T_{\mathbf{i}}^+$ is the number of outbound stubs that have been traversed up until the moment outbound stub $\mathbf{i}$ is about to be traversed, $\mathcal{I}_r(t) = 1$ (node $r$ is *inactive* after having traversed $t$ stubs), and

$$V_{\mathbf{i}}^- = L_n - \sum_{r=1}^n D_r^- \mathcal{I}_r(T_{\mathbf{i}}^+) - T_{\mathbf{i}}^+$$

is the number of unexplored inbound stubs belonging to active nodes at time $T_{\mathbf{i}}^+$. Note that $T_{\mathbf{i}}^+$ is also the number of inbound stubs that already belong to edges in the graph up until the moment outbound stub $\mathbf{i}$ is about to be explored. Symmetrically, we assign to inbound stub $\mathbf{i}$ a number of offspring $\chi_{\mathbf{i}}^-$ distributed according to

$$h_{\mathbf{i}}^-(t) = \begin{cases} \dfrac{1}{L_n - T_{\mathbf{i}}^-} \displaystyle\sum_{r=1}^n 1(D_r^- = t) D_r^+ \mathcal{I}_r(T_{\mathbf{i}}^-), \\ \quad t = 1, 2, \ldots, \\ \dfrac{1}{L_n - T_{\mathbf{i}}^-} \left\{ \displaystyle\sum_{r=1}^n 1(D_r^- = 0) D_r^+ \mathcal{I}_r(T_{\mathbf{i}}^-) + V_{\mathbf{i}}^+ \right\}, \\ \quad t = 0, \end{cases} \tag{4.2}$$

with $T_{\mathbf{i}}^-$ the number of inbound stubs that have been traversed up until the moment inbound stub $\mathbf{i}$ is about to be explored, and

$$V_{\mathbf{i}}^+ = L_n - \sum_{r=1}^n D_r^+ \mathcal{I}_r(T_{\mathbf{i}}^-) - T_{\mathbf{i}}^-$$

is the number of unexplored outbound stubs belonging to *active* nodes at time $T_{\mathbf{i}}^-$. As before, we have that $T_{\mathbf{i}}^-$ is also the number of outbound stubs that already belong to edges in the graph up until the moment inbound stub $\mathbf{i}$ is about to be explored.

Note that the number of outbound (inbound) stubs of the first node, that is, $Z_1^{\pm}$, is distributed according to $g_n^{\pm}$.

The key idea behind the coupling we will construct is that sampling from $h_{\mathbf{i}}^{\pm}$ and sampling from $f_n^{\pm}$ should be roughly equivalent as long as $T_{\mathbf{i}}^{\pm}$ is not too large. In turn, for large $n$, Assumption 2.1 implies that $f_n^{\pm}$ is very close to $f^{\pm}$. It follows that the process $\{Z_k^{\pm} : k \geq 0\}$ should be very close to a suitably constructed (delayed) branching process $\{\hat{Z}_k^{\pm} : k \geq 0\}$ having offspring distributions $(g^{\pm}, f^{\pm})$, where $g^{\pm}$ is the distribution of $\hat{Z}_1^{\pm}$ and all other nodes have offspring according to $f^{\pm}$.

To construct the coupling define $U = \bigcup_{k=0}^{\infty} \mathbb{N}_+^k$, with the convention that $\mathbb{N}_+^0 = \{\varnothing\}$, and let $\{U_{\mathbf{i}}\}_{\mathbf{i}\in U}$ be a sequence of i.i.d. Uniform(0, 1) random variables. For any nondecreasing function $F$, define $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$ to be its pseudo-inverse. Now set the number of outbound (inbound) stubs of $\mathbf{i}$ in the graph to be

$$\chi_{\mathbf{i}}^{\pm} = (H_{\mathbf{i}}^{\pm})^{-1}(U_{\mathbf{i}}), \qquad \mathbf{i} \neq \varnothing, \qquad \chi_{\varnothing}^{\pm} = (G_n^{\pm})^{-1}(U_{\varnothing}),$$

where $H_{\mathbf{i}}^{\pm}$ is the cumulative distribution function of $h_{\mathbf{i}}^{\pm}$, and the number of offspring of individual $\mathbf{i}$ in the outbound (inbound) branching process to be

$$\hat{\chi}_{\mathbf{i}}^{\pm} = (F^{\pm})^{-1}(U_{\mathbf{i}}), \qquad \mathbf{i} \neq \varnothing, \qquad \hat{\chi}_{\varnothing}^{\pm} = (G^{\pm})^{-1}(U_{\varnothing}).$$

In addition, we let $\hat{A}_r^{\pm}$ denote the set of individuals in the tree, corresponding to the process $\{\hat{Z}_k^{\pm} : k \geq 0\}$, at distance $r$ from the root.

Note that $\chi_{\mathbf{i}}$ and $\hat{\chi}_{\mathbf{i}}$ are now coupled through the same $U_{\mathbf{i}}$, and in view of the remarks following Definition 2.3, this coupling minimizes the Kantorovich–Rubinstein distance between the distributions $h_{\mathbf{i}}^{\pm}$ and $f^{\pm}$. Moreover, although the $\chi_{\mathbf{i}}^{\pm}$ are only defined for stubs $\mathbf{i}$ that have been created through the pairing process, the $\hat{\chi}_{\mathbf{i}}^{\pm}$ are well-defined regardless of whether $\mathbf{i}$ belongs to the tree or not. Furthermore, the sequence $\{U_{\mathbf{i}}\}_{\mathbf{i}\in U}$ defines the entire branching process $\{\hat{Z}_k^{\pm} : k \geq 0\}$, even after the graph has been fully explored.

The last thing we need to take care of is the observation that knowing $\chi_{\mathbf{i}}^{\pm}$ in the exploration of the graph does not necessarily tell us the identity of the node that stub $\mathbf{i}$ leads to, since there may be more than one node with $\chi_{\mathbf{i}}^{\pm}$ outbound (inbound) stubs, which is problematic if they do not also have the same number of inbound (outbound) stubs.The construction of the coupling requires that we keep track of both the inbound and outbound stubs discovered when a node first becomes *active*, since this information allows us to estimate the remaining number of unexplored stubs. To fix this problem, given $\chi_{\mathbf{i}}^{\pm} = t > 0$, pair outbound (inbound) stub $\mathbf{i}$ with an inbound (outbound) stub randomly chosen from the set of unpaired inbound (outbound) stubs belonging to *inactive* nodes and having exactly $t$ outbound (inbound) stubs; if $\chi_{\mathbf{i}}^{\pm} = 0$ sample the inbound (outbound) stub from the set of unpaired inbound (outbound) stubs belonging to either *inactive* nodes or *active* nodes having no outbound (inbound) stubs.

Summarizing the notation, we have:

- $A_r^+$ ($A_r^-$): set of outbound (inbound) stubs created during the $r$th step of the exploration process on the graph.
- $\hat{A}_r^+$ ($\hat{A}_r^-$): set of individuals in the outbound (inbound) tree at distance $r$ of the root.
- $Z_r^+$ ($Z_r^-$): number of outbound (inbound) stubs created during the $r$th step of the exploration process.
- $\hat{Z}_r^+$ ($\hat{Z}_r^-$): number of individuals in the $r$th generation of the outbound (inbound) tree.

The main observation upon which the analysis of the coupling is based is that if $|A|$ denotes the cardinality of set $A$, then

$$Z_k^{\pm} = |A_k^{\pm}| = |A_k^{\pm} \cap \hat{A}_k^{\pm}| + |A_k^{\pm} \cap (\hat{A}_k^{\pm})^c|$$

which implies that

$$\hat{Z}_k^{\pm} - |\hat{A}_k^{\pm} \cap (A_k^{\pm})^c| \le Z_k^{\pm} \le \hat{Z}_k^{\pm} + |A_k^{\pm} \cap (\hat{A}_k^{\pm})^c|. \tag{4.3}$$

4.3. *Coupling results*. We now present our main result on the coupling between the exploration process $\{Z_k^{\pm} : k \ge 1\}$ and the delayed branching process $\{\hat{Z}_k^{\pm} : k \ge 1\}$ described above. As mentioned earlier, the value of this new coupling is that it holds for a number of steps in the graph exploration process that is equivalent to having discovered $n^{1-\delta}$ number of nodes for arbitrarily small $0 < \delta < 1$; moreover, the coupled branching process is independent of the bi-degree sequence and of the number of nodes. Throughout the remainder of the paper, $\varepsilon > 0$ and $0 < \kappa \le 1$ are those from Assumption 2.1.

THEOREM 4.1. *Suppose that* $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ *satisfies Assumption* 2.1. *Then, for any* $0 < \delta < 1$, *any* $0 < \gamma < \min\{\delta\kappa, \varepsilon\}$, *there exist finite constants* $K, a > 0$ *such that for all* $1 \le k \le (1-\delta)\log_\mu n$,

$$\mathbb{P}_n\left(\bigcap_{m=1}^{k} \{|\hat{A}_m^{\pm} \cap (A_m^{\pm})^c| \le \hat{Z}_m^{\pm} n^{-\gamma}, |A_m^{\pm} \cap (\hat{A}_m^{\pm})^c| \le \hat{Z}_m^{\pm} n^{-\gamma}\}\right) \ge 1 - Kn^{-a}.$$

As an immediate corollary, relation (4.3) gives the following.

COROLLARY 4.2. *Suppose that* $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ *satisfies Assumption* 2.1. *Then, for any* $0 < \delta < 1$, *any* $0 < \gamma < \min\{\delta\kappa, \varepsilon\}$, *there exist finite constants* $K, a > 0$ *such that for all* $1 \le k \le (1-\delta)\log_\mu n$,

$$\mathbb{P}_n\left(\bigcap_{m=1}^{k} \{\hat{Z}_m^{\pm}(1 - n^{-\gamma}) \le Z_m^{\pm} \le \hat{Z}_m^{\pm}(1 + n^{-\gamma})\}\right) \ge 1 - Kn^{-\alpha}.$$

**5. Distances in the directed configuration model.** Having described the graph exploration process in the previous section, we are now ready to derive an expression for the hopcount between two randomly chosen nodes in a directed graph of size $n$ generated via the DCM. The main result of this section is Theorem 5.3, which expresses the tail distribution of the hopcount in terms of limiting random variables related to the branching processes $\{\hat{Z}_k^+ : k \geq 1\}$ and $\{\hat{Z}_k^- : k \geq 1\}$ introduced in the previous section. Although we will include some preliminary calculations here, we refer the reader to Section 6.3 for all other proofs.

As described in Section 4, we will compute the hopcount of a graph by selecting two nodes at random, say 1 and 2, and then start two independent breadth-first exploration processes. One will follow the outbound edges of node 1 while the other will use the inbound edges of node 2. At each step, we explore one generation of the out-component of node 1 and the corresponding generation of the in-component of node 2, starting with node 1.

In terms of the two nodes, $\{Z_k^+ : k \geq 1\}$ will denote the number of outbound stubs discovered during the $k$th step of the exploration of the out-component of node 1, while $\{Z_k^- : k \geq 1\}$ will denote the number of inbound stubs discovered during the $k$th step of the exploration of the in-component of node 2. An expression for the distribution of the hopcount is then obtained by computing the probability that there are no nodes in common given the current number of stubs explored so far in each of the two processes. We point out that the hopcount may be in fact infinite, which happens when node 2 is not in the out-component of node 1.

The first step in the analysis is a recursive relation for $\mathbb{P}_n(H_n > k)$. For this, we denote by $\mathscr{F}^{l,m} = \sigma(Z_i^-, Z_j^+ : 0 \leq i \leq l, 0 \leq j \leq m)$ the sigma algebra generated by the $Z_i^-$ and $Z_j^+$ of the first $l$ and $m$ generations, respectively. The next result follows from the analysis done in van der Hofstad, Hooghiemstra and Van Mieghem (2005) Lemma 4.1, which can be adapted to our case in a straight forward fashion:

$$\mathbb{P}_n(H_n > k) = \mathbb{E}_n\left[\prod_{i=2}^{k+1} \mathbb{P}_n\big(H_n > i-1 | H_n > i-2, \mathscr{F}^{\lceil i/2\rceil, \lfloor i/2\rfloor}\big)\right] \quad \text{for all } k \geq 1. \tag{5.1}$$

The presence of the ceiling and floor functions is due to the fact that we iteratively advance the exploration process alternating between nodes 1 and 2, starting with 1.

Let $p(A, B, L)$ denote the probability that none of the outbound stubs from a set of size $A$ connect to one of the inbound stubs from a set of size $B$, given that there are $L$ outbound/inbound stubs in total. Since we can only select $A$ inbound stubs outside of the set of size $B$ if $A + B \leq L$ and the probability of selecting the first such stub is $1 - B/L$, we get

$$p(A, B, L) = 1(A + B \leq L)\left(1 - \frac{B}{L}\right)p(A - 1, B, L - 1).$$

Continuing the recursion yields,

$$p(A,B,L) = 1(A+B \le L)\prod_{s=0}^{A-1}\left(1-\frac{B}{L-s}\right). \tag{5.2}$$

Next, observe that $H_n > 1$ holds if and only if none of the $Z_1^+$ outgoing edges points toward node 2. From the definition of the model, this occurs if and only if none of the $Z_1^+$ outbound stubs have been paired with one of the $Z_1^-$ inbound stubs. Hence,

$$\mathbb{P}_n\big(H_n > 1|\mathscr{F}^{1,1}\big) = p\big(Z_1^+, Z_1^-, L_n\big) = 1\big(Z_1^+ + Z_1^- \le L_n\big)\prod_{s=0}^{Z_1^+-1}\left(1-\frac{Z_1^-}{L_n - s}\right).$$

Similarly, we have

$$\mathbb{P}_n\big(H_n > 2|H_n > 1, \mathscr{F}^{2,1}\big) = 1\big(Z_2^+ + Z_1^- \le L_n - Z_1^+\big)\prod_{s=0}^{Z_2^+-1}\left(1-\frac{Z_1^-}{L_n - Z_1^+ - s}\right).$$

In order to write the full formula, we first define $\{\mathscr{S}_k\}_{k\ge 0}$ as follows:

$$\mathscr{S}_0 = 0, \qquad \mathscr{S}_1 = Z_1^+, \qquad \mathscr{S}_k = \sum_{j=1}^{\lceil k/2\rceil} Z_j^+ + \sum_{j=1}^{\lfloor k/2\rfloor} Z_j^- \qquad \text{for } k \ge 2. \tag{5.3}$$

We then obtain, for $i \ge 2$,

$$\begin{aligned}&\mathbb{P}_n\big(H_n > i-1|H_n > i-2, \mathscr{F}^{\lceil i/2\rceil,\lfloor i/2\rfloor}\big)\\ &\quad = 1(\mathscr{S}_i \le L_n)\prod_{s=0}^{Z_{\lceil i/2\rceil}^+-1}\left(1-\frac{Z_{\lfloor i/2\rfloor}^-}{L_n - \mathscr{S}_{i-2} - s}\right).\end{aligned}$$

Substituting this expression into (5.1) yields

$$\mathbb{P}_n(H_n > k) = \mathbb{E}_n\left[1(\mathscr{S}_{k+1} \le L_n)\prod_{i=2}^{k+1}\prod_{s=0}^{Z_{\lceil i/2\rceil}^+-1}\left(1-\frac{Z_{\lfloor i/2\rfloor}^-}{L_n - \mathscr{S}_{i-2} - s}\right)\right]. \tag{5.4}$$

The first result for the hopcount uses equation (5.4) combined with Corollary 4.2 to obtain an expression in terms of the branching processes $\{\hat{Z}_k^+ : k \ge 1\}$ and $\{\hat{Z}_k^- : k \ge 1\}$. We use the notation $g(x) = O(f(x))$ as $x \to \infty$ if $\limsup_{x\to\infty} g(x)/f(x) < \infty$.

PROPOSITION 5.1. *Suppose that* $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ *satisfies Assumption* 2.1. *Then, for any* $0 < \delta < 1$ *and for any* $0 \le k \le 2(1-\delta)\log_\mu n$, *there exists a constant* $a > 0$ *such that*

$$\left|\mathbb{P}_n(H_n > k) - E\left[\exp\left\{-\frac{1}{\nu n}\sum_{i=2}^{k+1}\hat{Z}_{\lceil i/2\rceil}^-\hat{Z}_{\lfloor i/2\rfloor}^+\right\}\right]\right| = O\big(n^{-a}\big), \qquad n \to \infty,$$

*where* $\{\hat{Z}_i^+ : i \geq 1\}$ *and* $\{\hat{Z}_i^- : i \geq 1\}$ *are independent delayed branching processes having offspring distributions* $(g^+, f^+)$ *and* $(g^-, f^-)$, *respectively.*

The next result shows a simplified expression for the limit in Proposition 5.1 in terms of the martingale limits $W^+$ and $W^-$. This result is independent of the coupling, and follows from the properties of the (delayed) branching processes $\{\hat{Z}_k^+ : k \geq 1\}$ and $\{\hat{Z}_k^- : k \geq 1\}$. We state it here since it plays an important role in establishing both Theorem 2.5 and Proposition 2.7.

PROPOSITION 5.2. *Suppose* $\{\hat{Z}_i^+ : i \geq 1\}$ *and* $\{\hat{Z}_i^- : i \geq 1\}$ *are independent delayed branching processes having offspring distributions* $(g^+, f^+)$ *and* $(g^-, f^-)$, *respectively. Suppose that* $f^+, f^-$ *have finite moments of order* $1 + \kappa \in (1, 2]$ *with common mean* $\mu > 1$, *and* $g^+, g^-$ *have common mean* $\nu$. *Then there exists* $b > 0$ *such that*

$$\left| E\left[\exp\left\{-\frac{1}{\nu n}\sum_{i=2}^{k+1} \hat{Z}^+_{\lceil i/2\rceil}\hat{Z}^-_{\lfloor i/2\rfloor}\right\} - \exp\left\{-\frac{\nu\mu^k}{(\mu-1)n}W^+W^-\right\}\right]\right|$$
$$= O(n^{-b}), \qquad n \to \infty,$$

*uniformly for all* $k \in \mathbb{N}_+$, *where* $W^\pm = \lim_{k\to\infty} \hat{Z}_k^\pm/(\nu\mu^{k-1})$.

Combining Propositions 5.1 and 5.2, we immediately obtain the following result.

THEOREM 5.3. *Suppose* $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ *satisfies Assumption* 2.1. *Then*, *for any* $0 < \delta < 1$ *and for any* $0 \leq k \leq 2(1-\delta)\log_\mu n$, *there exists a constant* $c > 0$ *such that*

$$\left|\mathbb{P}_n(H_n > k) - E\left[\exp\left\{-\frac{\nu\mu^k}{(\mu-1)n}W^-W^+\right\}\right]\right| = O(n^{-c}), \qquad n \to \infty,$$

*where* $W^\pm = \lim_{k\to\infty} \hat{Z}_k^\pm/(\nu\mu^{k-1})$, *with* $W^+$ *and* $W^-$ *independent of each other.*

As a corollary of Theorem 5.3, we obtain the following result for the probability that there exists a directed path between two randomly chosen nodes, which implies Proposition 2.7.

COROLLARY 5.4. *Suppose* $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ *satisfies Assumption* 2.1. *Then there exists a constant* $c > 0$ *such that*

$$\left|\mathbb{P}_n(H_n < \infty) - s^+s^-\right| = O(n^{-c}), \qquad n \to \infty,$$

*where* $s^\pm = P(W^\pm > 0)$.

Noting that

$$\mathbb{P}_n(H_n > k) = \mathbb{P}_n(H_n > k | H_n < \infty)\mathbb{P}_n(H_n < \infty) + \mathbb{P}_n(H_n = \infty),$$

defining $\mathcal{B} = \{W^+W^- > 0\}$, and using Theorem 5.3 and Corollary 5.4 gives

$$\begin{aligned}\mathbb{P}_n(H_n > k | H_n < \infty) &= \frac{\mathbb{P}_n(H_n > k) - \mathbb{P}_n(H_n = \infty)}{\mathbb{P}_n(H_n < \infty)} \\ &= \frac{1}{P(\mathcal{B})} E\left[\exp\left\{-\frac{\nu\mu^k}{(\mu-1)n} W^- W^+\right\}\right] - \frac{P(\mathcal{B}^c)}{P(\mathcal{B})} + O(n^{-c}) \\ &= E\left[\exp\left\{-\frac{\nu\mu^k}{(\mu-1)n} W^- W^+\right\}\Big| W^+W^- > 0\right] + O(n^{-c})\end{aligned}$$

as $n \to \infty$ and for the range of values of $k$ indicated in the theorems. Now define for $x \in \mathbb{R}$,

$$V_n(x) = 1 - E\left[\exp\left\{-\frac{\nu\mu^{\lfloor \log_\mu n\rfloor + \lfloor x\rfloor}}{(\mu-1)n} W^+W^-\right\}\Big| W^+W^- > 0\right].$$

That $V_n(x)$ is a cumulative distribution function for each fixed $n$ follows from noting that it is nondecreasing with $\lim_{x\to-\infty} V_n(x) = 0$ and $\lim_{x\to\infty} V_n(x) = 1$. Letting $\mathcal{H}_n$ be a random variable having distribution $V_n$ gives Theorem 2.5.

The remainder of the paper is devoted to the proofs of all the results presented in Sections 4 and 5.

**6. Proofs.** This section consists of four subsections. In Section 6.1, we prove some general results about delayed branching processes, including a bound for its minimum growth conditional on nonextinction. Section 6.2 contains the proof of Theorem 4.1, our main coupling theorem. The proofs of our results for the hopcount, Proposition 5.1, Proposition 5.2, Theorem 5.3 and Corollary 5.4, are given in Section 6.3. Finally, Section 6.4 contains the proof of Theorem 3.1, which shows that the i.i.d. algorithm given in Section 3.1 satisfies the main assumptions in the paper.

6.1. *Some results for delayed branching processes*. Our first result for a general delayed branching process is an expression for its extinction probability in terms of the probability of extinction of the corresponding nondelayed process, as well as for the distribution of its number of offspring conditional on extinction. Since these results are independent of the coupling with the graph, we do not use the $\pm$ notation.

LEMMA 6.1. *Let $\{\mathcal{Z}_k : k \geq 0\}$ denote a (nondelayed) branching process having offspring distribution $f$ and extinction probability $q$ and let $\{\hat{Z}_k : k \geq 1\}$ be a delayed branching process having offspring distributions $(g, f)$. Suppose $q > 0$.*

*Then, conditioned on extinction,* $\{\hat{Z}_k : k \geq 1\}$ *is a delayed branching process with offspring distributions* $(\tilde{g}, \tilde{f})$ *with*

$$\tilde{g}(i) = \frac{g(i)q^i}{\sum_{t=0}^{\infty} g(t)q^t} \quad \textit{and} \quad \tilde{f}(i) = f(i)q^{i-1}, \qquad i \geq 0.$$

*Moreover,* $P(\hat{Z}_k = 0$ *for some* $k \geq 1) = \sum_{t=0}^{\infty} g(t)q^t$.

PROOF. Let $\hat{\chi}_\varnothing$ have distribution $g$ and let $\{\mathcal{Z}_{k-1,i}\}_{i\geq 1}$ be a sequence of i.i.d. copies of $\mathcal{Z}_{k-1}$, independent of $\hat{\chi}_\varnothing$; set $\mu$ to be the mean of $f$. Computing the probability generating function of $\hat{Z}_k$, we obtain

$$P(W = 0)E\big[s^{\hat{Z}_k} | W = 0\big] = E\big[\big(E\big[s^{\mathcal{Z}_{k-1}} | \mathcal{W} = 0\big]q\big)^{\hat{\chi}_\varnothing}\big],$$

where $\mathcal{W}_i$ is the a.s. limit of the martingale $\{\mathcal{Z}_{k,i}/\mu^k : k \geq 0\}$ that has as root the $i$th individual in the first generation of $\{\hat{Z}_k : k \geq 1\}$. Also,

$$P(W = 0) = P(\hat{\chi}_\varnothing = 0) + P\left(\hat{\chi}_\varnothing \geq 1, \bigcap_{i=1}^{\hat{\chi}_\varnothing} \{\mathcal{W}_i = 0\}\right) = \sum_{j=0}^{\infty} g(j)q^j.$$

Hence,

$$E\big[s^{\hat{Z}_k} | W = 0\big] = \sum_{j=0}^{\infty} \big(E\big[s^{\mathcal{Z}_{k-1}} | \mathcal{W} = 0\big]\big)^j \tilde{g}(j),$$

where

$$\tilde{g}(j) = \frac{q^j g(j)}{\sum_{t=0}^{\infty} g(t)q^t}, \qquad j \geq 0.$$

Since conditionally on extinction $\{\mathcal{Z}_k : k \geq 0\}$ is a subcritical (nondelayed) branching process with offspring distribution $\tilde{f}(j) = f(j)q^{j-1}$, $j \geq 0$ [see, e.g., Athreya and Ney (2004), p. 52], the result follows. □

The second result we show is in some sense the counterpart of Doob's maximal martingale inequality, and it states that provided the limiting martingale is strictly positive, the branching process itself cannot grow too slowly. For this result and others in this section, we use the following version of Burkholder's inequality, which we state without proof.

LEMMA 6.2. *Let* $\{X_i\}_{i\geq 1}$ *be a sequence of i.i.d., mean zero random variables such that* $E[|X_1|^{1+\kappa}] < \infty$ *for some* $0 < \kappa \leq 1$. *Then*

$$P\left(\left|\sum_{i=1}^{n} X_i\right| > x\right) \leq \frac{1}{x^{1+\kappa}} E\left[\left|\sum_{i=1}^{n} X_i\right|^{1+\kappa}\right] \leq Q_{1+\kappa} E\big[|X_1|^{1+\kappa}\big] \frac{n}{x^{1+\kappa}},$$

*where* $Q_{1+\kappa}$ *is a constant that depends only on* $\kappa$.

LEMMA 6.3. *Suppose $\{\hat{Z}_k : k \geq 1\}$ is a delayed branching process with offspring distributions $(g, f)$, where $f$ has finite $1+\kappa \in (1,2]$ moment and mean $\mu > 1$, and $g$ has finite mean $\nu > 0$. Let $W = \lim_{k\to\infty} \hat{Z}_k/(\nu\mu^{k-1})$. Then, for any $1 < u < \mu$, there exists a constant $Q_1 < \infty$ such that for any $k \geq 1$,*

$$P\left(\inf_{r\geq k} \frac{\hat{Z}_r}{u^r} < 1, W > 0\right) \leq Q_1\big(u^{-\kappa k} + (u/\mu)^{\alpha k} 1(q > 0)\big),$$

*where $q$ is the extinction probability of a branching process having offspring distribution $f$, $\lambda = \sum_{i=1}^{\infty} f(i) i q^{i-1}$ and $\alpha = -\log\lambda/\log\mu > 0$ if $q > 0$.*

PROOF. We start by defining for $r \geq k$ the event $D_r = \{\min_{k\leq j\leq r} \hat{Z}_j/u^j \geq 1\}$ and letting $a_r = P(W > 0, (D_r)^c)$. Let $\{\hat{\chi}, \hat{\chi}_i\}$ be a sequence of i.i.d. random variables having distribution $f$ and use Lemma 6.2, applied conditionally on $\hat{Z}_{r-1}$, to obtain

$$\begin{aligned}
a_r &\leq P\big(D_{r-1}, \hat{Z}_r \leq u^r\big) + a_{r-1} \\
&\leq P\left(\hat{Z}_{r-1} \geq u^{r-1}, \sum_{i=1}^{\hat{Z}_{r-1}} \hat{\chi}_i \leq u\hat{Z}_{r-1}\right) + a_{r-1} \\
&\leq P\left(\hat{Z}_{r-1} \geq u^{r-1}, \sum_{i=1}^{\hat{Z}_{r-1}} (\mu - \hat{\chi}_i) \geq (\mu - u)\hat{Z}_{r-1}\right) + a_{r-1} \\
&\leq E\left[1\big(\hat{Z}_{r-1} \geq u^{r-1}\big) \frac{Q_{1+\kappa} E[|\hat{\chi} - \mu|^{1+\kappa}]}{(\mu - u)^{1+\kappa} (\hat{Z}_{r-1})^{\kappa}}\right] + a_{r-1} \\
&\leq Q u^{-\kappa(r-1)} + a_{r-1},
\end{aligned}$$

where $Q = Q_{1+\kappa} E[|\hat{\chi} - \mu|^{1+\kappa}]/(\mu - u)^{1+\kappa}$ and $E[(\hat{\chi})^{1+\kappa}] < \infty$ by Remark 2.4. It follows from iterating the inequality derived above that

$$a_r \leq Q \sum_{j=k}^{r-1} \frac{1}{u^{\kappa j}} + a_k \leq \frac{Q}{(u^{\kappa} - 1)u^{\kappa(k-1)}} + P\big(W > 0, \hat{Z}_k < u^k\big)$$

for all $r \geq k$. It remains to bound the last probability.

Let $\{\mathcal{Z}_k : k \geq 0\}$ be a (nondelayed) branching process having offspring distribution $f$, and let $\mathcal{W} = \lim_{k\to\infty} \mathcal{Z}_k/\mu^k$. It is well known [see Athreya and Ney (2004), p. 52], that conditional on nonextinction, $\mathcal{W}$ has an absolutely continuous distribution on $(0, \infty)$. Note also that for any $m \geq 1$ we have

$$W_{m+k} = \frac{\hat{Z}_{m+k}}{\nu\mu^{m+k-1}} = \frac{1}{\nu\mu^{m+k-1}} \sum_{\mathbf{i}\in\hat{A}_k} \mathcal{Z}_{m,\mathbf{i}},$$

where the $\{\mathcal{Z}_{m,\mathbf{i}}\}$ are i.i.d. copies of $\mathcal{Z}_m$ and $\hat{A}_k$ is the set of individuals in the $k$th generation of $\{\hat{Z}_k : k \geq 1\}$ and, therefore, for any $k \geq 1$,

$$W_{m+k} - W_k = \frac{1}{\nu\mu^{k-1}} \sum_{\mathbf{i}\in\hat{A}_k} \left(\frac{\mathcal{Z}_{m,\mathbf{i}}}{\mu^m} - 1\right).$$

Now define $\mathcal{W}_{\mathbf{i}} = \lim_{m\to\infty} \mathcal{Z}_{m,\mathbf{i}}/\mu^m$ to obtain that

$$W - W_k = \frac{1}{\nu\mu^{k-1}} \sum_{\mathbf{i}\in\hat{A}_k} (\mathcal{W}_{\mathbf{i}} - 1), \tag{6.1}$$

where the $\{\mathcal{W}_{\mathbf{i}}\}$ are i.i.d. copies of $\mathcal{W}$, independent of the history of the tree up to generation $k$. It follows that for $x_k = 2u^k/(\nu\mu^{k-1})$,

$$\begin{aligned}
&P(\hat{Z}_k < u^k, W > 0) \\
&\quad\leq P(\hat{Z}_k < u^k, W \geq x_k) + P(0 < W < x_k) \\
&\quad\leq P(\hat{Z}_k < u^k, W - W_k \geq x_k - u^k/(\nu\mu^{k-1})) + P(0 < W < x_k) \\
&\quad= E\left[1(\hat{Z}_k < u^k) P\left(\sum_{\mathbf{i}\in\hat{A}_k} (\mathcal{W}_{\mathbf{i}} - 1) \geq u^k \middle| \hat{Z}_k\right)\right] + P(0 < W < x_k).
\end{aligned}$$

Now note that $E[\hat{\chi}^{1+\kappa}] < \infty$ implies that $E[\mathcal{W}^{1+\kappa}] < \infty$. Then, by Lemma 6.2, applied conditionally on $\hat{Z}_k$, we obtain

$$P\left(\sum_{\mathbf{i}\in\hat{A}_k} (\mathcal{W}_{\mathbf{i}} - 1) \geq u^k \middle| \hat{Z}_k\right) \leq Q_{1+\kappa} E[|\mathcal{W} - 1|^{1+\kappa}] \cdot \frac{\hat{Z}_k}{u^{(1+\kappa)k}}.$$

Hence,

$$P(\hat{Z}_k < u^k, W > 0) \leq \frac{Q_{1+\kappa} E[|\mathcal{W} - 1|^{1+\kappa}]}{u^{\kappa k}} + P(0 < W < x_k).$$

Finally, to bound $P(0 < W < x_k)$, note that $W$ admits the representation

$$W = \frac{1}{\nu} \sum_{i=1}^{\hat{\chi}_\varnothing} \mathcal{W}_i,$$

where the $\{\mathcal{W}_i\}$ are i.i.d. copies of $\mathcal{W}$, independent of $\hat{\chi}_\varnothing$, with $\hat{\chi}_\varnothing$ distributed according to $g$. Note that $W > 0$ implies that at least one of the $\mathcal{W}_i$ is strictly positive. Let $N(t)$ be the number of nonzero random variables among $\{\mathcal{W}_1, \dots, \mathcal{W}_t\}$. It follows that if we let $\{V_i\}$ be i.i.d. random variables having the same distribution as $\mathcal{W}$ given $\mathcal{W} > 0$, then

$$P(0 < W < x_k) = P\left(\frac{1}{\nu} \sum_{i=1}^{N(\hat{\chi}_\varnothing)} V_i < x_k, N(\hat{\chi}_\varnothing) \geq 1\right) \leq P(V_1 < \nu x_k).$$

Hence, if $w(t)$ denotes the density of $\mathcal{W}$ conditional on nonextinction, we have that

$$P(V_1 < \nu x_k) = P(\mathcal{W} < \nu x_k | \mathcal{W} > 0) = \int_0^{\nu x_k} w(t)\,dt.$$

By Theorem 1 in Serge Dubuc (1971) [see also Theorem 4 in Biggins and Bingham (1993)], we have that if $\lambda = \sum_{i=1}^{\infty} f(i) i q^{i-1} > 0$, which under the assumptions of the lemma occurs whenever $q > 0$, then there exists a constant $C_0 < \infty$ such that

$$\int_0^{\nu x_k} w(t)\,dt \le C_0 (\nu x_k)^{\alpha}$$

for $\alpha = -\log \lambda / \log \mu$; whereas if $f(0) + f(1) = 0$, then Theorem 3 in Biggins and Bingham (1993) gives that for every $a > 0$ there exists a constant $C_a < \infty$ such that

$$\int_0^{\nu x_k} w(t)\,dt \le C_a (\nu x_k)^{a}.$$

We conclude that for $a^* = \kappa \log u / \log(\mu/u)$,

$$\begin{aligned} P\left(\min_{r \ge k} \frac{\hat{Z}_r}{u^r} < 1, W > 0\right) &\le \frac{Q}{(u^{\kappa} - 1) u^{\kappa(k-1)}} + \frac{Q_{1+\kappa} E[|\mathcal{W} - 1|^{1+\kappa}]}{u^{\kappa k}} \\ &\quad + C_0 (\nu x_k)^{\alpha} 1(q > 0) + C_{a^*} (\nu x_k)^{a^*} \\ &\le Q_1 \big(u^{-\kappa k} + (u/\mu)^{\alpha k} 1(q > 0)\big). \end{aligned}$$ $\square$

6.2. *Coupling with a branching process.* In this section, we prove Theorem 4.1. As mentioned in Section 4, the coupling we constructed is based on bounding the Kantorovich–Rubinstein distance between the distributions $H_{\mathbf{i}}^{\pm}$ and $F^{\pm}$, and the main difficulty lies in the fact that this distance grows as the number of explored stubs in the graph grows. The proof of the main theorem is based on four technical results, Lemmas 6.4, 6.6, 6.5 and Proposition 6.7, which we state and prove below.

Throughout this section, let

$$Y_k^{\pm} = \sum_{r=1}^{k} Z_r^{\pm}, \qquad k \ge 1; \qquad Y_0^{\pm} = 0.$$

The first of the technical lemmas gives us an upper bound for the Kantorovich–Rubinstein distance conditionally on the history of the graph exploration process and its coupled tree up to the moment that stub $\mathbf{i}$ is about to be traversed.

LEMMA 6.4. *Let $\mathcal{G}_{\mathbf{i}}$ denote the sigma-algebra generated by the bi-degree sequence $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ and the graph exploration process up to the time that outbound* (*inbound*) *stub* $\mathbf{i}$ *is about to be traversed. Then, provided* $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ *satisfies Assumption* 2.1, *for all* $n \geq (4/\nu)^{1/\varepsilon}$ *and* $T_{\mathbf{i}}^{\pm} \leq (\nu/2)n$, *we have*

$$\mathbb{E}_n\big[|\hat{\chi}_{\mathbf{i}}^{\pm} - \chi_{\mathbf{i}}^{\pm}||\mathcal{G}_{\mathbf{i}}\big] \leq \mathcal{E}(T_{\mathbf{i}}^{\pm}),$$

*where*

$$\mathcal{E}(t) = \frac{4}{\nu n} \sum_{r=1}^{n} \big(1 - \mathcal{I}_r(t)\big) D_r^+ D_r^- + \frac{4\mu t}{\nu n} + 3n^{-\varepsilon}, \tag{6.2}$$

*and* $0 < \varepsilon < 1$ *is the one from Assumption* 2.1.

PROOF. We first point out that for $\mathbf{i} = \varnothing$ the result holds trivially by Assumption 2.1, since

$$\mathbb{E}_n\big[|\hat{\chi}_{\varnothing}^{\pm} - \chi_{\varnothing}^{\pm}|\big] = d_1\big(G_n^{\pm}, G^{\pm}\big) \leq n^{-\varepsilon}.$$

For $\mathbf{i} \neq \varnothing$, we have

$$\mathbb{E}_n\big[|\hat{\chi}_{\mathbf{i}}^{\pm} - \chi_{\mathbf{i}}^{\pm}||\mathcal{G}_{\mathbf{i}}\big] = d_1\big(H_{\mathbf{i}}^{\pm}, F^{\pm}\big) \leq d_1\big(H_{\mathbf{i}}^{\pm}, F_n^{\pm}\big) + d_1\big(F_n^{\pm}, F^{\pm}\big).$$

Since by Assumption 2.1 we have that the second distance is smaller or equal than $n^{-\varepsilon}$, we only need to analyze the first one, which we do separately for the $+$ and $-$ cases. To this end, write

$$\begin{aligned}
d_1\big(H_{\mathbf{i}}^{+}, F_n^{+}\big) &= \sum_{k=0}^{\infty} \left| \sum_{j=0}^{k} \big(h_{\mathbf{i}}^{+}(j) - f_n^{+}(j)\big) \right| = \sum_{k=0}^{\infty} \left| \sum_{j=k+1}^{\infty} \big(f_n^{+}(j) - h_{\mathbf{i}}^{+}(j)\big) \right| \\
&= \sum_{k=0}^{\infty} \left| \sum_{j=k+1}^{\infty} \sum_{r=1}^{n} 1(D_r^+ = j) D_r^- \left( \frac{\mathcal{I}_r(T_{\mathbf{i}}^{+})}{L_n - T_{\mathbf{i}}^{+}} - \frac{1}{L_n} \right) \right| \\
&\leq \sum_{k=0}^{\infty} \sum_{r=1}^{n} \left| \frac{\mathcal{I}_r(T_{\mathbf{i}}^{+})}{L_n - T_{\mathbf{i}}^{+}} - \frac{1}{L_n} \right| D_r^- 1\big(D_r^+ > k\big) \\
&= \sum_{r=1}^{n} \left| \frac{(\mathcal{I}_r(T_{\mathbf{i}}^{+}) - 1)L_n + T_{\mathbf{i}}^{+}}{(L_n - T_{\mathbf{i}}^{+})L_n} \right| D_r^+ D_r^- \\
&\leq \frac{1}{L_n - T_{\mathbf{i}}^{+}} \sum_{r=1}^{n} \big(1 - \mathcal{I}_r\big(T_{\mathbf{i}}^{+}\big)\big) D_r^+ D_r^- + \frac{T_{\mathbf{i}}^{+}}{L_n - T_{\mathbf{i}}^{+}} \cdot \mu_n,
\end{aligned}$$

where $\mu_n = L_n^{-1} \sum_{r=1}^{n} D_r^+ D_r^-$ is the common mean of $F_n^+$ and $F_n^-$. Symmetrically,

$$d_1\big(H_{\mathbf{i}}^{-}, F_n^{-}\big) \leq \frac{1}{L_n - T_{\mathbf{i}}^{-}} \sum_{r=1}^{n} \big(1 - \mathcal{I}_r\big(T_{\mathbf{i}}^{-}\big)\big) D_r^+ D_r^- + \frac{T_{\mathbf{i}}^{-}}{L_n - T_{\mathbf{i}}^{-}} \cdot \mu_n.$$

Now note that

$$\mu_n \leq \mu + d_1\big(F_n^{\pm}, F^{\pm}\big),$$

and if $\nu_n$ denotes the common mean of $G_n^+$ and $G_n^-$, then

$$\frac{L_n}{n} = \nu_n \geq \nu - d_1(G_n^{\pm}, G^{\pm}),$$

which in turn implies that

$$(L_n - T_{\mathbf{i}}^{\pm})^{-1} \leq (\nu n - T_{\mathbf{i}}^{\pm} - n d_1(G_n^{\pm}, G^{\pm}))^{-1}.$$

We conclude that, under Assumption 2.1 and for $T_{\mathbf{i}}^{\pm} \leq (\nu/2)n$,

$$\begin{aligned} d_1(H_{\mathbf{i}}^{\pm}, F^{\pm}) &\leq \frac{1}{(\nu/2)n - n^{1-\varepsilon}} \sum_{r=1}^{n} (1 - \mathcal{I}_r(T_{\mathbf{i}}^{\pm})) D_r^+ D_r^- \\ &\quad + \frac{T_{\mathbf{i}}^{\pm}}{(\nu/2)n - n^{1-\varepsilon}} \cdot (\mu + n^{-\varepsilon}) + n^{-\varepsilon} \\ &\leq \frac{4}{\nu n} \sum_{r=1}^{n} (1 - \mathcal{I}_r(T_{\mathbf{i}}^{\pm})) D_r^+ D_r^- + \frac{4\mu T_{\mathbf{i}}^{\pm}}{\nu n} + 3n^{-\varepsilon} \end{aligned}$$

for all $n \geq (4/\nu)^{1/\varepsilon}$. $\square$

The second preliminary result provides an estimate for the expected value of the bound obtained in the previous lemma on the set where $\{\hat{Z}_k^{\pm} : k \geq 0\}$ behaves typically, that is, without exhibiting large deviations from its mean.

LEMMA 6.5. *Define $\mathcal{E}(t)$ according to (6.2), and for any fixed $0 < \eta < 1$ and all $m \geq 1$ define the event*

$$E_m = \bigcap_{r=1}^{m} \{\hat{Z}_r^{\pm}/\mu^r \leq n^{\eta}\}. \tag{6.3}$$

*Then, provided $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ satisfies Assumption 2.1, there exists a constant $Q_2 < \infty$ such that for any $0 \leq t \leq \nu n/2$ and any $k \geq 1$,*

$$\mathbb{E}_n[\mathcal{E}(t)] \leq Q_2\left(\frac{t^{\kappa}}{n^{\kappa}} + n^{-\varepsilon}\right) \quad \text{and} \quad \mathbb{E}_n[1(E_k)\hat{Z}_k^{\pm}\mathcal{E}(t)] \leq Q_2 \mu^k \left(\frac{t^{\kappa}}{n^{\kappa(1-\eta)}} + n^{-\varepsilon}\right),$$

*where $0 < \varepsilon < 1$ and $0 < \kappa \leq 1$ are those from Assumption 2.1.*

PROOF. We start by proving the bound for $\mathbb{E}_n[\mathcal{E}(t)]$. Let $X_r$ denote either $D_r^+$ or $D_r^-$ depending on whether we are exploring outbound stubs or inbound stubs, respectively. Recall that $\mathcal{I}_r(t)$ is the indicator of node $r$ being *inactive* at time $t$ in the graph exploration process. Next, note that $\mathcal{I}_r(0) = 1$, $\mathbb{E}_n[\mathcal{I}_r(1)] = 1 - 1/n$, and for any $2 \leq t < L_n$,

$$\mathbb{E}_n[\mathcal{I}_r(t)] = \left(1 - \frac{1}{n}\right) \prod_{s=1}^{t-1} \left(1 - \frac{X_r}{L_n - s}\right) \geq \left(1 - \frac{1}{n}\right)\left(1 - \frac{X_r}{L_n - t}\right)^{t-1},$$

from where it follows that $\mathbb{E}_n[1-\mathcal{I}_r(0)]=0$, $\mathbb{E}_n[1-\mathcal{I}_r(1)]=1/n$, and for $2\le t<L_n$,

$$\begin{aligned}\mathbb{E}_n\bigl[1-\mathcal{I}_r(t)\bigr] &\le 1-\left(1-\frac{1}{n}\right)\left(1-\frac{X_r}{L_n-t}\right)^{t-1}\\ &= \frac{1}{n}+\left(1-\frac{1}{n}\right)\frac{X_r}{L_n-t}\sum_{s=0}^{t-2}\left(1-\frac{X_r}{L_n-t}\right)^{s}\\ &\le \frac{1}{n}+\frac{(t-1)X_r}{L_n-t}.\end{aligned}$$

Now let $0<\kappa\le 1$ be the one from Assumption 2.1 and note that since $\mathbb{E}_n[1-\mathcal{I}_r(t)]\le 1$, then

$$\mathbb{E}_n\bigl[1-\mathcal{I}_r(t)\bigr]\le\bigl(\mathbb{E}_n\bigl[1-\mathcal{I}_r(t)\bigr]\bigr)^{\kappa}\le\left(\frac{1}{n}\right)^{\kappa}+\left(\frac{(t-1)X_r}{L_n-t}\right)^{\kappa},$$

where we used the inequality $(\sum_i y_i)^{\beta}\le\sum_i y_i^{\beta}$ for $y_i\ge 0$ and $0<\beta\le 1$. It follows that under Assumption 2.1 and for any $2\le t<\nu n-n^{1-\varepsilon}$,

$$\begin{aligned}&\frac{1}{n}\sum_{r=1}^{n}\bigl(\mathbb{E}_n\bigl[1-\mathcal{I}_r(t)\bigr]\bigr)^{\kappa}D_r^{+}D_r^{-}\\ &\quad\le\frac{1}{n^{1+\kappa}}\sum_{r=1}^{n}D_r^{+}D_r^{-}+\frac{(t-1)^{\kappa}}{n(L_n-t)^{\kappa}}\sum_{r=1}^{n}\bigl((D_r^{+})^{\kappa}+(D_r^{-})^{\kappa}\bigr)D_r^{+}D_r^{-}\\ &\quad\le\frac{\nu_n\mu_n}{n^{\kappa}}+\frac{K_\kappa t^{\kappa}}{(L_n-t)^{\kappa}}\\ &\quad\le\frac{(\nu+n^{-\varepsilon})(\mu+n^{-\varepsilon})}{n^{\kappa}}+\frac{K_\kappa t^{\kappa}}{(\nu n-n^{1-\varepsilon}-t)^{\kappa}}.\end{aligned}\tag{6.4}$$

It follows that $\mathbb{E}_n[\mathcal{E}(0)]\le 3n^{-\varepsilon}$, $\mathbb{E}_n[\mathcal{E}(1)]\le 4\mu_n/(\nu n)+3n^{-\varepsilon}$, and for any $2\le t\le\nu n/2$ we have

$$\begin{aligned}\mathbb{E}_n\bigl[\mathcal{E}(t)\bigr]&\le\frac{4}{\nu n}\sum_{r=1}^{n}\mathbb{E}_n\bigl[1-\mathcal{I}_r(t)\bigr]D_r^{+}D_r^{-}+\frac{4\mu t}{\nu n}+3n^{-\varepsilon}\\ &\le\frac{4\mu}{n^{\kappa}}\bigl(1+O\bigl(n^{-\varepsilon}\bigr)\bigr)+\frac{K_\kappa t^{\kappa}}{(\nu n/2)^{\kappa}}\bigl(1+O\bigl(n^{-\varepsilon}\bigr)\bigr)+\frac{4\mu t}{\nu n}+3n^{-\varepsilon}\\ &\le Q_0\left(\frac{t^{\kappa}}{n^{\kappa}}+n^{-\varepsilon}\right),\end{aligned}$$

for some constant $Q_0<\infty$.

Next, to compute a bound for $\mathbb{E}_n[1(E_k)\hat{Z}_k^{\pm}\mathcal{E}(t)]$ let $q = 1/\kappa$ and $p = q/(q-1)$, with $p = \infty$ if $q = 1$, and use Hölder's inequality to obtain, for $1 \le t \le \nu n/2$,

$$\begin{aligned}\mathbb{E}_n\big[1(E_k)\hat{Z}_k^{\pm}\mathcal{E}(t)\big] &= \frac{4}{\nu n}\sum_{r=1}^{n}\mathbb{E}_n\big[1(E_k)\hat{Z}_k^{\pm}\big(1-\mathcal{I}_r(t)\big)\big]D_r^+D_r^- \\ &\quad + \frac{4\mu t}{\nu n}E\big[1(E_k)\hat{Z}_k^{\pm}\big] + 3n^{-\varepsilon}E\big[1(E_k)\hat{Z}_k^{\pm}\big] \\ &\le \frac{4}{\nu n}\sum_{r=1}^{n}\big(E\big[1(E_k)\big(\hat{Z}_k^{\pm}\big)^p\big]\big)^{1/p}\big(\mathbb{E}_n\big[1-\mathcal{I}_r(t)\big]\big)^{1/q}D_r^+D_r^- \\ &\quad + \frac{4\mu t}{\nu n}E\big[\hat{Z}_k^{\pm}\big] + 3n^{-\varepsilon}E\big[\hat{Z}_k^{\pm}\big].\end{aligned}$$

Now note that

$$\begin{aligned}\big(E\big[1(E_k)\big(\hat{Z}_k^{\pm}\big)^p\big]\big)^{1/p} &\le \big(\big(\mu^k n^{\eta}\big)^{p-1}E\big[\hat{Z}_k^{\pm}\big]\big)^{1/p} = \mu^k n^{\eta}\left(\frac{\nu}{\mu n^{\eta}}\right)^{1/p} \\ &= \left(\frac{\mu}{\nu}\right)^{\kappa}E\big[\hat{Z}_k^{\pm}\big]n^{\kappa\eta}.\end{aligned}$$

Combining this inequality with (6.4) gives, for $1 \le t \le \nu n/2$,

$$\begin{aligned}&\mathbb{E}_n\big[1(E_k)\hat{Z}_k^{\pm}\mathcal{E}(t)\big] \\ &\quad \le E\big[\hat{Z}_k^{\pm}\big]\left(\frac{\mu^{\kappa}n^{\kappa\eta}}{\nu^{\kappa}}\cdot\frac{4}{\nu n}\sum_{r=1}^{n}\big(\mathbb{E}_n\big[1-\mathcal{I}_r(t)\big]\big)^{\kappa}D_r^+D_r^- + \frac{4\mu t}{\nu n} + 3n^{-\varepsilon}\right) \\ &\quad \le E\big[\hat{Z}_k^{\pm}\big]\left(\frac{\mu^{\kappa}n^{\kappa\eta}}{\nu^{\kappa}}\left\{\frac{4\mu}{n^{\kappa}}\big(1+O\big(n^{-\varepsilon}\big)\big) + \frac{K_{\kappa}t^{\kappa}}{(\nu n/2)^{\kappa}}\big(1+O\big(n^{-\varepsilon}\big)\big)\right\}\right. \\ &\qquad \left. + \frac{4\mu t}{\nu n} + 3n^{-\varepsilon}\right) \\ &\quad \le Q_0'\mu^k\left(\frac{t^{\kappa}}{n^{\kappa(1-\eta)}} + n^{-\varepsilon}\right),\end{aligned}$$

for some constant $Q_0' < \infty$. Noting that $\mathbb{E}_n[1(E_k)\hat{Z}_k^{\pm}\mathcal{E}(0)] \le E[\hat{Z}_k^{\pm}]3n^{-\varepsilon} = O(\mu^k n^{-\varepsilon})$ completes the proof. $\square$

The third technical lemma provides an estimate for the expected number of stubs that are discovered during step $k+1$ of the graph exploration process, on the set where the coupling holds uniformly well up to step $k$. This bound is the key component that will enable the induction step in the proof of Theorem 4.1.

LEMMA 6.6. *Let $E_k$ be defined according to* (6.3). *Fix* $0 < \delta < 1$, $0 < \gamma < \min\{\delta\kappa, \varepsilon\}$ *and define*

$$C_m = \bigcap_{r=1}^{m} \{ |\hat{A}_r^{\pm} \cap (A_r^{\pm})^c| \le \hat{Z}_r^{\pm} n^{-\gamma}, |A_r^{\pm} \cap (\hat{A}_r^{\pm})^c| \le \hat{Z}_r^{\pm} n^{-\gamma} \}. \tag{6.5}$$

*Then, provided* $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ *satisfies Assumption* 2.1, *there exists a constant* $Q_3 < \infty$ *such that for all* $1 \le k \le (1-\delta)\log n/\log\mu$,

$$\begin{aligned}&\mathbb{E}_n[1(C_k \cap E_k)(|\hat{A}_{k+1}^{\pm} \cap (A_{k+1}^{\pm})^c| + |A_{k+1}^{\pm} \cap (\hat{A}_{k+1}^{\pm})^c|)] \\ &\quad \le Q_3 \mu^k (\mu^{\kappa k} n^{-\kappa(1-2\eta)} + k n^{-\varepsilon}),\end{aligned}$$

*where* $0 < \varepsilon < 1$ *and* $0 < \kappa \le 1$ *are those from Assumption* 2.1.

PROOF. Let $\mathcal{F}_m$ denote the sigma-algebra generated by the bi-degree sequence and the history of the exploration process up until step $m-1$ is completed; note that this includes the value of $Z_m^{\pm}$. Define

$$u_{k+1} = \mathbb{E}_n[1(C_k \cap E_k)(|\hat{A}_{k+1}^{\pm} \cap (A_{k+1}^{\pm})^c| + |A_{k+1}^{\pm} \cap (\hat{A}_{k+1}^{\pm})^c|)]$$

and $\mathcal{E}(t)$ according to (6.2), then condition on $\mathcal{F}_k$ to obtain that

$$u_{k+1} = \mathbb{E}_n\left[1(C_k \cap E_k) E\left[\sum_{\mathbf{i} \in \hat{A}_k^{\pm} \cap A_k^{\pm}} (\hat{\chi}_{\mathbf{i}}^{\pm} - \chi_{\mathbf{i}}^{\pm})^+ + \sum_{\mathbf{i} \in \hat{A}_k^{\pm} \cap (A_k^{\pm})^c} \hat{\chi}_{\mathbf{i}}^{\pm} \Big| \mathcal{F}_k\right]\right] \tag{6.6}$$

$$+ \mathbb{E}_n\left[1(C_k \cap E_k) E\left[\sum_{\mathbf{i} \in A_k^{\pm} \cap \hat{A}_k^{\pm}} (\chi_{\mathbf{i}}^{\pm} - \hat{\chi}_{\mathbf{i}}^{\pm})^+ + \sum_{\mathbf{i} \in A_k^{\pm} \cap (\hat{A}_k^{\pm})^c} \chi_{\mathbf{i}}^{\pm} \Big| \mathcal{F}_k\right]\right]. \tag{6.7}$$

To analyze the conditional expectations, we first note that for $\mathbf{i} \in \hat{A}_k^{\pm} \cap A_k^{\pm}$ we must have that $T_{\mathbf{i}}^{\pm} \le Y_k^{\pm}$. Moreover, on the event $C_k \cap E_k$ we have that

$$Y_k^{\pm} = \sum_{i=1}^{k} Z_i^{\pm} \le (1 + n^{-\gamma}) \sum_{i=1}^{k} \hat{Z}_i^{\pm} \le 2n^{\eta} \sum_{i=1}^{k} \mu^i \le 2\mu^{k+1} n^{\eta}/(\mu-1) \triangleq y_k,$$

which for the range of values of $k$ in the lemma satisfies $y_k = o(n)$ as $n \to \infty$. It follows by Lemma 6.4 and the tower property that on the event $C_k \cap E_k$, the conditional expectation in (6.6) is bounded from above by

$$\begin{aligned}&\sum_{\mathbf{i} \in \hat{A}_k^{\pm} \cap A_k^{\pm}} \mathbb{E}_n[\mathcal{E}(T_{\mathbf{i}}^{\pm})|\mathcal{F}_k] + |\hat{A}_k^{\pm} \cap (A_k^{\pm})^c|\mu \\ &\quad \le \hat{Z}_k^{\pm} \mathbb{E}_n[\mathcal{E}(Y_k^{\pm})|\mathcal{F}_k] + |\hat{A}_k^{\pm} \cap (A_k^{\pm})^c|\mu \\ &\quad \le \hat{Z}_k^{\pm} \mathbb{E}_n[\mathcal{E}(y_k)|\mathcal{F}_k] + |\hat{A}_k^{\pm} \cap (A_k^{\pm})^c|\mu,\end{aligned}$$

where we used the observation that $\mathcal{E}(t)$ is nondecreasing and $T_{\mathbf{i}}^{\pm} \le Y_k^{\pm} \le y_k$.

Similarly, the conditional expectation in (6.7) is bounded from above by

$$\begin{aligned}
&\mathbb{E}_n\Big[\sum_{\mathbf{i}\in A_k^{\pm}}(\chi_{\mathbf{i}}^{\pm}-\hat{\chi}_{\mathbf{i}}^{\pm})^{+}+\sum_{\mathbf{i}\in A_k^{\pm}\cap(\hat{A}_k^{\pm})^c}\hat{\chi}_{\mathbf{i}}^{\pm}\Big|\mathcal{F}_k\Big]\\
&\quad\le\sum_{\mathbf{i}\in A_k^{\pm}}\mathbb{E}_n\big[\mathcal{E}(T_{\mathbf{i}}^{\pm})|\mathcal{F}_k\big]+\big|A_k^{\pm}\cap(\hat{A}_k^{\pm})^c\big|\mu\\
&\quad\le Z_k^{\pm}\mathbb{E}_n\big[\mathcal{E}(Y_k^{\pm})|\mathcal{F}_k\big]+\big|A_k^{\pm}\cap(\hat{A}_k^{\pm})^c\big|\mu\\
&\quad\le(1+n^{-\gamma})\hat{Z}_k^{\pm}\mathbb{E}_n\big[\mathcal{E}(y_k)|\mathcal{F}_k\big]+\big|A_k^{\pm}\cap(\hat{A}_k^{\pm})^c\big|\mu,
\end{aligned}$$

where in the last step we also used that $Z_k\le(1+n^{-\gamma})\hat{Z}_k$ on the event $C_k$.

Noting that $C_k\cap E_k\subseteq C_{k-1}\cap E_{k-1}$ gives

$$\begin{aligned}
u_{k+1}&\le\mathbb{E}_n\big[1(C_k\cap E_k)(2+n^{-\gamma})\hat{Z}_k^{\pm}\mathbb{E}_n[\mathcal{E}(y_k)|\mathcal{F}_k]\big]+\mu u_k\\
&\le 3\mathbb{E}_n\big[1(E_k)\hat{Z}_k^{\pm}\mathcal{E}(y_k)\big]+\mu u_k.
\end{aligned}$$

By Lemma 6.5, we have that

$$\begin{aligned}
3\mathbb{E}_n\big[1(E_k)\hat{Z}_k^{\pm}\mathcal{E}(y_k)\big]&\le 3Q_2\mu^k\left(\frac{y_k^{\kappa}}{n^{\kappa(1-\eta)}}+n^{-\varepsilon}\right)\\
&\le Q_2'\big(\mu^{(1+\kappa)k}n^{-\kappa(1-2\eta)}+\mu^k n^{-\varepsilon}\big)
\end{aligned}$$

for some constants $Q_2, Q_2'<\infty$. Let $r_k=Q_2'\mu^k(\mu^{\kappa k}n^{-\kappa(1-2\eta)}+n^{-\varepsilon})$ and iterate the inequality $u_{k+1}\le r_k+\mu u_k$ to obtain

$$\begin{aligned}
u_{k+1}&\le\sum_{j=1}^{k}\mu^{j-1}r_{k+1-j}+\mu^k u_1\\
&=\sum_{j=1}^{k}\mu^{j-1}Q_2'\mu^{k+1-j}\big(\mu^{\kappa(k+1-j)}n^{-\kappa(1-2\eta)}+n^{-\varepsilon}\big)+\mu^k u_1\\
&=Q_2'\mu^k\left(\sum_{j=1}^{k}\mu^{\kappa(k+1-j)}n^{-\kappa(1-2\eta)}+kn^{-\varepsilon}\right)+\mu^k u_1\\
&=Q_2'\mu^k\left(n^{-\kappa(1-2\eta)}\frac{\mu^{\kappa}(\mu^{\kappa k}-1)}{\mu^{\kappa}-1}+kn^{-\varepsilon}\right)+\mu^k u_1.
\end{aligned}$$

Noting that

$$u_1=\mathbb{E}_n\big[|\hat{\chi}_{\varnothing}^{\pm}-\chi_{\varnothing}^{\pm}|\big]\le n^{-\varepsilon}$$

completes the proof. $\square$

The last preliminary result before the proof of Theorem 4.1 is an analysis of the coupling when the branching process becomes extinct, which is most likely to occur when the out- or in-component of the node being explored in the graph is small.

PROPOSITION 6.7. *Fix $0 < \delta < 1$, $0 < \gamma < \min\{\delta\kappa, \varepsilon\}$ and define for $k \geq 1$ the event $C_k$ according to (6.5). Let $W^\pm = \lim_{k\to\infty} \hat{Z}_k^\pm/(\nu\mu^{k-1})$. Then, provided $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ satisfies Assumption 2.1, there exists a constant $Q_4 < \infty$ such that for all $1 \leq k \leq (1-\delta)\log n/\log\mu$,*

$$\mathbb{P}_n\big(C_k^c, W^\pm = 0\big) \leq Q_4 n^{-\varepsilon\kappa/(1+\kappa+\varepsilon)}.$$

PROOF. We start by pointing out that if $q^\pm = 0$ the probability that the branching process $\{\hat{Z}_k^\pm : k \geq 1\}$ becomes extinct is zero unless $\hat{Z}_1^\pm = 0$ [see Theorem 4 in Athreya and Ney (2004)]. Since in our construction of the coupling $Z_1^\pm = \hat{Z}_1^\pm$, we may assume from now on that $q^\pm > 0$.

Analogously to the events $E_m$ and $C_m$ defined in (6.3) and (6.5), we now define for $m \geq 1$

$$\begin{aligned} B_m &= \bigcap_{r=1}^{m} \big\{\big|\hat{A}_r^\pm \cap \big(A_r^\pm\big)^c\big| = 0, \big|A_r^\pm \cap \big(\hat{A}_r^\pm\big)^c\big| = 0\big\}, \\ I_m &= \bigcap_{r=1}^{m} \big\{\hat{Z}_r^\pm/\big(\lambda^\pm\big)^r \leq n^\tau\big\}, \end{aligned} \tag{6.8}$$

where $\lambda^\pm = \sum_{j=1}^\infty jf^\pm(j)(q^\pm)^{j-1} < 1$ and $\tau = \kappa/(1+\kappa+\varepsilon)$. Now note that

$$\mathbb{P}_n\big(C_k^c, W^\pm = 0\big) \leq \mathbb{P}_n\big(C_k^c \cap I_k\big) + P\big(I_k^c, W^\pm = 0\big),$$

where the last probability is independent of the bi-degree sequence. Next, use Lemma 6.1 to see that conditionally on $\{W^\pm = 0\}$, $\hat{Z}_k^\pm/(\nu^\pm(\lambda^\pm)^{k-1})$ is a mean one martingale, where $\nu^\pm$ is the mean of $\tilde{g}^\pm$ and $\lambda^\pm$ is the mean of $\tilde{f}^\pm$ ($\tilde{g}^\pm$ and $\tilde{f}^\pm$ defined according to Lemma 6.1 using $g^\pm$ and $f^\pm$, respectively). It then follows from Doob's inequality that

$$P\big(I_k^c, W^\pm = 0\big) \leq P\big(I_k^c | W^\pm = 0\big) \leq \big(\nu^\pm/\lambda^\pm\big)n^{-\tau}.$$

Next, write

$$\mathbb{P}_n\big(C_k^c \cap I_k\big) \leq \mathbb{P}_n\big(B_k^c \cap I_k\big) \leq \sum_{r=1}^{k} \mathbb{P}_n\big(B_{r-1} \cap B_r^c \cap I_k\big)$$

and note that the event $B_{r-1}$ implies that $Z_i^\pm = \hat{Z}_i^\pm$ for all $1 \leq i \leq r-1$. In addition, since $\lambda^\pm < 1$ we have $(\lambda^\pm)^{r_n+1} n^{-\tau} < 1$ for $r_n \triangleq \lfloor \tau \log n/|\log\lambda^\pm|\rfloor$. The observation that $\hat{Z}_i^{\pm}$ is integer valued then gives that the event $B_{r-1} \cap I_k$ implies that $Z_{r-1}^{\pm} = \hat{Z}_{r-1}^{\pm} = 0$ for all $r-1 > r_n$. Since for any $r \geq 1$, we have

$$
\begin{aligned}
|\hat{A}_r^{\pm} \cap (A_r^{\pm})^c| &= \sum_{\mathbf{i} \in \hat{A}_{r-1}^{\pm} \cap A_{r-1}^{\pm}} \sum_{j=1}^{\hat{\chi}_{\mathbf{i}}^{\pm}} 1\big((\mathbf{i}, j) \in (A_r^{\pm})^c\big) + \sum_{\mathbf{i} \in \hat{A}_{r-1}^{\pm} \cap (A_{r-1}^{\pm})^c} \hat{\chi}_{\mathbf{i}}^{\pm} \\
&= \sum_{\mathbf{i} \in \hat{A}_{r-1}^{\pm} \cap A_{r-1}^{\pm}} (\hat{\chi}_{\mathbf{i}}^{\pm} - \chi_{\mathbf{i}}^{\pm})^+ + \sum_{\mathbf{i} \in \hat{A}_{r-1}^{\pm} \cap (A_{r-1}^{\pm})^c} \hat{\chi}_{\mathbf{i}}^{\pm},
\end{aligned} \tag{6.9}
$$

and

$$
|A_r^{\pm} \cap (\hat{A}_r^{\pm})^c| = \sum_{\mathbf{i} \in A_{r-1}^{\pm} \cap \hat{A}_{r-1}^{\pm}} (\chi_{\mathbf{i}}^{\pm} - \hat{\chi}_{\mathbf{i}}^{\pm})^+ + \sum_{\mathbf{i} \in A_{r-1}^{\pm} \cap (\hat{A}_{r-1}^{\pm})^c} \chi_{\mathbf{i}}^{\pm} \tag{6.10}
$$

then $\mathbb{P}_n(B_{r-1} \cap B_r^c \cap I_k) = 0$ for all $r_n + 2 \leq r \leq k$. For $1 \leq r \leq r_n + 1$, we obtain using (6.9) and (6.10) that

$$
\begin{aligned}
\mathbb{P}_n\big(B_{r-1} \cap B_r^c \cap I_k\big) &\leq \mathbb{P}_n\big(|\hat{A}_r^{\pm} \cap (A_r^{\pm})^c| + |A_r^{\pm} \cap (\hat{A}_r^{\pm})^c| \geq 1, B_{r-1} \cap I_{r-1}\big) \\
&= \mathbb{P}_n\bigg(\sum_{\mathbf{i} \in \hat{A}_{r-1}^{\pm}} |\chi_{\mathbf{i}}^{\pm} - \hat{\chi}_{\mathbf{i}}^{\pm}| \geq 1, B_{r-1} \cap I_{r-1}\bigg).
\end{aligned}
$$

Now let $\mathcal{F}_r$ denote the sigma-algebra generated by the bi-degree sequence and the history of the exploration process up until step $r-1$ is completed; note that this includes the value of $Z_r^{\pm}$. Also define $\mathcal{G}_{\mathbf{i}}$ to be the sigma-algebra generated by the bi-degree sequence and the exploration process up to the time that outbound (inbound) stub $\mathbf{i}$ is about to be traversed; note that for $\mathbf{i} \in \hat{A}_{r-1}^{\pm}$ we have $\mathcal{F}_{r-1} \subseteq \mathcal{G}_{\mathbf{i}}$. Applying Markov's inequality conditionally on $\mathcal{F}_{r-1}$, we obtain

$$
\begin{aligned}
&\mathbb{P}_n\bigg(\sum_{\mathbf{i} \in \hat{A}_{r-1}^{\pm}} |\chi_{\mathbf{i}}^{\pm} - \hat{\chi}_{\mathbf{i}}^{\pm}| \geq 1, B_{r-1} \cap I_{r-1}\bigg) \\
&\qquad \leq \mathbb{E}_n\bigg[1(B_{r-1} \cap I_{r-1}) \sum_{\mathbf{i} \in \hat{A}_{r-1}^{\pm}} \mathbb{E}_n\big[|\chi_{\mathbf{i}}^{\pm} - \hat{\chi}_{\mathbf{i}}^{\pm}| | \mathcal{F}_{r-1}\big]\bigg].
\end{aligned}
$$

To analyze the conditional expectation, recall that $T_{\mathbf{i}}^{\pm}$ is the number of stubs (outbound for $+$ and inbound for $-$) that have been seen up until it is stub $\mathbf{i}$'s turn to be traversed, and use Lemma 6.4 to obtain

$$
\mathbb{E}_n\big[|\chi_{\mathbf{i}}^{\pm} - \hat{\chi}_{\mathbf{i}}^{\pm}| | \mathcal{F}_{r-1}\big] = \mathbb{E}_n\big[\mathbb{E}_n\big[|\chi_{\mathbf{i}}^{\pm} - \hat{\chi}_{\mathbf{i}}^{\pm}| | \mathcal{G}_{\mathbf{i}}\big] | \mathcal{F}_{r-1}\big] \leq \mathbb{E}_n\big[\mathcal{E}(T_{\mathbf{i}}^{\pm}) | \mathcal{F}_{r-1}\big].
$$

Recall that on the event $B_{r-1}$ we have $Z_i^{\pm} = \hat{Z}_i^{\pm}$ for all $1 \leq i \leq r-1$ and, therefore, for $\mathbf{i} \in \hat{A}_{r-1}^{\pm}$, $T_{\mathbf{i}}^{\pm} \leq Y_{r-1}^{\pm} = \hat{Y}_{r-1}^{\pm} = \sum_{i=1}^{r-1} \hat{Z}_i^{\pm}$. Moreover, on the event $I_{r-1}$

we have $\hat{Y}_{r-1}^{\pm} \le \sum_{i=1}^{r-1} (\lambda^{\pm})^{r-1} n^{\tau} \le n^{\tau}/(1-\lambda^{\pm})$, and it follows from Lemma 6.5 that

$$\begin{aligned}\mathbb{P}_n(B_{r-1} \cap B_r^c \cap I_k) &\le \mathbb{E}_n\Bigg[1(I_{r-1}) \sum_{\mathbf{i} \in \hat{A}_{r-1}^{\pm}} \mathbb{E}_n\big[\mathcal{E}\big(n^{\tau}/(1-\lambda^{\pm})\big) | \mathcal{F}_{r-1}\big]\Bigg] \\ &= \mathbb{E}_n\big[1(I_{r-1}) \hat{Z}_{r-1}^{\pm} \mathcal{E}\big(n^{\tau}/(1-\lambda^{\pm})\big)\big] \\ &\le (\lambda^{\pm})^{r-1} n^{\tau} \mathbb{E}_n\big[\mathcal{E}\big(n^{\tau}/(1-\lambda^{\pm})\big)\big] \\ &\le Q_2 (\lambda^{\pm})^{r-1} n^{\tau} \left( \frac{(n^{\tau}/(1-\lambda^{\pm}))^{\kappa}}{n^{\kappa}} + n^{-\varepsilon} \right) \\ &= O\big((\lambda^{\pm})^r n^{-\varepsilon\tau}\big),\end{aligned}$$

where the last equality is due to our choice of $\tau$. Thus, we have shown that

$$\mathbb{P}_n(C_k^c, W^{\pm} = 0) = O\left( n^{-\varepsilon\tau} \sum_{r=1}^{r_n+1} (\lambda^{\pm})^r + n^{-\tau} \right) = O(n^{-\varepsilon\tau}). \qquad \square$$

We are now ready to prove Theorem 4.1, which in view of Proposition 6.7, reduces to analyzing the event that the error in the coupling is large conditionally on the branching process surviving.

PROOF OF THEOREM 4.1. Let $W^{\pm} = \lim_{k\to\infty} \hat{Z}_k^{\pm}/(\nu\mu^{k-1})$ and for each $m \ge 1$ define the event $C_m$ according to (6.5). Now note that

$$\mathbb{P}_n(C_k^c) \le \mathbb{P}_n(C_k^c, W^{\pm} = 0) + \mathbb{P}_n(C_k^c, W^{\pm} > 0),$$

where by Proposition 6.7 we have

$$\mathbb{P}_n(C_k^c, W^{\pm} = 0) \le Q_4 n^{-\varepsilon\kappa/(1+\kappa+\epsilon)}$$

for some constant $Q_4 < \infty$.

To analyze $\mathbb{P}_n(C_k^c, W^{\pm} > 0)$, we proceed similarly to the proof of Proposition 6.7 by setting $\eta = (\delta\kappa - \gamma)/(4\kappa) \in (0, \delta/4)$ and defining the events $E_m$ and $B_m$, $m \ge 1$, according to (6.3) and (6.8), respectively. Define also $s_n = \min\{k, \lceil c \log n / \log \mu \rceil\}$, with $0 < c < \min\{\kappa(1-2\eta)/(1+\kappa), \varepsilon\}$, and note that $C_r^c \subset B_r^c$ for any $r \ge 1$. We start by deriving an upper bound as follows:

$$\begin{aligned}&\mathbb{P}_n(C_k^c, W^{\pm} > 0) \\ &\quad \le \mathbb{P}_n(C_k^c \cap E_k, W^{\pm} > 0) + P(E_k^c) \\ &\quad \le \mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap E_k, W^{\pm} > 0) + \mathbb{P}_n(C_{s_n-1}^c \cap E_k, W^{\pm} > 0) + P(E_k^c) \\ &\quad \le \mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap E_k, W^{\pm} > 0) + \mathbb{P}_n(B_{s_n-1}^c \cap E_k) + P(E_k^c) \\ &\quad \le \sum_{r=1}^{s_n-1} \mathbb{P}_n(B_{r-1} \cap B_r^c \cap E_k) + \mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap E_k, W^{\pm} > 0) + P(E_k^c).\end{aligned}$$

Note that Doob's inequality gives $P(E_k^c) \leq (\mu/\nu)n^{-\eta}$. Also, if we let $\mathcal{F}_r$ denote the sigma-algebra generated by the bi-degree sequence and the history of the exploration process up until step $r-1$ is completed, the same steps used in the proof of Proposition 6.7 give that for $1 \leq r \leq s_n - 1$,

$$\begin{aligned}
&\mathbb{P}_n(B_{r-1} \cap B_r^c \cap E_k) \\
&\quad \leq \mathbb{E}_n\biggl[1(E_{r-1}) \sum_{\mathbf{i} \in \hat{A}_{r-1}^{\pm}} \mathbb{E}_n\bigl[\mathcal{E}(\hat{Y}_{r-1}^{\pm}) \big| \mathcal{F}_{r-1}\bigr]\biggr] \\
&\quad = \mathbb{E}_n\bigl[1(E_{r-1})\hat{Z}_{r-1}^{\pm}\mathcal{E}(\hat{Y}_{r-1}^{\pm})\bigr] \leq \mathbb{E}_n\bigl[1(E_{r-1})\hat{Z}_{r-1}^{\pm}\mathcal{E}\bigl(n^{\eta}\mu^r/(\mu-1)\bigr)\bigr] \\
&\quad \leq Q_2\mu^{r-1}\biggl(\frac{(n^{\eta}\mu^r/(\mu-1))^{\kappa}}{n^{\kappa(1-\eta)}} + n^{-\varepsilon}\biggr),
\end{aligned}$$

where we used that on the event $E_{r-1}$ we have $\hat{Y}_{r-1}^{\pm} \leq \sum_{i=1}^{r-1} \mu^i n^{\eta} \leq n^{\eta}\mu^r/(\mu-1)$ and that $\mathcal{E}(t)$ is nondecreasing, followed by an application of Lemma 6.5. We then obtain that

$$\mathbb{P}_n(B_{r-1} \cap B_r^c \cap E_k) = O\biggl(\frac{\mu^{r(1+\kappa)}}{n^{\kappa(1-2\eta)}} + \mu^r n^{-\varepsilon}\biggr),$$

which implies that

$$\begin{aligned}
\sum_{r=1}^{s_n-1} \mathbb{P}_n(B_{r-1} \cap B_r^c \cap E_k) &= O\biggl(\frac{\mu^{s_n(1+\kappa)}}{n^{\kappa(1-2\eta)}} + \mu^{s_n} n^{-\varepsilon}\biggr) \\
&= O\bigl(n^{c(1+\kappa)-\kappa(1-2\eta)} + n^{c-\varepsilon}\bigr).
\end{aligned}$$

We have thus shown that, as $n \to \infty$,

$$\begin{aligned}
\mathbb{P}_n(C_k^c, W^{\pm} > 0) \leq{}& \mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap E_k, W^{\pm} > 0) \\
&+ O\bigl(n^{-\eta} + n^{c(1+\kappa)-\kappa(1-2\eta)} + n^{c-\varepsilon}\bigr),
\end{aligned}$$

with all the exponents of $n$ inside the big-O term strictly negative. To analyze the remaining probability, we first introduce one last conditioning event. Set $1 < u = \mu^{1-b}$ with $b = \min\{(1-\delta)/2, (\varepsilon-\gamma)/2, (\kappa\delta-\gamma)/4\}/(1-\delta) \in (0, 1/2)$, and define

$$J_{s_n} = \Bigl\{\inf_{r \geq s_n} \hat{Z}_r^{\pm}/u^r \geq 1\Bigr\}.$$

Now write

$$\begin{aligned}
&\mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap E_k, W^{\pm} > 0) \\
&\quad \leq \mathbb{P}_n(C_{s_n-1} \cap C_k^c \cap E_k \cap J_{s_n}) + P(J_{s_n}^c, W^{\pm} > 0) \\
&\quad \leq \sum_{r=s_n}^{k} \mathbb{P}_n(C_{r-1} \cap C_r^c \cap E_k \cap J_{s_n}) + P(J_{s_n}^c, W^{\pm} > 0).
\end{aligned}$$

By Lemma 6.3, we have

$$
\begin{aligned}
P\big(J_{s_n}^c, W^{\pm} > 0\big) &\le Q_1\big(u^{-\kappa s_n} + (u/\mu)^{\alpha^{\pm} s_n} 1\big(q^{\pm} > 0\big)\big) \\
&= O\big(n^{-\kappa c(1-b)} + n^{-\alpha^{\pm} cb} 1\big(q^{\pm} > 0\big)\big),
\end{aligned}
$$

where $\alpha^{\pm} = |\log \lambda^{\pm}|/\log \mu > 0$.

To bound each of the remaining probabilities, $\mathbb{P}_n(C_{r-1} \cap C_r^c \cap E_k \cap J_{s_n})$, use the union bound followed by Markov's inequality applied conditionally on $\mathcal{F}_{k-1}$, to obtain, for $s_n \le r \le k$,

$$
\begin{aligned}
&\mathbb{P}_n\big(C_{r-1} \cap C_r^c \cap E_k \cap J_{s_n}\big) \\
&\quad \le \mathbb{P}_n\big(\big|\hat{A}_r^{\pm} \cap \big(A_r^{\pm}\big)^c\big| > \hat{Z}_r^{\pm} n^{-\gamma}, C_{r-1} \cap E_k \cap J_{s_n}\big) \\
&\qquad + \mathbb{P}_n\big(\big|A_r^{\pm} \cap \big(\hat{A}_r^{\pm}\big)^c\big| > \hat{Z}_r^{\pm} n^{-\gamma}, C_{r-1} \cap E_k \cap J_{s_n}\big) \\
&\quad \le \mathbb{P}_n\big(\big|\hat{A}_r^{\pm} \cap \big(A_r^{\pm}\big)^c\big| > u^r n^{-\gamma}, C_{r-1} \cap E_{r-1}\big) \\
&\qquad + \mathbb{P}_n\big(\big|A_r^{\pm} \cap \big(\hat{A}_r^{\pm}\big)^c\big| > u^r n^{-\gamma}, C_{r-1} \cap E_{r-1}\big) \\
&\quad \le \frac{n^{\gamma}}{u^r} \mathbb{E}_n\big[1(C_{r-1} \cap E_{r-1}) \mathbb{E}_n\big[\big|\hat{A}_r^{\pm} \cap \big(A_r^{\pm}\big)^c\big| + \big|A_r^{\pm} \cap \big(\hat{A}_r^{\pm}\big)^c\big| \big| \mathcal{F}_{r-1}\big]\big].
\end{aligned}
$$

It follows from Lemma 6.6 that

$$
\mathbb{P}_n\big(C_{r-1} \cap C_r^c \cap E_k \cap J_{s_n}\big) \le Q_3 \frac{n^{\gamma}}{u^r} \cdot \mu^r \big(\mu^{\kappa r} n^{-\kappa(1-2\eta)} + r n^{-\varepsilon}\big),
$$

which in turn implies that, as $n \to \infty$,

$$
\begin{aligned}
&\sum_{r=s_n}^{k} \mathbb{P}_n\big(C_{r-1} \cap C_r^c \cap E_k \cap J_{s_n}\big) \\
&\qquad = O\left(n^{\gamma-\kappa(1-2\eta)} \sum_{r=s_n}^{k} \big(\mu^{1+\kappa}/u\big)^r + n^{\gamma-\varepsilon} \sum_{r=s_n}^{k} r(\mu/u)^r\right) \\
&\qquad = O\big(n^{\gamma-\kappa(1-2\eta)}\big(\mu^{1+\kappa}/u\big)^k + n^{\gamma-\varepsilon} k(\mu/u)^k\big) \\
&\qquad = O\big(n^{\gamma-\kappa(1-2\eta)+(1-\delta)(b+\kappa)} + n^{\gamma-\varepsilon+(1-\delta)b} \log n\big) \\
&\qquad = O\big(n^{(\gamma-\kappa\delta)/2+(1-\delta)b} + n^{\gamma-\varepsilon+(1-\delta)b} \log n\big).
\end{aligned}
$$

Since all the exponents inside the big-O term are strictly negative, this completes the proof. □

6.3. *Distances in the directed configuration model.* In this section, we prove Proposition 5.1, Proposition 5.2 and Corollary 5.4. As mentioned in Section 5, Propositions 5.1 and 5.2 together yield Theorem 5.3. As a preliminary result for the proof of Proposition 5.1, we first state and prove the following technical lemma.

Throughout the remainder of the section, we use $x \wedge y$ to denote the minimum of $x$ and $y$.

LEMMA 6.8. *For any nonnegative $x, x_0 > 0$, $y_i, z_i \geq 0$ with $z_i < x$ for all $i$, and any $m \geq 1$, we have*

$$-\frac{x_0}{x^2}(x_0 - x)^+ - \frac{x_0}{2} \max_{1 \leq i \leq m} \frac{z_i}{(x - z_i)^2} \leq \prod_{i=1}^{m} \left(1 - \frac{z_i}{x}\right)^{y_i} - \exp\left\{-\frac{1}{x_0} \sum_{i=1}^{m} y_i z_i\right\}$$
$$\leq \frac{|x - x_0|}{(x \wedge x_0)}.$$

PROOF. For the upper bound, note that

$$\prod_{i=2}^{m} \left(1 - \frac{z_i}{x}\right)^{y_i} = \exp\left\{\sum_{i=1}^{m} y_i \log\left(1 - \frac{z_i}{x}\right)\right\} \leq \exp\left\{-\frac{1}{x} \sum_{i=1}^{m} y_i z_i\right\}$$
$$\leq \exp\left\{-\frac{1}{x_0} \sum_{i=1}^{m} y_i z_i\right\} + \frac{|x - x_0|}{x_0 \wedge x},$$

where in the second step we used the inequality $\log(1 - t) \leq -t$ for $t \in [0, 1)$ and in the third step we used the inequality

$$\left|e^{-S/x} - e^{-S/x_0}\right| \leq \frac{S}{\xi^2} e^{-S/\xi} |x - x_0| \leq \sup_{t \geq 0} t e^{-t} \cdot \frac{|x - x_0|}{\xi} \leq \frac{|x - x_0|}{(x \wedge x_0)}$$

for any $S, x, x_0 \geq 0$ and some $\xi$ between $x$ and $x_0$.

Similarly, using the first-order Taylor expansion for $\log(1 - t)$, we obtain

$$\log(1 - c/x) = -\frac{c}{x} - \frac{c^2}{2x^2(1 - \xi')^2} = -\frac{c}{x_0} + \frac{c}{(\xi'')^2}(x - x_0) - \frac{c^2}{2x^2(1 - \xi')^2}$$

for any $c < x$, $0 < \xi' < c/x$, and $\xi''$ between $x$ and $x_0$, which in turn yields the inequality

$$\log(1 - c/x) \geq -\frac{c}{x_0} - \frac{c}{x^2}(x_0 - x)^+ - \frac{c^2}{2(x - c)^2}.$$

It follows that

$$\prod_{i=1}^{m} \left(1 - \frac{z_i}{x}\right)^{y_i} \geq \exp\left\{-\sum_{i=1}^{m} \left(\frac{y_i z_i}{x_0} + \frac{y_i z_i}{x^2}(x_0 - x)^+ + \frac{y_i z_i^2}{2(x - z_i)^2}\right)\right\}$$
$$\geq \exp\left\{-\frac{1}{x_0} \sum_{i=1}^{m} y_i z_i\right\}$$
$$- \exp\left\{-\frac{1}{x_0} \sum_{i=1}^{m} y_i z_i\right\} \sum_{i=1}^{m} \left(\frac{y_i z_i}{x^2}(x_0 - x)^+ + \frac{y_i z_i^2}{2(x - z_i)^2}\right)$$

$$\geq \exp\left\{-\frac{1}{x_0}\sum_{i=1}^{m} y_i z_i\right\} - \frac{x_0}{x^2}(x_0 - x)^+ - \frac{x_0}{2}\max_{1\leq i\leq m}\frac{z_i}{(x-z_i)^2}. \quad \square$$

We are now ready to prove Proposition 5.1.

PROOF OF PROPOSITION 5.1. Let $0 < \gamma < \min\{\kappa\delta, \varepsilon\}$, and construct the pairs of processes $\{(Z_i^+, \hat{Z}_i^+) : i \geq 0\}$ and $\{(Z_i^-, \hat{Z}_i^-) : i \geq 0\}$ according to the coupling described in Section 4.2, independently of each other. Now define the event

$$\mathcal{E}_k = \bigcap_{m=1}^{\lceil k/2\rceil+1} \{\hat{Z}_m^+(1-n^{-\gamma}) \leq Z_m^+ \leq \hat{Z}_m^+(1+n^{-\gamma}),$$
$$\hat{Z}_m^-(1-n^{-\gamma}) \leq Z_m^- \leq \hat{Z}_m^-(1+n^{-\gamma})\}$$

and note that since $\{(Z_i^+, \hat{Z}_i^+) : i \geq 0\}$ and $\{(Z_i^-, \hat{Z}_i^-) : i \geq 0\}$ are independent, then Corollary 4.2 gives $\mathbb{P}_n(\mathcal{E}_k^c) = O(n^{-a_1})$ for some $a_1 > 0$.

Next, use the triangle inequality to get

$$\left|\mathbb{P}_n(H_n > k) - E\left[\exp\left\{-\frac{\sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2\rceil}^+ \hat{Z}_{\lfloor i/2\rfloor}^-}{\nu n}\right\}\right]\right|$$
$$\leq \left|\mathbb{P}_n(H_n > k) - \mathbb{E}_n\left[\exp\left\{-\frac{\sum_{i=2}^{k+1} Z_{\lceil i/2\rceil}^+ Z_{\lfloor i/2\rfloor}^-}{\nu n}\right\}\right]\right| \tag{6.11}$$
$$+ \left|\mathbb{E}_n\left[\exp\left\{-\frac{\sum_{i=2}^{k+1} Z_{\lceil i/2\rceil}^+ Z_{\lfloor i/2\rfloor}^-}{\nu n}\right\}\right] - E\left[\exp\left\{-\frac{\sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2\rceil}^+ \hat{Z}_{\lfloor i/2\rfloor}^-}{\nu n}\right\}\right]\right|. \tag{6.12}$$

We start by bounding (6.12), for which we use the independence of $\{\hat{Z}_i^+\}$ and $\{\hat{Z}_i^-\}$ from the bi-degree sequence $(\mathbf{D}_n^-, \mathbf{D}_n^+)$, and the inequality $|e^{-x} - e^{-y}| \leq e^{-(x\wedge y)}|x - y|$ for $x, y \geq 0$ to obtain

$$\left|\mathbb{E}_n\left[\exp\left\{-\frac{\sum_{i=2}^{k+1} Z_{\lceil i/2\rceil}^+ Z_{\lfloor i/2\rfloor}^-}{\nu n}\right\}\right] - E\left[\exp\left\{-\frac{\sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2\rceil}^+ \hat{Z}_{\lfloor i/2\rfloor}^-}{\nu n}\right\}\right]\right|$$
$$\leq \mathbb{E}_n\left[\left|\exp\left\{-\frac{\sum_{i=2}^{k+1} Z_{\lceil i/2\rceil}^+ Z_{\lfloor i/2\rfloor}^-}{\nu n}\right\} - \exp\left\{-\frac{\sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2\rceil}^+ \hat{Z}_{\lfloor i/2\rfloor}^-}{\nu n}\right\}\right|\right]$$
$$\leq \frac{1}{n\nu}\mathbb{E}_n\left[1(\mathcal{E}_k)\exp\left\{-\frac{S_k(1-n^{-\gamma})^2}{\nu n}\right\}\left|\sum_{i=2}^{k+1}(Z_{\lceil i/2\rceil}^+ Z_{\lfloor i/2\rfloor}^- - \hat{Z}_{\lceil i/2\rceil}^+ \hat{Z}_{\lfloor i/2\rfloor}^-)\right|\right]$$
$$+ \mathbb{P}_n(\mathcal{E}_k^c),$$

where $S_k = \sum_{i=2}^{k+1} Z^+_{\lceil i/2\rceil} Z^-_{\lfloor i/2\rfloor}$. Since on the event $\mathcal{E}_k$ we have that for all $2 \le i \le k+1$,

$$(1-2n^{-\gamma}+n^{-2\gamma})\hat{Z}^+_{\lceil i/2\rceil}\hat{Z}^-_{\lfloor i/2\rfloor} \le Z^+_{\lceil i/2\rceil}Z^-_{\lfloor i/2\rfloor} \le (1+2n^{-\gamma}+n^{-2\gamma})\hat{Z}^+_{\lceil i/2\rceil}\hat{Z}^-_{\lfloor i/2\rfloor},$$

then for all $n \ge 1$,

$$|Z^+_{\lceil i/2\rceil}Z^-_{\lfloor i/2\rfloor} - \hat{Z}^+_{\lceil i/2\rceil}\hat{Z}^-_{\lfloor i/2\rfloor}| \le 3n^{-\gamma}\hat{Z}^+_{\lceil i/2\rceil}\hat{Z}^-_{\lfloor i/2\rfloor} \le \frac{3n^{-\gamma}}{(1-n^{-\gamma})^2} Z^+_{\lceil i/2\rceil} Z_{\lfloor i/2\rfloor}.$$

It follows that (6.12) is bounded from above by

$$\frac{3n^{-\gamma}}{(1-n^{-\gamma})^2}\mathbb{E}_n\left[\exp\left\{-\frac{S_k(1-n^{-\gamma})^2}{\nu n}\right\}\frac{1}{n\nu}S_k\right] + \mathbb{P}_n(\mathcal{E}_k^c) \le \frac{3e^{-1}n^{-\gamma}}{(1-n^{-\gamma})^4} + \mathbb{P}_n(\mathcal{E}_k^c),$$

where we used the observation that $\sup_{x\ge 0} e^{-x}x = e^{-1}$.

We now proceed to bound (6.11). From (5.4), we have that

$$\mathbb{P}_n(H_n > k) = \mathbb{E}_n\left[1(\mathscr{S}_{k+1} \le L_n)\prod_{i=2}^{k+1}\prod_{s=0}^{Z^+_{\lceil i/2\rceil}-1}\left(1 - \frac{Z^-_{\lfloor i/2\rfloor}}{L_n - \mathscr{S}_{i-2} - s}\right)\right],$$

where $\mathscr{S}_k$ is defined via (5.3). Recall that from Assumption 2.1 we have $|L_n - n\nu| \le n^{1-\varepsilon}$ and, therefore, $\{\mathscr{S}_{k+1} \le n^b\} \subseteq \{\mathscr{S}_{k+1} \le L_n\}$ for any $1-\delta < b < 1$ and all sufficiently large $n$. Now note that

$$\begin{aligned}\mathbb{E}_n\left[1(\mathscr{S}_{k+1} \le n^b)\prod_{i=2}^{k+1}\left(1-\frac{Z^-_{\lfloor i/2\rfloor}}{L_n - \mathscr{S}_{k+1}}\right)^{Z^+_{\lceil i/2\rceil}}\right] &\le \mathbb{P}_n(H_n > k)\\ &\le \mathbb{E}_n\left[\prod_{i=2}^{k+1}\left(1-\frac{Z^-_{\lfloor i/2\rfloor}}{L_n}\right)^{Z^+_{\lceil i/2\rceil}}\right].\end{aligned}$$

Using Lemma 6.8 with $x = L_n$ and $x_0 = \nu n$ gives

$$\mathbb{P}_n(H_n > k) \le \mathbb{E}_n[e^{-S_k/(\nu n)}] + \frac{|L_n - \nu n|}{L_n \wedge (\nu n)} = \mathbb{E}_n[e^{-S_k/(\nu n)}] + O(n^{-\varepsilon}).$$

Similarly, using Lemma 6.8 with $x = L_n - \mathscr{S}_{k+1}$ and $x_0 = \nu n$ we obtain

$$\begin{aligned}\mathbb{P}_n(H_n > k) &\ge \mathbb{E}_n[e^{-S_k/(\nu n)}1(\mathscr{S}_{k+1} \le n^b)]\\ &\quad - \mathbb{E}_n\left[1(\mathscr{S}_{k+1} \le n^b)\frac{\nu n}{(L_n - \mathscr{S}_{k+1})^2}(\nu n - L_n + \mathscr{S}_{k+1})^+\right]\\ &\quad - \mathbb{E}_n\left[1(\mathscr{S}_{k+1} \le n^b)\frac{\nu n}{2}\max_{2\le i\le k+1}\frac{Z^-_{\lfloor i/2\rfloor}}{(L_n - \mathscr{S}_{k+1} - Z^-_{\lfloor i/2\rfloor})^2}\right]\\ &\ge \mathbb{E}_n[e^{-S_k/(\nu n)}] - \mathbb{P}_n(\mathscr{S}_{k+1} > n^b)\end{aligned}$$

$$
\begin{aligned}
&- \frac{(n^{1-\varepsilon} + n^b)^+}{(L_n - n^b)^2} - \frac{\nu n}{2} \cdot \frac{n^b}{(L_n - 2n^b)^2} \\
&= \mathbb{E}_n\big[e^{-S_k/(\nu n)}\big] - \mathbb{P}_n\big(\mathscr{S}_{k+1} > n^b\big) - O\big(n^{-(1-b)}\big).
\end{aligned}
$$

Finally, note that using Markov's inequality we obtain

$$
\begin{aligned}
\mathbb{P}_n\big(\mathscr{S}_{k+1} > n^b\big) &\le \mathbb{P}_n\big(\mathcal{E}_k^c\big) + \mathbb{P}_n\big(\mathscr{S}_{k+1} > n^b, \mathcal{E}_k\big) \\
&\le \mathbb{P}_n\big(\mathcal{E}_k^c\big) + P\left(\sum_{j=1}^{\lceil k/2 \rceil} (\hat{Z}_j^+ + \hat{Z}_j^-)(1 + n^{-\gamma}) > n^b\right) \\
&\le \mathbb{P}_n\big(\mathcal{E}_k^c\big) + \frac{1 + n^{-\gamma}}{n^b} \sum_{j=1}^{\lceil k/2 \rceil} E[\hat{Z}_j^+ + \hat{Z}_j^-] \\
&= O\big(n^{-a_1} + \mu^{\lceil k/2 \rceil} n^{-b}\big),
\end{aligned}
$$

where in the last step we used the observation that $\sum_{j=1}^{\lceil k/2 \rceil} E[\hat{Z}_j^+ + \hat{Z}_j^-] = 2\sum_{j=1}^{\lceil k/2 \rceil} \nu \mu^{j-1} = O(\mu^{\lceil k/2 \rceil})$. Since $k \le 2(1-\delta)\log n/(\log \mu)$, then $\mu^{\lceil k/2 \rceil} = O(n^{1-\delta})$ and the result follows. □

We now proceed to prove Proposition 5.2, which shows that the expression derived for the hopcount in Proposition 5.1 can be closely approximated by an expression in terms of the limiting martingales of the branching processes $\{\hat{Z}_k^+ : k \ge 0\}$ and $\{\hat{Z}_k^- : k \ge 0\}$. Note that this result is independent of the bi-degree sequence $(\mathbf{D}_n^-, \mathbf{D}_n^+)$ since it involves only the coupled branching processes.

PROOF OF PROPOSITION 5.2. Fix $0 < \epsilon < \kappa/(2 + 2\kappa)$ and set $m_n = \lfloor (1-\epsilon)\log n/\log \mu \rfloor$. We start by noting that for $1 \le k \le m_n$ the inequality $|e^{-x} - e^{-y}| \le |x - y|$ for $x, y \ge 0$ gives

$$
\begin{aligned}
&\left| E\left[\exp\left\{-\frac{1}{\nu n} \sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2 \rceil}^+ \hat{Z}_{\lfloor i/2 \rfloor}^-\right\} - \exp\left\{-\frac{\nu \mu^k}{(\mu - 1)n} W^+ W^-\right\}\right]\right| \\
&\le E\left[\left|\frac{1}{\nu n} \sum_{i=2}^{k+1} \hat{Z}_{\lceil i/2 \rceil}^+ \hat{Z}_{\lfloor i/2 \rfloor}^- - \frac{\nu \mu^k}{(\mu - 1)n} W^- W^+\right|\right] \\
&\le \frac{1}{\nu n} \sum_{i=2}^{k+1} E[\hat{Z}_{\lceil i/2 \rceil}^+] E[\hat{Z}_{\lfloor i/2 \rfloor}^-] + \frac{\nu \mu^k}{(\mu - 1)n} E[W^+] E[W^-] \\
&= \frac{1}{\nu n} \sum_{i=2}^{k+1} \nu^2 \mu^{\lceil i/2 \rceil + \lfloor i/2 \rfloor - 2} + \frac{\nu \mu^k}{(\mu - 1)n} E[W^+] E[W^-]
\end{aligned}
$$

$$= \frac{\nu}{\mu^2 n}\left(\sum_{i=2}^{k+1} \mu^i + \frac{\mu^{k+2}}{\mu - 1}\right) \le \frac{2\nu\mu^k}{n(\mu-1)} \le \frac{2\nu\mu^{m_n}}{n(\mu-1)} = O(n^{-\epsilon})$$

as $n \to \infty$, where in the second equality we used the observation that $\lceil i/2\rceil + \lfloor i/2\rfloor = i$ for all $i \in \mathbb{N}$, and $E[W^-] = E[W^+] = 1$ (since $f^+$ and $f^-$ have finite moments of order $1+\kappa$). It remains to consider the case $k > m_n$.

Suppose from now on that $k > m_n$ and note that

$$\frac{\mu^k}{\mu-1} = \sum_{i=m_n}^{k+1} \mu^{i-2} + \frac{\mu^{m_n-2}}{\mu-1},$$

and, therefore, using $|e^{-x} - e^{-y}| \le |x-y|$ for $x, y \ge 0$ and the triangle inequality we obtain

$$\left| E\left[\exp\left\{-\frac{1}{\nu n}\sum_{i=2}^{k+1} \hat{Z}^+_{\lceil i/2\rceil}\hat{Z}^-_{\lfloor i/2\rfloor}\right\} - \exp\left\{-\frac{\nu\mu^k}{(\mu-1)n}W^+W^-\right\}\right]\right|$$

$$\le E\left[\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1} \hat{Z}^+_{\lceil i/2\rceil}\hat{Z}^-_{\lfloor i/2\rfloor}\right\}\right.\right.$$

$$\left.\left. - \exp\left\{-\frac{\nu}{n}\sum_{i=m_n}^{k+1} \mu^{i-2} W^+W^-\right\}\right|\right] \tag{6.13}$$

$$+ E\left[\frac{1}{\nu n}\sum_{i=2}^{m_n-1} \hat{Z}^+_{\lceil i/2\rceil}\hat{Z}^-_{\lfloor i/2\rfloor} + \frac{\nu\mu^{m_n-2}}{(\mu-1)n}W^+W^-\right], \tag{6.14}$$

where (6.14) is of order $O(n^{-\epsilon})$ as shown above. To bound (6.13), let $W_k^\pm = \hat{Z}_k^\pm/(\nu\mu^{k-1})$, set $\eta = \epsilon/2$, and define the events $\mathcal{B} = \{W^+W^- > 0\}$,

$$\mathcal{C}_r^\pm = \left\{\max_{\lfloor m_n/2\rfloor \le j \le r} 1(W_j^\pm > 0)|W_j^\pm - W^\pm|/W_j^\pm \le n^{-\eta}\right\}, \qquad \text{for } r \ge \lfloor m_n/2\rfloor,$$

and $\mathcal{C}_k = \mathcal{C}^+_{\lfloor (k+1)/2\rfloor} \cap \mathcal{C}^-_{\lceil (k+1)/2\rceil}$. Next, use the inequality $|e^{-y} - e^{-x}| \le e^{-(x\wedge y)}|x-y|$ for $x, y \ge 0$ to obtain

$$E\left[1(\mathcal{B}\cap\mathcal{C}_k)\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1} \hat{Z}^+_{\lceil i/2\rceil}\hat{Z}^-_{\lfloor i/2\rfloor}\right\} - \exp\left\{-\frac{\nu}{n}\sum_{i=m_n}^{k+1}\mu^{i-2}W^+W^-\right\}\right|\right]$$

$$\le E\left[1(\mathcal{B}\cap\mathcal{C}_k)\exp\left\{-\frac{\nu}{\mu^2 n}\sum_{i=m_n}^{k+1}\mu^i W^+_{\lceil i/2\rceil}W^-_{\lfloor i/2\rfloor}(1-n^{-\eta})\right\}\right.$$

$$\left.\times\left|\frac{\nu}{\mu^2 n}\sum_{i=m_n}^{k+1}\mu^i\left(W^+_{\lceil i/2\rceil}W^-_{\lfloor i/2\rfloor} - W^+W^-\right)\right|\right]$$

$$
\begin{aligned}
&\le \frac{3\nu n^{-\eta}}{\mu^2 n} E\left[\exp\left\{-\frac{\nu}{\mu^2 n}\sum_{i=m_n}^{k+1}\mu^i W^+_{\lceil i/2\rceil}W^-_{\lfloor i/2\rfloor}(1-n^{-\eta})\right\}\right. \\
&\qquad \left.\times \sum_{i=m_n}^{k+1}\mu^i W^+_{\lceil i/2\rceil}W^-_{\lfloor i/2\rfloor}\right] \\
&\le \frac{3n^{-\eta}}{\mu^2(1-n^{-\eta})}\sup_{x>0} e^{-x}x \le \frac{3n^{-\eta}}{\mu^2(1-n^{-\eta})},
\end{aligned}
$$

where we have used the observation that on the event $\mathcal{B}\cap\mathcal{C}_k$ we have

$$
\begin{aligned}
\left|W^+_{\lceil i/2\rceil}W^-_{\lfloor i/2\rfloor} - W^+W^-\right| &\le \left|W^+_{\lceil i/2\rceil} - W^+\right|W^-_{\lfloor i/2\rfloor} + W^+\left|W^-_{\lfloor i/2\rfloor} - W^-\right| \\
&\le W^+_{\lceil i/2\rceil}W^-_{\lfloor i/2\rfloor}n^{-\eta} + \left(1+n^{-\eta}\right)n^{-\eta}W^+_{\lceil i/2\rceil}W^-_{\lfloor i/2\rfloor} \\
&\le 3n^{-\eta}W^+_{\lceil i/2\rceil}W^-_{\lfloor i/2\rfloor},
\end{aligned}
$$

and that $\sup_{x\ge 0} xe^{-x} = e^{-1}$. Also, by using that $|e^{-x}-e^{-y}|\le 1$ for $x, y\ge 0$, we obtain

$$
\begin{aligned}
&E\left[1\left(\mathcal{B}\cap\mathcal{C}_k^c\right)\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1}\hat{Z}^+_{\lceil i/2\rceil}\hat{Z}^-_{\lfloor i/2\rfloor}\right\} - \exp\left\{-\frac{\nu}{\mu^2 n}\sum_{i=m_n}^{k+1}\mu^i W^+W^-\right\}\right|\right] \\
&\qquad \le P\left(W^+>0, \left(\mathcal{C}^+_{\lceil (k+1)/2\rceil}\right)^c\right) + P\left(W^->0, \left(\mathcal{C}^-_{\lfloor (k+1)/2\rfloor}\right)^c\right).
\end{aligned}
$$

To bound the last two probabilities, set $u=\mu^{\epsilon(2+\kappa)/(\kappa-\kappa\epsilon)}\in(1,\mu)$, define the event $\mathcal{D}^\pm = \{\inf_{j\ge\lfloor m_n/2\rfloor}\hat{Z}^\pm_j/u^j\ge 1\}$, and note that for any $r\ge\lfloor m_n/2\rfloor$,

$$
\begin{aligned}
P\left(W^\pm>0, \left(\mathcal{C}^\pm_r\right)^c\right) &\le P\left(W^\pm>0, \left(\mathcal{C}^\pm_r\right)^c\cap\mathcal{D}^\pm\right) + P\left(W^\pm>0, \left(\mathcal{D}^\pm\right)^c\right) \\
&\le \sum_{j=\lfloor m_n/2\rfloor}^{r} P\left(\left|W^\pm_j - W^\pm\right| > n^{-\eta}W^\pm_j, \mathcal{D}^\pm\right) \\
&\qquad + P\left(W^\pm>0, \left(\mathcal{D}^\pm\right)^c\right).
\end{aligned}
$$

By Lemma 6.3, we have that

$$
P\left(W^\pm>0, \left(\mathcal{D}^\pm\right)^c\right) \le Q_1\left(u^{-\kappa\lfloor m_n/2\rfloor} + (u/\mu)^{\alpha^\pm\lfloor m_n/2\rfloor}1\left(q^\pm>0\right)\right),
$$

for some constant $Q_1<\infty$, where $\lambda^\pm = \sum_{i=1}^\infty if^\pm(i)(q^\pm)^{i-1}\in[0,1)$ and $\alpha^\pm = |\log\lambda^\pm|/\log\mu>0$ if $q^\pm>0$; and for the remaining probabilities we use the representation (6.1) for $W^\pm - W^\pm_j$ and Lemma 6.2, applied conditionally on $\hat{Z}^\pm_j$, to obtain

$$
\begin{aligned}
&P\left(\left|W^\pm_j - W^\pm\right| > n^{-\eta}W^\pm_j, \mathcal{D}^\pm\right) \\
&\qquad \le P\left(\left|\sum_{\mathbf{i}\in\hat{A}^\pm_j}\left(\mathcal{W}^\pm_{\mathbf{i}} - 1\right)\right| > n^{-\eta}\hat{Z}^\pm_j, \hat{Z}^\pm_j\ge u^j\right)
\end{aligned}
$$

$$\le E\left[1(\hat{Z}_j^{\pm} \ge u^j) P\left(\left|\sum_{\mathbf{i}\in\hat{A}_j^{\pm}} (\mathcal{W}_{\mathbf{i}}^{\pm} - 1)\right| > n^{-\eta}\hat{Z}_j^{\pm}\middle|\hat{Z}_j^{\pm}\right)\right]$$

$$\le Q_{1+\kappa} E\big[|\mathcal{W}^{\pm} - 1|^{1+\kappa}\big] \cdot E\left[1(\hat{Z}_j^{\pm} \ge u^j)\frac{\hat{Z}_j^{\pm}}{(\hat{Z}_j^{\pm} n^{-\eta})^{1+\kappa}}\right]$$

$$\le Q_{1+\kappa} E\big[|\mathcal{W}^{\pm} - 1|^{1+\kappa}\big]\frac{n^{\eta(1+\kappa)}}{u^{\kappa j}},$$

where $\{\mathcal{W}_{\mathbf{i}}^{\pm}\}$ are i.i.d. random variables having the same distribution as $\mathcal{W}^{\pm} = \lim_{r\to\infty} \mathcal{Z}_r^{\pm}/\mu^r$, and $Q_{1+\kappa}$ is a finite constant that depends only on $\kappa$. It follows from our choice of $u$ and $\eta$ that

$$\begin{aligned}
&P(W^{\pm} > 0, (\mathcal{C}_r^{\pm})^c) \\
&\quad = O\left(\sum_{j=\lfloor m_n/2\rfloor}^{r} \frac{n^{\eta(1+\kappa)}}{u^{\kappa j}} + u^{-\kappa\lfloor m_n/2\rfloor} + (u/\mu)^{\alpha^{\pm}\lfloor m_n/2\rfloor} 1(q^{\pm} > 0)\right) \\
&\quad = O(n^{\eta(1+\kappa)} u^{-\kappa\lfloor m_n/2\rfloor} + (u/\mu)^{\alpha^{\pm}\lfloor m_n/2\rfloor} 1(q^{\pm} > 0)) \\
&\quad = O(n^{-\eta} + n^{-\alpha^{\pm}(1/2-\epsilon(1+\kappa)/\kappa)} 1(q^{\pm} > 0)).
\end{aligned}$$

Hence, as $n \to \infty$,

$$\begin{aligned}
&E\left[1(\mathcal{B})\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1} \hat{Z}_{\lceil i/2\rceil}^{+}\hat{Z}_{\lfloor i/2\rfloor}^{-}\right\} - \exp\left\{-\frac{\nu}{n}\sum_{i=m_n}^{k+1} \mu^i W^+ W^-\right\}\right|\right] \\
&\quad = O(n^{-\eta} + n^{-\alpha^{+}(1/2-\epsilon(1+\kappa)/\kappa)} 1(q^{+} > 0) + n^{-\alpha^{-}(1/2-\epsilon(1+\kappa)/\kappa)} 1(q^{-} > 0)).
\end{aligned}$$

Finally, on the event $\mathcal{B}^c$ we have

$$\begin{aligned}
&E\left[1(\mathcal{B}^c)\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1} \hat{Z}_{\lceil i/2\rceil}^{+}\hat{Z}_{\lfloor i/2\rfloor}^{-}\right\} - \exp\left\{-\frac{\nu}{n}\sum_{i=m_n}^{k+1} \mu^i W^- W^+\right\}\right|\right] \\
&\quad = E\left[1(\mathcal{B}^c, \hat{Z}_{\lceil m_n/2\rceil}^{+}\hat{Z}_{\lfloor m_n/2\rfloor}^{-} > 0)\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1} \hat{Z}_{\lceil i/2\rceil}^{+}\hat{Z}_{\lfloor i/2\rfloor}^{-}\right\} - 1\right|\right] \\
&\quad \le P(\mathcal{B}^c, \hat{Z}_{\lceil m_n/2\rceil}^{+}\hat{Z}_{\lfloor m_n/2\rfloor}^{-} > 0) \\
&\quad \le P(W^+ = 0, \hat{Z}_{\lceil m_n/2\rceil}^{+} > 0) + P(W^- = 0, \hat{Z}_{\lfloor m_n/2\rfloor}^{-} > 0).
\end{aligned} \tag{6.15}$$

By Lemma 6.1, conditionally on $W^{\pm} = 0$, $\{\hat{Z}_k^{\pm} : k \ge 1\}$ is a delayed branching process with offspring distributions $(\tilde{g}^{\pm}, \tilde{f}^{\pm})$, as defined in the lemma, having means $\nu^{\pm}$ and $\lambda^{\pm} < 1$, respectively. Moreover, by Theorem 4 in Athreya and Ney

(2004), $W^{\pm}=0$ implies that either $q^{\pm}>0$ or $\hat{Z}_1^{\pm}=0$. Therefore, from Markov's inequality we obtain

$$\begin{aligned}
&P\bigl(W^{\pm}=0, \hat{Z}^{\pm}_{\lfloor m_n/2\rfloor}>0\bigr)\\
&\quad\le E\bigl[\hat{Z}^{\pm}_{\lfloor m_n/2\rfloor}|W^{\pm}=0\bigr]1\bigl(q^{\pm}>0\bigr)=\nu^{\pm}\bigl(\lambda^{\pm}\bigr)^{\lfloor m_n/2\rfloor-1}1\bigl(q^{\pm}>0\bigr)\\
&\quad=O\bigl(n^{-(1-\epsilon)|\log\lambda^{\pm}|/\log\mu}1\bigl(q^{\pm}>0\bigr)\bigr)=O\bigl(n^{-(1-\epsilon)\alpha^{\pm}}1\bigl(q^{\pm}>0\bigr)\bigr).
\end{aligned}$$

We conclude that as $n\to\infty$,

$$\begin{aligned}
&E\left[\left|\exp\left\{-\frac{1}{\nu n}\sum_{i=m_n}^{k+1}\hat{Z}^{+}_{\lceil i/2\rceil}\hat{Z}^{-}_{\lfloor i/2\rfloor}\right\}-\exp\left\{-\frac{\nu}{n}\sum_{i=m_n}^{k+1}\mu^{i}W^{+}W^{-}\right\}\right|\right]\\
&\quad=O\bigl(n^{-\epsilon/2}+n^{-\alpha^{+}(1/2-\epsilon(1+\kappa)/\kappa)}1\bigl(q^{+}>0\bigr)\\
&\qquad+n^{-\alpha^{-}(1/2-\epsilon(1+\kappa)/\kappa)}1\bigl(q^{-}>0\bigr)\bigr). \qquad\square
\end{aligned}$$

The last proof in this section is that of Corollary 5.4, which computes an expression for the probability that two randomly chosen nodes are connected by a directed path.

PROOF OF COROLLARY 5.4. Fix $0<\delta<1/4$ and set $\omega_n=2(1-\delta)\log n/(\log\mu)$. Our analysis is based on splitting the probability that the hopcount is finite into two terms:

$$\mathbb{P}_n(H_n<\infty)=\mathbb{P}_n(H_n\le\omega_n)+\mathbb{P}_n(\omega_n<H_n<\infty),$$

where for the first probability we will use the approximation provided by Theorem 5.3. Intuitively, the second term corresponds to an event that should be negligible in the limit, since it is unlikely that if there exists a directed path between the two randomly chosen nodes it will not have been discovered after $\omega_n$ steps of the exploration process.

First, note that from Lemma 6.1 we have

$$s^{\pm}=1-\sum_{t=0}^{\infty}g^{\pm}(t)\bigl(q^{\pm}\bigr)^{t}\qquad\text{with } q^{\pm}=P\bigl(\mathcal{Z}_k^{\pm}=0\text{ for some }k\ge 1\bigr),$$

and since $W^{+}$ and $W^{-}$ are independent, $s^{+}s^{-}=P(W^{+}W^{-}>0)$. As in the previous proof, let $\mathcal{B}=\{W^{+}W^{-}>0\}$ and write

$$\begin{aligned}
&E\left[\exp\left\{-\frac{\nu\mu^{k}}{(\mu-1)n}W^{+}W^{-}\right\}\right]\\
&\quad=E\left[\exp\left\{-\frac{\nu\mu^{k}}{(\mu-1)n}W^{+}W^{-}\right\}1(\mathcal{B})\right]+P\bigl(\mathcal{B}^{c}\bigr).
\end{aligned}$$

Next, use the triangle inequality followed by an application of Theorem 5.3 to obtain

$$
\begin{aligned}
&\left|\mathbb{P}_n(H_n<\infty)-s^+s^-\right| \\
&\quad=\left|\mathbb{P}_n(H_n\le\omega_n)+\mathbb{P}_n(\omega_n<H_n<\infty)-P(\mathcal{B})\right| \\
&\quad\le\left|\mathbb{P}_n(H_n\le\omega_n)-P(\mathcal{B})\right|+\mathbb{P}_n(\omega_n<H_n<\infty) \\
&\quad\le\left|\mathbb{P}_n(H_n\le\omega_n)-1+E\left[\exp\left\{-\frac{\nu\mu^{\omega_n}}{(\mu-1)n}W^+W^-\right\}\right]\right| \\
&\qquad+\left|1-E\left[\exp\left\{-\frac{\nu\mu^{\omega_n}}{(\mu-1)n}W^+W^-\right\}\right]-P(\mathcal{B})\right|+\mathbb{P}_n(\omega_n<H_n<\infty) \\
&\quad\le E\left[\exp\left\{-\frac{\nu\mu^{\omega_n}}{(\mu-1)n}W^+W^-\right\}1(\mathcal{B})\right]+\mathbb{P}_n(\omega_n<H_n<\infty)+O(n^{-c_1})
\end{aligned}
$$

for some $c_1>0$, as $n\to\infty$.

To analyze $\mathbb{P}_n(\omega_n<H_n<\infty)$, use the expression in (5.4) to see that for any $k\ge 0$,

$$
\begin{aligned}
&\mathbb{P}_n(k<H_n<\infty) \\
&\quad\le\mathbb{E}_n\left[1(Z^+_{\lceil(k+1)/2\rceil}Z^-_{\lfloor(k+1)/2\rfloor}>0)\prod_{i=2}^{k+1}\left(1-\frac{Z^-_{\lfloor i/2\rfloor}}{L_n}\right)^{Z^+_{\lceil i/2\rceil}}\right].
\end{aligned}
\tag{6.16}
$$

Note that the same steps in the proofs of Propositions 5.1 and 5.2 give that (6.16) is equal to

$$
\begin{aligned}
&E\left[1(\hat{Z}^+_{\lceil(k+1)/2\rceil}\hat{Z}^-_{\lfloor(k+1)/2\rfloor}>0)\exp\left\{-\frac{\nu\mu^k}{(\mu-1)n}W^+W^-\right\}\right]+O(n^{-a}+n^{-b}) \\
&\quad\le E\left[\exp\left\{-\frac{\nu\mu^k}{(\mu-1)n}W^-W^+\right\}1(\mathcal{B})\right]+P(\hat{Z}^+_{\lceil(k+1)/2\rceil}\hat{Z}^-_{\lfloor(k+1)/2\rfloor}>0,\mathcal{B}^c) \\
&\qquad+O(n^{-a}+n^{-b})
\end{aligned}
$$

for some constants $a,b>0$ and all $0\le k\le\omega_n$. It follows that

$$
\begin{aligned}
\left|\mathbb{P}_n(H_n<\infty)-s^+s^-\right|&\le 2E\left[\exp\left\{-\frac{\nu n^{1-2\delta}}{\mu-1}W^+W^-\right\}1(\mathcal{B})\right] \\
&\quad+P(\hat{Z}^+_{\lceil(\omega_n+1)/2\rceil}\hat{Z}^-_{\lfloor(\omega_n+1)/2\rfloor}>0,\mathcal{B}^c) \\
&\quad+O(n^{-\min\{a,b,c_1\}}),
\end{aligned}
$$

where we have used the observation that $\mu^{\omega_n}\ge n^{2(1-\delta)}$.

Next, use Lemma 6.1 to see that conditionally on $W^\pm=0$ the process $\{\hat{Z}^\pm_k:k\ge 1\}$ is a subcritical delayed branching process with offspring distributions $(\tilde{g}^\pm,\tilde{f}^\pm)$, defined in the lemma, having means $\nu^{\pm}$ and $\lambda^{\pm} < 1$, respectively. It then follows from the union bound and Markov's inequality that

$$\begin{aligned}
&P\big(\hat{Z}^{+}_{\lceil(\omega_n+1)/2\rceil}\hat{Z}^{-}_{\lfloor(\omega_n+1)/2\rfloor} > 0, \mathcal{B}^c\big)\\
&\quad \le P\big(\hat{Z}^{+}_{\lceil(\omega_n+1)/2\rceil} \ge 1, W^+ = 0\big) + P\big(\hat{Z}^{-}_{\lfloor(\omega_n+1)/2\rfloor} \ge 1, W^- = 0\big)\\
&\quad \le \big(1-s^+\big)\nu^+\big(\lambda^+\big)^{\lceil(\omega_n+1)/2\rceil-1} + \big(1-s^-\big)\nu^-\big(\lambda^-\big)^{\lfloor(\omega_n+1)/2\rfloor-1}\\
&\quad = O\big(n^{-(1-\delta)\alpha^+}1\big(q^+>0\big) + n^{-(1-\delta)\alpha^-}1\big(q^->0\big)\big),
\end{aligned}$$

where $\alpha^{\pm} = |\log\lambda^{\pm}|/\log\mu$ provided $q^{\pm} > 0$.

Finally, to analyze the remaining expectation define the event $\mathcal{D} = \{W^+ > n^{-1/4}, W^- > n^{-1/4}\}$ and note that

$$\begin{aligned}
&E\left[\exp\left\{-\frac{\nu n^{1-2\delta}}{\mu-1}W^+W^-\right\}1(\mathcal{B})\right]\\
&\quad \le E\left[\exp\left\{-\frac{\nu n^{1-2\delta}}{\mu-1}W^+W^-\right\}1(\mathcal{D})\right] + P\big(\mathcal{B}\cap\mathcal{D}^c\big)\\
&\quad \le \exp\left\{-\frac{\nu}{(\mu-1)}\cdot n^{1/2-2\delta}\right\} + P\big(\mathcal{B}\cap\mathcal{D}^c\big)\\
&\quad \le P\big(0 < W^+ \le n^{-1/4}\big) + P\big(0 < W^- \le n^{-1/4}\big) + o\big(n^{-1}\big).
\end{aligned}$$

The proof of Lemma 6.3 gives that $P(0 < W^{\pm} \le n^{-1/4}) = O(n^{-1}1(q^{\pm} = 0) + n^{-\alpha^{\pm}/4}1(q^{\pm} > 0))$, which completes the proof. $\square$

6.4. *The i.i.d. algorithm.* This last section of the Appendix contains the proof of Theorem 3.1, which shows that the i.i.d. algorithm in Section 3.1 generates bi-degree sequences that satisfy Assumption 2.1 with high probability.

PROOF OF THEOREM 3.1. With some abuse of notation with respect to the proofs in the previous sections, define the events

$$B_n = \big\{d_1\big(G_n^+, G^+\big) \le n^{-\varepsilon}, d_1\big(G_n^-, G^-\big) \le n^{-\varepsilon}\big\} \quad \text{and}$$
$$E_n = \left\{\sum_{i=1}^n \big((D_i^-)^\kappa + (D_i^+)^\kappa\big)D_i^+D_i^- \le K_\kappa n\right\}.$$

Assume, without loss of generality, that $K_\kappa > E[((\mathscr{D}^-)^\kappa + (\mathscr{D}^+)^\kappa)\mathscr{D}^+\mathscr{D}^-] \triangleq H_\kappa$. Next, note that

$$\begin{aligned}
P\big(\Omega_n^c\big) &\le P\big(B_n^c\big) + P\big(E_n^c \cap B_n\big)\\
&\qquad + P\big(\max\big\{d_1\big(F_n^+, F^+\big), d_1\big(F_n^-, F^-\big)\big\} > n^{-\varepsilon}, B_n \cap E_n\big).
\end{aligned}$$

We start by showing that $P(B_n^c) \to 0$ as $n \to \infty$. To this end, let $\hat{G}_n^-$ and $\hat{G}_n^+$ denote the empirical distribution functions of $\mathscr{D}_1^-, \dots \mathscr{D}_n^-$ and $\mathscr{D}_1^+, \dots, \mathscr{D}_n^+$, respectively; note that although $\hat{G}_n^{\pm}$ is well-defined regardless of the value of $\Delta_n$, $G_n^{\pm}$, $F_n^{\pm}$ are only defined conditionally on the event $\Psi_n = \{|\Delta_n| \leq n^{1-\delta}\}$. Furthermore, since $E[|\mathscr{D}^- - \mathscr{D}^+|^{1/(1-\delta)}] \leq (E[|\mathscr{D}^- - \mathscr{D}^+|^{1+\kappa}])^{(1-\delta)/(1+\kappa)} < \infty$, the Kolmogorov–Marcinkiewicz–Zygmund strong law of large numbers gives

$$P\left(\lim_{n\to\infty} \frac{\Delta_n}{n^{1-\delta}} = 0\right) = 1,$$

and hence, $P(\Psi_n) \to 1$ as $n \to \infty$.

It follows from the triangle inequality and Theorem 2.2 in del Barrio, Giné and Matrán (1999) [see also Proposition 3 in Chen and Olvera-Cravioto (2015)], that

$$\begin{aligned} E\big[d_1(G_n^{\pm}, G^{\pm})\big] &\leq E\big[d_1(G_n^{\pm}, \hat{G}_n^{\pm})|\Psi_n\big] + E\big[d_1(\hat{G}_n^{\pm}, G^{\pm})|\Psi_n\big] \\ &\leq \frac{1}{P(\Psi_n)}\big(E\big[d_1(G_n^{\pm}, \hat{G}_n^{\pm})1(\Psi_n)\big] + K_\delta n^{-\delta}\big) \end{aligned}$$

for some finite constant $K_\delta$. Moreover, on the event $\Psi_n$,

$$\begin{aligned} d_1(G_n^+, \hat{G}_n^+) &= \frac{1}{n}\int_0^\infty \left|\sum_{i=1}^n \big(1(\mathscr{D}_i^- + \tau_i \leq x) - 1(\mathscr{D}_i^- \leq x)\big)\right| dx \\ &= \frac{1}{n}\int_0^\infty \sum_{i=1}^n 1\big(\mathscr{D}_i^- \leq x < \mathscr{D}_i^- + \tau_i\big)\, dx \\ &= \frac{1}{n}\sum_{i=1}^\infty \tau_i \leq \frac{|\Delta_n|}{n} \\ &\leq n^{-\delta}. \end{aligned}$$

Since the analysis of $d_1(G_n^-, \hat{G}_n^-)$ is the same, we obtain

$$E\big[d_1(G_n^{\pm}, G^{\pm})|\Psi_n\big] \leq \frac{1}{P(\Psi_n)}\big(n^{-\delta} + K_\delta n^{-\delta}\big),$$

from which it follows that as $n \to \infty$,

$$P(B_n^c) \leq n^{\varepsilon}\big(E\big[d_1(G_n^+, G^+)\big] + E\big[d_1(G_n^-, G^-)\big]\big) = O\big(n^{-\delta+\varepsilon}\big).$$

Next, to analyze $P(E_n^c \cap B_n)$, note that $\tau_i \chi_i = 0$ and, therefore,

$$\begin{aligned} &\sum_{i=1}^n \big((D_i^-)^\kappa + (D_i^+)^\kappa\big) D_i^+ D_i^- \\ &\qquad \leq \sum_{i=1}^n \big((\mathscr{D}_i^- + \tau_i)^\kappa + (\mathscr{D}_i^+ + \chi_i)^\kappa\big)(\mathscr{D}_i^+ + \chi_i)(\mathscr{D}_i^- + \tau_i) \end{aligned}$$

$$\leq \sum_{i=1}^{n} \bigl((\mathscr{D}_i^-)^\kappa + \tau_i + (\mathscr{D}_i^+)^\kappa + \chi_i\bigr)\bigl(\mathscr{D}_i^+\mathscr{D}_i^- + \mathscr{D}_i^+\tau_i + \mathscr{D}_i^-\chi_i\bigr)$$

$$= \sum_{i=1}^{n} \bigl\{\bigl((\mathscr{D}_i^-)^\kappa + (\mathscr{D}_i^+)^\kappa\bigr)\mathscr{D}_i^+\mathscr{D}_i^- + \tau_i\bigl(\mathscr{D}_i^+\mathscr{D}_i^- + \mathscr{D}_i^+\bigr) + \chi_i\bigl(\mathscr{D}_i^+\mathscr{D}_i^- + \mathscr{D}_i^-\bigr)\bigr\}.$$

Now set $Y_i = ((\mathscr{D}^-)_i^\kappa + (\mathscr{D}_i^+)^\kappa)\mathscr{D}_i^+\mathscr{D}_i^- - H_\kappa$ and $W_i = \tau_i(\mathscr{D}_i^+\mathscr{D}_i^- + \mathscr{D}_i^+) + \chi_i(\mathscr{D}_i^+\mathscr{D}_i^- + \mathscr{D}_i^-)$ to obtain

$$\begin{aligned}
P\bigl(E_n^c \cap B_n\bigr) &\leq P\left(\sum_{i=1}^n Y_i + \sum_{i=1}^n W_i > (K_\kappa - H_\kappa)n,\, B_n \middle| \Psi_n\right) \\
&\leq \frac{1}{P(\Psi_n)} P\left(\sum_{i=1}^n Y_i > n(K_\kappa - H_\kappa)/2\right) \\
&\quad + P\left(\sum_{i=1}^n W_i > n(K_\kappa - H_\kappa)/2,\, B_n \middle| \Psi_n\right).
\end{aligned}$$

Since $n^{-1}\sum_{i=1}^n Y_i \to 0$ almost surely by the strong law of large numbers, the first probability converges to zero as $n \to \infty$. For the second probability, use Markov's inequality to obtain

$$P\left(\sum_{i=1}^n W_i > n(K_\kappa - H_\kappa)/2,\, B_n \middle| \Psi_n\right) \leq \frac{2E[W_1 1(\Psi_n \cap B_n)]}{P(\Psi_n)(K_\kappa - H_\kappa)}.$$

Now note that

$$\begin{aligned}
E\bigl[W_1 1(\Psi_n \cap B_n)\bigr] &= E\bigl[E\bigl[\tau_1 | \bigl\{(\mathscr{D}_i^-, \mathscr{D}_i^+)\bigr\}_{i=1}^n\bigr](\mathscr{D}_1^-\mathscr{D}_1^+ + \mathscr{D}_1^+)1(\Psi_n \cap B_n)\bigr] \\
&\quad + E\bigl[E\bigl[\chi_1 | \bigl\{(\mathscr{D}_i^-, \mathscr{D}_i^+)\bigr\}_{i=1}^n\bigr](\mathscr{D}_1^-\mathscr{D}_1^+ + \mathscr{D}_1^-)1(\Psi_n \cap B_n)\bigr] \\
&= E\left[\frac{\Delta_n^-}{L_n}(\mathscr{D}_1^-\mathscr{D}_1^+ + \mathscr{D}_1^+)1(\Psi_n \cap B_n)\right] \\
&\quad + E\left[\frac{\Delta_n^+}{L_n}(\mathscr{D}_1^-\mathscr{D}_1^+ + \mathscr{D}_1^-)1(\Psi_n \cap B_n)\right] \\
&\leq n^{1-\delta} E\left[\frac{\mathscr{D}_1^-\mathscr{D}_1^+ + \mathscr{D}_1^+ + \mathscr{D}_1^-}{L_n} 1(B_n)\right] \\
&\leq \frac{n^{-\delta}}{\nu - n^{-\varepsilon}} E\bigl[\mathscr{D}_1^-\mathscr{D}_1^+ + \mathscr{D}_1^+ + \mathscr{D}_1^-\bigr],
\end{aligned}$$

where in the last step we used the observation that on the event $B_n$ we have $L_n = n\nu_n \geq n(\nu - n^{-\varepsilon})$, since $|\nu_n - \nu| \leq d_1(G_n^+, G^+)$. Hence, we have shown that $P(E_n^c \cap B_n) \to 0$ as $n \to \infty$.

For the size-biased distributions, note that

$$
\begin{aligned}
d_1(F_n^+, F^+) &\le \int_0^\infty \left|\frac{1}{L_n} - \frac{1}{\nu n}\right| \sum_{i=1}^n 1(D_i^- > x) D_i^+ \, dx \\
&\quad + \int_0^\infty \left| \frac{1}{\nu n} \sum_{i=1}^n 1(\mathscr{D}_i^- > x)\mathscr{D}_i^+ - 1 + F^+(x) \right| dx \\
&\quad + \frac{1}{n\nu} \int_0^\infty \left| \sum_{i=1}^n \bigl(1(D_i^- > x) D_i^+ - 1(\mathscr{D}_i^- > x)\mathscr{D}_i^+\bigr) \right| dx \\
&= \left| \frac{\nu - \nu_n}{\nu \nu_n} \right| \frac{1}{n} \sum_{i=1}^n D_i^- D_i^+ + \int_0^\infty \left| \frac{1}{n} \sum_{i=1}^n X_i^+(x) \right| dx \\
&\quad + \frac{1}{\nu n} \int_0^\infty \sum_{i=1}^n \bigl(1(\mathscr{D}_i^- \le x < \mathscr{D}_i^- + \tau_i)\mathscr{D}_i^+ + 1(\mathscr{D}_i^- > x)\chi_i\bigr) \, dx,
\end{aligned}
$$

where $X_i^+(x) = 1(\mathscr{D}_i^- > x)\mathscr{D}_i^+/\nu - 1 + F^+(x)$. Now recall that $|\nu_n - \nu| \le d_1(G_n^+, G^+)$ and note that $n^{-1} \sum_{i=1}^n D_i^0 D_i^+ \le K_\kappa$, which yields that on the event $B_n \cap E_n$,

$$
\begin{aligned}
d_1(F_n^+, F^+) &\le \left| \frac{\nu - \nu_n}{\nu(\nu - n^{-\varepsilon})} \right| K_\kappa + \frac{L_n}{n^2 \nu(\nu - n^{-\varepsilon})} \sum_{i=1}^n (\tau_i \mathscr{D}_i^+ + \chi_i \mathscr{D}_i^-) \\
&\quad + \int_0^\infty \left| \frac{1}{n} \sum_{i=1}^n X_i^+(x) \right| dx.
\end{aligned}
$$

Since the case $d_1(F_n^-, F^-)$ is symmetric by setting $X_i^-(x) = 1(\mathscr{D}_i^+ > x)\mathscr{D}_i^-/\nu - 1 + F^-(x)$, it follows that

$$
\begin{aligned}
&P\bigl(d_1(F_n^\pm, F^\pm) > n^{-\varepsilon}, B_n \cap E_n\bigr) \\
&\quad \le \frac{1}{P(\Psi_n)} P\Biggl( \left| \frac{\nu - \nu_n}{\nu(\nu - n^{-\varepsilon})} \right| K_\kappa + \frac{1(\Psi_n) L_n}{n^2 \nu(\nu - n^{-\varepsilon})} \sum_{i=1}^n (\tau_i \mathscr{D}_i^+ + \chi_i \mathscr{D}_i^-) \\
&\qquad + \int_0^\infty \left| \frac{1}{n} \sum_{i=1}^n X_i^\pm(x) \right| dx > n^{-\varepsilon} \Biggr) \\
&\quad \le \frac{n^\varepsilon}{P(\Psi_n)} \Biggl( \frac{K_\kappa E[d_1(G_n^+, G^+)]}{\nu(\nu - n^{-\varepsilon})} + \frac{E[1(\Psi_n) L_n (\tau_1 \mathscr{D}_1^+ + \chi_1 \mathscr{D}_1^-)]}{n\nu(\nu - n^{-\varepsilon})} \\
&\qquad + \frac{1}{n} \int_0^\infty E\left[ \left| \sum_{i=1}^n X_i^\pm(x) \right| \right] dx \Biggr)
\end{aligned}
$$

$$= O\left(n^{-\delta+\varepsilon} + \frac{E[1(\Psi_n)L_n(\tau_1\mathscr{D}_1^+ + \chi_1\mathscr{D}_1^-)]}{n^{1-\varepsilon}}\right.$$
$$\left. + \frac{1}{n^{1-\varepsilon}}\int_0^\infty E\left[\left|\sum_{i=1}^n X_i^\pm(x)\right|\right] dx\right).$$

To bound the middle term in the last expression, note that

$$\begin{aligned}
&E\big[1(\Psi_n)L_n\big(\tau_1\mathscr{D}_1^+ + \chi_1\mathscr{D}_1^-\big)\big]\\
&\quad = E\big[1(\Psi_n)L_n\big(E\big[\tau_1\big|\{\mathscr{D}_i^-, \mathscr{D}_i^+\}\big\}_{i=1}^n\big]\mathscr{D}_1^+ + E\big[\chi_1\big|\{\mathscr{D}_i^-, \mathscr{D}_i^+)\}_{i=1}^n\big]\mathscr{D}_1^-\big)\big]\\
&\quad = E\big[1(\Psi_n)\big(|\Delta_n|\mathscr{D}_1^+ + |\Delta_n|\mathscr{D}_1^-\big)\big] \le n^{1-\delta}E\big[\mathscr{D}^+ + \mathscr{D}^-\big].
\end{aligned}$$

To complete the proof, choose $1/(1-\varepsilon) < p < 1+\kappa$, use the monotonicity of the norm $\|X\|_p = (E[|X|^p])^{1/p}$ and apply Lemma 6.2 to obtain

$$\begin{aligned}
\frac{1}{n^{1-\varepsilon}}\int_0^\infty E\left[\left|\sum_{i=1}^n X_i^\pm(x)\right|\right] dx &\le \frac{1}{n^{1-\varepsilon}}\int_0^\infty \left\|\sum_{i=1}^n X_i^\pm(x)\right\|_p dx\\
&\le \frac{1}{n^{1-\varepsilon}}\int_0^\infty \big(Q_p n E\big[|X_1^\pm(x)|^p\big]\big)^{1/p} dx\\
&= \frac{(Q_p)^{1/p}}{\nu} n^{1/p-1+\varepsilon}\int_0^\infty \|\nu X_1^\pm(x)\|_p\, dx,
\end{aligned}$$

for some finite constant $Q_p$ that depends only on $p$; note that our choice of $p$ implies that $n^{1/p-1+\varepsilon} \to 0$. It only remains to verify that the integral is finite. To do this, first use Minkowski's inequality to get

$$\begin{aligned}
\|\nu X_1^+(x)\|_p &= \big\|1\big(\mathscr{D}^- > x\big)\mathscr{D}^+ - E\big[1\big(\mathscr{D}^- > x\big)\mathscr{D}^+\big]\big\|_p\\
&\le \big\|1\big(\mathscr{D}^- > x\big)\mathscr{D}^+\big\|_p + E\big[1\big(\mathscr{D}^- > x\big)\mathscr{D}^+\big].
\end{aligned}$$

Furthermore, for $x \ge 1$,

$$\begin{aligned}
\big\|1\big(\mathscr{D}^- > x\big)\mathscr{D}^+\big\|_p &= \big(E\big[1\big(\mathscr{D}^- > x\big)\big(\mathscr{D}^+\big)^p\big]\big)^{1/p} \le \big(E\big[\big(\mathscr{D}^-/x\big)^{1+\kappa}\big(\mathscr{D}^+\big)^p\big]\big)^{1/p}\\
&= \big(E\big[\big(\mathscr{D}^-\big)^{1+\kappa}\big(\mathscr{D}^+\big)^p\big]\big)^{1/p} x^{-(1+\kappa)/p},
\end{aligned}$$

while for $0 < x < 1$,

$$\big\|1\big(\mathscr{D}^- > x\big)\mathscr{D}^+\big\|_p = \big\|1\big(\mathscr{D}^- \ge 1\big)\mathscr{D}^+\big\|_p \le \big(E\big[\mathscr{D}^-\big(\mathscr{D}^+\big)^p\big]\big)^{1/p}.$$

It follows that

$$\begin{aligned}
\int_0^\infty \|\nu X_1^+(x)\|_p\, dx &\le \big(E\big[\mathscr{D}^-\big(\mathscr{D}^+\big)^p\big]\big)^{1/p} + E\big[\mathscr{D}^-\mathscr{D}^+\big]\\
&\quad + \int_1^\infty \big(E\big[\big(\mathscr{D}^-\big)^{1+\kappa}\big(\mathscr{D}^+\big)^p\big]\big)^{1/p} x^{-(1+\kappa)/p}\, dx < \infty.
\end{aligned}$$

The proof for $X_1^-(x)$ is obtained by exchanging the roles of $\mathscr{D}^-$ and $\mathscr{D}^+$. $\square$

## REFERENCES

ARRATIA, R. and LIGGETT, T. M. (2005). How likely is an i.i.d. degree sequence to be graphical? *Ann. Appl. Probab.* **15** 652–670. MR2114985

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

BIGGINS, J. D. and BINGHAM, N. H. (1993). Large deviations in the supercritical branching process. *Adv. in Appl. Probab.* **25** 757–772. MR1241927

BOLDI, P., ROSA, M. and VIGNA, S. (2011). HyperANF: Approximating the neighbourhood function of very large graphs on a budget. In *Proceedings of the* 20*th International Conference on World Wide Web* 625–634. ACM, New York.

BOLDI, P. and VIGNA, S. (2004). The webgraph framework I: Compression techniques. In *Proceedings of the* 13*th International Conference on World Wide Web* 595–602. ACM, New York.

BOLDI, P. and VIGNA, S. (2013). In-core computation of geometric centralities with hyperball: A hundred billion nodes and beyond. In *Data Mining Workshops* (*ICDMW*), 2013 *IEEE* 13*th International Conference on* 621–628. IEEE, New York.

BOLLOBÁS, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.* **1** 311–316.

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

BRODER, A., KUMAR, R., MAGHOUL, F., RAGHAVAN, P., RAJAGOPALAN, S., STATA, R., TOMKINS, A. and WIENER, J. (2000). Graph structure in the web. *Comput. Netw.* **33** 309–320.

CHEN, N., LITVAK, N. and OLVERA-CRAVIOTO, M. (2017). Generalized PageRank on directed configuration networks. *Random Structures Algorithms* **51** 237–274.

CHEN, N. and OLVERA-CRAVIOTO, M. (2013). Directed random graphs with given degree distributions. *Stoch. Syst.* **3** 147–186. MR3353470

CHEN, N. and OLVERA-CRAVIOTO, M. (2015). Efficient simulation for branching linear recursions. In *Proceedings of the* 2015 *Winter Simulation Conference* 1–12.

COOPER, C. and FRIEZE, A. (2004). The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combin. Probab. Comput.* **13** 319–337.

DEL BARRIO, E., GINÉ, E. and MATRÁN, C. (1999). Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.* **27** 1009–1071. MR1698999

DURRETT, R. (2010). *Random Graph Dynamics*, 1st ed. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, Cambridge.

GOH, K.-I., OH, E., KAHNG, B. and KIM, D. (2003). Betweenness centrality correlation in social networks. *Phys. Rev. E* **67** 017101.

GRANDELL, J. (1997). *Mixed Poisson Processes*. Springer, Berlin.

MILLER, J. C. (2009). Percolation and epidemics in random clustered networks. *Phys. Rev. E* (3) **80** 020901, 4. MR2593889

NEWMAN, M. E. J. (2002). Spread of epidemic disease on networks. *Phys. Rev. E* (3) **66** 016128, 11. MR1919737

NORROS, I., REITTU, H. et al. (2006). On a conditionally Poissonian graph process. *Adv. in Appl. Probab.* **38** 59–75.

PENROSE, M. D. (2016). The strong giant in a random digraph. *J. Appl. Probab.* **53** 57–70. MR3540780

SERGE DUBUC, M. (1971). La densité de la loi-limite d'un processus en cascade expansif. *Probab. Theory Related Fields* **19** 281–290.

TIMÁR, G., GOLTSEV, A., DOROGOVTSEV, S. and MENDES, J. (2017). Mapping the structure of directed networks: Beyond the "bow-tie" diagram. *Physics Review Letters* **118**.

VAN DER HOFSTAD, R. (2016). *Random Graphs and Complex Networks* **1**. Cambridge Univ. Press, Cambridge.

VAN DEN ESKER, H., VAN DER HOFSTAD, R. and HOOGHIEMSTRA, G. (2008). Universality for the distance in finite variance random graphs. *J. Stat. Phys.* **133** 169–202. MR2438903

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.

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.

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.

VILLANI, C. (2008). *Optimal Transport: Old and New* **338**. Springer Science & Business Media, Berlin.

WORMALD, N. C. (1978). Some problems in the enumeration of labelled graphs Ph.D. thesis Newcastle Univ.

NETWORK SCIENCE INSTITUTE
NORTHEASTERN UNIVERSITY
210-177
360 HUNTINGTON AVENUE
BOSTON, MASSACHUSETTS 02115
USA
E-MAIL: w.vanderhoorn@northeastern.edu

DEPARTMENT OF INDUSTRIAL ENGINEERING AND OPERATIONS RESEARCH
UNIVERSITY OF CALIFORNIA, BERKELEY
4125 ETCHEVERRY HALL
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: molvera@berkeley.edu