Universality for the distance in finite variance random graphs

Henri van den Esker[*] Remco van der Hofstad [†] Gerard Hooghiemstra[*]

June 16, 2008

Abstract

We generalize the asymptotic behavior of the graph distance between two uniformly chosen nodes in the configuration model to a wide class of random graphs. Among others, this class contains the Poissonian random graph, the expected degree random graph and the generalized random graph (including the classical Erdős-Rényi graph).

In the paper we assign to each node a deterministic capacity and the probability that there exists an edge between a pair of nodes is equal to a function of the product of the capacities of the pair divided by the total capacity of all the nodes. We consider capacities which are such that the degrees of a node have uniformly bounded moments of order strictly larger than two, so that, in particular, the degrees have finite variance. We prove that the graph distance grows like $\log_\nu N$, where the ν depends on the capacities and N denotes the size of the graph. In addition, the random fluctuations around this asymptotic mean $\log_\nu N$ are shown to be tight. We also consider the case where the capacities are independent copies of a positive random Λ with $\mathbb{P}(\Lambda > x) \leq cx^{1-\tau}$, for some constant c and $\tau > 3$, again resulting in graphs where the degrees have finite variance.

The method of proof of these results is to couple each member of the class to the Poissonian random graph, for which we then give the complete proof by adapting the arguments of [13].

1 Introduction

Various papers (see e.g., [4, 7, 13, 18, 20]) study properties of random graphs with a given degree sequence. Among such properties as connectivity, cluster size and diameter, the graph distance between two uniformly chosen nodes is an important one. For two connected nodes the graph distance is defined as the minimum number of edges of a path that connects these nodes. If the nodes are not connected, then the graph distance is put equal to infinity.

For the configuration model (see Section 1.4 for a definition) a distance result appeared in [13], when the distribution of the i.i.d. degrees $D^{(C)}$ satisfies

$$\mathbb{P}(D^{(C)} > x) \leq cx^{1-\tau}, \qquad x \geq 0, \tag{1.1}$$

for some constant c and with $\tau > 3$. We use the superscript (C) to differentiate between models. The result in [13] states that with probability converging to 1 (**whp**), the average distance between nodes in the giant component has, for

$$\tilde{\nu} = \frac{\mathbb{E}[D^{(C)}(D^{(C)} - 1)]}{\mathbb{E}[D^{(C)}]} > 1, \tag{1.2}$$

bounded fluctuations around $\log_{\tilde{\nu}} N$. The condition $\tilde{\nu} > 1$ corresponds to the supercritical case of an associated branching process, and is the condition under which a giant component exists.

[*]Delft University of Technology, Electrical Engineering, Mathematics and Computer Science, P.O. Box 5031, 2600 GA Delft, The Netherlands. E-mail: H.vandenEsker@ewi.tudelft.nl and G.Hooghiemstra@ewi.tudelft.nl

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

In this paper we extend the above distance result to a wide class of random graphs. Models which fall in this class are the generalized random graph (GRG), the expected degree random graph (EDRG) and the Poissonian random graph (PRG). All three models will be introduced in more detail below.

The method of proof is coupling. It is shown that the distance result holds for all models in the general class if and only if the result holds for the PRG (Section 2). In Section 4 we prove the distance result for the PRG. This proof is parallel to that in [13] for the configuration model. Therefore, most of the proofs of the auxiliary propositions and lemmas are left out, as they are similar to those in [13]. Since details of these proofs are different, we included full proofs in an extended version of this paper. The extended version will not be published, but is available on the web ([9]).

1.1 Model assumptions

The graph models considered here are *static* models, meaning that the number of nodes is fixed. The graph G_N has N nodes, numbered $1, 2, \ldots, N$. Associated with the nodes is a sequence $\{\lambda_i\}_{i=1}^N$ of positive reals. We call λ_i the *capacity* of node i; nodes with a large capacity will obtain a high degree, whereas nodes with small capacity will only be incident to a limited number of nodes. Furthermore, we define

$$l_N = \lambda_1 + \lambda_2 + \cdots + \lambda_N, \tag{1.3}$$

i.e., l_N is the total capacity of all nodes of the graph G_N.

The binary random variables $\{X_{ij}\}_{1\le i<j\le N}$, are defined by setting $X_{ij} = 1$, if the edge between node i and node j in the graph G_N is present, otherwise we set $X_{ij} = 0$. If $i > j$, then by convention $X_{ji} = X_{ij}$. We call $\{X_{ij}\}$ the *connection variables* and $p_{ij} = \mathbb{P}(X_{ij} = 1)$ the *connection probability of edge* ij. In this paper we consider graphs G_N that satisfy the following two assumptions:

A1: The connection variables $\{X_{ij}\}_{1\le i<j\le N}$, are independent.

A2: The connection probability p_{ij}, for $1 \le i < j \le N$, can be written as $p_{ij} = h(\lambda_i\lambda_j/l_N)$, for some function $h : [0,\infty) \to [0,1]$ satisfying

$$h(x) - x = \mathcal{O}\big(x^2\big), \text{ for } x \downarrow 0. \tag{1.4}$$

1.2 Three special cases

We give three examples of random graph models that satisfy assumptions **A1** and **A2**, and hence fall in the class of models considered here.

The first example is the Poissonian random graph (PRG), which is a variant of a model introduced by Norros and Reittu in [20]. The second random graph model, which we call the expected degree random graph (EDRG), was introduced by Chung and Lu in [7, 8]. The third and last example is the generalized random graph (GRG), which was introduced by Britton, Deijfen and Martin-Löf in [6].

We now define the three models and verify that each of them satisfy the conditions **A1** and **A2** above.

- **Poissonian random graph:** In [20], the PRG was introduced. The main feature of the PRG $G_N^{(P)}$ is that, conditionally on the capacities, the number of edges between any pair of nodes i and j is a Poisson random variable. The model in [20] is introduced as a growth model, but as a consequence of [20, Proposition 2.1], it can be formulated as a static model, and we will do so. Start with the graph $G_N^{(P)}$ consisting of N nodes and capacities $\{\lambda_i\}_{i=1}^N$.

The number of edges between two different nodes i and j is given by an independent Poisson random variable $E_{ij}^{(P)}$ with parameter

$$\lambda_i \lambda_j / l_N. \tag{1.5}$$

Denote by $\mathbf{1}_A$ the indicator of the set A. The connection variables are then $X_{ij}^{(P)} = \mathbf{1}_{\{E_{ij}^{(P)}>0\}}$, so that, for $1 \le i < j \le N$, the connection probabilities are given by

$$p_{ij}^{(P)} = \mathbb{P}(X_{ij}^{(P)} = 1) = \mathbb{P}(E_{ij}^{(P)} > 0) = 1 - \exp(-\lambda_i\lambda_j/l_N) = h^{(P)}(\lambda_i\lambda_j/l_N),$$

where $h^{(P)}(x) = 1 - e^{-x}$. Obviously, $h^{(P)}(x) - x = \mathcal{O}\big(x^2\big)$, for $x \downarrow 0$. Since, by definition, the random variables $\{X_{ij}^{(P)}\}_{1\le i<j\le N}$ are independent, we conclude that the assumptions **A1** and **A2** are satisfied.

It should be noted that in this paper we define the PRG using a deterministic sequence of capacities, while Norros and Reittu start with an i.i.d. sequence of random capacities. The case where the capacities $\{\Lambda_i\}_{i=1}^N$ are i.i.d. random variables, satisfying certain tail estimates, is a special case of our set up and is studied in more detail in Theorem 1.4 below.

- **Expected degree random graph:** In [7, 8] a random graph model is introduced starting from a sequence of deterministic capacities $\{\lambda_i\}_{i=1}^N$. We construct the EDRG $G_N^{(E)}$ as follows. Let $\{X_{ij}^{(E)}\}_{1\le i<j\le N}$ be a sequence of independent Bernoulli random variables with success probability

$$p_{ij}^{(E)} = \mathbb{P}\left(X_{ij}^{(E)} = 1\right) = (\lambda_i\lambda_j/l_N) \wedge 1, \quad \text{for } 1 \le i < j \le N,$$

where $x \wedge y$ denotes the minimum of x and y. This minimum is to ensure that the result is a probability.

Assumption **A1** is satisfied by definition, since the connection variables are independent Bernoulli variables, and assumption **A2** is also satisfied if we pick $h^{(E)}(x) = x \wedge 1$.

If we assume that $\lambda_i\lambda_j/l_N < 1$ for all $1 \le i < j \le N$, then the expected degree of a node i is given by λ_i:

$$\mathbb{E}\left[\sum_{j=1}^N X_{ij}^{(E)}\right] = \sum_{j=1}^N \lambda_i\lambda_j/l_N = \lambda_i.$$

The Erdős-Rényi random graph, usually denoted by $G(N,p)$, is a special case of the EDRG. In the graph $G(N,p)$, an edge between a pair of nodes is present with probability $p \in [0,1]$, independently of the other edges. When $p = \lambda/N$, for some constant $\lambda > 0$, then we obtain the graph $G(N, \lambda/N)$ from the EDRG by picking $\lambda_i = \lambda$ for all $i \in \{1, \dots, N\}$, since then $p_{ij}^{(E)} = \lambda_i\lambda_j/l_N = \lambda/N = p$, for all $1 \le i < j \le N$.

- **Generalized random graph:** The generalized random graph (GRG) is an adapted version of the EDRG, see the previous example. We define $G_N^{(G)}$ with N nodes as follows. The sequence of connection variables is again given by a sequence of independent Bernoulli random variables $\{X_{ij}^{(G)}\}_{1\le i<j\le N}$ with

$$\mathbb{P}\left(X_{ij}^{(G)} = 1\right) = p_{ij}^{(G)} = \frac{\lambda_i\lambda_j/l_N}{1 + \lambda_i\lambda_j/l_N}.$$

In [6] the edge probabilities are given by $(\lambda_i\lambda_j/N)/(1 + \lambda_i\lambda_j/N)$, so that we have replaced $\lambda_i/N^{1/2}$ by $\lambda_i/l_N^{1/2}$, $1 \le i \le N$. This makes hardly any difference.

Again, the assumptions **A1** and **A2** are satisfied. To satisfy assumption **A2** we pick $h^{(G)}(x) = x/(1+x) = x + \mathcal{O}\big(x^2\big)$.

1.3 Main results

We state conditions on the capacities $\{\lambda_i\}_{i=1}^N$, under which our main result will hold. We shall need three conditions, which we denote by (**C1**), (**C2**) and (**C3**), respectively.

(C1) Convergence of means: Define

$$\mu_N = \frac{1}{N}\sum_{i=1}^N \lambda_i \qquad \text{and} \qquad \nu_N = \frac{\sum_{i=1}^N \lambda_i^2}{\sum_{i=1}^N \lambda_i}, \tag{1.6}$$

then there exist constants $\mu \in (0,\infty)$, $\nu \in (1,\infty)$ and $\alpha_1 > 0$ such that

$$|\mu_N - \mu| = \mathcal{O}\big(N^{-\alpha_1}\big) \qquad \text{and} \qquad |\nu_N - \nu| = \mathcal{O}\big(N^{-\alpha_1}\big). \tag{1.7}$$

(C2) Convergence of branching processes: Define

$$f_n^{(N)} = \frac{1}{N}\sum_{i=1}^N e^{-\lambda_i}\frac{\lambda_i^n}{n!} \qquad \text{and} \qquad g_n^{(N)} = \frac{1}{N\mu_N}\sum_{i=1}^N e^{-\lambda_i}\frac{\lambda_i^{n+1}}{n!}, \tag{1.8}$$

then there exist sequences $\{f_n\}_{n\geq 0}$ and $\{g_n\}_{n\geq 0}$, independent of N, and $\alpha_2 > 0$ such that

$$\mathrm{d}_{\mathrm{TV}}(f^{(N)}, f) = \mathcal{O}\big(N^{-\alpha_2}\big) \qquad \text{and} \qquad \mathrm{d}_{\mathrm{TV}}(g^{(N)}, g) = \mathcal{O}\big(N^{-\alpha_2}\big), \tag{1.9}$$

where $\mathrm{d}_{\mathrm{TV}}(\cdot,\cdot)$ is the total variance distance, i.e., for probability mass functions $p = \{p_j\}$ and $q = \{q_j\}$:

$$\mathrm{d}_{\mathrm{TV}}(p, q) = \tfrac{1}{2}\sum_j |p_j - q_j|. \tag{1.10}$$

(C3) Moment and maximal bound on capacities: There exists a $\tau > 3$ such that for every $\varepsilon > 0$,

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \lambda_i^{\tau-1-\varepsilon} < \infty, \qquad \text{and} \qquad \lambda_N^{(N)} = \max_{1\leq i\leq N} \lambda_i \leq N^{\gamma}, \tag{1.11}$$

where

$$\gamma = \frac{1}{\tau - 1} + \varepsilon < 1/2. \tag{1.12}$$

It is not hard to see that μ_N and ν_N in (1.6) are the means of the probability mass functions $\{f_n^{(N)}\}_{n\geq 0}$, $\{g_n^{(N)}\}_{n\geq 0}$, respectively. Thus, (**C1**) is equivalent to the fact that the means of $\{f_n^{(N)}\}_{n\geq 0}$, $\{g_n^{(N)}\}_{n\geq 0}$ converge. It turns out that (**C1**) is equivalent to the convergence of the first and second moment of the degree of a uniform node.

Condition (**C2**) says that the laws $\{f_n^{(N)}\}_{n\geq 0}$, $\{g_n^{(N)}\}_{n\geq 0}$ are close to certain limiting laws $\{f_n\}_{n\geq 0}$, $\{g_n\}_{n\geq 0}$, which shall turn out to be crucial in our proofs, since it allows us to use a coupling to branching processes.

The second bound in Condition (**C3**) gives an upper bound on the maximal capacity of a node, while it can be seen that the first inequality is equivalent to the statement that a uniform node has a uniformly bounded moment of order at least $\tau - 1 - \varepsilon$. Since $\tau > 3$, we can pick $\varepsilon > 0$ so small that $\tau - 1 - \varepsilon > 2$, so that the degrees have *finite variances*.

We shall prove our main results in the generality of Conditions (**C1**)–(**C3**), but shall give special cases where Conditions (**C1**)–(**C3**) are satisfied following our main results.

In order to be able to state our main results, we define the process $\{\mathcal{Z}_t\}_{t\geq 0}$ as a branching process (BP) starting from $\mathcal{Z}_0 = 1$, where in the first generation the offspring distribution is equal to $\{f_n\}$, whereas in the second and further generations the offspring is chosen in accordance to $\{g_n\}$.

We define the average graph distance or hopcount H_N between two different randomly chosen nodes A_1 and A_2 in the graph G_N as the minimum number of edges that form a path from the node A_1 to node A_2 where, by convention, the distance equals ∞ if the nodes A_1 and A_2 are not connected.

Theorem 1.1 (Fluctuations of the graph distance) *Assume that the capacities $\{\lambda_i\}_{i=1}^N$ satisfy (**C1**)–(**C3**), and let the graph G_N, with capacities $\{\lambda_i\}_{i=1}^N$ satisfy **A1** and **A2**, for some function $h : [0,\infty) \mapsto [0,1]$. Let $\sigma_N = \lfloor \log_\nu N \rfloor$ and $a_N = \sigma_N - \log_\nu N$. There exist random variables $(R_a)_{a\in(-1,0]}$ such that, as $N \to \infty$,*

$$\mathbb{P}\left(H_N = \sigma_N + k \,|\, H_N < \infty\right) = \mathbb{P}\left(R_{a_N} = k\right) + o(1), \qquad k = 0, \pm 1, \pm 2, \dots. \tag{1.13}$$

We identify the random variables $(R_a)_{a\in(-1,0]}$ in Theorem 1.3 below. Before doing so, we state one of the consequences of Theorem 1.1:

Corollary 1.2 (Concentration of the graph distance) *Under the given assumptions of Theorem 1.1,*

- *with probability $1 - o(1)$ and conditionally on $H_N < \infty$, the random variable H_N is in between $(1 \pm \varepsilon)\log_\nu N$ for any $\varepsilon > 0$;*
- *conditionally on $H_N < \infty$, the sequence of random variables $H_N - \log_\nu N$ is tight, i.e.,*

$$\lim_{K\to\infty} \limsup_{N\to\infty} \mathbb{P}\left(|H_N - \log_\nu N| \le K \big| H_N < \infty\right) = 1. \tag{1.14}$$

We use a limit result from branching process theory to identify the limiting random variables $(R_a)_{a\in(-1,0]}$. It is well known, see [11, p. 244], that the process $\{\mathcal{Z}_t/\mu\nu^{t-1}\}_{t\ge 1}$ is a non-negative martingale and consequently converges almost surely to a limit $\mathcal{W}$:

$$\lim_{t\to\infty} \frac{\mathcal{Z}_t}{\mu\nu^{t-1}} = \mathcal{W}, \text{ a.s.} \tag{1.15}$$

Let $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ be two independent copies of $\mathcal{W}$ in (1.15), then we can identify the limit random variables $(R_a)_{a\in(-1,0]}$ as follows:

Theorem 1.3 *For $a \in (-1, 0]$,*

$$\mathbb{P}\left(R_a > j\right) = \mathbb{E}\big[\exp\left\{-\kappa\nu^{a+j}\mathcal{W}^{(1)}\mathcal{W}^{(2)}\right\} |\mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0\big],$$

where $\kappa = \mu(\nu - 1)^{-1}$.

A special case of the above theorems is the case where

$$\lambda_i = (1 - F)^{-1}(i/N), \qquad i = 1, 2, \dots, N, \tag{1.16}$$

where F is a distribution function of a positive random variable, i.e., $F(0) = 0$, F is non-decreasing and $\lim_{x\to\infty} F(x) = 1$, see (B.1) below for a definition of $(1 - F)^{-1}$. In Appendix B, we will formulate quite general conditions on a distribution function F such that (**C1**)–(**C3**) hold with $\{\lambda_i\}_{i=1}^N$ defined by (1.16). The special cases

$$F(x) = 1 - \frac{c}{x^{\tau-1}}, \qquad \text{for which} \qquad (1 - F)^{-1}(y) = \left(\frac{c}{y}\right)^{1/(\tau-1)},$$

where $\tau > 3$, extends the results obtained by Chung and Lu [7] from $H_N / \log_\nu N \xrightarrow{\mathbb{P}} 1$ to the study of the fluctuations of H_N.

Instead of assigning *deterministic* capacities to the nodes one can also assign *random* capacities. Associated with the nodes is a sequence $\{\Lambda_i\}_{i=1}^N$ of positive i.i.d. random variables, with distribution

$$F_\Lambda(x) = \mathbb{P}(\Lambda \le x).$$

Then, we set, for $1 \le i \le N$,

$$\lambda_i = \Lambda_i.$$

For the i.i.d. case, we can identify μ, ν, f and g appearing in conditions (**C1**)–(**C3**) as

$$\mu = \mathbb{E}[\Lambda], \qquad \nu = \frac{\mathbb{E}[\Lambda^2]}{\mathbb{E}[\Lambda]}, \qquad f_n = \mathbb{E}\left[e^{-\Lambda}\frac{\Lambda^n}{n!}\right] \qquad \text{and} \qquad g_n = \frac{1}{\mu}\mathbb{E}\left[e^{-\Lambda}\frac{\Lambda^{n+1}}{n!}\right], \tag{1.17}$$

for $n \ge 0$.

The next theorem states that results of the deterministic capacities carry over to the case of random capacities:

Theorem 1.4 *Given an i.i.d. sequence of random variables $\{\Lambda_i\}_{i=1}^N$, with common distribution function F_Λ. If there exist constants $c > 0$ and $\tau > 3$ such that*

$$1 - F_\Lambda(x) \le c x^{1-\tau}, \text{ for all } x \ge 0, \tag{1.18}$$

*and with ν, given by (1.17), satisfying $\nu > 1$, then there exists an event $\mathcal{J}$ with $\mathbb{P}(\mathcal{J}) = 1 + o(1)$ such that, conditionally on $\{\Lambda_i\}_{i=1}^N$ satisfying $\mathcal{J}$, the conditions (**C1**)–(**C3**) hold.*

The proof of Theorem 1.4 is given in Appendix C. We have the following corollary to Theorem 1.4:

Corollary 1.5 *In the case of i.i.d. capacities, with common distribution function F_Λ satisfying (1.18) and with $\nu > 1$, the results of Theorem 1.1, Corollary 1.2 and Theorem 1.3 hold with high probability. More precisely, for every $k = 0, \pm 1, \pm 2, \ldots$, the random variable*

$$\frac{\mathbb{P}\left(H_N = \sigma_N + k \,\middle|\, \{\Lambda_i\}_{i=1}^N\right)}{\mathbb{P}\left(H_N < \infty \,\middle|\, \{\Lambda_i\}_{i=1}^N\right)} - \mathbb{P}(R_{a_N} = k) \tag{1.19}$$

converges in probability to zero.

We demonstrate Theorem 1.1 for the i.i.d. case using Corollary 1.5. Assume that $F_\Lambda(x) = 0$ for $x \le x_0$ and

$$F_\Lambda(x) = 1 - c x^{1-\tau}, \qquad x > x_0,$$

with $\tau = 3.5$, $c = 2.5981847$ and $x_0 = 0.7437937$, then $\nu = \mathbb{E}[\Lambda^2]/\mathbb{E}[\Lambda] \approx 2.231381$. We can pick different values of the size of the simulated graph, so that for each two simulated values N and M we have $a_N = a_M$. As an example, we take $N = M\nu^{2k}$, for some integer k. This induces, starting from $N_0 = M = 5000$, by taking for k the successive values 0,1,2,3

$$N_0 = M = 5000, \qquad N_1 = 24895, \qquad N_2 = 123955, \qquad \text{and} \qquad N_3 = 617181. \tag{1.20}$$

Observe that $a_{N_k} = 0.6117...$ for $k = 0, 1, 2, 3$. According to Corollary 1.5, the survival functions of the average graph distance H_{N_k}, run approximately parallel on distance 2 in the limit for $N \to \infty$, since $\log_\nu N_k = \log_\nu M + 2k$ for $k = 0, 1, 2, 3$. In Figure 1 we have simulated the survival function of the average graph distance for the graphs with sizes N_k with $k = 0, 1, 2, 3$, and, indeed, the plots are approximately parallel on distance 2.

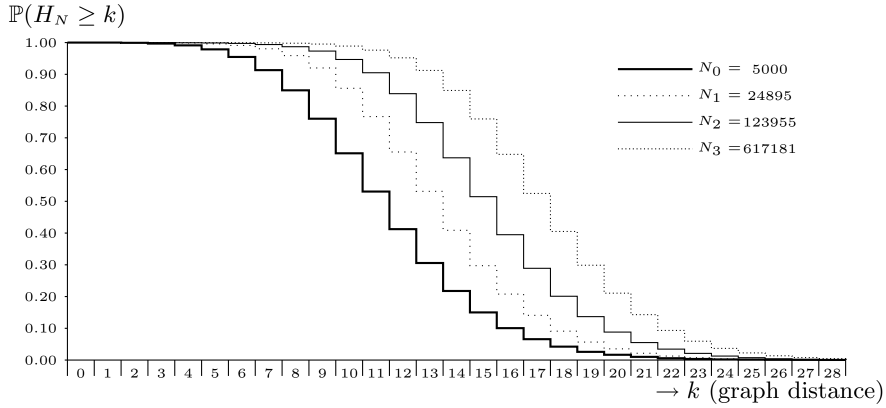

Figure 1: Empirical survival functions of the graph distance for $\tau = 3.5$ and for four values of N. Each plot is averaged over 1000 samples.

A general version of inhomogeneous random graphs with independent edges is presented in [5]. Our choice of graphs is a special case of the rank-1 case treated in [5, Section 16.4]. In the general setting of [5], the vertices are given by $\{x_i\}_{i=1}^N \subset \mathcal{X}$, for some state space $\mathcal{X}$, and there is an edge between vertices x_i and x_j with probability

$$p_{ij} = \min\{\kappa(x_i, x_j)/N, 1\}, \tag{1.21}$$

where $\kappa\colon \mathcal{X} \times \mathcal{X} \to [0, \infty)$ is a suitable kernel. The rank-1 case of [5] is the case where κ is of product form, i.e., $\kappa(x_i, x_j) = \psi(x_i)\psi(x_j)$, for some $\psi\colon \mathcal{X} \to [0, \infty)$. In fact, it is even possible that κ depends on N, i.e., $\kappa = \kappa_N$, such that κ_N converges to some limit as $N \to \infty$ in a suitable way. This allows one to simultaneously treat the cases where $p_{ij} = 1 - e^{-\kappa(x_i,x_j)/N}$ or $p_{ij} = \kappa(x_i, x_j)/(N + \kappa(x_i, x_j))$ (recall the special cases in Section 1.2). In [5], various results are proved in this generality, such as the phase transition of the resulting graphs, and distance results, such as average distances and the diameter. The main tool is a deep comparison to multitype Poisson branching processes.

In particular, [5, Theorem 3.14] states that for suitable *bounded* κ_N, the average distance between uniform pairs of connected nodes is equal to $(1 + o(1)) \log_\nu N$, for a certain $\nu > 1$. The condition $\nu > 1$ corresponds exactly to the random graph having a giant connected component of size proportional to the size of the graph (see [5, Theorem 3.1]). This average distance result generalizes the first result in Corollary 1.2, apart from the fact that in our setting we do not assume that κ_N is bounded. In fact, in certain cases, κ_N can be unbounded. In our paper, we state conditions on ψ in the rank-1 case of [5] under which we can identify the *fluctuations* of the average distances. It would be interesting to generalize our main results to the general setting of [5], i.e., to study the fluctuations of H_N in the general setting of [5]. However, it is unclear to us how we should generalize the tail assumption in (1.18) to this inhomogeneous setting.

1.4 Relations with the configuration model

The configuration model (CM) appeared in the context of random regular graphs as early as 1978 (see [2, 16]). Molloy and Reed [17] were the first to use the configuration model with specified degree sequences. Here we consider the CM as introduced in [13]. Start with an i.i.d. sequence $\{D_i^{(C)}\}_{i=1}^N$ of positive integer valued random variables , where $D_i^{(C)}$ will denote the degree of node i. To built a graph it is mandatory that $D_1^{(C)} + D_2^{(C)} + \ldots + D_N^{(C)}$ is even, so if $D_1^{(C)} + D_2^{(C)} + \ldots + D_N^{(C)}$ is odd we increase $D_N^{(C)}$ by one, which will have little effect. We build the graph model by attaching $D_i^{(C)}$ stubs or half edges to node i and pair the stubs at random, so that two half edges will form one edge.

In [13], the authors prove a version of Theorem 1.1-1.3 for the configuration model. Theorems 1.1-1.3 hold for the configuration model with only two changes:

1. Replace the condition $\nu > 1$ in Theorem 1.1, which is implicitly contained in (**C1**) by the condition $\tilde{\nu} > 1$, defined in (1.2).
2. Replace the offspring distributions of the BP $\{\mathcal{Z}_t\}_{t\geq 0}$, by

   (a)
   $$\tilde{f}_n = \mathbb{P}(D^{(C)} = n), \quad n \geq 1,$$

   (b)
   $$\tilde{g}_n = \frac{(n+1)\tilde{f}_{n+1}}{\mathbb{E}[D^{(C)}]}, \quad n \geq 0.$$

For the configuration model, the setting is as in Theorem 1.4, where, for the CM, the *degrees* are chosen in an i.d.d. fashion. The result in [13] proves that when we pick two uniform nodes, that their distance, when averaged out over the randomness in the i.i.d. degrees, satisfies (1.13). The convergence result in (1.19), which holds *conditionally* on the random degrees, and is stronger than (1.13), is *not* proved in [13]. We conjecture that a version of (1.19) also holds in the CM, when the degrees satisfy appropriate conditions that would take the place of (**C1**)–(**C3**).

One wonders why a result like Theorems 1.1-1.3 holds true for the class of models introduced in Section 1.1, especially if one realizes that in the CM the degrees are *independent*, and the edges are not, whereas for instance in the GRG (and in the other two examples) precisely the opposite is true, i.e., in the GRG the edges are independent and the degrees are not. To understand at least at an intuitive level why the distance result holds true, we compare the configuration model with the generalized random graph.

By construction the degree sequence $D_1^{(C)}, D_2^{(C)}, \ldots, D_N^{(C)}$ of the CM is an i.i.d. sequence, and conditionally on $\mathcal{D} = \{D_1^{(C)} = d_1, D_2^{(C)} = d_2, \ldots, D_N^{(C)} = d_N\}$, the graph configuration is uniform over all configurations satisfying $\mathcal{D}$, because the pairing is at random. Hence if we condition on both the event $\mathcal{D}$ and the event $\mathcal{S} = \{$the resulting graph has no self-loops and no multiple edges$\}$, then the CM renders a simple graph, which is picked uniformly from all possible *simple* configurations with degree sequence satisfying $\mathcal{D}$. Since for $N \to \infty$ the probability of the event $\mathcal{S}$ converges to $\exp(-\tilde{\nu}/2 - \tilde{\nu}^2/4) > 0$ (see [3, p. 51]), it follows from [16, Theorem 9.9] that properties that hold **whp** in the CM also hold **whp** in the conditioned simple graph. Hence a property as tightness of the graph distance $H_N^{(C)}$ in the CM is inherited by the conditioned simple graph, with the same degree sequence. This suggests that also the limiting distribution of the fluctuations of the graph distance in the CM conditioned on $\mathcal{S}$ is the same as the one in the CM as identified in [13]. A direct proof of this claim is missing though.

On the other hand the GRG with given degree sequence $d_1, d_2, \ldots, d_N$ is also uniform over all possible (simple) configurations. Moreover [6, Theorem 3.1] shows that the degree sequence $D_1^{(G)}, D_2^{(G)}, \ldots, D_N^{(G)}$ of the GRG is asymptotically independent with marginal distribution a mixed Poisson distribution:

$$\mathbb{P}(D^{(G)} = k) = \int_0^\infty e^{-x} \frac{x^k}{k!}\, dF_\Lambda(x), \quad k = 0, 1, 2 \ldots, \tag{1.22}$$

where F_Λ is the capacity distribution. Hence starting from $D_1^{(G)}, D_2^{(G)}, \ldots, D_N^{(G)}$ as an i.i.d. sequence with common distribution given by (1.22), the (conditioned) CM with these degrees is close to the GRG, at least in an asymptotic sense, so that one expects that the asymptotic fluctuations of the graph distance of the CM also hold for the generalized random graph. Also note from the mixed Poisson distribution (1.22), that

$$\tilde{\nu} = \frac{\mathbb{E}[D^{(C)}(D^{(C)} - 1)]}{\mathbb{E}[D^{(C)}]} = \frac{\mathbb{E}[\Lambda^2]}{\mathbb{E}[\Lambda]},$$

which is equal to ν, according to (1.17). As said earlier, a proof of this intuitive reasoning is missing, and our method of proof is by coupling each random graph satisfying **A1** and **A2** to the Poisson random graph (PRG), and by giving a separate proof of Theorem 1.1-1.3 for the PRG.

We finish this section by giving an overview of different distance results in random graphs. Let τ denote the exponent of the probability mass function of the degree distribution. In this paper and in [7, 13] the case $\tau > 3$ is studied. Results for $2 < \tau < 3$ for various models appeared in [7, 14, 19, 20]. Typically in that case, the distance fluctuates around a constant times $2 \log\log N/|\log(\tau-2)|$. For $1 < \tau < 2$, there exists a subset of nodes with a high degree, called the *core* (see [10]). The core forms a complete graph and almost every node is attached to the core and, thus, the graph distance is with high probability at most 3.

1.5 Organization of the paper

The coupling argument that ties the fluctuations of the graph distance $H_N^{(P)}$ in the PRG to the fluctuations of the graph distance in random graphs satisfying assumptions **A1** and **A2** is treated in Section 2. In Section 4 we show that the fluctuations of the graph distance $H_N^{(P)}$ is given by Theorem 1.1. The derivation of the fluctuations of the graph distance $H_N^{(P)}$ is similar to the derivation of the fluctuations of the graph distance $H_N^{(C)}$ in the configuration model, see [13]. The proof in [13] is more complicated than the proof presented here for the PRG model, mainly because in the latter the expansion of a given node e.g., the nodes on a given distance, can be described by means of the so-called Reittu-Norros process, a marked branching process. This branching process will be introduced in Section 3.

Most of the proofs of the auxiliary propositions and lemmas introduced in Sections 3 and 4 are left out as they are similar to those in [13]. Since details of these proofs are different, we included full proofs in Appendix D of the extended version of this paper (see [9]).

2 Coupling

In this section we denote by G_N the PRG and by G'_N some other random graph satisfying the assumptions **A1** and **A2**, given in Section 1.1. We number the nodes of both G_N and G'_N from 1 to N and we assign the capacity λ_i, for $1 \le i \le N$, to node i in each graph. We denote by H_N and H'_N the graph distance between two randomly chosen nodes A_1 and A_2, such that $A_1 \neq A_2$, in G_N and G'_N, respectively. We will show that for $N \to \infty$,

$$\mathbb{P}\left(H_N \neq H'_N\right) = o(1). \tag{2.1}$$

The above implies that **whp** the coupling of the graph distances is successful. Therefore, given the succesful coupling (2.1), it is sufficient to show Theorem 1.1 for the PRG.

The coupling bound in (2.1) has since been considerably strengthened by Janson in [15], who studies when two random graphs with independent edges are asymptotically equivalent. By [15, Example 3.6] and under the assumptions **A1** and **A2**, we have that the probability of *any* event A_N is asymptotically equal for G_N and G'_N when $N \to \infty$.

2.1 Coupling of G_N and G'_N

We next describe the coupling of the connection variables of the graphs G_N and G'_N. A classical coupling is used, see e.g. [21]. Denote by $\{X_{ij}\}_{1\le i<j\le N}$ and $\{X'_{ij}\}_{1\le i<j\le N}$ the connection variables of the graphs G_N and G'_N, and, similarly, denote the connection probabilities by $\{p_{ij}\}_{1\le i<j\le N}$ and $\{p'_{ij}\}_{1\le i<j\le N}$. For the coupling we introduce independent random variables $\{K_{ij}\}_{1\le i<j\le N}$. Set

$\underline{p}_{ij} = \min\{p_{ij}, p'_{ij}\}$ and $\overline{p}_{ij} = \max\{p_{ij}, p'_{ij}\}$, and define random variables $\hat{X}_{ij}$ and $\hat{X}'_{ij}$ with

$$\mathbb{P}\left(\hat{X}_{ij} = 1, \hat{X}'_{ij} = 1, K_{ij} = 0\right) = \underline{p}_{ij}, \qquad \mathbb{P}\left(\hat{X}_{ij} = 1, \hat{X}'_{ij} = 0, K_{ij} = 1\right) = p_{ij} - \underline{p}_{ij},$$
$$\mathbb{P}\left(\hat{X}_{ij} = 0, \hat{X}'_{ij} = 1, K_{ij} = 1\right) = \overline{p}_{ij} - p_{ij}, \qquad \mathbb{P}\left(\hat{X}_{ij} = 0, \hat{X}'_{ij} = 0, K_{ij} = 0\right) = 1 - \overline{p}_{ij},$$

whereas all other combinations have probability 0. Then the laws of $\hat{X}_{ij}$ and $\hat{X}'_{ij}$ are the same as the laws of X_{ij} and X'_{ij}, respectively. Furthermore, K_{ij} assumes the value 1 with probability $|p_{ij} - p'_{ij}|$, and is 0 otherwise. Note that we do abuse the notation in the above display. We should replace the probability measure $\mathbb{P}$ in the above display by some other probability measure $\mathbb{Q}$, because the probability space is defined by the graphs G_N and G'_N, instead of only the graph G_N. Since the graphs, conditioned on the capacities, are constructed independently from each other, this abuse of notation is not a problem.

Consider the nodes i and j, $1 \le i < j \le N$, in the graphs G_N and G'_N simultaneously. Then the event $\{K_{ij} = 0\} = \{\hat{X}_{ij} = \hat{X}'_{ij}\}$ corresponds to the event that in both graphs there exists a connection between nodes i and j, or that in both graphs there is no connection between nodes i and j. The event $\{K_{ij} = 1\} = \{\hat{X}_{ij} \neq \hat{X}'_{ij}\}$ corresponds with the event that there exists a connection in one of the graphs, but not in the other one. We call the event $\{K_{ij} = 1\}$ a *mismatch* between the nodes i and j.

Assumption **A2** implies that for some constant $C' > 0$,

$$\mathbb{P}(K_{ij} = 1) = \left|p_{ij} - p'_{ij}\right| \le |p_{ij} - \lambda_i\lambda_j/l_N| + \left|p'_{ij} - \lambda_i\lambda_j/l_N\right| \le C'\frac{\lambda_i^2\lambda_j^2}{l_N^2}, \tag{2.2}$$

for all $1 \le i < j \le N$. The number of mismatches due to all the nodes incident to node i, $1 \le i \le N$, is given by

$$K_i = \sum_{j\neq i} K_{ij}. \tag{2.3}$$

Obviously, we cannot couple all the connections in the graphs G_N and G'_N successfully, but the total number of mismatches due to all the nodes can be bounded from above by any positive power of N. To this end, we define the event $\mathcal{A}_N$ as

$$\mathcal{A}_N = \bigcap_{i=1}^{N} \left\{K_i \mathbf{1}_{\{\lambda_i > c_N\}} = 0\right\} = \left\{\sum_{i=1}^{N} K_i \mathbf{1}_{\{\lambda_i > c_N\}} = 0\right\}, \tag{2.4}$$

where $c_N = N^{\xi}$, with $\xi > 0$. Then, on the event $\mathcal{A}_N$, all nodes with capacity greater than c_N are successfully coupled.

Lemma 2.1 *For each $\xi > 0$ there exists a constant $\theta > 0$ such that*

$$\mathbb{P}(\mathcal{A}_N^c) = \mathcal{O}\big(N^{-\theta}\big). \tag{2.5}$$

Proof. We bound $\mathbb{P}(\mathcal{A}_N^c)$ using Boole's inequality, and the Markov inequality:

$$\mathbb{P}(\mathcal{A}_N^c) \le \sum_{i=1}^{N} \mathbb{P}\left(K_i \mathbf{1}_{\{\lambda_i > c_N\}} \ge 1\right) \le \sum_{i=1}^{N} \mathbb{E}[K_i]\, \mathbf{1}_{\{\lambda_i > c_N\}}. \tag{2.6}$$

Then, using (2.2), (1.6) and (1.7), the expectation $\mathbb{E}[K_i]$ can be bounded by,

$$\mathbb{E}[K_i] = \sum_{j\neq i} \mathbb{E}[K_{ij}] = \sum_{j\neq i} \mathbb{P}(K_{ij} = 1) \le \frac{C'\lambda_i^2}{l_N^2} \sum_{j=1}^{N} \lambda_j^2 = \frac{C'\lambda_i^2}{N} \frac{\nu_N}{\mu_N} = \mathcal{O}\big(\lambda_i^2 N^{-1}\big) \tag{2.7}$$

Using the above and (2.6), we have that

$$\mathbb{P}\left(\mathcal{A}_N^c\right) \leq \sum_{i=1}^N \mathbb{E}[K_i]\, \mathbf{1}_{\{\lambda_i > c_N\}} = \mathcal{O}\big(\frac{1}{N}\sum_{i=1}^N \lambda_i^2 \mathbf{1}_{\{\lambda_i > c_N\}}\big). \tag{2.8}$$

Observe that,

$$\frac{1}{N}\sum_{i=1}^N \lambda_i^2 \mathbf{1}_{\{\lambda_i > c_N\}} = \frac{1}{N}\sum_{i=1}^N \lambda_i^{\tau-1-\varepsilon}\lambda_i^{-(\tau-3-\varepsilon)}\mathbf{1}_{\{\lambda_i > c_N\}} \leq c_N^{-(\tau-3-\varepsilon)}\frac{1}{N}\sum_{i=1}^N \lambda_i^{\tau-1-\varepsilon} = \mathcal{O}\big(c_N^{-(\tau-3-\varepsilon)}\big),$$

where we applied (1.11) in the last step. Pick $\theta = \xi(\tau - 3 - \varepsilon)$, then combining the above display and (2.8), to conclude that

$$\mathbb{P}\left(\mathcal{A}_N^c\right) = \mathcal{O}\big(c_N^{-(\tau-3-\varepsilon)}\big) = \mathcal{O}\big(N^{-\theta}\big).$$

□

2.2 Coupling the graph distances of G_N and G'_N

In this subsection we couple the graph distance of the PRG with any random graph satisfying assumptions **A1** and **A2**.

Theorem 2.2 *Let G_N be a PRG and let and G'_N be a random graph satisfying assumption* **A1** *and* **A2***. Let H_N and H'_N be the graph distances between two different uniformly chosen nodes A_1 and A_2 in, respectively, the graphs G_N and G'_N. Then*

$$\mathbb{P}\left(H_N \neq H'_N\right) = o(1). \tag{2.9}$$

In order to prove Theorem 2.2, we will use the following strategy. We know that for the PRG the random variable H_N is concentrated around $\log_\nu N$. Hence if we take $t_N = \lfloor(\frac{1}{2}+\eta)\log_\nu N\rfloor$, with $\eta > 0$, then $\mathbb{P}\left(H_N > 2t_N\right)$ is small and in order that $H_N \neq H'_N$ on the set where $\{H_N \leq 2t_N\}$, there must have been at least one mismatch between two nodes i and j, both being not on a greater graph distance from our random node A than $2t_N$.

We define the neighborhood shells of a uniformly chosen node $A \in \{1, 2, \ldots, N\}$, i.e., all nodes on a fixed graph distance of node A, by

$$\partial\mathcal{N}_0 = \{A\} \qquad \text{and} \qquad \partial\mathcal{N}_k = \{1 \leq j \leq N : d(A, j) = k\}, \tag{2.10}$$

where $d(i, j)$ denotes the graph distance between nodes i and j, i.e., the minimum number of edges in a path between the nodes i and j. Furthermore, define the set of nodes reachable in at most t steps from root A by

$$\mathcal{N}_t = \{1 \leq j \leq N : d(A, j) \leq t\} = \bigcup_{k=0}^t \partial\mathcal{N}_k. \tag{2.11}$$

Proposition 2.3 *For N sufficiently large, $t \in \mathbb{N}$, and every $b \in (0, 1)$,*

$$\mathbb{P}\left(H_N \neq H'_N\right) \leq \mathbb{P}\left(\mathcal{A}_N^c\right) + \mathbb{P}\left(H_N > 2t\right) + 2t\mathbb{P}\left(|\mathcal{N}_{t-1}| > N^b\right) + \mathcal{O}\big(tN^{-1+b}c_N^4\big). \tag{2.12}$$

Before giving a proof, we show that Theorem 2.2 is a consequence of Proposition 2.3.

Proof of Theorem 2.2. By Lemma 2.1, we have that, $\mathbb{P}\left(\mathcal{A}_N^c\right) \leq N^{-\theta}$. From Corollary 1.2, applied to the PRG model, we obtain that $\mathbb{P}\left(H_N > 2t_N\right) = o(1)$. The third term on the right side of (2.12) can be bounded using the following lemma:

Lemma 2.4 *Let $\{\mathcal{N}_t\}_{t\geq 0}$ be the reachable sets of a uniformly chosen node A in the PRG G_N. Then for $\eta, \delta \in (-1/2, 1/2)$ and all $t \leq (1/2+\eta)\log_\nu N$, there exists a constant $\beta_1 > 0$ such that*

$$\mathbb{P}\left(|\mathcal{N}_t| > N^{1/2+\delta}\right) = \mathcal{O}\big((\log_\nu N)N^{-(\delta-\eta)}\big). \tag{2.13}$$

Proof. The full proof of this lemma is given in the extended version of this paper [9, Lemma D.2]. We now give a heuristic derivation. We will couple $|\partial\mathcal{N}_t|$ to $\mathcal{Z}_t$ so that $|\partial\mathcal{N}_t| \approx \mathcal{Z}_t$, **whp**. If we replace $|\mathcal{N}_t|$ by $\sum_{k=0}^{t}\mathcal{Z}_k$ in (2.13), then it is easy to verify that, using the Markov inequality,

$$\mathbb{P}\left(\sum_{k=1}^{t}\mathcal{Z}_k > N^{1/2+\delta}\right) \leq N^{-1/2-\delta}\sum_{k=1}^{t}\mathbb{E}[\mathcal{Z}_k] = N^{-1/2-\delta}\sum_{k=1}^{t}\mu\nu^{k-1} = \mathcal{O}\big((\log_\nu N)N^{-(\delta-\eta)}\big),$$

which then implies the result. □

We now prove that all terms in the right hand of (2.12) are $o(1)$ for an appropriate choice of b. Lemma 2.4 implies that $2t_N\mathbb{P}\left(|\mathcal{N}_t| > N^b\right) = o(1)$ for some appropriately chosen $b > \frac{1}{2}$. Then, provided that $b < 1$, we see that $t_N N^{b-1}c_N^4 = t_N N^{4\xi+b-1} = o(1)$, where we substitute $c_N = N^\xi$, and picking $\xi \in (0, (1-b)/4)$. Hence, by Proposition 2.3, $\mathbb{P}(H_N \neq H'_N) = o(1)$, which is precisely the content of Theorem 2.2. □

Proof of Proposition 2.3. We use that

$$\mathbb{P}\left(H_N \neq H'_N\right) \leq \mathbb{P}\left(\mathcal{A}_N^c\right) + \mathbb{P}\left(H_N > 2t\right) + \mathbb{P}\left(\{H_N \leq 2t\} \cap \mathcal{A}_N \cap \{H_N \neq H'_N\}\right). \tag{2.14}$$

Let $\mathcal{N}_t^{(i)}$ and $\mathcal{N}_t'^{(i)}$, for $i = 1, 2$, be the union of neighborhood shells of the nodes A_i in G_N and G'_N, respectively. Now, we use the fact that if $H_N \leq 2t$ and if $H_N \neq H'_N$, then $\mathcal{N}_t^{(1)} \neq \mathcal{N}_t'^{(1)}$ and/or $\mathcal{N}_t^{(2)} \neq \mathcal{N}_t'^{(2)}$. Since A_1, A_2 are chosen uniformly from $\{1, \ldots, N\}$, we have

$$\mathbb{P}\left(\{H_N \leq 2t\} \cap \mathcal{A}_N \cap \{H_N \neq H'_N\}\right) \leq 2\mathbb{P}\left(\{\mathcal{N}_t \neq \mathcal{N}'_t\} \cap \mathcal{A}_N\right). \tag{2.15}$$

If $\mathcal{N}_t \neq \mathcal{N}'_t$, then there must be a $k \in \{1, \ldots, t\}$ for which $\mathcal{N}_k \neq \mathcal{N}'_k$, but $\mathcal{N}_{k-1} = \mathcal{N}'_{k-1}$. Thus,

$$\mathbb{P}\left(\{H_N \leq 2t\} \cap \mathcal{A}_N \cap \{H_N \neq H'_N\}\right) \leq 2\sum_{k=1}^{t}\mathbb{P}\left(\{\mathcal{N}_k \neq \mathcal{N}'_k\} \cap \{\mathcal{N}_{k-1} = \mathcal{N}'_{k-1}\} \cap \mathcal{A}_N\right). \tag{2.16}$$

In turn, the event $\{\mathcal{N}_k \neq \mathcal{N}'_k\} \cap \{\mathcal{N}_{k-1} = \mathcal{N}'_{k-1}\}$ implies that one of the edges from $\partial\mathcal{N}_{k-1}$ must be miscoupled, thus $K_{ij} = 1$ for some $i \in \partial\mathcal{N}_{k-1}$ and $j \in \mathcal{N}_{k-1}^c$, where $\mathcal{N}_{k-1}^c = \{1, 2, \ldots, N\}\backslash\mathcal{N}_{k-1}$. The event $\mathcal{A}_N$ implies that $\lambda_i, \lambda_j \leq c_N$. Therefore, we bound

$$\begin{aligned}\mathbb{P}\left(\{\mathcal{N}_k \neq \mathcal{N}'_k\} \cap \{\mathcal{N}_{k-1} = \mathcal{N}'_{k-1}\} \cap \mathcal{A}_N\right) &\leq \mathbb{P}\left(|\mathcal{N}_{k-1}| > N^b\right) \\ &+ \sum_{i,j}\mathbb{P}\left(\{i \in \partial\mathcal{N}_{k-1}, j \in \mathcal{N}_{k-1}^c, K_{ij} = 1\} \cap \{|\mathcal{N}_{k-1}| \leq N^b\}\right)\mathbf{1}_{\{\lambda_i,\lambda_j \leq c_N\}}.\end{aligned} \tag{2.17}$$

Since $i \in \mathcal{N}_{k-1}^c$ and $j \in \partial\mathcal{N}_{k-1}$, the event $\{K_{ij} = 1\}$ is independent of $\mathcal{N}_{k-1}$ and, therefore, from $\partial\mathcal{N}_{k-1}$ as $\partial\mathcal{N}_{k-1} \subset \mathcal{N}_{k-1}$. The edge between the nodes i and j points out of $\mathcal{N}_{k-1}$, while $\mathcal{N}_{k-1}$ is determined by the occupation status of edges that are between nodes in $\mathcal{N}_{k-2}$ or pointing out of $\partial\mathcal{N}_{k-2}$. Thus, we can replace each term in the sum of (2.17) by

$$\mathbb{P}\left(K_{ij} = 1\right)\mathbb{P}\left(\{i \in \partial\mathcal{N}_{k-1}, j \in \mathcal{N}_{k-1}^c\} \cap \{|\mathcal{N}_{k-1}| \leq N^b\}\right)\mathbf{1}_{\{\lambda_i,\lambda_j \leq c_N\}}. \tag{2.18}$$

Since by (2.2), we have

$$\mathbb{P}\left(K_{ij} = 1\right)\mathbf{1}_{\{\lambda_i,\lambda_j \leq c_N\}} \leq C'\frac{\lambda_i^2\lambda_j^2}{l_N^2}\mathbf{1}_{\{\lambda_i,\lambda_j \leq c_N\}} = \mathcal{O}\big(c_N^4 N^{-2}\big), \tag{2.19}$$

we can bound the right side of (2.17) from above by

$$\mathbb{P}\left(|\mathcal{N}_{k-1}| > N^b\right) + \mathcal{O}\big(c_N^4 N^{-2}\big) \sum_{i,j} \mathbb{P}\left(\{i \in \partial\mathcal{N}_{k-1}, j \in \mathcal{N}_{k-1}^c\} \cap \{|\mathcal{N}_{k-1}| \leq N^b\}\right).$$

Finally, we bound the sum on the right side by

$$\sum_{i,j} \mathbb{P}\left(\{i \in \partial\mathcal{N}_{k-1}, j \in \mathcal{N}_{k-1}^c\} \cap \{|\mathcal{N}_{k-1}| \leq N^b\}\right) \leq N\mathbb{E}\Big[|\partial\mathcal{N}_{k-1}|\mathbf{1}_{\{|\mathcal{N}_{k-1}| \leq N^b\}}\Big] \leq N^{1+b}.$$

Therefore, we can bound each term in the sum of (2.16) by

$$\mathbb{P}\left(\{\mathcal{N}_k \neq \mathcal{N}_k'\} \cap \{\mathcal{N}_{k-1} = \mathcal{N}_{k-1}'\} \cap \mathcal{A}_N\right) \leq \mathbb{P}\left(|\mathcal{N}_{k-1}| > N^b\right) + \mathcal{O}\big(c_N^4 N^{-1+b}\big).$$

Since, for $k \leq t$, we have that $\mathbb{P}\left(|\mathcal{N}_{k-1}| > N^b\right) \leq \mathbb{P}\left(|\mathcal{N}_{t-1}| > N^b\right)$, by summing over $k = 1, \ldots, t$ in (2.16), we arrive at

$$\mathbb{P}\left(\{H_N \leq 2t\} \cap \mathcal{A}_N \cap \{H_N \neq H_N'\}\right) \leq 2t\mathbb{P}\left(|\mathcal{N}_{t-1}| > N^b\right) + \mathcal{O}\big(tc_N^4 N^{-(1-b)}\big).$$

Together with (2.14) this proves the proposition. □

3 The Poissonian random graph model

The proof of the fluctuations of the graph distance in the CM in [13] has been done in a number of steps. One of the most important steps is the coupling of the expansion of the neighborhood shells of a node to a branching process (BP). For the PRG, we follow the same strategy as in [13], although the details differ substantially.

The first step is to introduce the $\overline{\text{NR}}$-process, which is a marked BP and was introduced by Norros and Reittu in [20]. We can thin the $\overline{\text{NR}}$-process in such a way that the resulting process, the $\underline{\text{NR}}$-process, can be coupled to the expansion of the neighborhood shells of a randomly chosen node in the PRG. Finally, we introduce capacities for the $\overline{\text{NR}}$-process and the $\underline{\text{NR}}$-process.

3.1 The NR-process and its thinned version

The $\overline{\text{NR}}$-process is a marked delayed BP denoted by $\{\overline{Z}_t, \overline{\boldsymbol{M}}_t\}_{t\geq 0}$, where $\overline{Z}_t$ denotes the number of individuals of generation t, and where the vector

$$\overline{\boldsymbol{M}}_t = (\overline{M}_{t,1}, \overline{M}_{t,2}, \ldots, \overline{M}_{t,\overline{Z}_t}) \in \{1, 2, \ldots, N\}^{\overline{Z}_t},$$

denotes the marks of the individuals in generation t. We now give a more precise definition of the $\overline{\text{NR}}$-process and describe its connection with the PRG. We define $\overline{Z}_0 = 1$ and take $\overline{M}_{0,1}$ randomly from the set $\{1, 2, \ldots, N\}$, corresponding to the choice of A_1. The offspring of an individual with mark $m \in \{1, 2, \ldots, N\}$ is as follows: the total number of children has a Poisson distribution with parameter λ_m, of which, for each $i \in \{1, 2, \ldots, N\}$, a Poisson distributed number with parameter

$$\frac{\lambda_i \lambda_m}{l_N}, \tag{3.1}$$

bears mark i, independently of the other individuals. Since $\sum_{i=1}^N \lambda_i\lambda_m/l_N = \lambda_m$, and sums of independent Poisson random variables are again Poissonian, we may take the number of children with different marks mutually independent. As a result of this definition, the marks of the children of an individual in $\{\overline{Z}_t, \overline{\boldsymbol{M}}_t\}_{t\geq 0}$ can be seen as independent realizations of a random variable M, with distribution

$$\mathbb{P}(M = m) = \frac{\lambda_m}{l_N}, \quad 1 \leq m \leq N, \tag{3.2}$$

and, consequently,

$$\mathbb{E}[\lambda_M] = \sum_{m=1}^{N} \lambda_m \mathbb{P}(M = m) = \frac{1}{l_N} \sum_{m=1}^{N} \lambda_m^2. \tag{3.3}$$

For the definition of the $\underline{\text{NR}}$-process we start with a copy of the $\overline{\text{NR}}$-process $\{\overline{Z}_t, \overline{\boldsymbol{M}}_t\}_{t\geq 0}$, and reduce this process generation by generation, i.e., in the order

$$\overline{M}_{0,1}, \overline{M}_{1,1}, \ldots \overline{M}_{1,\overline{Z}_1}, \overline{M}_{2,1}, \ldots \tag{3.4}$$

by discarding each individual and all its descendants whose mark has appeared before. The process obtained in this way is called the $\underline{\text{NR}}$-process and is denoted by the sequence $\{\underline{Z}_t, \underline{\boldsymbol{M}}_t\}_{t\geq 0}$. One of the main results of [20] is [20, Proposition 3.1]:

Proposition 3.1 *Let $\{\underline{Z}_t, \underline{\boldsymbol{M}}_t\}_{t\geq 0}$ be the $\underline{\text{NR}}$-process and let $\underline{\boldsymbol{M}}_t$ be the set of marks in the t^{th} generation, then the sequence of sets $\{\underline{\boldsymbol{M}}_t\}_{t\geq 0}$ has the same distribution as the sequence $\{\partial\mathcal{N}_t\}_{t\geq 0}$ given by* (2.10).

Proof. In [20], it is assumed that the sequence of capacities is random, but in most proofs, including [20, Proposition 3.1], Norros and Reittu condition on the capacities, and therefore consider the capacities as being deterministic. Thus, the proof of [20, Proposition 3.1] holds verbatim. □

As a consequence of the previous proposition, we can couple the $\underline{\text{NR}}$-process to the neighborhood shells of a uniformly chosen node $A \in \{1, 2, \ldots, N\}$, i.e., all nodes on a fixed graph distance of A, see (2.10) and note that $A \sim \overline{M}_{0,1}$. Thus, using the the above proposition, we can couple the expansion of the neighborhood shells and the $\underline{\text{NR}}$-process in such a way that

$$\underline{\boldsymbol{M}}_t = \partial\mathcal{N}_t \qquad \text{and} \qquad \underline{Z}_t = |\partial\mathcal{N}_t|, \ t \geq 0. \tag{3.5}$$

Furthermore, we see that an individual with mark m, $1 \leq m \leq N$, in the $\underline{\text{NR}}$-process is identified with node m in the graph G_N, whose capacity is given by λ_m.

The offspring distribution $f^{(N)}$ of $\overline{Z}_1$, i.e., the first generation of $\{\overline{Z}_t\}_{t\geq 0}$, is given by

$$f_n^{(N)} = \mathbb{P}(Poi(\lambda_A) = n) = \frac{1}{N} \sum_{m=1}^{N} e^{-\lambda_m} \frac{\lambda_m^n}{n!}, \qquad n \geq 0. \tag{3.6}$$

Recall that individuals in the second and further generations have a random mark distributed as M, given by (3.2). Hence, if we denote the offspring distribution of the second and further generations by $g_n^{(N)}$, then we obtain

$$g_n^{(N)} = \mathbb{P}(Poi(\lambda_M) = n) = \sum_{m=1}^{N} e^{-\Lambda_m} \frac{\lambda_m^n}{n!} \frac{\lambda_m}{l_N} = \frac{1}{l_N} \sum_{m=1}^{N} e^{-\lambda_m} \frac{\lambda_m^{n+1}}{n!}, \qquad n \geq 0. \tag{3.7}$$

Furthermore, we can relate $g_n^{(N)}$ and $f_n^{(N)}$ by

$$g_n^{(N)} = \frac{(n+1)}{l_N/N} \frac{1}{N} \sum_{m=1}^{N} e^{-\lambda_m} \frac{\lambda_m^{n+1}}{(n+1)!} = \frac{(n+1) f_{n+1}^{(N)}}{\mu_N}. \tag{3.8}$$

It follows from condition (**C2**) that $f_n^{(N)} \to f_n$ and $g_n^{(N)} \to g_n$.

3.2 Coupling with a delayed BP

In this subsection we will introduce a coupling between the $\underline{\text{NR}}$-process and the delayed BP $\{\mathcal{Z}_t\}_{t\geq 0}$, which is defined by condition (**C2**) in Section 1.3. This coupling is used in the proof of Theorem 1.1 and 1.3 for the PRG, to express the probability distribution of H_N in terms of the BP $\{\mathcal{Z}_t\}_{t\geq 0}$. The full proof of these propositions are given in the extended version of this paper [9].

Introduce the total capacity of the t^{th} generation of the $\overline{\text{NR}}$-process $\{\overline{Z}_t, \overline{\boldsymbol{M}}_t\}_{t\geq 0}$ and the $\underline{\text{NR}}$-process $\{\underline{Z}_t, \underline{\boldsymbol{M}}_t\}_{t\geq 0}$ as, respectively,

$$\overline{C}_{t+1} = \sum_{i=1}^{\overline{Z}_t} \lambda\big(\overline{M}_{t,i}\big) \qquad \text{and} \qquad \underline{C}_{t+1} = \sum_{i=1}^{\underline{Z}_t} \lambda\big(\underline{M}_{t,i}\big)\,,\ t\geq 0, \tag{3.9}$$

where, to improve readability, we write $\lambda(A) = \lambda_A$. Using the coupling given by (3.5), we can rewrite the capacity $\underline{C}_{t+1}$ as

$$\underline{C}_{t+1} = \sum_{i\in\partial\mathcal{N}_t} \lambda_i. \tag{3.10}$$

For the proof of Theorem 1.1 and 1.3, in the case of the PRG, we need to control the difference between $\overline{C}_t$ and $\underline{C}_t$ for fixed t. For this we will use the following proposition:

Proposition 3.2 *There exist constants $\alpha_2, \beta_2 > 0$, such that for all $0 < \eta < \alpha_2$ and all $t \leq t_N = (1/2+\eta)\log_\nu N$,*

$$\mathbb{P}\left(\sum_{k=1}^{t}(\overline{C}_k - \underline{C}_k) > N^{1/2-\alpha_2}\right) \leq N^{-\beta_2}. \tag{3.11}$$

Proof. Notice that $\underline{C}_k \leq \overline{C}_k$ holds trivially, because $\underline{Z}_k$ is obtained from $\overline{Z}_k$ by thinning. The full proof of Proposition 3.2 is given in [9, Section D.1]. The proof consists of several steps. Denote by a *duplicate* an individual in the $\overline{\text{NR}}$-process whose mark has appeared previously. (See [9] for a formal definition of a duplicate.) Firstly, we have to keep track of all the duplicates. We show that **whp** duplicates do not appear in the first t_N generations and that the number of duplicates in the first t_N generations can be bounded from above by some small power of N. Then, secondly, we bound the total progeny of each duplicate. Combining these results gives the claim of this proposition. □

In order to prove Theorem 1.1 and Theorem 1.3 we will grow two $\underline{\text{NR}}$-processes $\{\underline{Z}_t^{(i)}, \underline{\boldsymbol{M}}_t^{(i)}\}_{t\geq 0}$, for $i = 1,2$. The root of $\{\underline{Z}_t^{(i)}, \underline{\boldsymbol{M}}_t^{(i)}\}_{t\geq 0}$ starts from a uniformly chosen node or mark $A_i \in \{1,2,\ldots,N\}$. These two nodes are different **whp**, because

$$\mathbb{P}(A_1 = A_2) = \frac{1}{N}.$$

By (3.5), the $\underline{\text{NR}}$-process can be coupled to the neighborhood expansion shells $\{\mathcal{N}_t^{(1)}\}_{t\geq 0}$ and $\{\mathcal{N}_t^{(2)}\}_{t\geq 0}$. In the following lemma we compute the distribution of the number of edges between two shells with different subindeces, i.e., $\mathcal{N}_k^{(1)}$ and $\mathcal{N}_t^{(2)}$.

Lemma 3.3 *Fix integers k and t. Then conditionally on $\mathcal{N}_k^{(1)}$ and $\mathcal{N}_t^{(2)}$ and given that $\mathcal{N}_k^{(1)} \cap \mathcal{N}_t^{(2)} = \varnothing$ the number of edges between the nodes in $\mathcal{N}_k^{(1)}$ and $\mathcal{N}_t^{(2)}$ is distributed as a Poisson random variable with mean*

$$\frac{\underline{C}_{k+1}^{(1)}\underline{C}_{t+1}^{(2)}}{l_N}. \tag{3.12}$$

Proof. Conditionally on $\mathcal{N}_k^{(1)}$, $\mathcal{N}_t^{(2)}$ and $\mathcal{N}_k^{(1)} \cap \mathcal{N}_t^{(2)} = \varnothing$, the number of edges between $\mathcal{N}_k^{(1)}$ and $\mathcal{N}_t^{(2)}$ is given by

$$\sum_{i \in \partial \mathcal{N}_k^{(1)}} \sum_{j \in \partial \mathcal{N}_t^{(2)}} E_{ij}^{(P)}, \tag{3.13}$$

where $E_{ij}^{(P)}$ are independent Poisson random variables with mean $\lambda_i \lambda_j / l_N$, see (1.5). Sums of independent Poisson random variables are again Poissonian, thus (3.13) is a Poisson random variable with mean the expected value of (3.13) given by

$$\sum_{i \in \partial \mathcal{N}_k^{(1)}} \sum_{j \in \partial \mathcal{N}_t^{(2)}} \frac{\lambda_i \lambda_j}{l_N} = \frac{\underline{C}_{k+1}^{(1)} \underline{C}_{t+1}^{(2)}}{l_N},$$

where we have used (3.10) in the last step. □

The further proof of Theorems 1.1-1.3 crucially relies on the following technical claim:

Proposition 3.4 *There exist constants $u_2, v_2, \eta > 0$ such that for all $t \le t_N = (1 + 2\eta) \log_\nu N$, as $N \to \infty$,*

$$\mathbb{P}\Big(\frac{1}{N} \Big| \sum_{k=2}^{t+1} \mathcal{Z}_{\lceil k/2 \rceil}^{(1)} \mathcal{Z}_{\lfloor k/2 \rfloor}^{(2)} - \sum_{k=2}^{t+1} \underline{C}_{\lceil k/2 \rceil}^{(1)} \underline{C}_{\lfloor k/2 \rfloor}^{(2)} \Big| > N^{-u_2} \Big) = \mathcal{O}\big(N^{-v_2}\big). \tag{3.14}$$

Proof. The proof of Proposition 3.4 is given in the extended version of this paper, see [9]. We will intuitively explain how $\underline{C}_k^{(i)}$, for $i = 1, 2$ and $1 \le k \le t$, can be replaced by $\mathcal{Z}_k^{(i)}$. Firstly, whenever $k \le (1/2 + \eta) \log_\nu N$, we can neglect the influence of the thinning. Thus, we can replace $\underline{C}_k^{(i)}$ by $\overline{C}_k^{(i)}$. The capacity $\overline{C}_k^{(i)}$ is the sum of the capacities of the $\overline{Z}_{k-1}^{(i)}$ nodes of the $(k-1)^{\rm st}$ generation. Conditionally on $\overline{Z}_{k-1}^{(i)}$, the value of $\mathbb{E}\Big[\overline{C}_k^{(i)} | \overline{Z}_{k-1}^{(i)} \Big]$ is given by $\nu \overline{Z}_{k-1}^{(i)}$, thus the next step is to replace $\overline{C}_k^{(i)}$ by $\nu \overline{Z}_{k-1}^{(i)}$. Finally, by condition (**C2**), the offspring distributions of the BP $\{\overline{Z}_k\}_{k \ge 0}$ converges to those of $\{\mathcal{Z}_k\}_{k \ge 0}$. Therefore, we can replace $\nu \overline{Z}_{k-1}^{(i)}$ by $\nu \mathcal{Z}_{k-1}^{(i)}$, which we, finally, replace by $\mathcal{Z}_k^{(i)}$.
□

In the next section we will use this proposition in combination with Lemma 3.3 to replace sums over capacities, which do not depend on N, by sums over sizes of a BP, which no longer depend on N.

4 Proof of Theorem 1.1 and 1.3 for the PRG

In this section, we prove Theorem 1.1 and 1.3 for the PRG model. Using the coupling result in Proposition 2.3 we obtain Theorem 1.1 and 1.3 for all random graphs satisfying the assumptions **A1** and **A2**. As in the previous section, we denote by G_N a PRG.

We grow two NR-processes. Each NR-process starts from a uniformly chosen node $A_i \in \{1, 2, \ldots, N\}$, $i = 1, 2$, such that $A_1 \ne A_2$, **whp**.

Step 1: Expressing $\mathbb{P}(H_N > t)$ in capacities. We have $H_N > 1$ iff (if and only if) there are no edges between the nodes A_1 and A_2. Given the capacities $\underline{C}_1^{(1)}$ and $\underline{C}_1^{(2)}$, the number of edges between the nodes A_1 and A_2 has, according to Lemma 3.3, a Poisson distribution with mean

$$\frac{\underline{C}_1^{(1)} \underline{C}_1^{(2)}}{l_N}. \tag{4.1}$$

Therefore,

$$\mathbb{P}(H_N > 1) = \mathbb{E}\Big[\exp\Big\{-\frac{\underline{C}_1^{(1)}\underline{C}_1^{(2)}}{l_N}\Big\}\Big]. \tag{4.2}$$

We next inspect the capacity of the first generation of $\underline{Z}_1^{(1)}$, which is given by $\underline{C}_2^{(1)}$. Given $\underline{C}_2^{(1)}$ and $\underline{C}_1^{(2)}$, that is, the total capacity of the nodes in $\underline{Z}_1^{(1)}$ and the capacity of node A_2, we again have a Poisson number of edges between node A_2 and the nodes in $\underline{Z}_1^{(1)}$, however, this time with parameter

$$\frac{\underline{C}_2^{(1)}\underline{C}_1^{(2)}}{l_N}. \tag{4.3}$$

In order to compute the survival probability $\mathbb{P}(H_N > t)$ we need more notation. We write $\mathbb{Q}_C^{(t_1,t_2)}$ for the conditional probabilities given $\{\underline{C}_k^{(1)}\}_{k=1}^{t_1}$ and $\{\underline{C}_k^{(2)}\}_{k=1}^{t_2}$. We further write $\mathbb{E}_C^{(t_1,t_2)}$ for the expectation with respect to $\mathbb{Q}_C^{(t_1,t_2)}$. For $t_2 = 0$, we only condition on $\{\underline{C}_k^{(1)}\}_{k=1}^{t_1}$. Lemma 3.3 implies that

$$\mathbb{Q}_C^{(k,t)}(H_N > k+t-1|H_N > k+t-2) = \exp\Big\{-\frac{\underline{C}_k^{(1)}\underline{C}_t^{(2)}}{l_N}\Big\}.$$

Indeed, the event $\{H_N > k+t-2\}$ implies that $\mathcal{N}_k^{(1)} \cap \mathcal{N}_t^{(2)} = \varnothing$. From (4.1) and the above statement,

$$\begin{aligned}
\mathbb{P}(H_N > 2) &= \mathbb{E}\big[\mathbb{Q}_C^{(1,1)}(H_N > 1)\,\mathbb{Q}_C^{(1,1)}(H_N > 2|H_N > 1)\big] \\
&= \mathbb{E}\big[\mathbb{Q}_C^{(1,1)}(H_N > 1)\,\mathbb{E}_C^{(1,1)}\big[Q_C^{(2,1)}(H_N > 2|H_N > 1)\big]\big] \\
&= \mathbb{E}\big[\mathbb{E}_C^{(1,1)}\big[\mathbb{Q}_C^{(1,1)}(H_N > 1)\,\mathbb{Q}_C^{(2,1)}(H_N > 2|H_N > 1)\big]\big] \\
&= \mathbb{E}\Big[\exp\Big\{-\frac{\underline{C}_1^{(1)}\underline{C}_1^{(2)}}{l_N}\Big\}\cdot\exp\Big\{-\frac{\underline{C}_2^{(1)}\underline{C}_1^{(2)}}{l_N}\Big\}\Big] \\
&= \mathbb{E}\Big[\exp\Big\{-\frac{\sum_{k=2}^{3}\underline{C}_{\lceil k/2\rceil}^{(1)}\underline{C}_{\lfloor k/2\rfloor}^{(2)}}{l_N}\Big\}\Big].
\end{aligned}$$

By induction we obtain as in [13, Lemma 4.1],

$$\mathbb{P}(H_N > t) = \mathbb{E}\Big[\exp\Big\{-\frac{\sum_{k=2}^{t+1}\underline{C}_{\lceil k/2\rceil}^{(1)}\underline{C}_{\lfloor k/2\rfloor}^{(2)}}{l_N}\Big\}\Big]. \tag{4.4}$$

Note that (4.4) is an equality, while in [13] an error appeared.

Step 2: Coupling with the delayed BP. In this step we replace $\underline{C}_t^{(1)}$ and $\underline{C}_t^{(2)}$ by $\mathcal{Z}_t^{(1)}$ and $\mathcal{Z}_t^{(2)}$.

For each event $\mathcal{B}$, and any two nonnegative random variables V and W,

$$\big|\mathbb{E}\big[e^{-V}\big] - \mathbb{E}\big[e^{-W}\big]\big| \le \big|\mathbb{E}\big[(e^{-V} - e^{-W})\mathbf{1}_{\mathcal{B}}\big]\big| + \mathbb{P}\left(\mathcal{B}^c\right).$$

Now take

$$\mathcal{B} = \Big\{\frac{1}{N}\Big|\sum_{k=2}^{t+1}\mathcal{Z}_{\lceil k/2\rceil}^{(1)}\mathcal{Z}_{\lfloor k/2\rfloor}^{(2)} - \sum_{k=2}^{t+1}\underline{C}_{\lceil k/2\rceil}^{(1)}\underline{C}_{\lfloor k/2\rfloor}^{(2)}\Big| \le N^{-u_2}\Big\},$$

and the random variables V and W as

$$V = \frac{1}{N}\sum_{k=2}^{t+1}\mathcal{Z}_{\lceil k/2\rceil}^{(1)}\mathcal{Z}_{\lfloor k/2\rfloor}^{(2)}, \qquad W = \frac{1}{N}\sum_{k=2}^{t+1}\underline{C}_{\lceil k/2\rceil}^{(1)}\underline{C}_{\lfloor k/2\rfloor}^{(2)}.$$

Then, for $t \le t_N$, Proposition 3.4 implies that $\mathbb{P}(\mathcal{B}^c) = \mathcal{O}\big(N^{-v_2}\big)$, whereas on the event $\mathcal{B}$ we have $|V - W| \le N^{-u_2}$. Hence, using that $e^{-v} - e^{-w} = \mathcal{O}\big(v - w\big)$ when v, w are small, and that $e^{-v} \le 1$, $v \ge 0$, we obtain

$$\big|\mathbb{E}\big[e^{-V}\big] - \mathbb{E}\big[e^{-W}\big]\big| \le \mathcal{O}\big(N^{-u_2}\big)\mathbb{P}(\mathcal{B}) + \mathbb{P}(\mathcal{B}^c) = \mathcal{O}\big(N^{-u_2}\big) + \mathcal{O}\big(N^{-v_2}\big). \tag{4.5}$$

It is now clear from *Step* 1, the above result and by replacing l_N by $N\mu$, using (1.7), that for some $\beta > 0$ and all $t \leq t_N$,

$$\mathbb{P}(H_N > t) = \mathbb{E}\left[\exp\left\{-\frac{\sum_{k=2}^{t+1} \mathcal{Z}^{(1)}_{\lceil k/2\rceil}\mathcal{Z}^{(2)}_{\lfloor k/2\rfloor}}{\mu N}\right\}\right] + \mathcal{O}\big(N^{-\beta}\big). \tag{4.6}$$

Step 3: Evaluation of the limit points. From this step on, the remainder of the proof of our main theorem is identical to the proof of [13, Theorem 1.1]. To keep the proof self-contained, we finish the main line of the argument. Starting from (4.6) we replace t by $\sigma_N + t$ and assume that $\sigma_N + t \leq t_N$, where, as before, $\sigma_N = \lfloor \log_\nu N \rfloor$, to obtain

$$\mathbb{P}(H_N > \sigma_N + t) = \mathbb{E}\left[\exp\left\{-\frac{\sum_{k=2}^{\sigma_N+t+1} \mathcal{Z}^{(1)}_{\lceil k/2\rceil}\mathcal{Z}^{(2)}_{\lfloor k/2\rfloor}}{\mu N}\right\}\right] + \mathcal{O}\big(N^{-\beta}\big). \tag{4.7}$$

We write $N = \nu^{\log_\nu N} = \nu^{\sigma_N - a_N}$, where we recall that $a_N = \lfloor \log_\nu N \rfloor - \log_\nu N$. Then

$$\frac{\sum_{k=2}^{\sigma_N+t+1} \mathcal{Z}^{(1)}_{\lceil k/2\rceil}\mathcal{Z}^{(2)}_{\lfloor k/2\rfloor}}{\mu N} = \mu\nu^{a_N+t}\frac{\sum_{k=2}^{\sigma_N+t+1} \mathcal{Z}^{(1)}_{\lceil k/2\rceil}\mathcal{Z}^{(2)}_{\lfloor k/2\rfloor}}{\mu^2\nu^{\sigma_N+t}}.$$

In the above expression, the factor ν^{a_N} prevents proper convergence.

Without the factor $\mu\nu^{a_N+t}$, we obtain from [13, Appendix A.4] that, with probability 1,

$$\lim_{N\to\infty} \frac{\sum_{k=2}^{\sigma_N+t+1} \mathcal{Z}^{(1)}_{\lceil k/2\rceil}\mathcal{Z}^{(2)}_{\lfloor k/2\rfloor}}{\mu^2\nu^{\sigma_N+t}} = \frac{\mathcal{W}^{(1)}\mathcal{W}^{(2)}}{\nu - 1}, \tag{4.8}$$

We now use the speed of convergence result of [1], which was further developed in [13, Section 2] and which reads that there exists a positive β such that:

$$\mathbb{P}\Big(|\mathcal{W} - \mathcal{W}_k| > (\log N)^{-\alpha}\Big) = \mathcal{O}\big(N^{-\beta}\big), \qquad k \leq \lfloor \frac{1}{2}\log_\nu N \rfloor. \tag{4.9}$$

for each $\alpha > 0$. Combining (4.7) and (4.9) we obtain that for each $\alpha > 0$ and for $t \leq 2\eta \log_\nu N$,

$$\mathbb{P}(H_N > \sigma_N + t) = \mathbb{E}\big[\exp\{-\kappa\nu^{a_N+t}\mathcal{W}^{(1)}\mathcal{W}^{(2)}\}\big] + \mathcal{O}\big((\log N)^{-\alpha}\big). \tag{4.10}$$

From (4.10) one can finally derive as in [13], that, asymptotically as $N \to \infty$, the probability $\mathbb{P}(H_N < \infty)$ is equivalent to the probability $q^2 = \mathbb{P}(\mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0)$, where q is the survival probability of the branching process $\{\mathcal{Z}_t\}_{t\geq 0}$, so that (4.10) induces for $t \leq 2\eta \log_\nu N$,

$$\mathbb{P}(H_N \leq \sigma_N + t | H_N < \infty) = \mathbb{E}\big[1 - \exp\{-\kappa\nu^{a_N+t}\mathcal{W}^{(1)}\mathcal{W}^{(2)}\} | \mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0\big] + o(1). \tag{4.11}$$

A Appendix: Mixed Poisson distributions

In this section, we study mixed Poisson distributions.

Lemma A.1 *Consider a mixed Poisson distribution with mixing random variable Θ_f. Let, for $n = 0, 1, \ldots$,*

$$f_n = \mathbb{P}\left(Poi(\Theta_f) = n\right),$$

and consider the sequence $\{g_n\}_{n\geq 0}$, where

$$g_n = \frac{(n+1)f_{n+1}}{\mathbb{E}[\Theta_f]},$$

then $\{g_n\}_{n\geq 0}$ is a mixed Poisson distribution with mixing random variable Θ_g, where

$$\mathbb{P}\left(\Theta_g \leq x\right) = \frac{1}{\mathbb{E}[\Theta_f]}\mathbb{E}\Big[\Theta_f \mathbf{1}_{\{\Theta_f \leq x\}}\Big]. \tag{A.1}$$

Proof. We will assume that (A.1) holds and show that $\mathbb{P}\left(Poi(\Theta_g) = n\right) = g_n$. Let F_f and F_g be the distribution functions of Θ_f and Θ_g, respectively. Observe that

$$\frac{\mathrm{d}F_g(x)}{\mathrm{d}x} = \frac{x}{\mathbb{E}[\Theta_f]}\frac{\mathrm{d}F_f(x)}{\mathrm{d}x}.$$

Therefore,

$$\begin{aligned}\mathbb{P}\left(Poi(\Theta_g) = n\right) &= \int_0^\infty e^{-x}\frac{x^n}{n!}\,\mathrm{d}F_g(x) = \frac{1}{\mathbb{E}[\Theta_f]}\int_0^\infty e^{-x}\frac{x^{n+1}}{n!}\,\mathrm{d}F_f(x)\\ &= \frac{n+1}{\mathbb{E}[\Theta_f]}\int_0^\infty e^{-x}\frac{x^{n+1}}{(n+1)!}\,\mathrm{d}F_f(x) = \frac{(n+1)f_{n+1}}{\mathbb{E}[\Theta_f]} = g_n,\end{aligned}$$

which proves the claim. □

As an example, we consider the random variable $\Theta_f^{(N)}$ that takes the value λ_i, $i = 1, 2, \ldots, N$, with probability p_i, then, using Lemma A.1, $\Theta_g^{(N)}$ takes the value λ_i, $i = 1, 2, \ldots, N$, with probability

$$\mathbb{P}\left(\Theta_g^{(N)} = \lambda_i\right) = \frac{p_i\lambda_i}{\sum_{j=1}^N p_j\lambda_j}. \tag{A.2}$$

We close this section with a lemma that relates condition (**C3**) to a condition on the capacities $\{\lambda_i\}_{i=1}^N$:

Lemma A.2 (Moments of $\{g_n^{(N)}\}_{n\geq 0}$ and $\{\lambda_i\}_{i=1}^N$) *Let $q \geq 1$, and assume that (**C1**) holds. Then,*

$$\limsup_{N\to\infty}\sum_{n=0}^\infty n^q g_n^{(N)} < \infty \qquad \textit{if and only if} \qquad \limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \lambda_i^{q+1} < \infty. \tag{A.3}$$

Proof. Observe from (1.8), that

$$\sum_{n=0}^\infty n^q g_n^{(N)} = \frac{1}{N\mu_N}\sum_{i=1}^N \lambda_i \sum_{n=0}^\infty n^q e^{-\lambda_i}\frac{\lambda_i^n}{n!} = \frac{1}{N\mu_N}\sum_{i=1}^N \lambda_i \mathbb{E}\big[Poi(\lambda_i)^q\big]. \tag{A.4}$$

Since $q \geq 1$, by the Hölder inequality,

$$\mathbb{E}\big[Poi(\lambda_i)^q\big] \geq \mathbb{E}\big[Poi(\lambda_i)\big]^q = \lambda_i^q. \tag{A.5}$$

As a result, we obtain that

$$\frac{1}{N}\sum_{i=1}^{N}\lambda_i^{q+1} \leq \mu_N \sum_{n=0}^{\infty} n^q g_n^{(N)}. \tag{A.6}$$

By (**C1**), $\limsup_{N\to\infty}\mu_N = \mu < \infty$, so that indeed $\limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^{q+1} < \infty$ follows when $\limsup_{N\to\infty}\sum_{n=0}^{\infty} n^q g_n^{(N)} < \infty$. This proves the first implication.

For the second implication, we make use of the fact that there exists a constant $c > 0$ such that, uniformly for all $\lambda \geq 0$, $\mathbb{E}\big[Poi(\lambda)^q\big] \leq c(\lambda^q + 1)$. Then, by (A.4), we have that

$$\sum_{n=0}^{\infty} n^q g_n^{(N)} \leq \frac{c}{N\mu_N}\sum_{i=1}^{N}\lambda_i(\lambda_i^q + 1) = c + \frac{c}{\mu_N}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^{q+1}. \tag{A.7}$$

By (**C1**), $\liminf_{N\to\infty}\mu_N = \mu > 0$, so that when $\limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^{q+1} < \infty$, we obtain

$$\limsup_{N\to\infty}\sum_{n=0}^{\infty} n^q g_n^{(N)} \leq c + \frac{c}{\mu}\limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^{q+1} < \infty. \tag{A.8}$$

This proves the second implication, and hence completes the proof. □

B Appendix: Deterministic capacities

Let F be any distribution function, and write $\bar{F}(x) = 1 - F(x)$ for the survival function of F. Define the inverse $\bar{F}^{-1}(u)$, for $u \in (0,1)$, by

$$\bar{F}^{-1}(u) = \inf\{s : \bar{F}(s) \leq u\}, \tag{B.1}$$

where the definition is chosen such that

$$\bar{F}^{-1}(1-u) = F^{-1}(u) = \inf\{s : F(s) \geq u\}. \tag{B.2}$$

We shall often make use of (B.2), in particular since it implies that $\bar{F}^{-1}(U)$ has distribution function F when U is uniform on $(0,1)$. Throughout this section, we will use the abbreviations $u_N = \lceil uN\rceil/N$ and $U_N = \lceil UN\rceil/N$.

In this section, we let the capacities $\{\lambda_i\}_{i=1}^N$ be given by

$$\lambda_i = \bar{F}^{-1}(i/N), \qquad i = 1, 2, \ldots, N, \tag{B.3}$$

where $F\colon [0,\infty) \mapsto [0,1]$ is a distribution function satisfying (1.18). We shall verify that conditions (**C1**)–(**C3**) hold under appropriate conditions on F, the main condition being (1.18), but we shall need one more technical assumption.

We let the random variable $\Theta_f^{(N)}$ takes the value λ_i, $i = 1, 2, \ldots, N$, with probability $p_i = \frac{1}{N}$, and note that $\Theta_f^{(N)}$ is equal in distribution to $\bar{F}^{-1}(U_N)$, where U has a uniform $(0,1)$ distribution. We start by identifying $f, f^{(N)}$, g and $g^{(N)}$. We observe that, in this setting,

$$f_n^{(N)} = \mathbb{E}[\mathbb{P}(Poi(\Theta_f^{(N)}) = n)] = \mathbb{E}\left[e^{-\bar{F}^{-1}(U_N)}\frac{\bar{F}^{-1}(U_N)^n}{n!}\right].$$

We define

$$f_n = \mathbb{E}\left[e^{-\bar{F}^{-1}(U)}\frac{\bar{F}^{-1}(U)^n}{n!}\right] = \mathbb{E}[\mathbb{P}(Poi(\Theta_f) = n)], \tag{B.4}$$

where $\Theta_f = \bar{F}^{-1}(U)$. Observe that we can use the same U as in the definition of $\Theta_f^{(N)}$, and doing so, implicitly introduces a coupling between the random variables Θ_f and $\Theta_f^{(N)}$. Then, using that $\Theta_f \geq \Theta_f^{(N)}$ a.s., since $u \mapsto \bar{F}^{-1}(u)$ is non-increasing,

$$\begin{aligned}\frac{1}{2}\sum_{n=1}^{\infty} |f_n^{(N)} - f_n| &= \mathrm{d}_{\mathrm{TV}}(Poi(\Theta_f^{(N)}), Poi(\Theta_f)) \leq \mathbb{P}(Poi(\Theta_f^{(N)}) \neq Poi(\Theta_f)) \\ &= \mathbb{P}(Poi(\Theta_f - \Theta_f^{(N)}) \geq 1) \leq \mathbb{E}[\Theta_f - \Theta_f^{(N)}] = \mathbb{E}\left[\bar{F}^{-1}(U) - \bar{F}^{-1}\big(U_N\big)\right], \end{aligned} \tag{B.5}$$

where the first inequality in the above chain is the coupling inequality, the second the Markov inequality.

We first investigate the convergence of the integrals $\mathbb{E}\big[(\Theta_f^{(N)})^i\big]$ to $\mathbb{E}\big[(\Theta_f)^i\big]$, for $i = 1, 2$, and then prove that this is sufficient for conditions (**C1**)–(**C3**).

Lemma B.1 (Convergence of moments of $\Theta_f^{(N)}$) *Let $\Theta_f^{(N)} = \bar{F}^{-1}\big(U_N\big)$, and assume that F satisfies* (1.18). *Furthermore, assume that F is a continuous distribution function with density $f\colon [0,\infty) \to [0,\infty)$, and that there exist an $a > 0$ such that, as $y \downarrow 0$,*

$$\int_y^1 \frac{\bar{F}^{-1}(u)}{\underline{f}(\bar{F}^{-1}(u))}\,\mathrm{d}u = \mathcal{O}\big(y^{-a}\big), \tag{B.6}$$

where $\underline{f}(x) = \inf_{0\leq y\leq x} f(y)$. Then, there exists an $\alpha > 0$ such that

$$\int_0^1 \bar{F}^{-1}(u)\left(\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\right)\,\mathrm{d}u = \mathcal{O}\big(N^{-\alpha}\big), \tag{B.7}$$

and, for $i = 1, 2$,

$$\int_0^1 \bar{F}^{-1}(u)^i - \bar{F}^{-1}(u_N)^i\,\mathrm{d}u = \mathcal{O}\big(N^{-\alpha}\big). \tag{B.8}$$

The proof of Lemma B.1 is followed by an example of cases where the conditions are satisfied.

Proof of Lemma B.1: We first claim that (1.18) implies that for $y \in (0, 1)$,

$$\bar{F}^{-1}(y) \leq c^b y^{-b}, \tag{B.9}$$

where $b = 1/(\tau - 1)$. We show the claim (B.9) by contradiction. Suppose that there is $y \in (0, 1)$ satisfying $\bar{F}^{-1}(y) > c^b y^{-b}$. Observe that if $\bar{F}^{-1}(y) = w$, then $\bar{F}(x) > y$ for each $x < w$, by definition (B.1). Then, with $x = c^b y^{-b} < \bar{F}^{-1}(y) = w$,

$$\bar{F}(x) > y = cx^{1-\tau}, \tag{B.10}$$

which contradicts (1.18). This proves (B.9).

We shall now simultaneously prove (B.7), and (B.8) for $i = 2$. The claim in (B.8) for $i = 1$ is similar, and in fact easier to prove. When $i = 2$, we can rewrite and bound

$$\int_0^1 \bar{F}^{-1}(u)^2 - \bar{F}^{-1}(u_N)^2\,\mathrm{d}u \leq 2\int_0^1 \bar{F}^{-1}(u)\big(\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\big)\,\mathrm{d}u, \tag{B.11}$$

where we have used the special product $x^2 - y^2 = (x - y)(x + y)$ and that $\bar{F}^{-1}(u_N) \leq \bar{F}^{-1}(u)$, for all $u \in (0, 1)$. Thus, (B.8) for $i = 2$ follows from (B.7). We split the integral into $u \in (0, N^{-d})$ and $u \in [N^{-d}, 1)$, for an appropriate $d > 0$. For the former, we bound using (B.9),

$$\int_0^{N^{-d}} \bar{F}^{-1}(u)\big(\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\big)\,\mathrm{d}u \leq \int_0^{N^{-d}} \bar{F}^{-1}(u)^2\,\mathrm{d}u \leq c^{2b}\int_0^{N^{-d}} u^{-2b}\,\mathrm{d}u = \mathcal{O}\big(N^{-d(1-2b)}\big),$$

where we used that $b = 1/(\tau - 1) < 1/2$, since $\tau > 3$. For the integral over $u \in [N^{-d}, 1)$, we note by Taylor's theorem that

$$\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N) = -(u - u_N)\frac{\mathrm{d}}{\mathrm{d}u}\bar{F}^{-1}(u_N^*) = \frac{u_N - u}{f(\bar{F}^{-1}(u_N^*))}, \tag{B.12}$$

for some $u_N^* \in [u, u_N]$. Since $u \mapsto \bar{F}^{-1}(u)$ is non-increasing, we have that $\bar{F}^{-1}(u_N^*) \leq \bar{F}^{-1}(u)$, so that

$$\frac{u_N - u}{f(\bar{F}^{-1}(u_N^*))} \leq \frac{1}{N\underline{f}(\bar{F}^{-1}(u))}. \tag{B.13}$$

Thus, we arrive at

$$\int_{N^{-d}}^{1} \bar{F}^{-1}(u)\big(\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\big)\,\mathrm{d}u \leq \frac{1}{N}\int_{N^{-d}}^{1} \frac{\bar{F}^{-1}(u)}{\underline{f}(\bar{F}^{-1}(u))}\,\mathrm{d}u = \mathcal{O}\big(N^{-1+ad}\big), \tag{B.14}$$

where, in the last inequality, we have made use of (B.6). We conclude that

$$\int_{0}^{1} \bar{F}^{-1}(u)\big(\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\big)\,\mathrm{d}u \leq \mathcal{O}\big(N^{-d(1-2b)}\big) + \mathcal{O}\big(N^{ad-1}\big) = \mathcal{O}\big(N^{-\alpha}\big), \tag{B.15}$$

where $\alpha = \min\{d(1-2b), 1-ad\} > 0$, provided that we pick $d > 0$ so small that $1 - ad > 0$. □

We now present the promised example. Consider the class of continuous distribution functions F satisfying (1.18) for which $f(x) = F'(x)$ is positive and decreasing on $[0, \infty)$. We shall prove that such F indeed satisfy (B.6). We will use (B.9) to prove (B.6) when $f(x) = F'(x)$ is positive and decreasing on $[0, \infty)$. Then, for such f, $\underline{f}(x) = f(x)$. In this case, we note that

$$2\frac{\bar{F}^{-1}(u)}{f(\bar{F}^{-1}(u))} = -\frac{\mathrm{d}}{\mathrm{d}u}\bar{F}^{-1}(u)^2, \tag{B.16}$$

so that, by (B.9) with $b = 1/(\tau - 1)$,

$$2\int_{y}^{1} \frac{\bar{F}^{-1}(u)}{\underline{f}(\bar{F}^{-1}(u))}\,\mathrm{d}u = \bar{F}^{-1}(y)^2 \leq c^{2b}y^{-2b} = \mathcal{O}\big(y^{-2b}\big), \tag{B.17}$$

so that (B.6) is satisfied with $a = 2b$. This shows in particular that for

$$\bar{F}_\Lambda(x) = x^{-\tau+1}, \quad x \geq 1, \tag{B.18}$$

the assumptions of Lemma B.1 are fulfilled.

We will now show that the conditions (**C1**)–(**C3**) do hold for the weights $\{\lambda_i\}_{i=1}^N = \bar{F}^{-1}(i/N)$ with F satisfying (1.18) and (B.6). We start by validating the conditions (**C1**) and (**C3**), since for condition (**C2**) more effort is needed.

Condition (C1): Combining (1.6) with (B.3), we rewrite μ_N as

$$\mu_N = \frac{1}{N}\sum_{i=1}^{N} \bar{F}^{-1}(i/N) = \int_0^1 \bar{F}^{-1}\left(u_N\right)\,\mathrm{d}u, \tag{B.19}$$

where, as before, $u_N = \frac{\lceil Nu \rceil}{N}$. Similarly, one can show that

$$\nu_N = \frac{1}{\mu_N}\int_0^1 \bar{F}^{-1}\left(u_N\right)^2\,\mathrm{d}u. \tag{B.20}$$

Therefore, we choose

$$\mu = \int_0^1 \bar{F}^{-1}(u)\, \mathrm{d}u \qquad\qquad \text{and} \qquad\qquad \nu = \frac{1}{\mu}\int_0^1 \bar{F}^{-1}(u)^2\, \mathrm{d}u. \tag{B.21}$$

Effectively, we have replaced u_N in (B.19) and (B.20) by u. Observe that the following also hold

$$\mu = \mathbb{E}\big[\Theta_f\big], \quad \mu_N = \mathbb{E}\big[\Theta_f^{(N)}\big], \quad \nu = \mathbb{E}\big[\Theta_f^2\big]/\mathbb{E}\big[\Theta_f\big] \quad \text{and} \quad \nu_N = \mathbb{E}\big[(\Theta_f^{(N)})^2\big]/\mathbb{E}\big[\Theta_f^{(N)}\big]. \tag{B.22}$$

Condition (**C1**) is satisfied, since Lemma B.1, where we take $i = 1$, implies the existence of a constant $\alpha_1 > 0$ such that

$$|\mu_N - \mu| \le \int_0^1 \left|\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\right| \mathrm{d}u = \mathcal{O}\big(N^{-\alpha_1}\big). \tag{B.23}$$

Using (B.20) and (B.21), we obtain that

$$\begin{aligned}
|\nu_N - \nu| &= \left|\frac{1}{\mu_N}\int_0^1 \bar{F}^{-1}(u_N)^2\, \mathrm{d}u - \frac{1}{\mu}\int_0^1 \bar{F}^{-1}(u)^2\, \mathrm{d}u\right| \\
&\le \frac{1}{\mu_N}\int_0^1 \left|\bar{F}^{-1}(u_N)^2 - F^{-1}(u_N)^2\right| \mathrm{d}u + \left(\frac{1}{\mu_N} - \frac{1}{\mu}\right)\int_0^1 \bar{F}^{-1}(u)^2\, \mathrm{d}u.
\end{aligned}$$

We bound the above display, using Lemma B.1 and (B.23), by

$$|\nu_N - \nu| \le \frac{1}{\mu - N^{-\alpha_1}} N^{-\alpha_1} + \left(\frac{1}{\mu - N^{-\alpha_1}} - \frac{1}{\mu}\right)\int_0^1 \bar{F}^{-1}(u)^2\, \mathrm{d}u = \mathcal{O}\big(N^{-\alpha_1}\big),$$

since

$$\int_0^1 \bar{F}^{-1}(u)^2\, \mathrm{d}u = \mathbb{E}\Big[\bar{F}^{-1}(U)^2\Big] = \mathbb{E}\big[\Theta_f^2\big] < \infty, \tag{B.24}$$

where $\Theta_f = \bar{F}^{-1}(U)$ is a random variable with distribution F satisfying (1.18).

Condition (C3): Observe that, again using that $u \mapsto \bar{F}^{-1}(u)$ is non-increasing,

$$\frac{1}{N}\sum_{i=1}^N \lambda_i^{\tau-1-\varepsilon} = \frac{1}{N}\sum_{i=1}^N \bar{F}^{-1}(i/N)^{\tau-1-\varepsilon} = \mathbb{E}\big[\bar{F}^{-1}(U_N)^{\tau-1-\varepsilon}\big] \le \mathbb{E}\big[\bar{F}^{-1}(U)^{\tau-1-\varepsilon}\big],$$

and $\bar{F}^{-1}(U)$ is a random variable with distribution F which satisfies (1.18), compare with (B.24). Therefore $\mathbb{E}\big[\bar{F}^{-1}(U)^{\tau-1-\varepsilon}\big] < \infty$, which in turn implies that

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \lambda_i^{\tau-1-\varepsilon} < \infty.$$

Since, for fixed N, we have that $\bar{F}^{-1}(i/N)$ is non-increasing in i, we obtain from (B.9),

$$\lambda_N^{(N)} = \max_{1\le i\le N} \bar{F}^{-1}(i/N) = \bar{F}^{-1}(1/N) = \mathcal{O}\big(N^{1/(1-\tau)}\big) \le N^{\gamma}.$$

Condition (C2): We bound (B.5) by Lemma B.1: there exists a constant $\alpha_2 > 0$ such that

$$\mathrm{d}_{\mathrm{TV}}(Poi(\Theta_f^{(N)}), Poi(\Theta_f)) \le \mathbb{E}\left[\bar{F}^{-1}(U) - \bar{F}^{-1}(U_N)\right] = \mathcal{O}\big(N^{-\alpha_2}\big).$$

Similarly to (B.5), we obtain

$$\frac{1}{2}\sum_{n=0}^{\infty}|g_n^{(N)}-g_n| = \mathrm{d}_{\mathrm{TV}}(Poi\big(\Theta_g^{(N)}\big)\,, Poi(\Theta_g))$$
$$\leq \mathbb{E}[|\Theta_g-\Theta_g^{(N)}|] = \mathbb{E}\left[\left|\bar{F}_{\Theta_g}^{-1}(U)-\bar{F}_{\Theta_g^{(N)}}^{-1}(U)\right|\right]. \tag{B.25}$$

We claim that for two distribution functions $G\colon \mathbb{R}\mapsto[0,1]$ and $H\colon \mathbb{R}\mapsto[0,1]$ with $G(0)=H(0)=0$, we have

$$\int_0^1 \left|\bar{G}^{-1}(u)-\bar{H}^{-1}(u)\right|\,\mathrm{d}u = \int_0^\infty \left|\bar{G}(x)-\bar{H}(x)\right|\,\mathrm{d}x. \tag{B.26}$$

To see this, note that when $\bar{G}(x)\geq\bar{H}(x)$ for every $x\in\mathbb{R}$, so that $\bar{G}(x)-\bar{H}(x)\geq 0, \bar{G}^{-1}(u)-\bar{H}^{-1}(u)\geq 0$, then both sides of (B.26) are equal to $\mathbb{E}[X]-\mathbb{E}[Y]$, where X and Y have distribution functions G and H, respectively. This proves the claim for G and H for which $\bar{G}(x)\geq\bar{H}(x)$ for every $x\in\mathbb{R}$. For other G and H, we write, for $x\in[0,\infty)$,

$$(\bar{G}\vee\bar{H})(x)=\max\{\bar{G}(x),\bar{H}(x)\}, \qquad (\bar{G}\wedge\bar{H})(x)=\min\{\bar{G}(x),\bar{H}(x)\}, \tag{B.27}$$

so that

$$\int_0^\infty \left|\bar{G}(x)-\bar{H}(x)\right|\,\mathrm{d}x = \int_0^\infty (\bar{G}\vee\bar{H})(x)-(\bar{G}\wedge\bar{H})(x)\,\mathrm{d}x. \tag{B.28}$$

Then, both $\bar{G}\wedge\bar{H}$ and $\bar{G}\vee\bar{H}$ are survival functions for which $(\bar{G}\vee\bar{H})(x)\geq(\bar{G}\wedge\bar{H})(x)$ for every $x\in\mathbb{R}$. Thus, by the above proof, (B.26) holds for them. The claim is then completed by noting that

$$(\bar{G}\vee\bar{H})^{-1}(u)=\inf\{s:\max\{\bar{G}(s),\bar{H}(s)\}\leq u\}=\max\{\bar{G}^{-1}(u),\bar{H}^{-1}(u)\}, \tag{B.29}$$

and

$$(\bar{G}\wedge\bar{H})^{-1}(u)=\inf\{s:\min\{\bar{G}(s),\bar{H}(s)\}\leq u\}=\min\{\bar{G}^{-1}(u),\bar{H}^{-1}(u)\} \tag{B.30}$$

so that

$$(\bar{G}\vee\bar{H})^{-1}(u)-(\bar{G}\wedge\bar{H})^{-1}(u)=\left|\bar{G}^{-1}(u)-\bar{H}^{-1}(u)\right|. \tag{B.31}$$

This completes the proof of (B.26).

Using (B.26), we can rewrite display (B.25) as

$$\mathrm{d}_{\mathrm{TV}}(Poi(\Theta_g^{(N)}), Poi(\Theta_g)) \leq \int_0^1 \left|\bar{F}_{\Theta_g}^{-1}(u)-\bar{F}_{\Theta_g^{(N)}}^{-1}(u)\right|\mathrm{d}u = \int_0^\infty \left|\bar{F}_{\Theta_g}(x)-\bar{F}_{\Theta_g^{(N)}}(x)\right|\,\mathrm{d}x. \tag{B.32}$$

Recall that $\mu=\mathbb{E}[\Theta_f]$ and $\mu_N=\mathbb{E}\left[\Theta_f^{(N)}\right]$, then, using Lemma A.1, the integrand can be rewritten as

$$\begin{aligned}\left|\bar{F}_{\Theta_g}(x)-\bar{F}_{\Theta_g^{(N)}}(x)\right| &= \left|\mathbb{E}\left[\frac{\Theta_f}{\mu}\mathbf{1}_{\{\Theta_f>x\}}-\frac{\Theta_f^{(N)}}{\mu_N}\mathbf{1}_{\left\{\Theta_f^{(N)}>x\right\}}\right]\right|\\ &\leq \mathbb{E}\left[\left|\frac{\bar{F}^{-1}(U)}{\mu}\mathbf{1}_{\{\bar{F}^{-1}(U)>x\}}-\frac{\bar{F}^{-1}(U_N)}{\mu_N}\mathbf{1}_{\{\bar{F}^{-1}(U_N>x\}}\right|\right]\\ &= \int_0^1\left|\frac{\bar{F}^{-1}(u)}{\mu}\mathbf{1}_{\{\bar{F}^{-1}(u)>x\}}-\frac{\bar{F}^{-1}(u_N)}{\mu_N}\mathbf{1}_{\{\bar{F}^{-1}(u_N)>x\}}\right|\,\mathrm{d}u. \end{aligned}\tag{B.33}$$

Using (B.32), (B.33), and the triangle inequality twice, we obtain

$$\begin{aligned} \mathrm{d}_{\mathrm{TV}}(Poi(\Theta_g^{(N)}), Poi(\Theta_g)) \leq & \left|\frac{1}{\mu} - \frac{1}{\mu_N}\right| \int_0^\infty \int_0^1 \bar{F}^{-1}(u) \mathbf{1}_{\{\bar{F}^{-1}(u) > x\}} \,\mathrm{d}u\,\mathrm{d}x \\ & + \frac{1}{\mu_N} \int_0^\infty \int_0^1 \bar{F}^{-1}(u) \left|\mathbf{1}_{\{\bar{F}^{-1}(u)>x\}} - \mathbf{1}_{\{\bar{F}^{-1}(u_N)>x\}}\right| \mathrm{d}u\,\mathrm{d}x \\ & + \frac{1}{\mu_N} \int_0^\infty \int_0^1 \left|\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\right| \mathbf{1}_{\{\bar{F}^{-1}(u_N)>x\}} \,\mathrm{d}u\,\mathrm{d}x. \end{aligned} \tag{B.34}$$

Next, we will bound each double integral in display (B.34). The first double integral is finite, since

$$\int_0^\infty \int_0^1 \bar{F}^{-1}(u) \mathbf{1}_{\{\bar{F}^{-1}(u)>x\}} \,\mathrm{d}u\,\mathrm{d}x = \int_0^1 \bar{F}^{-1}(u)^2 \,\mathrm{d}u = \mathbb{E}\big[\Theta_f^2\big] < \infty, \tag{B.35}$$

where we used in the last step that the distribution F satisfies (1.18). Observe that $\mathbf{1}_{\{\bar{F}^{-1}(u)>x\}} \neq \mathbf{1}_{\{\bar{F}^{-1}(u_N)>x\}}$ if and only if $\bar{F}^{-1}(u_N) < x < \bar{F}^{-1}(u)$, thus

$$\int_0^\infty \int_0^1 \bar{F}^{-1}(u) \left|\mathbf{1}_{\{\bar{F}^{-1}(u)>x\}} - \mathbf{1}_{\{\bar{F}^{-1}(u_N)>x\}}\right| \mathrm{d}u\,\mathrm{d}x = \int_0^1 \bar{F}^{-1}(u) \left(\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\right) \mathrm{d}u. \tag{B.36}$$

Using (B.7) once more, it follows that the right hand side of (B.36) is $\mathcal{O}\big(N^{-\alpha}\big)$. Finally, using again (B.7),

$$\begin{aligned} \int_0^\infty \int_0^1 \left(\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\right) \mathbf{1}_{\{\bar{F}^{-1}(u_N)>x\}} \,\mathrm{d}u\,\mathrm{d}x &= \int_0^1 \left(\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\right) \bar{F}^{-1}(u_N) \,\mathrm{d}u \\ &\leq \int_0^1 \bar{F}^{-1}(u) \left(\bar{F}^{-1}(u) - \bar{F}^{-1}(u_N)\right) \mathrm{d}u = \mathcal{O}\big(N^{-\alpha}\big). \end{aligned} \tag{B.37}$$

Combining (B.35)–(B.37) with (B.34), we obtain that

$$\mathrm{d}_{\mathrm{TV}}(Poi(\Theta_g^{(N)}), Poi(\Theta_g)) = \mathcal{O}\big(\frac{1}{\mu} - \frac{1}{\mu_N}\big) + \frac{1}{\mu_N}\mathcal{O}\big(N^{-\alpha}\big) + \frac{1}{\mu_N}\mathcal{O}\big(N^{-\alpha}\big) = \mathcal{O}\big(N^{-\alpha_3}\big),$$

where we used that $|\mu - \mu_N| = \mathcal{O}\big(N^{-\alpha_1}\big)$ and take $\alpha_3 = \min\{\alpha, \alpha_1\}$.

C Appendix: i.i.d. capacities

In this section, we prove Theorem 1.4.

We associate a sequence $\{\Lambda_i\}_{i=1}^N$ of positive i.i.d. random variables to the nodes, where the random capacities $\{\Lambda_i\}_{i=1}^N$ have distribution function $F_\Lambda(x) = \mathbb{P}(\Lambda \leq x)$. Furthermore, let μ, ν, f_n, g_n be as defined in (1.17), and $\mu_N, \nu_N, f_n^{(N)}, g_n^{(N)}$ as in definition (C.5), (C.6), below. Define

$$\mathcal{J} = \mathcal{J}(\alpha) = \mathcal{J}_1(\alpha) \cap \mathcal{J}_2(\alpha) \cap \mathcal{J}_3, \tag{C.1}$$

where

$$\mathcal{J}_1(\alpha) = \big\{|\mu_N - \mu| < N^{-\alpha}\big\} \cap \big\{|\nu_N - \nu| < N^{-\alpha}\big\}, \tag{C.2}$$

$$\mathcal{J}_2(\alpha) = \big\{\mathrm{d}_{\mathrm{TV}}(f^{(N)}, f) < N^{-\alpha}\big\} \cap \big\{\mathrm{d}_{\mathrm{TV}}(g^{(N)}, g) < N^{-\alpha}\big\}, \tag{C.3}$$

and, for some $B > 0$ sufficiently large,

$$\mathcal{J}_3 = \left\{\sum_{n=0}^\infty n^{\tau-2-\varepsilon} g_n^{(N)} \leq B\right\} \cap \big\{\Lambda_N^{(N)} \leq N^\gamma\big\}, \tag{C.4}$$

and where the exponent γ is equal to $1/(\tau-1)+\varepsilon$ (recall (1.12)).

In Sections C.1–C.3 below, we will show that, when (1.18) holds, there exist constants

$$\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3 > 0,$$

such that the events $\mathcal{J}_1(\alpha_1)$, $\mathcal{J}_2(\alpha_2)$, $\mathcal{J}_3$ occur with probability exceeding $1-\mathcal{O}(N^{-\beta_1})$, $1-\mathcal{O}(N^{-\beta_2})$ and $1-\mathcal{O}(N^{-\beta_3})$, respectively. Then, by taking $\alpha=\min\{\alpha_1,\alpha_2\}$ and $\beta=\min\{\beta_1,\beta_2,\beta_3\}$, the event $\mathcal{J}(\alpha)$ occurs with probability exceeding $1-\mathcal{O}(N^{-\beta})$. On the event $\mathcal{J}(\alpha)$ the conditions (**C1**)–(**C3**) do hold, and this then proves Theorem 1.4.

For the sake of completeness, we give the counterparts of (1.6) and (1.8) using the random capacities. Display (1.6) becomes

$$\mu_N = \frac{1}{N}\sum_{i=1}^{N} \Lambda_i \qquad \text{and} \qquad \nu_N = \frac{1}{\mu_N N}\sum_{i=1}^{N}\Lambda_i^2, \tag{C.5}$$

Using the law of large number the values μ_N and ν_N converge a.s. to μ and ν in (1.17), respectively.

The offspring distribution of the first and second generation is given by (compare (1.8)),

$$f_n^{(N)} = \frac{1}{N}\sum_{i=1}^{N} e^{-\Lambda_i}\frac{\Lambda_i^n}{n!} \qquad \text{and} \qquad g_n^{(N)} = \frac{1}{\mu_N N}\sum_{i=1}^{N} e^{-\Lambda_i}\frac{\Lambda_i^{n+1}}{n!} = \frac{(n+1)f_{n+1}^{(N)}}{\mu_N}, \tag{C.6}$$

which, by the strong law of large numbers, converge a.s. to f and g in (1.17), respectively.

C.1 Convergence of means (C1):

In this section we will show that (**C1**) holds **whp**, i.e, that there exist constants $\alpha_1,\beta_1>0$, such that the event

$$\mathcal{J}_1(\alpha_1) = \left\{|\mu_N-\mu| < N^{-\alpha_1}\right\} \cap \left\{|\nu_N-\nu| < N^{-\alpha_1}\right\},$$

occurs with probability exceeding $1-\mathcal{O}(N^{-\beta_1})$, where μ and ν are defined in (1.17).

The next lemma is the crucial estimate of the proof:

Lemma C.1 *Let $q\in(0,\tau-1)$, and define*

$$S_{N,q} = \frac{1}{N}\sum_{i=1}^{N}\Lambda_i^q \qquad \text{and} \qquad \mathcal{S}_{N,q}(u) = \left\{|S_{N,q}-\mathbb{E}[\Lambda^q]| \le N^{-u}\right\},$$

then there exist constants $u,v>0$ such that

$$\mathbb{P}\left(\mathcal{S}_{N,q}(u)\right) = 1-\mathcal{O}(N^{-v}). \tag{C.7}$$

Proof. Apply the Marcienkiewicz-Zygmund inequality [12, Chapter 3; Corollary 8.2], with $X_i = \Lambda_i^q - \mathbb{E}[\Lambda_i^q]$ and $r\in(1,2)$, such that $qr<\tau-1$. Then,

$$\begin{aligned}
\mathbb{P}\left(\mathcal{S}_{N,q}(u)^c\right) &= \mathbb{P}\left(\Big|\frac{1}{N}\sum_{i=1}^{N}X_i\Big| > N^{-u}\right) = \mathbb{P}\left(\Big|\frac{1}{N}\sum_{i=1}^{N}X_i\Big|^r > N^{-ur}\right)\\
&= \mathbb{P}\left(\Big|\sum_{i=1}^{N}X_i\Big|^r > N^{r(1-u)}\right) \le N^{-r(1-u)}c_r N\mathbb{E}[|X_1|^r] = \mathcal{O}(N^{1-r(1-u)}),
\end{aligned}$$

since $\mathbb{E}[|X_1|^r] \le \mathbb{E}[|\Lambda_1|^{qr}] < \infty$. Now recall that $r>1$ and fix $u>0$ so small that $v=r(1-u)-1>0$. □

Using Lemma C.1, we pick the constants $u, v > 0$ such that

$$\mathbb{P}\left(\mathcal{S}_{N,1}(u) \cap \mathcal{S}_{N,2}(u)\right) = 1 - \mathcal{O}\big(N^{-v}\big).$$

Observe that $\mu_N = S_{N,1}$ and $\nu_N = S_{N,2}/S_{N,1}$. On the event $\mathcal{S}_{N,1}(u) \cap \mathcal{S}_{N,2}(u)$, we have that

$$|\mu_N - \mu| = |S_{N,1} - \mathbb{E}[\Lambda]\,| = \mathcal{O}\big(N^{-u}\big),$$

and

$$\begin{aligned}|\nu_N - \nu| &= \left|\frac{S_{N,2}}{S_{N,1}} - \frac{\mathbb{E}[\Lambda^2]}{\mathbb{E}[\Lambda]}\right| \leq S_{N,2}\left|\frac{1}{S_{N,1}} - \frac{1}{\mathbb{E}[\Lambda]}\right| + \frac{|S_{N,2} - \mathbb{E}[\Lambda^2]\,|}{\mathbb{E}[\Lambda]} = \frac{S_{N,2}|\mathbb{E}[\Lambda] - S_{N,1}|}{\mathbb{E}[\Lambda]\, S_{N,1}} + \mathcal{O}\big(N^{-u}\big) \\ &= \frac{\big(\mathbb{E}[\Lambda^2] + \mathcal{O}\big(N^{-u}\big)\big)\,\mathcal{O}\big(N^{-u}\big)}{\mathbb{E}[\Lambda]\,(\mathbb{E}[\Lambda] - \mathcal{O}\big(N^{-u}\big))} + \mathcal{O}\big(N^{-u}\big) = \mathcal{O}\big(N^{-u}\big).\end{aligned} \tag{C.8}$$

Therefore, we can pick $\alpha_1 = u$ and $\beta_1 = v$. □

C.2 Convergence of branching processes (C2):

In this section we will show that there exist constants $\alpha_2, \beta_2 > 0$, such that

$$\mathbb{P}\left(\mathrm{d_{TV}}(g^{(N)}, g) > N^{-\alpha_2}\right) = \mathcal{O}\big(N^{-\beta_2}\big). \tag{C.9}$$

The proof of the existence of constants $\alpha_2, \beta_2 > 0$, such that $\mathbb{P}\left(\mathrm{d_{TV}}(f^{(N)}, f) > N^{-\alpha_2}\right) = \mathcal{O}\big(N^{-\beta_2}\big)$ is less difficult and therefore omitted.

In this section, we add a subscript N to $\mathbb{P}$ and $\mathbb{E}$, to make clear that we condition on the i.i.d. capacities $\{\Lambda_i\}_{i=1}^N$. Thus,

$$\mathbb{P}_N(\,\cdot\,) = \mathbb{P}\left(\,\cdot\,|\Lambda_1, \ldots, \Lambda_N\right) \qquad \text{and} \qquad \mathbb{E}_N[\,\cdot\,] = \mathbb{E}[\,\cdot\,|\Lambda_1, \ldots, \Lambda_N]\,. \tag{C.10}$$

Using Lemma A.1 with $\{f_n^{(N)}\}_{n\geq 0}$ and $\{f_n\}_{n\geq 0}$, it follows that $g_n^{(N)} = \mathbb{P}_N\big(Poi\big(\Theta_g^{(N)}\big) = n\big)$, where, for $i = 1, 2, \ldots, N$,

$$\mathbb{P}_N\big(\Theta_g^{(N)} = \Lambda_i\big) = \frac{\Lambda_i}{N\mu_N},$$

and $g_n = \mathbb{P}\left(Poi(\Theta_g) = n\right)$, where Θ_g is given by

$$F_{\Theta_g}(x) = \frac{1}{\mu}\mathbb{E}\big[\Lambda \mathbf{1}_{\{\Lambda \leq x\}}\big]\,. \tag{C.11}$$

Let U be uniformly distributed over $(0, 1)$, then we couple $\Theta_g^{(N)}$ and $Poi(\Theta_g)$ by letting $\Theta_g = \bar{F}_{\Theta_g}^{-1}(U)$ and $\Theta_g^{(N)} = \bar{F}_{\Theta_g^{(N)}}^{-1}(U)$. Arguing as in (B.5), we find that,

$$\mathrm{d_{TV}}(g^{(N)}, g) = \mathrm{d_{TV}}(Poi\big(\Theta_g^{(N)}\big)\,, Poi(\Theta_g)) \leq \mathbb{E}_N\big[\big|\Theta_g^{(N)} - \Theta_g\big|\big]\,.$$

Then, using the coupling and (B.26), we can write

$$\begin{aligned}\mathrm{d_{TV}}(g^{(N)}, g) &\leq \mathbb{E}_N\big[\big|\Theta_g^{(N)} - \Theta_g\big|\big] = \mathbb{E}_N\left[\left|\bar{F}_{\Theta_g^{(N)}}^{-1}(U) - \bar{F}_{\Theta_g}^{-1}(U)\right|\right] \\ &= \int_0^1 \left|\bar{F}_{\Theta_g^{(N)}}^{-1}(u) - \bar{F}_{\Theta_g}^{-1}(u)\right| \mathrm{d}u = \int_0^\infty \left|\bar{F}_{\Theta_g^{(N)}}(x) - \bar{F}_{\Theta_g}(x)\right| \mathrm{d}x,\end{aligned} \tag{C.12}$$

where F_{Θ_g} is given by (C.11) and $F_{\Theta_g^{(N)}}$ by

$$F_{\Theta_g^{(N)}}(x) = \frac{1}{\mu_N N}\sum_{i=1}^N \Lambda_i \mathbf{1}_{\{\Lambda_i \leq x\}}. \tag{C.13}$$

Before we prove (C.9), we state two claims. The proofs of these claims are deferred to the end of this section. The first claim is that

$$\mathrm{d}_{\mathrm{TV}}(g^{(N)}, g) \le \left|1 - \frac{\mu_N}{\mu}\right| \nu_N + \frac{E_N}{\mu}, \tag{C.14}$$

where

$$E_N = \int_0^\infty \left| \frac{1}{N} \sum_{i=1}^N \Lambda_i \mathbf{1}_{\{\Lambda_i > x\}} - \mathbb{E}\big[\Lambda \mathbf{1}_{\{\Lambda > x\}}\big] \right| \mathrm{d}x. \tag{C.15}$$

Secondly, we claim that there exist constants $\alpha, \beta > 0$ such that

$$\mathbb{P}\left(E_N > N^{-\alpha}\right) = \mathcal{O}\big(N^{-\beta}\big). \tag{C.16}$$

We finish the proof of (C.9), given the claims (C.14) and (C.16). Indeed, from Section C.1 we know that there exist constants $\alpha_1, \beta_1 > 0$ such that $\mathbb{P}\left(\mathcal{J}_1(\alpha_1)\right) \ge 1 - \mathcal{O}\big(N^{-\beta_1}\big)$.

Combining (C.14) and (C.16), we can bound the total variation by

$$\begin{aligned} \mathbb{P}\left(\mathrm{d}_{\mathrm{TV}}(g^{(N)}, g) > N^{-\alpha_2}\right) &\le N^{\alpha_2} \mathbb{E}\left[\left(\left|1 - \frac{\mu_N}{\mu}\right| \nu_N + E_N/\mu\right) \mathbf{1}_{\{E_N \le N^{-\alpha}\} \cap \mathcal{J}_1(\alpha_1)}\right] \\ &\quad + \mathbb{P}\left(E_N > N^{-\alpha}\right) + \mathcal{O}\big(N^{-\beta_1}\big) = \mathcal{O}\big(N^{-\min\{\alpha_1 - \alpha_2, \alpha - \alpha_2, \beta, \beta_1\}}\big). \end{aligned} \tag{C.17}$$

Pick $\alpha_2 \in (0, \min\{\alpha_1, \alpha\})$, then the right-hand side in the above display tends to zero as a negative power of N, and (C.9) follows, when (C.14) and (C.16) hold. □

Finally, we complete the proof by proving, in turn, the claims (C.14) and (C.16).

Proof of (C.14)**:** Using (C.11), (C.12) and (C.13) we bound $\mathrm{d}_{\mathrm{TV}}(g^{(N)}, g)$ by

$$\begin{aligned} \mathrm{d}_{\mathrm{TV}}(g^{(N)}, g) &= \int_0^\infty \left| \frac{1}{N\mu_N} \sum_{i=1}^N \Lambda_i \mathbf{1}_{\{\Lambda_i > x\}} - \frac{1}{\mu} \mathbb{E}\big[\Lambda \mathbf{1}_{\{\Lambda > x\}}\big] \right| \mathrm{d}x \\ &\le \left|1 - \frac{\mu_N}{\mu}\right| \int_0^\infty \frac{1}{N\mu_N} \sum_{i=1}^N \Lambda_i \mathbf{1}_{\{\Lambda_i > x\}} \mathrm{d}x \\ &\quad + \frac{1}{\mu} \int_0^\infty \left| \frac{1}{N} \sum_{i=1}^N \Lambda_i \mathbf{1}_{\{\Lambda_i > x\}} - \mathbb{E}\big[\Lambda \mathbf{1}_{\{\Lambda > x\}}\big] \right| \mathrm{d}x \\ &= \left|1 - \frac{\mu_N}{\mu}\right| \nu_N + E_N/\mu, \end{aligned} \tag{C.18}$$

where in the last step we have used that

$$\int_0^\infty \frac{1}{N\mu_N} \sum_{i=1}^N \Lambda_i \mathbf{1}_{\{\Lambda_i > x\}} \mathrm{d}x = \frac{1}{N\mu_N} \sum_{i=1}^N \Lambda_i \int_0^\infty \mathbf{1}_{\{\Lambda_i > x\}} \mathrm{d}x = \frac{1}{N\mu_N} \sum_{i=1}^N \Lambda_i^2 = \nu_N.$$

This proves (C.14). □

Proof of (C.16)**:** For $a > 1$, we define the density f_a by

$$f_a(x) = \frac{a-1}{(x+1)^a}, \qquad x > 0. \tag{C.19}$$

Then, for any $\delta, a > 1$, by subsequently using the Markov inequality, then Jensen's inequality on the convex function $w \mapsto w^\delta$, and finally Fubini's theorem,

$$\begin{aligned}
\mathbb{P}\left(E_N > N^{-\alpha}\right) &= \mathbb{P}\left(E_N^\delta > N^{-\alpha\delta}\right) \\
&\le N^{\alpha\delta}(a-1)^{-\delta}\mathbb{E}\left[\left(\int_0^\infty (1+x)^a \left|\frac{1}{N}\sum_{i=1}^N \Lambda_i \mathbf{1}_{\{\Lambda_i > x\}} - \mathbb{E}\big[\Lambda \mathbf{1}_{\{\Lambda > x\}}\big]\right| f_a(x)\,\mathrm{d}x\right)^\delta\right] \\
&\le N^{\alpha\delta}(a-1)^{-\delta+1}\mathbb{E}\left[\int_0^\infty (1+x)^{a(\delta-1)} \left|\frac{1}{N}\sum_{i=1}^N \Lambda_i \mathbf{1}_{\{\Lambda_i > x\}} - \mathbb{E}\big[\Lambda \mathbf{1}_{\{\Lambda > x\}}\big]\right|^\delta \mathrm{d}x\right] \\
&= \mathcal{O}\big(N^{\alpha\delta}\big) \cdot \int_0^\infty (1+x)^{a(\delta-1)}\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N \Lambda_i \mathbf{1}_{\{\Lambda_i > x\}} - \mathbb{E}\big[\Lambda \mathbf{1}_{\{\Lambda > x\}}\big]\right|^\delta\right] \mathrm{d}x. \qquad \text{(C.20)}
\end{aligned}$$

In order to bound (C.20), we use the Marcinkiewicz-Zygmund inequality [12], on the variables $X_i(x) = \Lambda_i \mathbf{1}_{\{\Lambda_i > x\}}$, $1 \le i \le N$. Then, (C.20) can be bounded from above by

$$\begin{aligned}
\mathbb{P}\left(E_N > N^{-\alpha}\right) &\le \mathcal{O}(N^{\alpha\delta}) \int_0^\infty (1+x)^{a(\delta-1)}\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N X_i(x) - \mathbb{E}[X(x)]\right|^\delta\right] \mathrm{d}x \\
&\le \mathcal{O}(N^{\alpha\delta+1-\delta}) \int_0^\infty (1+x)^{a(\delta-1)}\mathbb{E}\Big[|X(x) - \mathbb{E}[X(x)]|^\delta\Big] \mathrm{d}x. \qquad \text{(C.21)}
\end{aligned}$$

Since, for any $x > 0$,

$$\mathbb{E}\Big[|X(x) - \mathbb{E}[X(x)]|^\delta\Big] \le \mathbb{E}\Big[X(x)^\delta\Big] = \int_x^\infty y^\delta \,\mathrm{d}F_\Lambda(y), \qquad \text{(C.22)}$$

we obtain for (1.18),

$$\begin{aligned}
\int_0^\infty (1+x)^{a(\delta-1)}\mathbb{E}\Big[|X(x) - \mathbb{E}[X(x)]|^\delta\Big] \mathrm{d}x &\le \int_0^\infty (1+x)^{a(\delta-1)} \int_x^\infty y^\delta \,\mathrm{d}F_\Lambda(y)\,\mathrm{d}x \\
&\le \int_0^\infty y^\delta (1+y)^{a(\delta-1)+1}\,\mathrm{d}F_\Lambda(y) < \infty, \qquad \text{(C.23)}
\end{aligned}$$

when we first pick $a > 1$ and next $\delta > 1$ so small that $\delta + a(\delta-1) + 1 < \tau - 1$.

Combining (C.21) and (C.23), we obtain

$$\mathbb{P}\left(E_N > N^{-\alpha}\right) = \mathcal{O}\big(N^{\alpha\delta+1-\delta}\big) = \mathcal{O}\big(N^{-(\delta(1-\alpha)-1)}\big), \qquad \text{(C.24)}$$

The right hand side of (C.24) tends to zero if $\delta(1-\alpha) - 1 > 0$, which is the case if we choose α small and δ slightly bigger than 1. □

C.3 Moment and maximal bound on capacities (C3):

In this subsection we show that $\mathbb{P}(\mathcal{J}_3) = 1 - \mathcal{O}\big(N^{-\beta_3}\big)$.

By Boole's inequality, (1.18), and the definition $\gamma = 1/(\tau-1) + \varepsilon$,

$$\mathbb{P}\left(\Lambda_N^{(N)} \ge N^\gamma\right) \le N\mathbb{P}\left(\Lambda_1 \ge N^\gamma\right) = \mathcal{O}\big(N^{-(\tau-1)\varepsilon}\big).$$

Apply Lemma C.1, with $q = \tau - 1 - \varepsilon$, to obtain for $u, v > 0$,

$$1 - \mathbb{P}\left(\mathcal{S}_{N,\tau-1-\varepsilon}(u)\right) = \mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^N \Lambda_i^{\tau-1-\varepsilon} - \mathbb{E}\big[\Lambda^{\tau-1-\varepsilon}\big]\right| > N^{-u}\right) = \mathcal{O}\big(N^{-v}\big).$$

Furthermore, (A.7), with $q = \tau - 2 + \varepsilon > 1$, shows that there exist constants $C_1, C_2 < \infty$ such that, uniformly in $\{\Lambda_i\}_{i=1}^N$,

$$\sum_{n=0}^{\infty} n^{\tau-2-\varepsilon} g_n^{(N)} \le C_1 + \frac{C_2}{N} \sum_{i=1}^{N} \Lambda_i^{\tau-1-\varepsilon}. \tag{C.25}$$

Therefore, for N sufficiently large and $B > C_1 + C_2 \mathbb{E}\big[\Lambda^{\tau-1-\varepsilon}\big]$,

$$\begin{aligned} 1 - \mathbb{P}(\mathcal{J}_3) &\le \mathbb{P}\left(\Lambda_N^{(N)} \ge N^{\gamma}\right) + \mathbb{P}\left(\left\{\sum_{n=0}^{\infty} n^{\tau-2-\varepsilon} g_n^{(N)} > B\right\} \cap \mathcal{S}_{N,\tau-1-\varepsilon}(u)\right) + \mathcal{O}\big(N^{-v}\big) \\ &\le \mathcal{O}\big(N^{-(\tau-1)\varepsilon}\big) + 0 + \mathcal{O}\big(N^{-v}\big) = \mathcal{O}\big(N^{-\min\{(\tau-1)\varepsilon, v\}}\big). \end{aligned} \tag{C.26}$$

Thus, we may take $\beta_3 = \min\{(\tau-1)\varepsilon, u\}$.

Acknowledgements.

The work of HvdE and RvdH was supported in part by Netherlands Organisation for Scientific Research (NWO). We thank Michel Dekking for carefully reading the manuscript and many suggestions, which resulted in a more accessible paper. We thank Chris Klaassen for shedding light on equation (B.26). Finally, we thank Olaf Wittich for discussions on Lemma B.1.

References

[1] S. Asmussen. Some martingale methods in the limit theory of supercritical branching processes. In *Branching processes*, pages 1–26. Advances in Probability and related Topics, 1978.

[2] E.A. Bender and E.R. Canfield. The asymptotic number of labelled graphs with given degree sequences. *Journal of Combinatorial Theory*, A24(3):296–307, 1978.

[3] B. Bollobás. *Random Graphs*. Cambridge University Press, 2001.

[4] B. Bollobás, C. Borgs, J.T. Chayes, and O. Riordan. Directed scale-free graphs. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 132–139. Symposium on Discrete Algorithms, 2003.

[5] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms*, 31(1):3–122, 2007.

[6] T. Britton, M. Deijfen, and A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. *Journal of Statistical Physics*, 124(6):1377–1397, 2006.

[7] F. Chung and L. Lu. The average distances in random graphs with expected degrees. *PNAS*, 99(25):15879–15882, 2002.

[8] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6(2):125–145, 2002.

[9] H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. Universality for the distance in finite variance random graphs: Extended version. `http://arxiv.org/abs/math.PR/0605414`.

[10] H. van den Esker, R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with infinite mean degrees. *Extremes*, 8:111–140, 2006.

[11] W. Feller. *An Introduction to Probability Theory and Its Applications Volume II, 2nd edition*. John Wiley and Sons, New York, 1971.

[12] A. Gut. *Probability: a graduate course.* Springer, Berlin, 2005.

[13] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures and Algorithms*, 27(2):76–123, 2005.

[14] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electronic Journal of Probability*, 12:703–766, 2007.

[15] S. Janson. Asymptotic equivalence and contiguity of some random graphs. Preprint 2008.

[16] S. Janson, T. Luczak, and A. Ruciński. *Random Graphs.* John Wiley & Sons, New York, 2000.

[17] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. In *Proceedings of the sixth international seminar on Random graphs and probabilistic methods in combinatorics and computer science*, pages 161–179, 1995.

[18] M.E.J. Newman, S.H. Strogatz, and D.J. Watts. Random graphs with arbitrary degree distribution and their application. *Physical Review E*, 64:026118, 2001.

[19] I. Norros and H. Reittu. On the power-law random graph model of massive data networks. *Performance Evaluation*, 55(1-2):3–23, 2004.

[20] I. Norros and H. Reittu. On a conditionally Poissonian graph process. *Advances in Applied Probability*, 38(1):59–75, 2006.

[21] H. Thorisson. *Coupling, Stationarity and Regeneration.* Springer, Berlin, 2000.