Distances in random graphs with finite variance degrees

Remco van der Hofstad*
Gerard Hooghiemstra and Piet Van Mieghem†

October 13, 2004

Abstract

In this paper we study a random graph with N nodes, where node j has degree D_j and $\{D_j\}_{j=1}^N$ are i.i.d. with $\mathbb{P}(D_j \leq x) = F(x)$. We assume that $1 - F(x) \leq cx^{-\tau+1}$ for some $\tau > 3$ and some constant $c > 0$. This graph model is a variant of the so-called configuration model, and includes heavy tail degrees with finite variance.

The minimal number of edges between two arbitrary connected nodes, also known as the graph distance or the hopcount, is investigated when $N \to \infty$. We prove that the graph distance grows like $\log_\nu N$, when the base of the logarithm equals $\nu = \mathbb{E}[D_j(D_j - 1)]/\mathbb{E}[D_j] > 1$. This confirms the heuristic argument of Newman, Strogatz and Watts [35]. In addition, the random fluctuations around this asymptotic mean $\log_\nu N$ are characterized and shown to be uniformly bounded. In particular, we show convergence in distribution of the centered graph distance along exponentially growing subsequences.

1 Introduction

The study of complex networks plays an increasingly important role in science. Examples of such networks are electrical power grids and telephony networks, social relations, the World-Wide Web and Internet, co-authorship and citation networks of scientists, etc. The structure of these networks affects their performance. For instance, the topology of social networks affects the spread of information and disease (see e.g., [37]). The rapid evolution in, and the success of, the Internet have incited fundamental research on the topology of networks.

Different scientific disciplines report their own viewpoints and new insights in the broad area of networking. In computer science and electrical engineering, massive Internet measurements have lead to fundamental questions in the modelling and characterization of the Internet topology [22, 38]. These modelling questions drive the understanding of the Internet's complex behavior and allow to plan and to control end-to-end communication. The pioneering work of Strogatz and Watts (see e.g. [37, 41] and the references therein) have triggered an immense number of research papers in the field of theoretical physics. Strogatz and Watts proposed 'small world networks' and illustrated how such small worlds can arise due to underlying mechanisms in different practical networks such as social networks, growing structures in nature, the Web, etc.

Albert and Barabási in [3] showed that preferential attachment of nodes gives rise to a class of graphs often called 'scale free networks'. See also [4, 8] and the references therein. Scale free networks seem to explain the structure of the World-Wide Web, the autonomous domain structure of Internet, citation graphs and many other complex networks (see e.g., [4, 33]). The essence of scale free networks is that the nodal degree is a power law, or, alternatively, heavy-tailed, meaning that the number of nodes with degree equal to k is proportional to $k^{-\tau}$ for some power exponent $\tau > 1$.

*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

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

On the World-Wide Web, it has indeed been shown that there are power law degree sequences, both for the in- and out degrees (see [16, 29]). The work of Albert and Barabási have inspired substantial work on scale-free graphs and can be seen as a way to understand the emergence of power law degree sequences. In the model by Albert and Barabási [3], this power exponent is restricted to $\tau = 3$ [14], but in refinements of the model, different values of τ can be obtained. See, e.g., [2, 10, 19, 30] and the references therein. We will comment on the relations between our work and preferential attachment models in Section 1.4 below. For an overview of the extensive field of random graphs, we refer to the books of Bollobás [9] and Janson *et al.* [28].

The current paper presents a rigorous mathematical derivation for the random fluctuations of the graph distance between two arbitrary nodes in a graph with finite variance degrees. These finite variance degrees include power laws with power exponent $\tau > 3$. We consider the configuration model with power law degree sequences, a variation on a model originally proposed by Newman, Strogatz and Watts [35], prove their conjecture and proceed beyond their results by combining coupling theory, branching processes and shortest path graphs.

1.1 Model definition

Fix an integer N. Consider an i.i.d. sequence $D_1, D_2, \ldots, D_N$. We will construct an undirected graph with N nodes where node j has degree D_j. We will assume that $L_N = \sum_{j=1}^{N} D_j$ is even. If L_N is odd, then we add a stub to the N^{th} node, so that D_N is increased by 1. This single stub will make hardly any difference in what follows, and we will ignore this effect. We will later specify the distribution of D_1.

To construct the graph, we have N separate nodes and incident to node j, we have D_j stubs. All stubs need to be connected to build the graph. The stubs are numbered in a given order from 1 to L_N. We start by connecting at random the first stub with one of the $L_N - 1$ remaining stubs. Once paired, two stubs form a single edge of the graph. Hence, a stub can be seen as the left or the right half of an edge. We continue the procedure of randomly choosing and pairing the stubs until all stubs are connected. Unfortunately, nodes having self-loops may occur. However, self-loops are scarce when $N \to \infty$.

We now specify the degree distribution we will investigate in this paper. The probability mass function and the distribution function of the nodal degree D are denoted by

$$\mathbb{P}(D = j) = f_j, \quad j = 0, 1, 2, \ldots, \quad \text{and} \quad F(x) = \sum_{j=0}^{\lfloor x \rfloor} f_j, \tag{1.1}$$

where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. Our main assumption is that for some $\tau > 3$ and some positive constant c,

$$1 - F(x) \leq c x^{-\tau+1}, \qquad (x > 0). \tag{1.2}$$

This condition implies that the second moment of D is finite. The often used condition that $1 - F(x) = x^{-\gamma+1} L(x)$, $\gamma > 3$, with L a slowly varying function is covered by (1.2), because by Potter's Theorem [23, Lemma 2, p. 277], any slowly varying function $L(x)$ can be bounded above and below by an arbitrary small power of x, so that (1.2) holds for any $\tau < \gamma$.

The above model is closely related to the so-called *configuration model*, in which the degrees of the nodes are often assumed to be fixed (rather than i.i.d.). See [33, Section 4.2.1] and the references therein. We will review some results proved for the configuration model in Section 1.4 below.

1.2 Main results

We denote

$$\mu = \mathbb{E}[D], \qquad \nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}, \tag{1.3}$$

and we define the distance or hopcount H_N between the nodes 1 and 2 as the minimum number of edges that form a path from 1 to 2 where, by convention, the distance equals ∞ if nodes 1 and 2 are not connected. Since the nodes are exchangeable, the distance between two randomly chosen nodes is equal in distribution to H_N. Our main result is the following theorem:

Theorem 1.1 (Limit law for the typical nodal distance) *Assume that $\tau > 3$ in (1.2) and that $\nu > 1$. For $k \geq 1$, let $a_k = \lfloor \log_\nu k \rfloor - \log_\nu k \in (-1, 0]$. There exist random variables $(R_a)_{a\in(-1,0]}$ such that as $N \to \infty$,*

$$\mathbb{P}\big(H_N - \lfloor \log_\nu N \rfloor = k \big| H_N < \infty\big) = \mathbb{P}(R_{a_N} = k) + o(1), \qquad k \in \mathbb{Z}. \tag{1.4}$$

In words, Theorem 1.1 states that for $\tau > 3$, the graph distance H_N between two randomly chosen connected nodes grows like the $\log_\nu N$, where N is the size of the graph, and that the fluctuations around this mean remain uniformly bounded in N. Theorem 1.1 proves the conjecture in Newman, Strogatz and Watts [35, Section II.F, (54)], where a heuristic is given that the number of edges between arbitrary nodes grows like $\log_\nu N$. In addition, Theorem 1.1 improves upon that conjecture by specifying the fluctuations around the value $\log_\nu N$.

We will identify the laws of $(R_a)_{a\in(-1,0]}$ in Theorem 1.4 below. Before doing so, we state two consequences of the above theorem:

Corollary 1.2 (Convergence in distribution along subsequences) *Fix an integer N_1. Under the assumptions in Theorem 1.1, and conditionally on $H_N < \infty$, along the subsequence $N_k = \lfloor N_1 \nu^{k-1} \rfloor$, the sequence of random variables $H_{N_k} - \lfloor \log_\nu N_k \rfloor$ converges in distribution to $R_{a_{N_1}}$ as $k \to \infty$.*

Simulations illustrating the convergence in Corollary 1.2 are discussed in Section 1.5.

Corollary 1.3 (Concentration of the hopcount) *Under the assumptions in Theorem 1.1,*

(i) *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 $\varepsilon > 0$;*

(ii) *conditionally on $H_N < \infty$, the random variables $H_N - \log_\nu N$ form a tight sequence, i.e.,*

$$\lim_{K\to\infty} \limsup_{N\to\infty} \mathbb{P}\big(|H_N - \log_\nu N| \leq K \big| H_N < \infty\big) = 1. \tag{1.5}$$

We need a limit result from branching process theory before we can identify the limiting random variables $(R_a)_{a\in(-1,0]}$. In Section 2 below, we introduce a *delayed* branching process $\{\mathcal{Z}_k\}$, where in the first generation, the offspring distribution is chosen according to (1.1) and in the second and further generations, the offspring is chosen in accordance to g given by

$$g_j = \frac{(j+1)f_{j+1}}{\mu}, \quad j = 0, 1, \ldots. \tag{1.6}$$

The process $\{\mathcal{Z}_k/\mu\nu^{k-1}\}$ is a martingale with uniformly bounded expectation and consequently converges almost surely to a limit:

$$\lim_{n\to\infty} \frac{\mathcal{Z}_n}{\mu\nu^{n-1}} = \mathcal{W} \qquad a.s. \tag{1.7}$$

In the theorem below we need two independent copies $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ of $\mathcal{W}$.

Theorem 1.4 (The limit laws) *Under the assumptions in Theorem 1.1, and for $a \in (-1, 0]$,*

$$\mathbb{P}(R_a > k) = \mathbb{E}\big[\exp\{-\kappa\nu^{a+k}\mathcal{W}^{(1)}\mathcal{W}^{(2)}\}\big|\mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0\big], \tag{1.8}$$

where $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ are independent limit copies of $\mathcal{W}$ in (1.7) and where $\kappa = \mu(\nu - 1)^{-1}$.

We will also provide an error bound of the convergence stated in Theorem 1.1. Indeed, we show that for any $\alpha > 0$, and for all $k \leq \eta \log_\nu N$ for some $\eta > 0$ sufficiently small,

$$\mathbb{P}(H_N > \lfloor \log_\nu N \rfloor + k) = \mathbb{E}\big(\exp\{-\kappa \nu^{a_N+k} \mathcal{W}^{(1)} \mathcal{W}^{(2)}\}\big) + O((\log N)^{-\alpha}). \tag{1.9}$$

Unfortunately, due to the conditioning in Theorem 1.1, it is hard to obtain an explicit error bound in (1.4).

The law of R_a is involved, and can in most cases not be computed exactly. The reason for this is the fact that the random variables $\mathcal{W}$ that appear in its statement are hard to compute explicitly. For example, for the power-law degree graph with $\tau > 3$, we do not know what the law of $\mathcal{W}$ is. See also Section 2. There are two examples where the law of $\mathcal{W}$ is known. The first is when all degrees in the graph are equal to some $r > 2$, and we obtain the r-regular graph (see also [15], where the diameter of this graph is studied). In this case, we have that $\mu = r, \nu = r - 1$, and $\mathcal{W} = 1$ a.s. In particular, $\mathbb{P}(H_N < \infty) = 1 + o(1)$. Therefore, we obtain that

$$\mathbb{P}(R_a > k) = \exp\{-\frac{r}{r-2}(r-1)^{a+k}\}, \tag{1.10}$$

and H_N is asymptotically equal to $\log_{r-1} N$. The second example is when the law g is geometric, in which case the branching process with offspring g conditioned to be positive converges to an exponential random variable with parameter 1. This example corresponds to

$$g_j = p(1-p)^{j-1}, \quad \text{so that} \quad f_j = \frac{1}{jc_p} p(1-p)^{j-2}, \quad \forall j \geq 1, \tag{1.11}$$

and c_p is the normalizing constant. For $p > \frac{1}{2}$, the law of $\mathcal{W}$ has the same law as the sum of D_1 copies of a random variable $\mathcal{Y}$, where $\mathcal{Y} = 0$ with probability $\frac{1-p}{p}$ and equal to an exponential random variable with parameter 1 with probability $\frac{2p-1}{p}$. Even in this simple case, the computation of the exact law of R_a is non-trivial. Although the laws R_a are hard to compute exactly, Theorems 1.1 and 1.4 make it possible to simulate the hopcount in random graphs of arbitrary size since the law of $\mathcal{W}$ is simple to approximate numerically, for example using Fast Fourier Transforms.

In [27], the expected value of the random variable R_a is computed numerically, by comparing it to $\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0]$, where $\mathcal{W}$ is the almost sure limit of the associated branching process in (1.7) . One would expect that, with α as in (1.9), there exists some β with $0 < \beta < \alpha$, satifying

$$\mathbb{E}[H_N | H_N < \infty] = \lfloor \log_\nu N \rfloor + \mathbb{E}[R_a] + O((\log N)^{-\beta}). \tag{1.12}$$

If so, an accurate computation of $\mathbb{E}[R_a]$ would yield the fine asymptotics of the expected hopcount, and this would yield an extension of the conjectured results in [35, (54)]. Our methods stop short of proving (1.12), and this remains an interesting question.

Our final result describes the size of the largest connected component and the maximal size of all other connected components. In its statement, we write G for the random graph with degree distribution given by (1.1), and we write q for the survival probability of the delayed branching process $\{\mathcal{Z}_k\}$ described above. Thus, $1 - q$ is the extinction probability of the branching process.

Theorem 1.5 (The sizes of the connected components) *With probability* $1-o(1)$*, the largest connected component in* G *has* $qN(1+o(1))$ *nodes, and there exists* $\gamma < \infty$ *such that all other connected components have at most* $\gamma \log N$ *nodes.*

1.3 Methodology and heuristics

One can understand Theorems 1.1 and 1.4 intuitively as follows. Denote by $Z_k^{(1)}$, respectively, $Z_k^{(2)}$ the number of stubs of nodes at distance $k-1$ from node 1, respectively, node 2 (see Section 3 for the precise definitions). Then for $N \to \infty$, the random process $Z_1^{(i)}, Z_2^{(i)}, \ldots, Z_k^{(i)}$, which will be called shortest path graphs (SPG's), behave as a delayed branching process as long as $Z_k^{(i)}$ is of small order compared to N. Thus, the local neighborhood of the node i is close in distribution to a branching process.

We sample the stubs uniformly from all stubs and thus, for large N, we attach the stubs to the SPG proportionally to jf_j. Moreover, when a new stub is attached to the SPG, the chosen stub is used to attach the new node and forms an edge together with the present stub. Therefore, the number of stubs of the freshly chosen node decreases by one and is equal to j if the number of stubs of the chosen node was originally equal to $j+1$. This motivates (1.6).

The offspring of the node 1 is distributed as D_1, whereas the offspring distribution of $Z_2^{(1)}, Z_3^{(1)}, \ldots$ has (for $N \to \infty$) probability mass function (1.6). Consequently, as noted in [35, (51)], the mean number of free stubs at distance k is close to $\mu\nu^{k-1}$, where $\nu = \sum_{j=1}^{\infty} jg_j$ is defined in (1.3). Moreover, a stub in $Z_k^{(1)}$ is attached with a positive probability to a stub in $Z_k^{(2)}$ whenever $Z_k^{(1)} Z_k^{(2)}$ is of order L_N. The total degree L_N is proportional to N by the law of large numbers, because $\mu = \mathbb{E}[D_1] < \infty$. Since both sets grow at the same rate, each has to be of order $\sqrt{N}$. Therefore, k is typically $\frac{1}{2}\log_\nu N$, and the typical distance between 1 and 2 is of order $2k = \log_\nu N$. This can be made precise by coupling $Z_1^{(1)}, Z_2^{(1)}, \ldots$ to a branching process $\hat{Z}_1^{(1)}, \hat{Z}_2^{(1)}, \ldots$ having offspring distribution $g_j^{(N)}$ given by

$$g_j^{(N)} = \sum_{i=1}^{N} I[D_i = j+1]\frac{D_i}{L_N} = \frac{j+1}{L_N}\sum_{i=1}^{N} I[D_i = j+1], \tag{1.13}$$

where $I[E]$ is the indicator of the event E. This coupling will be described in Section 3.1. In turn, the branching process $\hat{Z}_1^{(1)}, \hat{Z}_2^{(1)}, \ldots$ will be coupled, in a conventional way, to a branching process $\mathcal{Z}_1^{(1)}, \mathcal{Z}_2^{(1)}, \ldots$ with offspring distribution $\{g_j\}$ defined in (1.6). The limit result of Theorem 1.1 and Theorem 1.4 depends on the martingale limit for super-critical branching processes with finite mean.

The proof of Theorems 1.1 and 1.4 are based upon a comparison of the local neighborhoods of nodes to branching processes. Such techniques are used extensively in random graph theory. An early example is in [15], where the diameter of a random regular graph was investigated. See also [5, Chapter 10], where comparisons to branching processes are used to describe the phase transition and the birth of the giant component for the random graph $G(p, N)$.

The proof of Theorem 1.5 makes essential use of results by Molloy and Reed [31, 32] for the usual configuration model. We will now describe their result. When the number of nodes with degree i in the graph of size N equals $d_i(N)$ where $\lim_{N\to\infty} d_i(N)/N = Q(i)$, Molloy and Reed [31, 32] identify the condition $\sum_{i=1}^{\infty} i(i-2)Q(i) > 0$ as the necessary and sufficient condition to ensure that a 'giant component' proportional to the size of the graph exists. By rewriting the condition $\nu > 1$ in Theorem 1.1 as $\mathbb{E}[D^2] - 2\mathbb{E}[D] > 0$, we see that a similar condition as in the model of Molloy and Reed is needed here. To prove Theorem 1.5, we need to check that the technical conditions in [31, 32] are satisfied in our model. In fact, we need to alter the graph G a little bit in order to apply their results, since in [31] it is assumed that no nodes of degree larger than $N^{\frac{1}{4}-\epsilon}$ exist for some $\epsilon > 0$.

The novelty of our results is that we investigate *typical distances* in random graphs. In random graph theory, it is more customary to investigate the *diameter* in the graph, and in fact, this would also be an interesting problem. The research question investigated in this paper is inspired by the Internet. In a seminal paper [22], Faloutsos *et al.* have shown that the degree distribution of autonomous systems in Internet follows a power law with power exponent $\tau \approx 2.2$. Thus, the power law random graph with this value of τ can possibly lead to a good Internet model on the autonomous systems (AS) level (see [22, 38]). For the Internet on the more detailed router level, extensive measurements exist for the hopcount, which is the number of routers traversed between two typical routers, as well as for the AS-count, which is the number of autonomous systems traversed between two typical routers. To validate the configuration model with i.i.d. degrees, we intend to compare the distribution of the distance between pairs of nodes to these measurements in Internet. For this, a good understanding of the typical distances between nodes in the degree random graph are necessary, which formed the main motivation for our work. The hopcount in Internet seems to be close to a Poisson random variable with a fairly large parameter. In turn, a Poisson random variable with large parameter can be approximated by a normal random variable with equal expectation and variance. See e.g. [34, 40] for data of the hopcount in Internet.

From a practical point of view, there are good reasons to study the typical distances in random graphs rather than the diameter. For one, typical distances are simpler to measure, and thus allow for a simpler validation of the model. Also, the diameter is a number, while the *distribution* of the typical distances contains substantially more information. Finally, the diameter is rather sensitive to small changes to a graph. For instance, when adding a string of a few nodes, one can dramatically alter the diameter, while the typical distances in the graph hardly change. Thus, typical distances in the graph are more robust to modelling discrepancies.

1.4 Related work

There is a wealth of related work which we will now summarize. The model investigated here was also studied in [36], with $1-F(x) = x^{-\tau+1}L(x)$, where $\tau \in (2,3)$ and L denotes a slowly varying function. It was shown in [36] that the average distance is bounded from above by $2\frac{\log\log N}{|\log(\tau-2)|}(1+o(1))$. We plan to return to the question of average distances and connected component sizes when $\tau < 3$ in three future publications [24, 25, 26].

There is substantial work on random graphs that are, although different from ours, still similar in spirit. In [1], random graphs were considered with a degree sequence that is *precisely* equal to a power law, meaning that the number of nodes with degree k is precisely proportional to $k^{-\tau}$. Aiello *et al.* [1] show that the largest connected component is of the order of the size of the graph when $\tau < \tau_0 = 3.47875\ldots$, where τ_0 is the solution of $\zeta(\tau-2) - 2\zeta(\tau-1) = 0$, and where ζ is the Riemann Zeta function. When $\tau > \tau_0$, the largest connected component is of smaller order than the size of the graph and more precise bounds are given for the largest connected component. When $\tau \in (1,2)$, the graph is with high probability connected. The proofs of these facts use couplings with branching processes and strengthen previous results due to Molloy and Reed [31, 32] described above. For this same model, Dorogovtsev *et al.* [20, 21] investigate the leading asymptotics and the fluctuations around the mean of the distance between arbitrary nodes in the graph from a theoretical physics point of view, using mainly generating functions.

A second related model can be found in [17] and [18], where edges between nodes i and j are present with probability equal to $w_i w_j / \sum_l w_l$ for some 'expected degree vector' $w = (w_1, \ldots, w_N)$. Chung and Lu [17] show that when w_i is proportional to $i^{-\frac{1}{\tau-1}}$ the average distance between pairs of nodes is $\log_\nu N(1+o(1))$ when $\tau > 3$, and $2\frac{\log\log N}{|\log(\tau-2)|}(1+o(1))$ when $\tau \in (2,3)$. The difference between this model and ours is that the nodes are not exchangeable in [17], but the observed phenomena are similar. This result can be heuristically understood as follows. Firstly, the actual degree vector in [17] should be close to the expected degree vector. Secondly, for the expected degree vector, we can compute that the number of nodes for which the degree is less than or equal to k equals

$$|\{i : w_i \leq k\}| \propto |\{i : i^{-\frac{1}{\tau-1}} \leq k\}| \approx k^{-\tau+1}.$$

Thus, one expects that the number of nodes with degree at most k decreases as $k^{-\tau+1}$, similarly as in our model. In [18], Chung and Lu study the sizes of the connected components in the above model. The advantage of this model is that the edges are *independently* present, which makes the resulting graph closer to a traditional random graph.

All the models described above are *static*, i.e., the size of the graph is *fixed*, and we have not modeled the *growth* of the graph. As described in the introduction, there is a large body of work investigating *dynamical* models for complex networks, often in the context of the World-Wide Web. In various forms, preferential attachment has been shown to lead to power law degree sequences. Therefore, such models intend to *explain* the occurrence of power law degree sequences in random graphs. See [2, 3, 4, 10, 11, 12, 13, 14, 19, 30] and the references therein. In the preferential attachment model, nodes with a fixed degree m are added sequentially. Their stubs are attached to a receiving node with a probability proportionally to the degree of the receiving node, thus favoring nodes with large degrees. For this model, it is shown that the number of nodes with degree k decays proportionally to k^{-3} [14], the diameter is of order $\frac{\log N}{\log \log N}$ when $m \geq 2$ [11], and couplings to a classical random graph $G(N,p)$ are given for an appropriately chosen p in [13]. See also [12] for a survey.

It can be expected that our model is a snapshot of the above models, i.e., a realization of the graph growth processes at the time instant that the graph has a certain prescribed size. Thus, rather than to describe the growth of the model, we investigate the properties of the model at a given time instant. This is suggested in [4, Section VII.D], and it would be very interesting indeed to investigate this further mathematically, i.e., to investigate the relation between the configuration and the preferential attachment models.

The reason why we study the random graphs at a given time instant is that we are interested in the topology of the random graph. In [38], and inspired by the observed power law degree sequence in [22], the configuration model with i.i.d. degrees is proposed as a model for the AS-graph in Internet, and it is argued on a qualitative basis that this simple model serves as a better model for the Internet topology than currently used topology generators. Our results can be seen as a step towards the quantitative understanding of whether the hopcount in Internet is described well by the average graph distance in the configuration model.

In [33, Table II], many more examples are given of real networks that have power law degree sequences. Interestingly, there are also many examples where power laws are *not* observed, and often the degree law falls off faster than a power law. These observed degrees can be described by a degree distribution as in (1.1) with $1-F(x)$ smaller than any power, and the results in this paper thus apply. Such examples are described in more detail in [4, Section II]. Examples where the tails of the degree distribution are lighter than power laws are power and neural networks [4, Section II.K], where the tails are observed to be exponential, and protein folding [4, Section II.L], where the tails are observed to be Gaussian. In other examples, a degree distribution is found that for small values is a power law, but has an exponential cut off. An example of such a degree distribution is

$$f_k = Ck^{-\gamma}e^{-k/\kappa}, \tag{1.14}$$

for some $\kappa > 0$ and $\gamma \in \mathbb{R}$. The size of κ indicates up to what degree the power law still holds, and where the exponential cut off starts to set in. For this example, our results apply since the exponential tail ensures that (1.2) holds for *any* $\tau > 3$ by picking $c > 0$ large enough. Thus, we prove the conjectures on the expected path lengths in [35, (55), (56)] and [4, Section V.C, (63) and (64)] for this particular model.

1.5 Simulation for illustration of the main results

To illustrate Theorem 1.1, we have simulated the random graph with degree distribution $D = \lceil U^{-\frac{1}{\tau-1}} \rceil$, where U is uniformly distributed over $(0,1)$ and where for $x \in \mathbb{R}$, $\lceil x \rceil$ is the smallest integer greater than or equal to x. Thus,

$$1 - F(k) = \mathbb{P}(U^{-\frac{1}{\tau-1}} > k) = k^{1-\tau}, \quad k = 1, 2, 3, \ldots,$$

for which $\mu = 1 + \zeta(\tau-1)$ and $\nu = 2\zeta(\tau-2)/\mu$.

We observe that for $\tau = 3.5$ and $N = 25,000$ and $N = 125,000$, the values $a_N = -0.62\ldots$ are identical up to two decimals. We hence expect, on the basis of our main theorem, that the survival functions $\mathbb{P}(H_N > k)$ for these two cases are similar. Because $\lfloor \log_\nu 25,000 \rfloor = 12$ and $\lfloor \log_\nu 125,000 \rfloor = 14$, we expect that the empirical survival function for $N = 125,000$ is a shift of the empirical survival function for $N = 25,000$, over the horizontal distance $14 - 12 = 2$. Figure 1 supports this claim, given the statistical inaccuracy. In Figure 1 we have also included the empirical

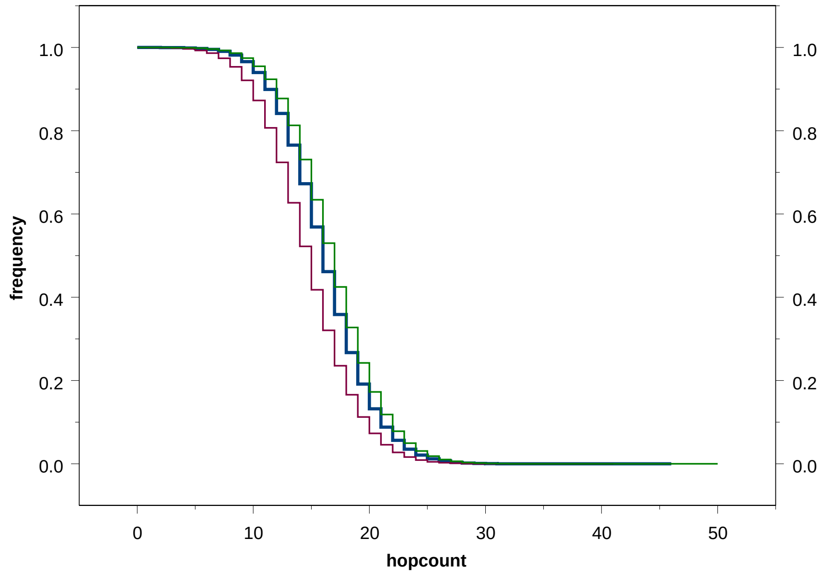

Figure 1: Empirical survival functions of the hopcount for $\tau = 3.5$ and the values $N = 25,000$, $N = 75,000$ (bold) and $N = 125,000$, based on samples of size $1,000$.

survival function for $N = 75,000$, for which $a_N = -0.99\ldots$, as the bold line. This empirical survival function clearly has a different shape. Thus, the empirical survival function for $N = 75,000$ is not a shift of the empirical survival function for $N = 25,000$ or $N = 125,000$.

We finally demonstrate Corollary 1.2 for $\tau = 3.5$ in Figure 2. In this case $\nu^2 \approx 5$ and $N_k = N_1\nu^{2k}$, $k = 0, 1, 2, 3$. We take $N_1 = 5,000$, and so $N_2 = 25,000$, $N_3 = 125,000$, $N_4 = 625,000$. For these values of $N_1, \ldots, N_4$, we have simulated the hopcount with $1,000$ replications and we expect from Corollary 1.2 that the survival functions run parallel at mutual distance 2.

1.6 Organization of the paper

We will first review the relevant literature on branching processes in Section 2. We will then explain how we can couple our degree model to independent branching processes in Section 3. This section is also valuable for our coming paper [24], where we study the case $\tau \in (2, 3)$. In particular, in [24], we will use Lemmas A.2.2 and A.2.8 and Proposition A.3.1. The bounds for the coupling are formulated in Sections 3.1, 3.2 and 3.3. In these sections, we will state the results on the coupling that are needed in the proof of the main results, Theorems 1.1 and 1.4. Parts of this section apply more generally, i.e., to $\tau \in (2, 3)$. We prove Theorems 1.1 and 1.4 in Section 4 and Theorem 1.5 in Section 5. The technical details of the coupling of $\{\hat{Z}_k^{(i)}\}$ to $\{\mathcal{Z}_k^{(i)}\}$ for $i = 1, 2$ are contained in Section A.1, while the details of the coupling of $\{Z_k^{(i)}\}$ to $\{\hat{Z}_k^{(i)}\}$ for $i = 1, 2$ are in Section A.2. Finally, we prove that at any fixed time m, with probability converging to 1, $Z_m^{(i)} = \mathcal{Z}_m^{(i)}$ for $i = 1, 2$ in Section A.3.

2 Review of branching process theory with finite mean

Since we rely heavily on the theory of branching processes, we will briefly review this theory in the case where the expected value of the offspring distribution is finite. The theory of branching processes is well understood (see e.g. [7]).

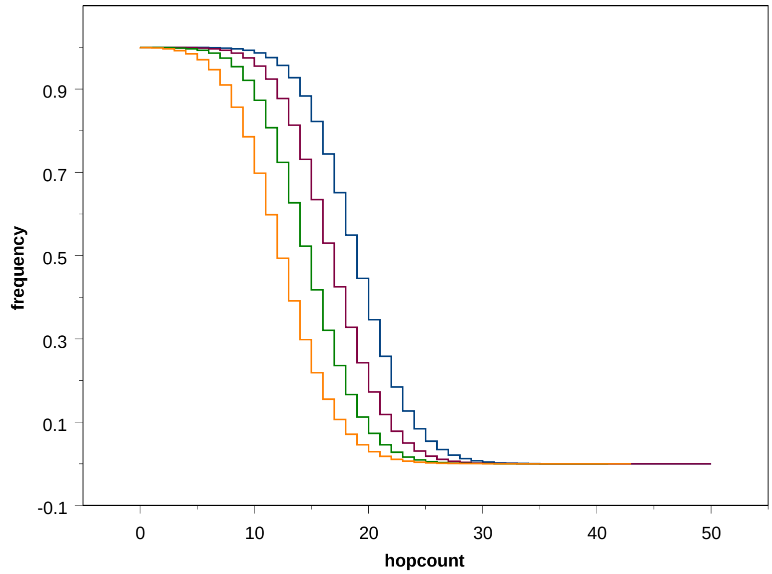

Figure 2: Empirical survival functions of the hopcount for $\tau = 3.5$ and the four values $N_k = 5,000\nu^{2k}$, $k = 0, 1, 2, 3$, based on $1,000$ runs.

For the formal definition of the delayed branching process (BP) that we consider here, we define a double sequence $\{X_{n,i}\}_{n\geq 1, i\geq 1}$ of i.i.d. random variables each with distribution equal to the offspring distribution $\{g_j\}_{j=0}^{\infty}$, where we recall

$$g_j = \frac{(j+1)f_{j+1}}{\mu}, \quad j = 0, 1, \ldots. \tag{2.1}$$

We further let $X_{0,1}$ have probability mass function f in (1.1), independently from $\{X_{n,i}\}_{n\geq 1, i\geq 1}$. The BP $\{\mathcal{Z}_n\}$ is now defined by $\mathcal{Z}_0 = 1$ and

$$\mathcal{Z}_{n+1} = \sum_{i=1}^{\mathcal{Z}_n} X_{n,i}, \quad n \geq 0.$$

Because $\tau > 3$, we have that both $\mathbb{E}[\mathcal{Z}_1] = \mathbb{E}[X_{0,1}] = \mu < \infty$ and $\nu = \mathbb{E}[X_{1,1}] < \infty$. We further assume that $\nu = \mathbb{E}[X_{1,1}] > 1$, so that the BP is super-critical. Given that the $(n-1)^{\text{st}}$ generation consists of m individuals, the conditional expectation of $\mathcal{Z}_n$ equals $m\nu$, independently of the size of the preceding generations, so that for $n \geq 1$, we have $\mathbb{E}[\mathcal{Z}_n|\mathcal{Z}_{n-1}] = \mathcal{Z}_{n-1}\nu$. Hence, $\mathcal{W}_n = \frac{\mathcal{Z}_n}{\mu\nu^{n-1}}$, is a martingale. Since $\mathbb{E}[|\mathcal{W}_n|] = \mathbb{E}[\mathcal{W}_n] = 1$, the sequence $\mathbb{E}[|\mathcal{W}_n|]$ is uniformly bounded by 1 and so by Doob's martingale convergence theorem [42, p. 58] the sequence $\mathcal{W}_n$ converges almost surely. If we denote the a.s. limit by a proper random variable $\mathcal{W}$, we obtain (1.7).

There are only few examples where the limit random variable $\mathcal{W}$ is known. It is known that $\mathcal{W}$ has an atom at 0 of size $p \geq 0$, equal to the extinction probability of the (delayed-)BP ($q = 1 - p$). Conditioned on non-extinction the limit $\mathcal{W}$ has an absolute continuous density on $(0, \infty)$.

We need a result that follows from [6] concerning the speed of convergence of $\mathcal{W}_n$ to $\mathcal{W}$. Define

$$\mathcal{R}_n = \frac{\mathcal{W}_n}{\nu} \int_{\nu^n/n^\alpha}^{\infty} x\, dG(x), \quad \alpha > 0,$$

where G is the distribution function of the offspring with probabilities $\{g_j\}$. Since

$$\mu_\alpha = \int_0^\infty x[\log^+ x]^\alpha \, dG(x) < \infty, \qquad (\log^+ x = \max(0, \log x)),$$

for each $\alpha > 0$, it follows from ([6, page 8, line 4]) that with probability 1,

$$\mathcal{W} - \mathcal{W}_k + \sum_{n=k}^\infty \mathcal{R}_n = o(k^{-\alpha}). \tag{2.2}$$

An immediate consequence of (2.2) is that if $|\mathcal{W} - \mathcal{W}_k| > k^{-\alpha}$, then $\sum_{n=k}^\infty \mathcal{R}_n > k^{-\alpha}$. Hence, using $\mathbb{E}[\mathcal{W}_n] = 1$ and partial integration,

$$\begin{aligned}\mathbb{P}(|\mathcal{W} - \mathcal{W}_k| > k^{-\alpha}) \quad &\leq \mathbb{P}\left(\sum_{n=k}^\infty \mathcal{R}_n > k^{-\alpha}\right) \leq k^\alpha \sum_{n=k}^\infty \mathbb{E}[\mathcal{R}_n] = -\sum_{n=k}^\infty \frac{k^\alpha}{\nu} \int_{\nu^n/n^\alpha}^\infty x \, d\,[1 - G(x)] \\ &= \sum_{n=k}^\infty \frac{k^\alpha}{\nu}[1 - G(\nu^n/n^\alpha)] + \sum_{n=k}^\infty \frac{k^\alpha}{\nu} \int_{\nu^n/n^\alpha}^\infty [1 - G(x)] \, dx.\end{aligned}$$

Since $1 - F(x) \leq c \cdot x^{1-\tau}$ (see (1.2)), we find $1 - G(x) \leq c' \cdot x^{2-\tau}$ so that for each $\alpha > 0$, and with $k = \lfloor \frac{1}{2} \log_\nu N \rfloor$,

$$\mathbb{P}\left(|\mathcal{W} - \mathcal{W}_k| > (\log N)^{-\alpha}\right) \leq O((\log N)^\alpha) \sum_{n=k}^\infty (\nu^n/n^\alpha)^{3-\tau} = O(e^{-\beta \log N}) = O(N^{-\beta}), \tag{2.3}$$

for some positive β, because $\tau > 3$ and $\nu > 1$.

3 Graph construction and coupling with a BP

In this section, we will describe how the shortest path graph (SPG) from node 1 can be obtained, and we will couple it to a BP. This coupling works for any degree distribution. In Sections 3.2 and 3.3 below, we will obtain bounds on the coupling.

The SPG from node 1 is the random graph as observed from node 1, and consists of the shortest paths between node 1 and all other nodes $\{2, \ldots, N\}$. As will be shown below, it is not necessarily a tree because cycles may occur. Recall that two stubs together form an edge. We define $Z_1^{(1)} = D_1$, and for $k \geq 2$, we denote by $Z_k^{(1)}$ the number of stubs attached to nodes at distance $k-1$ from node 1, but are not part of an edge connected to a node at distance $k-2$. We will refer to such stubs as 'free stubs'. Thus, $Z_k^{(1)}$ is the number of outgoing stubs from nodes at distance $k-1$.

In Section 3.1 we will describe a coupling that, conditionally on $D_1, \ldots, D_N$, couples $\{Z_k^{(1)}\}$ to a BP $\{\hat{Z}_k^{(1)}\}$ with the *random* offspring distribution

$$\begin{aligned}g_j^{(N)} \quad &= \quad \sum_{i=1}^N I[D_i = j+1] \mathbb{P}(\text{a stub from node } i \text{ is sampled} | D_1, \ldots, D_N) \\ &= \quad \sum_{i=1}^N I[D_i = j+1] \frac{D_i}{L_N} = \frac{j+1}{L_N} \sum_{i=1}^N I[D_i = j+1],\end{aligned} \tag{3.1}$$

where as before $L_N = D_1 + D_2 + \ldots + D_N$. By the strong law of large numbers, for $N \to \infty$,

$$\frac{L_N}{N} \to \mathbb{E}[D], \quad \text{and} \quad \frac{1}{N} \sum_{i=1}^N I[D_i = j+1] \to \mathbb{P}(D = j+1), \qquad a.s.$$

so that a.s.,

$$g_j^{(N)} \to (j+1)\mathbb{P}(D=j+1)/\mathbb{E}[D] = g_j, \quad N \to \infty. \tag{3.2}$$

Therefore, the BP $\{\hat{Z}_k^{(1)}\}$ with offspring distribution $\{g_j^{(N)}\}$ is expected to be close to a BP with offspring distribution $\{g_j\}$ given in (1.6). Consequently, in Section 3.3, we will couple the BP $\{\hat{Z}_k^{(1)}\}$ to a BP $\{\mathcal{Z}_k^{(1)}\}$ with offspring distribution $\{g_j\}$. This will allow us to prove Theorems 1.1 and 1.4 in Section 4.

Throughout the paper we use the following lemma. It shows that L_N is close to $\mathbb{E}[L_N] = \mu N$.

Lemma 3.1 (Concentration of L_N) *For each $0 < a < \frac{1}{2}$, $b = 1-2a$ and some constant $c > 0$,*

$$\mathbb{P}\left(\left|\frac{L_N}{\mathbb{E}[L_N]} - 1\right| \geq N^{-a}\right) \leq cN^{-b}. \tag{3.3}$$

Proof. The proof is immediate from the Chebychev inequality, since

$$\mathbb{P}\left(\left(\frac{L_N}{\mathbb{E}[L_N]} - 1\right)^2 \geq N^{-2a}\right) \leq \frac{N^{2a}}{(N\mu)^2}\mathrm{Var}(L_N) = \frac{\mathrm{Var}(D)}{\mu^2}N^{2a-1},$$

so that $b = 1-2a > 0$ and $c = \frac{\mathrm{Var}(D)}{\mu^2} < \infty$. $\square$

3.1 Coupling with a branching process with offspring $g^{(N)}$

We will construct the SPG in such a way that we simultaneously construct a BP with offspring distribution $\{g_j^{(N)}\}$ in (3.1). This BP is of course purely imaginary. The BP is coupled with the SPG such that it enables us to control their difference.

As above, we will use the notation $Z_k^{(1)}$ and $Z_k^{(2)}$ to denote the number of stubs attached to nodes at distance $k-1$ from node 1, respectively, node 2, but not part of an edge connected to a node at distance $k-2$. For $k = 1$, $Z_k^{(i)} = D_i$. We start with a description of the coupling of the SPG with root 1, and a BP with offspring distribution $g^{(N)}$ given in (3.1). The first stages of the generation of the SPG are drawn in Figure 3. We will explain the meaning of the labels 1, 2 and 3 below.

We draw repeatedly and independently from the distribution $\{g_j^{(N)}\}$. This is done conditionally given $D_1, D_2, \ldots, D_N$, so that we draw from the *random* distribution (3.1). After each draw we will update the realization of the SPG and the BP, and classify the stubs according to three categories, which will be labelled 1, 2 and 3. These labels will be updated as the growth of the SPG proceeds. The labels have the following meaning:

1. Stubs with label 1 are stubs belonging to a node that is not yet attached to the SPG.
2. Stubs with label 2 are attached to the SPG (because the corresponding node has been chosen), but not yet paired with another stub. These are called 'free stubs'.
3. Stubs with label 3 in the SPG are paired with another stub to form an edge in the SPG.

The growth process as depicted in Figure 3 starts by giving all stubs label 1. Then, because we construct the SPG starting from node 1, we relabel the D_1 stubs of node 1 with the label 2. We note that $Z_1^{(1)}$ is equal to the number of stubs connected to node 1, and thus $Z_1^{(1)} = D_1$. We next identify $Z_j^{(1)}$ for $j > 1$. $Z_j^{(1)}$ is obtained by sequentially growing the SPG from the free stubs in generation $Z_{j-1}^{(1)}$. When all free stubs in generation $j-1$ have chosen their connecting stub, $Z_j^{(1)}$ is equal to the number of stubs labelled 2 (i.e., free stubs) attached to the SPG. Note that not necessarily each stub of $Z_{j-1}^{(1)}$ contributes to stubs of $Z_j^{(1)}$, because a cycle may 'swallow' two free stubs in generation $j-1$. This is the case precisely when a stub with label 2 is chosen.

For the BP, we start with $\hat{Z}_1^{(1)} = D_1$, and grow from the free stubs available in the BP tree by sequentially growing from the stubs (alike for the SPG). For the coupling, as long as there are free

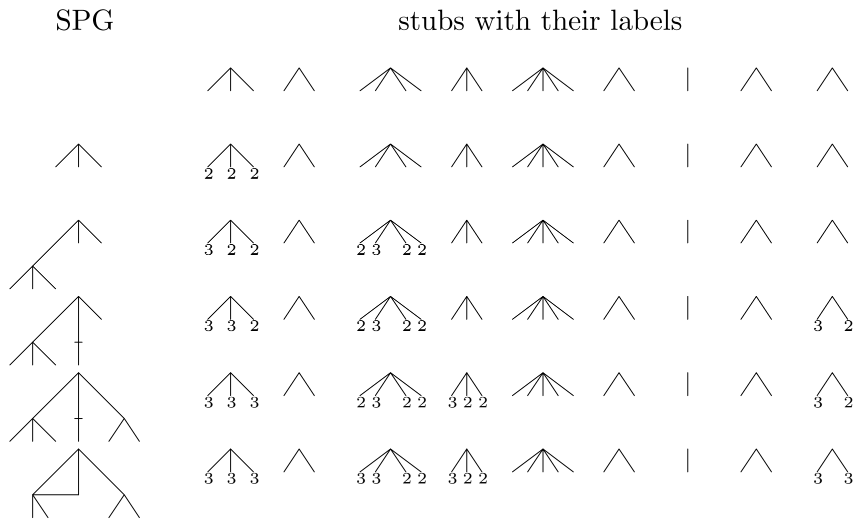

Figure 3: Schematic drawing of the growth of the SPG from the node 1 with $N = 9$ and the updating of the labels. The stubs without labels have label 1. The first line shows the N different degrees. The growth process starts by choosing the first stub of node 1 whose stubs are labeled by 2 as illustrated in the second line, while all the other stubs maintain the label 1. Next, we uniformly choose a stub with label 1 or 2. In the example in line 3, this is the second stub from node 3, whose stubs are labeled by 2 except for the second stub which is labeled 3. The left hand side column visualizes growth of the SPG by the attachment of stub 2 of node 3 to the first stub of node 1. Once an edge is established the paired stubs are labeled 3. In the next step, the next stub of node one is again matched to a uniform stub out of those with label 1 or 2. In the example in line 4, it is the first stub of the last node that will be attached to the second stub of node 1, the next in sequence to be paired. The last line exhibits the result of creating a cycle when the first stub of node 3 is chosen to be attached to the last stub of node 9 (the last node). This process is continued until there are no more stubs with labels 1 or 2. In this example, we have $Z_1^{(1)} = 3$ and $Z_2^{(1)} = 6$.

stubs in *both* the BP and the SPG in a given generation, we couple the BP and SPG in the following way. At each step we will take an independent draw from all stubs, according to the distribution (3.1). Since the stubs are specified by their label (1, 2 or 3), we can now present the construction rules for the BP and the SPG.

1. If the chosen stub has label 1, then in both the BP and the SPG we will connect the present stub to the chosen stub to form an edge and attach the remaining stubs of the chosen node as children. We update the labels as follows. The present and chosen stub melt together to form an edge and both are assigned label 3. All 'brother' stubs (except for the chosen stub) belonging to the same node of the chosen stub receive label 2.

2. In this case we choose a stub with label 2, which is already connected to the SPG. For the BP, the chosen stub is simply connected to the stub which is grown, and the number of free stubs is the number of 'brother stubs' of the chosen stub. For the SPG, a self-loop is created when the chosen stub and present stub are 'brother' stubs which belong to the same node. When they are not 'brother' stubs, then a cycle is formed. Neither a self-loop nor a cycle changes the distances in the SPG. Note that for the SPG *two* free stubs are used, while for the BP only *one* stub is used. This is illustrated in Figure 4.

   The updating of the labels solely consists of changing the label of the present and the chosen stub from 2 to 3.

3. A stub with label 3 is chosen. This case is illustrated in Figure 5. This possibility of choosing an already matched stub with label 3 must be included for the BP which relies on the property

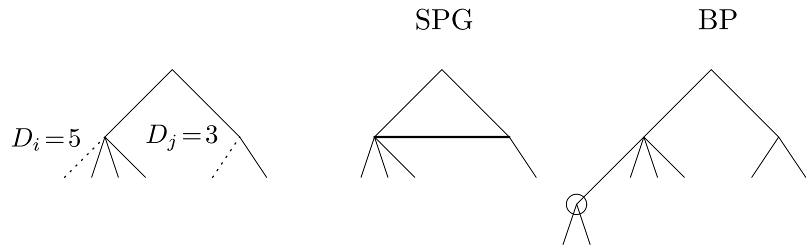

Figure 4: Example of the coupling when a cycle occurs. Edges have twice the length of stubs. In the SPG the two dotted stubs in the left picture are to be connected. The middle picture gives the result of creating the cycle in the SPG where the bold line is the edge creating the cycle. The third figure draws the BP where the cycle is removed and the degree of the circled node is 3.

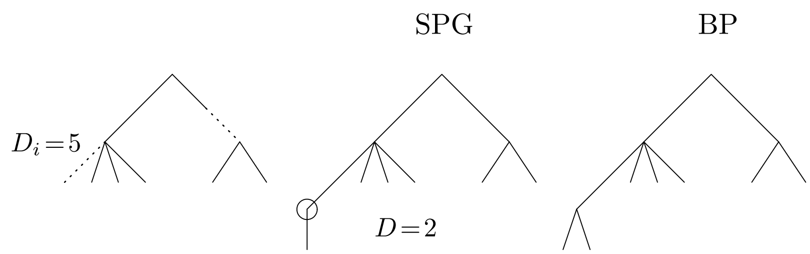

Figure 5: An example of the coupling where we need to perform a redraw. In the draw from $g^{(N)}$, we draw the dotted stub in the SPG with degree 3. In the BP, we keep this degree, while in the SPG we draw again from the conditional distribution given that we do not draw a stub with label 3. In this example, this redraw gives the value $D = 2$.

that all subsequent iterations in the process are i.i.d. Note that this includes the case where we draw the present stub, which of course is impossible for the SPG.

The rule now for the BP is that the corresponding node with the prescribed number of stubs is simply attached. Since for the SPG, we sample *without* replacement, we have to resample from distribution (3.1), until we draw a stub with label 1 or 2. This procedure is referred to as *a redraw*. Since we sample uniformly from all stubs, the conditional sampling until we hit a stub with label 1 or 2 is also uniform out of the set of all stubs with labels 1 and 2, so that it has the correct distribution. Obviously there are two cases: either we draw a stub with label 1 or one with label 2. When we draw a stub with label 1 in the SPG then we update as under rule 1 above, while when we draw a stub having label 2 in the SPG, we update as under rule 2 above.

Clearly, the redraws and the cycles cause possible differences between the BP and the SPG: the degrees of the chosen node are possibly different. We will need to show that the above difference only leads to an error term.

The above process stops in the j^{th} generation when there are no more free stubs in generation $j-1$ for either the BP or for the SPG. When there are no more free stubs for the SPG, we complete the j^{th} generation for the BP by drawing from distribution (3.1) for all the remaining free stubs. The labels of the stubs remain unchanged. When there are no more free stubs for the BP, we complete the j^{th} generation for the SPG by drawing from distribution (3.1) iteratively until we draw a stub with label 1 or 2. This is done for all the remaining free stubs in the j^{th} generation of the SPG. The labels are updated as under 1 and 2 above.

We continue the above process of drawing stubs until there are no more stubs having label 1 or 2, so that all stubs have label 3. Then, the construction is finalized, and we have generated the SPG as seen from node 1. We have thus obtained the structure of the SPG, and know how many nodes there are at a given distance from node 1.

The above construction will be performed similarly from node 2. This construction is close to being independent as long as the SPG's from the roots 1 and 2 do not share any nodes. More precisely, the corresponding BP's are independent. Thus, we have now constructed the SPG's and BP's from both node 1 and node 2.

3.2 Coupling with a BP with offspring distribution $\{g_j^{(N)}\}$

In the previous section, we have obtained a coupling of the SPG and the BP with offspring distribution $\{g_j^{(N)}\}$. In this and the next section, we will summarize bounds on the couplings that we need for the proof of Theorems 1.1 and 1.4. These results will be repeated in the appendix together with a full proof. We start with the coupling of the number of stubs $Z_j^{(1)}$ in the SPG and the number of children $\hat{Z}_j^{(1)}$ in the j^{th} generation of the BP with offspring distribution $\{g_j^{(N)}\}$.

Proposition 3.2 (Coupling SPG with the BP with random offspring distribution) *There exist $\eta, \beta > 0$, $\alpha > \frac{1}{2} + \eta$ and a constant C, such that for all $j \leq (\frac{1}{2} + \eta)\log_\nu N$,*

$$\mathbb{P}\Big((1 - N^{-\alpha}\nu^j)\hat{Z}_j^{(1)} \leq Z_j^{(1)} \leq (1 + N^{-\alpha}\nu^j)\hat{Z}_j^{(1)}\Big) \geq 1 - CjN^{-\beta}. \tag{3.4}$$

3.3 Coupling with a BP with offspring distribution $\{g_j\}$

We next describe the coupling with the BP with offspring distribution $\{g_j\}$ and their bounds. A classical coupling argument is used (see e.g. [39]). Let $X^{(N)}$ have law $\{g_j^{(N)}\}$ and X have law $\{g_j\}$. We define $Y^{(N)}$ by

$$\mathbb{P}(Y^{(N)} = n) = \min(g_n^{(N)}, g_n), \qquad \mathbb{P}(Y^{(N)} = \infty) = 1 - \sum_{n=0}^{\infty} \min(g_n^{(N)}, g_n) = \frac{1}{2}\sum_{n=0}^{\infty} |g_n^{(N)} - g_n|. \tag{3.5}$$

Let $\hat{X}^{(N)} = Y^{(N)}$ when $Y^{(N)} < \infty$, and $\mathbb{P}(X^{(N)} = n, Y^{(N)} = \infty) = g_n^{(N)} - \min(g_n^{(N)}, g_n)$, whereas $\hat{X} = X$ when $Y^{(N)} < \infty$, and $\mathbb{P}(X = n, Y^{(N)} = \infty) = g_n - \min(g_n^{(N)}, g_n)$. Then $\hat{X}^{(N)}$ has law $g^{(N)}$, and $\hat{X}$ has law g. Moreover, with large probability, $\hat{X}^{(N)} = \hat{X}$ due to Proposition 3.4 below.

This coupling argument is applied to each node in the BP $\{\hat{Z}_i^{(1)}\}_{i\geq 0}$ and $\{\hat{Z}_i^{(2)}\}_{i\geq 0}$. The BP's with offspring distribution $\{g_j\}$ will be denoted by $\{\mathcal{Z}_i^{(1)}\}_{i\geq 0}$ and $\{\mathcal{Z}_i^{(2)}\}_{i\geq 0}$. We can interpret this coupling as follows. Each node has an i.i.d. indicator variable which equals one with probability

$$p_N = \frac{1}{2}\sum_{n=0}^{\infty} |g_n^{(N)} - g_n|. \tag{3.6}$$

When at a certain node this indicator variable is 0, then the offspring in $\{\hat{Z}_i^{(1)}\}_{i\geq 0}$ or $\{\hat{Z}_i^{(2)}\}_{i\geq 0}$ equals the one in $\{\mathcal{Z}_i^{(1)}\}_{i\geq 0}$ or $\{\mathcal{Z}_i^{(2)}\}_{i\geq 0}$, and the node is successfully coupled. When the indicator is 1, then an error has occurred, and the coupling is not successful. In this case, the laws of the offspring of $\{\hat{Z}_i^{(1)}\}_{i\geq 0}$ or $\{\hat{Z}_i^{(2)}\}_{i\geq 0}$ is different from the one in $\{\mathcal{Z}_i^{(1)}\}_{i\geq 0}$ or $\{\mathcal{Z}_i^{(2)}\}_{i\geq 0}$, and we record an error. Below we will use the notation $\mathbb{P}_N$ to denote the conditional expectation given $D_1, D_2, \ldots, D_N$ and $\mathbb{E}_N$ to denotes the expectation with respect to the probability measure $\mathbb{P}_N$. Finally, we write

$$\nu_N = \sum_{n=0}^{\infty} n g_n^{(N)}. \tag{3.7}$$

In the following proposition, we prove that at any fixed time, we can couple the SPG to the delayed BP with law $\{g_j\}$:

Proposition 3.3 (Coupling at fixed time) *For any $m \in \mathbb{N}$ fixed, there exist independent branching processes $\mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}$, such that*

$$\lim_{N\to\infty} \mathbb{P}(Z_m^{(i)} = \mathcal{Z}_m^{(i)}) = 1. \tag{3.8}$$

In the course of the proof we will also rely on the following more technical claims:

Proposition 3.4 (Convergence in total variation distance) *There exist $\alpha_2, \beta_2 > 0$ such that*

$$\mathbb{P}\Big(\sum_{n=0}^{\infty}(n+1)|g_n^{(N)} - g_n| \geq N^{-\alpha_2}\Big) \leq N^{-\beta_2}. \tag{3.9}$$

Consequently,

$$\mathbb{P}(|\nu_N - \nu| > N^{-\alpha_2}) \leq N^{-\beta_2}, \tag{3.10}$$

and

$$\mathbb{P}(p_N > N^{-\alpha_2}) \leq N^{-\beta_2}. \tag{3.11}$$

Proposition 3.5 (Coupling of sums) *There exist $\varepsilon, \beta, \eta > 0$ such that for all $j \leq (1+2\eta)\log_\nu N$, as $N \to \infty$,*

$$\mathbb{P}\Big(\frac{1}{N}\Big|\sum_{i=1}^{j} \mathcal{Z}^{(1)}_{\lceil i/2\rceil}\mathcal{Z}^{(2)}_{\lfloor i/2\rfloor} - \sum_{i=1}^{j} \hat{Z}^{(1)}_{\lceil i/2\rceil}\hat{Z}^{(2)}_{\lfloor i/2\rfloor}\Big| > N^{-\varepsilon}\Big) = O(N^{-\beta}). \tag{3.12}$$

4 Proof of Theorem 1.1 and 1.4

The proof consists of four steps.

1. We first express the survival probability $\mathbb{P}(H_N > j)$ in the number of stubs $\{Z_i^{(k)}\}, k = 1, 2$, of the SPG's. For $j \leq (1+2\eta)\log_\nu N$, where η is specified in Proposition 3.2, we will show that

$$\mathbb{P}(H_N > j) = \mathbb{E}\left[\exp\left\{\frac{-\sum_{i=2}^{j+1} Z^{(1)}_{\lceil i/2\rceil} Z^{(2)}_{\lfloor i/2\rfloor}}{L_N}\right\} + O\left(RM_N(j)\right)\right], \tag{4.1}$$

   with

$$RM_N(j) = \sum_{i=2}^{j+1} \frac{Z^{(1)}_{\lceil i/2\rceil} Z^{(2)}_{\lfloor i/2\rfloor} \sum_{k=1}^{\lceil i/2\rceil}(Z_k^{(1)} + Z_k^{(2)})}{L_N^2}.$$

2. We use Proposition 3.2 to show that in (4.1) we can replace $\{Z_k^{(i)}\}, i = 1, 2$ by the BP $\{\hat{Z}_k^{(i)}\}, i = 1, 2$. The error term $\mathbb{E}[|RM_N(j)|]$ and the error involved in replacing the SPG by the BP is bounded by a constant times $N^{-\beta}$, for some $\beta > 0$, uniformly in $j \leq (1+2\eta)\log_\nu N$.

3. In this step we show that there exists $\beta > 0$ such that for all $j \leq (1+2\eta)\log_\nu N$, as $N \to \infty$,

$$\mathbb{P}(H_N > j) = \mathbb{E}\left[\exp\left\{\frac{-\sum_{i=2}^{j+1} \mathcal{Z}^{(1)}_{\lceil i/2\rceil} \mathcal{Z}^{(2)}_{\lfloor i/2\rfloor}}{\mu N}\right\}\right] + O(N^{-\beta}), \tag{4.2}$$

   where $\mathcal{Z}_k^{(i)}$, $i = 1, 2$, denotes the delayed BP with offspring distribution (1.6).

4. We complete the proof of Theorem 1.1 and 1.4, using *step 3*, and the almost sure limit in (1.7) applied to $\mathcal{Z}_n^{(1)}$ and $\mathcal{Z}_n^{(2)}$. We finally use the speed of convergence of the above martingale limit result to obtain (1.9).

Step 1: A formula for $\mathbb{P}(H_N > j)$. The following lemma expresses $\mathbb{P}(H_N > j)$ in terms of $\mathbb{Q}_Z^{(k,l)}$, the conditional probabilities given $\{Z_s^{(1)}\}_{s=1}^k$ and $\{Z_s^{(2)}\}_{s=1}^l$. For $l = 0$, we only condition on $\{Z_s^{(1)}\}_{s=1}^k$.

Lemma 4.1 *For $j \geq 1$,*

$$\mathbb{P}(H_N > j) = \mathbb{E}\Big[\prod_{i=2}^{j+1} \mathbb{Q}_Z^{(\lceil i/2\rceil,\lfloor i/2\rfloor)}(H_N > i-1 | H_N > i-2)\Big]. \tag{4.3}$$

Proof. We first compute that

$$\mathbb{P}(H_N > j) = \mathbb{E}\big[\mathbb{Q}_Z^{(1,1)}(H_N > j)\big] = \mathbb{E}\big[\mathbb{Q}_Z^{(1,1)}(H_N > 1)\mathbb{Q}_Z^{(1,1)}(H_N > j | H_N > 1)\big].$$

Continuing this further, and writing $\mathbb{E}_Z^{(k,l)}$ for the expectation with respect to $\mathbb{Q}_Z^{(k,l)}$,

$$\begin{aligned}\mathbb{Q}_Z^{(1,1)}(H_N > j|H_N > 1) &= \mathbb{E}_Z^{(1,1)}\big[\mathbb{Q}_Z^{(2,1)}(H_N > j|H_N > 1)\big] \\ &= \mathbb{E}_Z^{(1,1)}\big[\mathbb{Q}_Z^{(2,1)}(H_N > 2|H_N > 1)\mathbb{Q}_Z^{(2,1)}(H_N > j|H_N > 2)\big].\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}(H_N > j) &= \mathbb{E}\big[\mathbb{Q}_Z^{(1,1)}(H_N > 1)\mathbb{E}_Z^{(1,1)}\big[\mathbb{Q}_Z^{(2,1)}(H_N > 2|H_N > 1)\mathbb{Q}_Z^{(2,1)}(H_N > j|H_N > 2)\big]\big] \\ &= \mathbb{E}\big[\mathbb{E}_Z^{(1,1)}\big[\mathbb{Q}_Z^{(1,1)}(H_N > 1)\mathbb{Q}_Z^{(2,1)}(H_N > 2|H_N > 1)\mathbb{Q}_Z^{(2,1)}(H_N > j|H_N > 2)\big]\big] \\ &= \mathbb{E}\big[\mathbb{Q}_Z^{(1,1)}(H_N > 1)\mathbb{Q}_Z^{(2,1)}(H_N > 2|H_N > 1)\mathbb{Q}_Z^{(2,1)}(H_N > j|H_N > 2)\big],\end{aligned}$$

where, in the second equality, we use that $\mathbb{Q}_Z^{(1,1)}(H_N > 1)$ is measurable with respect to the σ-algebra generated by $Z_1^{(1,N)}$. This proves the claim for $j = 2$.

More generally, we obtain that for k, l such that $k + l \leq j - 1$,

$$\begin{aligned}&\mathbb{Q}_Z^{(k,l)}(H_N > j|H_N > k+l-1) = \mathbb{E}_Z^{(k,l)}\big[\mathbb{Q}_Z^{(k,l+1)}(H_N > j|H_N > k+l-1)\big] \\ &\quad = \mathbb{E}_Z^{(k,l)}\big[\mathbb{Q}_Z^{(k,l+1)}(H_N > k+l|H_N > k+l-1)\mathbb{Q}_Z^{(k,l+1)}(H_N > j|H_N > k+l)\big],\end{aligned}$$

and, similarly,

$$\mathbb{Q}_Z^{(k,l)}(H_N > j|H_N > k+l-1) = \mathbb{E}_Z^{(k,l)}\big[\mathbb{Q}_Z^{(k+1,l)}(H_N > k+l|H_N > k+l-1)\mathbb{Q}_Z^{(k+1,l)}(H_N > j|H_N > k+l)\big].$$

In the above formulas, we can choose to increase k or l by one depending on $\{Z_s^{(1,N)}\}_{s=1}^k$ and $\{Z_s^{(2,N)}\}_{s=1}^l$. We will iterate the above recursions, until $k + l = j - 1$, when the last term becomes 1. This yields that

$$\mathbb{P}(H_N > j) = \mathbb{E}\Big[\prod_{i=1}^{j} \mathbb{Q}_Z^{(\lfloor i/2\rfloor+1,\lceil i/2\rceil)}(H_N > i|H_N > i-1)\Big]. \tag{4.4}$$

Renumbering gives the final result. □

We will next prove (4.1). In order to do so, we start by proving upper and lower bounds on the probabilities of not connecting two sets of stubs to each other. For this, suppose we have two disjoint sets of stubs A with $|A| = n$ and B with $|B| = m$ out of a total of L stubs. We match stubs at random, in such a way that two stubs form one edge, as in the construction of the SPG. In particular, loops are possible.

Let $p(n, m, L)$ denote the probability that none of the n stubs in A attaches to one of the m stubs in B. Then, by conditioning on whether we choose a stub in A or not, we obtain the recursion

$$p(n, m, L) = \frac{n-1}{L-1}p(n-2, m, L-2) + \left(1 - \frac{m+n-1}{L-1}\right)p(n-1, m, L-2) \tag{4.5}$$

Since $p(n-2,m,L-2) \geq p(n-1,m,L-2)$, because we have to match one additional stub, we obtain

$$p(n,m,L) \geq \left(1-\frac{m}{L-1}\right) p(n-1,m,L-2) \geq \prod_{i=0}^{n-1}\left(1-\frac{m}{L-2i-1}\right). \tag{4.6}$$

On the other hand, we can rewrite (4.5) as

$$p(n,m,L) = \left(1-\frac{m}{L-1}\right) p(n-1,m,L-2) + \frac{n-1}{L-1}\left(p(n-2,m,L-2) - p(n-1,m,L-2)\right). \tag{4.7}$$

We claim that

$$p(n-2,m,L-2) - p(n-1,m,L-2) = \frac{m}{L-3} p(n-2,m-1,L-2) \leq \frac{m}{L-3}. \tag{4.8}$$

Indeed, the difference $p(n-2,m,L-2) - p(n-1,m,L-2)$ is equal to the probability of the event that the first $n-2$ stubs do not connect to B, while the last one does. By exchangeability of the stubs, this probability equals the probability that the first stub is attached to a stub in B, and the remaining $n-2$ stubs are not. This latter probability is equal to $\frac{m}{L-3}p(n-2,m-1,L-2)$.

The equations (4.7) and (4.8) yield

$$p(n,m,L) \leq \left(1-\frac{m}{L-1}\right) p(n-1,m,L-2) + \frac{n-1}{(L-1)}\frac{m}{(L-3)}.$$

Iteration gives the upper bound

$$p(n,m,L) \leq \left[\prod_{i=0}^{n-1}\left(1-\frac{m}{L-2i-1}\right)\right] + \frac{n^2 m}{(L-2n)^2}. \tag{4.9}$$

Since the event $\{H_N > 1\}$ holds if and only if no stubs of root 1 attaches to one of those of root 2, we obtain, using (4.6) and (4.9), that

$$\prod_{i=0}^{Z_1^{(1)}-1}\left(1-\frac{Z_1^{(2)}}{L_N-2i-1}\right) \leq \mathbb{Q}_Z^{(1,1)}(H_N>1) \leq \left[\prod_{i=0}^{Z_1^{(1)}-1}\left(1-\frac{Z_1^{(2)}}{L_N-2i-1}\right)\right] + \frac{(Z_1^{(1)})^2 Z_1^{(2)}}{(L_N-2Z_1^{(1)})^2}. \tag{4.10}$$

Similarly,

$$\mathbb{Q}_Z^{(2,1)}(H_N>2 | H_N>1) \geq \prod_{i=0}^{Z_1^{(2)}-1}\left(1-\frac{Z_2^{(1)}}{L_N-2Z_1^{(1)}-2i-1}\right), \tag{4.11}$$

with a matching upper bound with an error term bounded by $\frac{(Z_1^{(2)})^2 Z_2^{(1)}}{(L_N-2Z_1^{(1)}-2Z_1^{(2)})^2}$.

We use that, for natural numbers n,m,M with $M+n+m = o(L)$,

$$\prod_{i=0}^{n-1}\left(1-\frac{m}{L-M-2i-1}\right) = e^{-\frac{nm}{L}} + O\left(\frac{nm(M+n+m)}{L^2}\right), \quad L\to\infty. \tag{4.12}$$

Using (4.12), the bounds in (4.10) yield

$$\mathbb{Q}_Z^{(1,1)}(H_N>1) = \exp\left\{-\frac{Z_1^{(1)}Z_1^{(2)}}{L_N}\right\} + O\left(\frac{Z_1^{(1)}Z_1^{(2)}(Z_1^{(1)}+Z_1^{(2)})}{L_N^2}\right).$$

Similarly, we can conclude that, as long as $\sum_{k=1}^{\lceil i/2\rceil}(Z_k^{(1)}+Z_k^{(2)})=o(L_N)$, we have

$$\begin{aligned}&\mathbb{Q}_Z^{(\lceil i/2\rceil,\lfloor i/2\rfloor)}(H_N>i-1|H_N>i-2)\\&=\exp\left\{-\frac{Z_{\lceil i/2\rceil}^{(1)}Z_{\lfloor i/2\rfloor}^{(2)}}{L_N}\right\}+O\left(\frac{Z_{\lceil i/2\rceil}^{(1)}Z_{\lfloor i/2\rfloor}^{(2)}(\sum_{k=1}^{\lceil i/2\rceil}(Z_k^{(1)}+Z_k^{(2)}))}{L_N^2}\right).\end{aligned}\tag{4.13}$$

Recall that

$$RM_N(j)=\sum_{i=2}^{j+1}\frac{Z_{\lceil i/2\rceil}^{(1)}Z_{\lfloor i/2\rfloor}^{(2)}\sum_{k=1}^{\lceil i/2\rceil}(Z_k^{(1)}+Z_k^{(2)})}{L_N^2}.$$

With the assumption that $\sum_{k=1}^{\lceil i/2\rceil}(Z_k^{(1)}+Z_k^{(2)})=o(L_N)$, the bound in (4.1) becomes evident by applying (4.3). We will show at the end of step 2 that for all $j<(1+2\eta)\log N$, we have $\sum_{k=1}^{\lceil j/2\rceil}(Z_k^{(1)}+Z_k^{(2)})=o(L_N)$ and that there exists a $\beta>0$ such that

$$\mathbb{E}[RM_N(j)]=O(N^{-\beta}).\tag{4.14}$$

Step 2: Coupling of SPG to the BP with offspring $\{g_j^{(N)}\}$. We start by showing that for some $\beta>0$ and uniformly in $j\le(1+2\eta)\log_\nu N$, the main term in (4.1) satisfies

$$\mathbb{E}\left[\exp\left\{\frac{-\sum_{i=2}^{j+1}Z_{\lceil i/2\rceil}^{(1)}Z_{\lfloor i/2\rfloor}^{(2)}}{L_N}\right\}\right]=\mathbb{E}\left[\exp\left\{\frac{-\sum_{i=2}^{j+1}\hat{Z}_{\lceil i/2\rceil}^{(1)}\hat{Z}_{\lfloor i/2\rfloor}^{(2)}}{L_N}\right\}\right]+O(N^{-\beta}).\tag{4.15}$$

We will deal with the error term (4.14) at the end of this step. Bound

$$|\sum_{i=2}^{j+1}Z_{\lceil i/2\rceil}^{(1)}Z_{\lfloor i/2\rfloor}^{(2)}-\hat{Z}_{\lceil i/2\rceil}^{(1)}\hat{Z}_{\lfloor i/2\rfloor}^{(2)}|\le\sum_{i=2}^{j+1}Z_{\lceil i/2\rceil}^{(1)}|Z_{\lfloor i/2\rfloor}^{(2)}-\hat{Z}_{\lfloor i/2\rfloor}^{(2)}|+\sum_{i=2}^{j+1}\hat{Z}_{\lfloor i/2\rfloor}^{(2)}|Z_{\lceil i/2\rceil}^{(1)}-\hat{Z}_{\lceil i/2\rceil}^{(1)}|.$$

By Proposition 3.2 and uniformly in $j\le(1+2\eta)\log_\nu N$, we have, with probability exceeding $1-O(N^{-\beta}\log_\nu N)$, that

$$\max\left(\sum_{i=2}^{j+1}Z_{\lceil i/2\rceil}^{(1)}|Z_{\lfloor i/2\rfloor}^{(2)}-\hat{Z}_{\lfloor i/2\rfloor}^{(2)}|,\sum_{i=2}^{j+1}\hat{Z}_{\lfloor i/2\rfloor}^{(2)}|Z_{\lceil i/2\rceil}^{(1)}-\hat{Z}_{\lceil i/2\rceil}^{(1)}|\right)=O(\nu^{(\frac{1}{2}+\eta)\log_\nu N}N^{-\alpha})\sum_{i=2}^{j+1}\hat{Z}_{\lceil i/2\rceil}^{(1)}\hat{Z}_{\lfloor i/2\rfloor}^{(2)}.$$

Since $\alpha>\frac{1}{2}+\eta$, we have $\nu^{(\frac{1}{2}+\eta)\log_\nu N}N^{-\alpha}=N^{\frac{1}{2}+\eta-\alpha}=N^{-\alpha_1}$, for some $\alpha_1>0$. Hence, for any ε with $0<\varepsilon<\alpha_1$, where as before $\mathbb{P}_N$ denotes the conditional probability given the degrees $D_1,D_2,\ldots,D_N$, and $\mathbb{E}_N$ the expectation with respect to $\mathbb{P}_N$, we have

$$\begin{aligned}&\mathbb{P}_N\left(\frac{1}{N}|\sum_{i=2}^{j+1}Z_{\lceil i/2\rceil}^{(1)}Z_{\lfloor i/2\rfloor}^{(2)}-\hat{Z}_{\lceil i/2\rceil}^{(1)}\hat{Z}_{\lfloor i/2\rfloor}^{(2)}|>N^{-\varepsilon}\right)\\&\le O(N^{-\beta}\log_\nu N)+\mathbb{P}_N\left(\frac{1}{N}\sum_{i=2}^{j+1}\hat{Z}_{\lfloor i/2\rfloor}^{(1)}\hat{Z}_{\lceil i/2\rceil}^{(2)}>O\left(N^{\alpha_1-\varepsilon}\right)\right)\\&\le O(N^{-\beta}\log_\nu N)+O\left(N^{\varepsilon-\alpha_1-1}\right)\sum_{i=2}^{j+1}\mathbb{E}_N[\hat{Z}_{\lceil i/2\rceil}^{(1)}\hat{Z}_{\lfloor i/2\rfloor}^{(2)}],\end{aligned}$$

where we have applied the Markov inequality in the last line. The involved conditional expectation can be computed explicitly and we obtain

$$\sum_{i=2}^{j+1}\mathbb{E}_N[\hat{Z}_{\lceil i/2\rceil}^{(1)}\hat{Z}_{\lfloor i/2\rfloor}^{(2)}]=D_1D_2\sum_{i=2}^{j+1}\nu_N^{\lceil i/2\rceil-1}\nu_N^{\lfloor i/2\rfloor-1}=D_1D_2\sum_{i=0}^{j-1}\nu_N^i\le cD_1D_2\nu_N^j,$$

for some constant c. Proposition 3.4 implies that we can bound ν_N^j by $\nu^j(1+N^{-\alpha_2})^j$, with probability exceeding $1-N^{-\beta_2}$, for some $\alpha_2, \beta_2 > 0$, whereas Lemma 3.1 implies L_N^{-1} can be replaced by $(\mu N)^{-1}$ with probability exceeding $1 - N^{-\beta_3}$, for some $\beta_3 > 0$. Putting this together we obtain after taking the expectation with respect to $D_1, D_2, \ldots, D_N$,

$$\mathbb{P}\left(\frac{1}{L_N}\Big|\sum_{i=2}^{j+1} Z^{(1)}_{\lfloor i/2\rfloor} Z^{(2)}_{\lceil i/2\rceil} - \hat{Z}^{(1)}_{\lfloor i/2\rfloor} \hat{Z}^{(2)}_{\lceil i/2\rceil}\Big| > N^{-\varepsilon}\right)$$
$$\leq O(N^{-\beta}\log_\nu N) + O(N^{-\beta_1}) + O(N^{-\beta_2}) + O(N^{-\beta_3}) + O\left(\frac{\nu^j(1+O(\log_\nu N/N^{\alpha_2}))}{N^{1+\alpha_1-\varepsilon}}\right).$$

Since $\nu^j \leq N^{1+2\eta}$ for $j \leq (1+2\eta)\log_\nu N$, we obtain

$$\mathbb{P}\left(\frac{1}{L_N}\Big|\sum_{i=2}^{j+1} Z^{(1)}_{\lceil i/2\rceil} Z^{(2)}_{\lfloor i/2\rfloor} - \hat{Z}^{(1)}_{\lceil i/2\rceil} \hat{Z}^{(2)}_{\lfloor i/2\rfloor}\Big| > N^{-\varepsilon}\right) = O(N^{-\beta}), \tag{4.16}$$

for some $\beta > 0$ by taking $\beta, \beta_2, \beta_3, \eta$ and ε sufficiently small. For $x - y$ small, and $x, y \geq 0$, we find $e^{-y} = e^{-x} + O(x-y)$, so that

$$\exp\left\{-\frac{\sum_{i=2}^{j+1} Z^{(1)}_{\lceil i/2\rceil} Z^{(2)}_{\lfloor i/2\rfloor}}{L_N}\right\} - \exp\left\{-\frac{\sum_{i=2}^{j+1} \hat{Z}^{(1)}_{\lfloor i/2\rfloor} \hat{Z}^{(2)}_{\lceil i/2\rceil}}{L_N}\right\} = O(N^{-\varepsilon}),$$

with probability exceeding $1 - O(N^{-\beta})$. In combination with the inequality $e^{-x} \leq 1$ for $x \geq 0$, we obtain (4.15).

We turn to the proof of (4.14). In the course of the proof of (4.14) we have to verify that the assumption $\sum_{k=1}^{\lceil j/2\rceil}(Z_k^{(1)} + Z_k^{(2)}) = o(L_N)$ holds for all $j \leq (1+2\eta)\log_\nu N$. From Proposition 3.2 and, uniformly in $j \leq (1+2\eta)\log_\nu N$, we have with probability exceeding $1 - O(N^{-\beta}\log_\nu N)$ that

$$\sum_{k=1}^{\lceil j/2\rceil}(Z_k^{(1)} + Z_k^{(2)}) \leq (1 + O(N^{\frac{1}{2}+\eta-\alpha})) \sum_{k=1}^{\lceil j/2\rceil}(\hat{Z}_k^{(1)} + \hat{Z}_k^{(2)}). \tag{4.17}$$

so that, for all $i \leq j$,

$$\mathbb{P}_N\left(\frac{\sum_{k=1}^{\lceil i/2\rceil}(Z_k^{(1)} + Z_k^{(2)})}{L_N^{3/4}} > N^{-\varepsilon}\right) \leq O(N^{-\beta}\log_\nu N) + (1 + O(N^{\frac{1}{2}+\eta-\alpha}))\mathbb{E}_N\left[\frac{\sum_{k=1}^{\lceil j/2\rceil}(\hat{Z}_k^{(1)} + \hat{Z}_k^{(2)})}{N^{-\varepsilon}L_N^{3/4}}\right].$$

Bounding the expectation of $\hat{Z}_k^{(i)}$, we find for $0 < \varepsilon < 1/4$ and for all $i \leq j \leq (1+2\eta)\log_\nu N$,

$$\mathbb{P}\left(\frac{\sum_{k=1}^{\lceil j/2\rceil}(Z_k^{(1)} + Z_k^{(2)})}{L_N^{3/4}} > N^{-\varepsilon}\right) \leq N^{-\beta} + (1 + O(N^{-\alpha_1}))\frac{N^{\frac{1}{2}+\eta}}{N^{\frac{3}{4}-\varepsilon}} = O(N^{-\beta}),$$

for some $\beta > 0$. In particular, we find that $\sum_{k=1}^{\lceil j/2\rceil}(Z_k^{(1)} + Z_k^{(2)}) = o(L_N)$ with probability exceeding $1 - O(N^{-\beta})$. Hence, for $\varepsilon_1 > 0$,

$$\mathbb{P}\left(\sum_{i=2}^{j+1}\frac{Z^{(1)}_{\lceil i/2\rceil} Z^{(2)}_{\lfloor i/2\rfloor}\sum_{k=1}^{\lceil i/2\rceil}(Z_k^{(1)} + Z_k^{(2)})}{L_N^2} > N^{-\varepsilon_1}\right) \leq O(N^{-\beta}) + \mathbb{P}\left(\sum_{i=2}^{j+1}\frac{Z^{(1)}_{\lceil i/2\rceil} Z^{(2)}_{\lfloor i/2\rfloor}}{L_N^{5/4}} > N^{\varepsilon-\varepsilon_1}\right).$$

By Proposition 3.2, the product $Z^{(1)}_{\lceil i/2\rceil} Z^{(2)}_{\lfloor i/2\rfloor}$ can be bounded by $(1 + O(N^{\frac{1}{2}+\eta-\alpha}))\hat{Z}^{(1)}_{\lceil i/2\rceil}\hat{Z}^{(2)}_{\lfloor i/2\rfloor}$ and $\mathbb{E}[\sum_{i=2}^{j+1}\hat{Z}^{(1)}_{\lceil i/2\rceil}\hat{Z}^{(2)}_{\lfloor i/2\rfloor}] \leq N^{1+2\eta}$, while $L_N^{5/4}$ is of order $N^{5/4}$. Therefore, we obtain from the Markov inequality that

$$\mathbb{P}\left(\sum_{i=2}^{j+1}\frac{Z^{(1)}_{\lceil i/2\rceil} Z^{(2)}_{\lfloor i/2\rfloor}\sum_{k=1}^{\lceil i/2\rceil}(Z_k^{(1)} + Z_k^{(2)})}{L_N^2} > N^{-\varepsilon_1}\right) \leq O(N^{-\beta}),$$

for some $\beta > 0$. Since $RM_N(j)$ is the difference of two numbers between 0 and 1 and hence $|RM_N(j)| \leq 1$, we obtain that, when $\varepsilon_1 \geq \beta$,

$$\mathbb{E}[RM_N(j)] \leq N^{-\varepsilon_1} + \mathbb{P}\left(\frac{1}{L_N^2}\sum_{i=2}^{j+1} Z^{(1)}_{\lceil i/2\rceil} Z^{(2)}_{\lfloor i/2\rfloor} \sum_{k=1}^{\lceil i/2\rceil}(Z^{(1)}_k + Z^{(2)}_k) > N^{-\varepsilon_1}\right) \leq O(N^{-\beta}). \tag{4.18}$$

This proves (4.14).

Step 3: Coupling to the BP with offspring $\{g_j\}$. Proposition 3.5 combined with Lemma 3.1 yields

$$\mathbb{P}\Big(\frac{1}{L_N}\Big|\sum_{i=2}^{j+1} \mathcal{Z}^{(1)}_{\lceil i/2\rceil}\mathcal{Z}^{(1)}_{\lfloor i/2\rfloor} - \sum_{i=2}^{j+1} \hat{Z}^{(1)}_{\lceil i/2\rceil}\hat{Z}^{(2)}_{\lfloor i/2\rfloor}\Big| > N^{-\varepsilon}\Big) = O(N^{-\beta}).$$

From this result we obtain, as in the first half of *step 2*,

$$\mathbb{E}\left[\exp\left\{-\frac{\sum_{i=2}^{j+1}\hat{Z}^{(1)}_{\lfloor i/2\rfloor}\hat{Z}^{(2)}_{\lceil i/2\rceil}}{L_N}\right\}\right] = \mathbb{E}\left[\exp\left\{-\frac{\sum_{i=2}^{j+1}\mathcal{Z}^{(1)}_{\lceil i/2\rceil}\mathcal{Z}^{(2)}_{\lfloor i/2\rfloor}}{L_N}\right\}\right] + O(N^{-\beta}),$$

where, as before, β is a generic small positive number. Using (4.1) and the result of *step 2*, it follows that

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

To obtain (4.2), we finally replace, again at the cost of an additional term $O(N^{-\beta})$, the random number L_N by $\mu N(1 + O(N^{-a}))$.

Step 4: Evaluation of the limit points. We start from (4.2) with $j = k + \sigma_N \leq (1+2\eta)\log_\nu N$, where $\sigma_N = \lfloor \log_\nu N\rfloor$, to obtain

$$\mathbb{P}(H_N > \sigma_N + k) = \mathbb{E}\left[\exp\left\{\frac{-\sum_{i=2}^{\sigma_N+k+1}\mathcal{Z}^{(1)}_{\lceil i/2\rceil}\mathcal{Z}^{(2)}_{\lfloor i/2\rfloor}}{\mu N}\right\}\right] + O(N^{-\beta}). \tag{4.19}$$

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

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

In the above expression, the factor ν^{a_N} prevents proper convergence. Without the factor $\mu\nu^{a_N+k}$, we obtain from (1.7), with probability 1,

$$\lim_{N\to\infty}\frac{\sum_{i=2}^{\sigma_N+k+1}\mathcal{Z}^{(1)}_{\lceil i/2\rceil}\mathcal{Z}^{(2)}_{\lfloor i/2\rfloor}}{\mu^2\nu^{\sigma_N+k}} = \frac{\mathcal{W}^{(1)}\mathcal{W}^{(2)}}{\nu-1}. \tag{4.20}$$

A proof of this result is given at the end of the appendix. Using (2.3) we conclude that for each $\alpha > 0$, there is a $\beta > 0$ such that

$$\mathbb{P}\left(\left|\frac{\sum_{i=2}^{\sigma_N+k+1}\mathcal{Z}^{(1)}_{\lceil i/2\rceil}\mathcal{Z}^{(2)}_{\lfloor i/2\rfloor}}{\mu^2\nu^{\sigma_N+k}} - \frac{\mathcal{W}^{(1)}\mathcal{W}^{(2)}}{\nu-1}\right| > O((\log N)^{-\alpha})\right) = O(N^{-\beta}).$$

Hence, for $k \leq 2\eta\log_\nu N$ and each $\alpha > 0$,

$$\mathbb{P}(H_N > \sigma_N + k) = \mathbb{E}\big(\exp\{-\kappa\nu^{a_N+k}\mathcal{W}^{(1)}\mathcal{W}^{(2)}\}\big) + O((\log N)^{-\alpha}), \tag{4.21}$$

where $\kappa = \mu/(\nu-1)$. This proves (1.9).

We proceed by proving (1.4), with R_a given in (1.8). For this, we need to condition on node 1 and node 2 being connected. Node 1 and node 2 are connected if and only if $H_N < \infty$. Using (4.21), for (1.4), it suffices to prove that

$$\mathbb{P}(H_N < \infty) = q^2 + o(1), \quad \text{where} \quad q = \mathbb{P}(\mathcal{W}^{(1)} > 0). \tag{4.22}$$

We prove (4.22) using upper and lower bounds. We note that, with $k = \eta \log_\nu N$,

$$\mathbb{P}(H_N < \infty) \geq \mathbb{P}(H_N \leq \sigma_N + k) = \mathbb{E}\big(1 - \exp\{-\kappa \nu^{a_N+k} \mathcal{W}^{(1)}\mathcal{W}^{(2)}\}\big) + O((\log N)^{-\alpha}). \tag{4.23}$$

Therefore,

$$\mathbb{P}(H_N < \infty) \geq q^2 \mathbb{E}\big(1 - \exp\{-\kappa \nu^{a_N+k} \mathcal{W}^{(1)}\mathcal{W}^{(2)}\}\big|\mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0\big) + O((\log N)^{-\alpha}). \tag{4.24}$$

By dominated convergence, for $k = 2\eta \log_\nu N$, the conditional expectation converges to 1, so that indeed $\mathbb{P}(H_N < \infty) \geq q^2 + o(1)$. For the upper bound, we rewrite, for any m,

$$\mathbb{P}(H_N < \infty) = \mathbb{P}(H_N < \infty, Z_m^{(1)} Z_m^{(2)} = 0) + \mathbb{P}(H_N < \infty, Z_m^{(1)} Z_m^{(2)} > 0). \tag{4.25}$$

The second term is bounded from above by

$$\mathbb{P}(H_N < \infty, Z_m^{(1)} Z_m^{(2)} > 0) \leq \mathbb{P}(Z_m^{(1)} Z_m^{(2)} > 0) = \mathbb{P}(\mathcal{Z}_m^{(1)} \mathcal{Z}_m^{(2)} > 0) + o(1) = q_m^2 + o(1), \tag{4.26}$$

where we use Proposition 3.3, and we write $q_m = \mathbb{P}(\mathcal{Z}_m^{(1)} > 0)$. When $m \to \infty$, we have that $q_m \to q$, so that we are done when we can show that for any m fixed, $\mathbb{P}(H_N < \infty, Z_m^{(1)} Z_m^{(2)} = 0) = o(1)$. We note that if $Z_m^{(1)} Z_m^{(2)} = 0$ and $H_N < \infty$, then $H_N \leq m-1$. Therefore, using (4.21) with $k = m - \sigma_N - 1$, we conclude

$$\mathbb{P}(H_N < \infty, Z_m^{(1)} Z_m^{(2)} = 0) \leq \mathbb{P}(H_N \leq m-1) = \mathbb{E}\big(1 - \exp\{-\kappa \nu^{a_N+k} \mathcal{W}^{(1)}\mathcal{W}^{(2)}\}\big) + o(1) = o(1). \tag{4.27}$$

This completes the proof of (4.22). We finally complete the proof of Theorems 1.1 and 1.4 using (4.22), which, together with (4.21), implies that, for $k \leq 2\eta \log_\nu N$,

$$\mathbb{P}(H_N \leq \sigma_N + k | H_N < \infty) = \mathbb{E}\big(1 - \exp\{-\kappa \nu^{a_N+k} \mathcal{W}^{(1)}\mathcal{W}^{(2)}\}|\mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0\big) + o(1). \tag{4.28}$$

□

5 On the connected components

In this section, we will investigate the sizes of the connected components and prove Theorem 1.5.

Proof of Theorem 1.5. In the proof, we will make essential use of the results in [31, 32], where the statement in Theorem 1.5 is proved for certain degree sequences. Indeed, denote by

$$d_i(N) = \sum_{j=1}^{N} I[D_j = i], \qquad i = 0, 1, \ldots, \tag{5.1}$$

the degree sequence of our random graph G, where $D_1, D_2, \ldots, D_N$ is the i.i.d. sequence with distribution F introduced in (1.1) and satisfying (1.2). In [31], the bounds on the connected components in Theorem 1.5 are proved with only a lower bound on the largest connected component size, while in [32], the asymptotic size of the largest connected component is determined. Both papers assume a number of hypotheses on the degree sequence $\{d_i(N)\}_{i \geq 0}$. Thus, Theorem 1.5 follows when we can show that the probability that our degree sequences in (5.1) satisfy the restrictions is at least $1 - o(1)$. In fact, we need to alter the random graph G in a certain way to meet the conditions of

Molloy and Reed, and subsequently need to prove that the alteration does not affect the results. We now go over their conditions and definitions.

Firstly, the degree sequence needs to be **feasible**, meaning that there exists at least one graph with the degree sequence. This is true, since L_N is even and we have that

$$\sum_{i=1}^{\infty} i d_i(N) = \sum_{i=1}^{\infty} i \sum_{j=1}^{N} I[D_j = i] = \sum_{j=1}^{N}\sum_{i=1}^{\infty} i I[D_j = i] = \sum_{j=1}^{N} D_j = L_N.$$

Secondly, the degree sequence needs to be **smooth**, meaning that for some sequence λ_i, we have

$$\lim_{N\to\infty} \frac{d_i(N)}{N} = \lambda_i.$$

In our setting, this follows almost surely from the law of large numbers, with $\lambda_i = f_i = \mathbb{P}(D = i)$.

Thirdly, and this is the most serious condition, the degree sequence needs to be **well-behaved**, meaning that it is smooth, feasible, and that for every ϵ', there exists $N' = N'(\epsilon')$, such that for all $N > N'$, we have that

1.
$$\sup_i \left| i(i-2)\frac{d_i(N)}{N} - i(i-2)\lambda_i \right| < \epsilon'; \tag{5.2}$$

2. there exists i^*
$$\Big| \sum_{i=1}^{i^*} i(i-2)\frac{d_i(N)}{N} - \sum_{i=1}^{\infty} i(i-2)\lambda_i \Big| \leq \epsilon'; \tag{5.3}$$

3. there exists an $\epsilon > 0$ such that $d_i(N) = 0$ for all $i \geq \lceil N^{\frac{1}{4}-\epsilon}\rceil$.

We start with the last assumption, which is not satisfied by our graph. Indeed, the last restriction means that all nodes have degree at most $\lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$. We will first alter the graph, and thus the degree sequences, in the following way. Fix $\epsilon > 0$ small. For nodes j with $D_j \geq \lceil N^{\frac{1}{4}-\epsilon}\rceil$, we remove $D_j - \lceil N^{\frac{1}{4}-\epsilon}\rceil + 1$ edges. We do this by first removing at random edges between pairs i, j where the degrees of D_i and D_j *both* exceed $\lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$, and stop when the degree of D_i or D_j (or both) is (for the first time) smaller than $\lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$ (removing at random here means with equal probabilities for all edges between i and j). When there are no more edges between pairs of nodes with both degrees exceeding $\lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$, we remove edges from the nodes with degrees exceeding $\lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$. We do this by deleting at random the necessary number of edges so that the degree becomes $\lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$. Thus, we end up with a graph G' such that all degrees are at most $\lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$. Moreover each node j for which $D_j \geq \lceil N^{\frac{1}{4}-\epsilon}\rceil$ has degree equal to $\lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$ in the altered graph G'. This will be the graph to which we apply the results of Molloy and Reed. Let D_j' be the degree of the node j in G', and write $d_i'(N)$ for the number of nodes with degree equal to i in G'. Then $d_i'(N) = 0$ for $i \geq \lceil N^{\frac{1}{4}-\epsilon}\rceil$, as required.

We first compute the number of removed edges, which we denote by R_N. Its expectation is bounded above by

$$\begin{aligned}\mathbb{E}[R_N] &\leq \mathbb{E}\Big[\sum_{j=1}^{N} (D_j + 1 - \lceil N^{\frac{1}{4}-\epsilon}\rceil)^+ I[D_j \geq \lceil N^{\frac{1}{4}-\epsilon}\rceil]\Big] \leq N \sum_{l \geq \lceil N^{\frac{1}{4}-\epsilon}\rceil - 1} \mathbb{P}(D_1 > l) \\ &\leq cN \sum_{l \geq \lceil N^{\frac{1}{4}-\epsilon}\rceil - 1} l^{-\tau+1} = CN^{1-(\tau-2)(\frac{1}{4}-\epsilon)} < N^{\frac{3}{4}},\end{aligned}$$

for $\tau > 3$ and ϵ sufficiently small. We are hence removing only a fraction of the L_N available edges (see Lemma 3.1 that L_N is close to μN). Moreover, with probability converging to one, we have that $R_N \leq 2N^{\frac{3}{4}}$, since by a computation analogous to the one given above for $\mathbb{E}[R_N]$, we have $\mathrm{Var}(R_N) \leq CN^{1-(\tau-3)(\frac{1}{4}-\epsilon)}$, so that by the Chebychev inequality,

$$\mathbb{P}(R_N > 2N^{\frac{3}{4}}) \leq \mathbb{P}(|R_N - \mathbb{E}[R_N]| > N^{\frac{3}{4}}) \leq N^{-\frac{3}{2}}\mathrm{Var}(R_N) \leq CN^{-\frac{1}{2}-(\tau-3)(\frac{1}{4}-\epsilon)} \leq CN^{-\frac{1}{2}}. \qquad (5.4)$$

We start by checking (5.2) for the graph G', with $\lambda_i = f_i$ in (1.1). For this, we will use the following bound from [9, Corollary 1.4(i)], which states that if S_N is binomial with parameters N and p, and if $x = (Np(1-p))^{1/2} \geq 1$, then

$$\mathbb{P}\big(|S_N - Np| \geq x(Np(1-p))^{1/2}\big) \leq \frac{1}{x}e^{-x^2/2}. \qquad (5.5)$$

We first check condition (5.2) for $i = \lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$. By construction, we have that for $i = \lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$,

$$d_i'(N) = \sum_{j\geq i} d_j(N). \qquad (5.6)$$

Hence, $d_i'(N)$ is a binomial random variable with parameters N and $p = 1 - F(\lceil N^{\frac{1}{4}-\epsilon}\rceil - 2)$, where F is the distribution function in (1.1). Thus, by (5.5), with $x = C\sqrt{\log N}$, we have that

$$\mathbb{P}\big(|d_i'(N) - Np| \geq C((\log N)Np(1-p))^{1/2}\big) \leq \frac{1}{C}N^{-C^2/2} = o(1). \qquad (5.7)$$

Thus, we have with probability $1 - o(1)$ that for $i = \lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$, $\lambda_i = f_i$ and $p = 1 - F(\lceil N^{\frac{1}{4}-\epsilon}\rceil - 2)$,

$$\begin{aligned} i(i-2)\Big|\frac{d_i'(N)}{N} - \lambda_i\Big| &\leq i^2\Big|\frac{d_i'(N)}{N} - p + p - \lambda_i\Big| \leq i^2\Big|\frac{d_i'(N)}{N} - p\Big| + i^2|p - \lambda_i| \\ &\leq i^2\frac{C\sqrt{\log N}}{\sqrt{N}}[1 - F(\lceil N^{\frac{1}{4}-\epsilon}\rceil - 2)]^{1/2} + i^2 f_i + i^2[1 - F(\lceil N^{\frac{1}{4}-\epsilon}\rceil - 2)] \\ &\leq CN^{(\frac{1}{2}-2\epsilon)} \cdot \frac{\log N}{\sqrt{N}} \cdot N^{\frac{1}{2}(1-\tau)(\frac{1}{4}-\epsilon)} + 2N^{\frac{1}{2}-2\epsilon} \cdot c\lceil N^{\frac{1}{4}-\epsilon}\rceil^{1-\tau} < \epsilon', \end{aligned}$$

for $\tau > 3$. This proves (5.2) for $i = \lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$.

We next prove (5.2) for $i < \lceil N^{\frac{1}{4}-\epsilon}\rceil - 1$. For this, we use the triangle inequality

$$i(i-2)\Big|\frac{d_i'(N)}{N} - \lambda_i\Big| \leq i^2\Big|\frac{d_i'(N)}{N} - \frac{d_i(N)}{N}\Big| + i^2\Big|\frac{d_i(N)}{N} - \lambda_i\Big|, \qquad (5.8)$$

and we bound these two terms separately.

We start with the second term, and use (5.5), which gives that

$$\mathbb{P}\big(|d_i(N) - Nf_i| \geq C(f_i N\log N)^{1/2}\big) \leq N^{-C^2/2}. \qquad (5.9)$$

We will take $C > 2$, so that

$$\mathbb{P}\big(\exists i < \lceil N^{\frac{1}{4}-\epsilon}\rceil - 1 : |d_i(N) - Nf_i| \geq C(f_i N\log N)^{1/2}\big) \leq \sum_{i=1}^{N} N^{-C^2/2} = N^{1-C^2/2} = o(1). \qquad (5.10)$$

On the complementary event, we have that

$$\sup_{i<\lceil N^{\frac{1}{4}-\epsilon}\rceil-1}\Big|i(i-2)\frac{d_i(N)}{N} - i(i-2)\lambda_i\Big| \leq C\sup_{i<\lceil N^{\frac{1}{4}-\epsilon}\rceil-1} i^2\Big(\frac{f_i\log N}{N}\Big)^{1/2} = o(1). \qquad (5.11)$$

Thus, we have bounded the second term in (5.8). We next turn to the first term in (5.8). First, we clearly have that $|d_i'(N) - d_i(N)| \leq R_N$. Thus, since $R_N \leq 2N^{\frac{3}{4}}$,

$$i^2\Big|\frac{d_i'(N)}{N} - \frac{d_i(N)}{N}\Big| \leq i^2\frac{R_N}{N} \leq 2i^2N^{-\frac{1}{4}} \leq 2N^{-\frac{1}{4}+2(\frac{1}{8}-\epsilon)} \leq \epsilon',$$

for $i \leq N^{\frac{1}{8}-\epsilon}$. For $i > N^{\frac{1}{8}-\epsilon}$, we bound $d_i'(N) \leq \sum_{j\geq i} d_j(N)$, so that, again using (5.6–5.7),

$$i^2\Big|\frac{d_i'(N)}{N} - \frac{d_i(N)}{N}\Big| \leq \frac{2i^2}{N}\sum_{j\geq i} d_j(N) = 2i^2(1-F(i-1))(1+o(1)) \leq 2cN^{(\frac{1}{8}-\varepsilon)(3-\tau)} \to 0.$$

To check (5.3), we first take i^* fixed so that

$$\sum_{i=i^*+1}^{\infty} i(i-2)\lambda_i \leq \epsilon'/2. \tag{5.12}$$

This is possible, since $\mathbb{E}[D^2] < \infty$. Thus, we are left to show that

$$\sum_{i=1}^{i^*} i(i-2)\Big|\frac{d_i(N)}{N} - \lambda_i\Big| \leq \epsilon'/2. \tag{5.13}$$

In order to do so, we use the bound in (5.10) to obtain that

$$\sum_{i=1}^{i^*} i(i-2)\Big|\frac{d_i(N)}{N} - \lambda_i\Big| \leq C\sum_{i=1}^{i^*} i^2\Big(\frac{f_i \log N}{N}\Big)^{1/2} \leq C(i^*)^3\Big(\frac{\log N}{N}\Big)^{1/2} \leq \epsilon'/2, \tag{5.14}$$

whenever N is sufficiently large. The same result applies to $d_i'(N)$, since $|d_i'(N) - d_i(N)| \leq R_N$, and $R_N = o(N)$, so that

$$\sum_{i=1}^{i^*} i(i-2)\Big|\frac{d_i(N)}{N} - \frac{d_i'(N)}{N}\Big| \leq (i^*)^3\frac{R_N}{N} = o(1).$$

Therefore, we have proved all conditions for the graph G', and thus obtain the result in Theorem 1.5 for G'. To complete the proof, we need to show that the result for G' implies the result for G.

This implication is proved in several small steps. First, denote the largest connected components of G and G' by LC_G and $LC_{G'}$. Since G can be obtained from G' by adding the removed edges back, we obtain that (since we put back at most R_N connected components of size at most $\gamma \log N$),

$$|LC_{G'}| \leq |LC_G| \leq |LC_{G'}| + R_N \cdot \gamma \log N. \tag{5.15}$$

Thus, since $|LC_{G'}| = qN(1+o(1))$ and $R_N \leq 2N^{\frac{3}{4}}$ with probability $1+o(1)$, we obtain that

$$qN(1+o(1)) \leq |LC_G| \leq qN(1+o(1)) + O(N^{\frac{3}{4}}\log N) = qN(1+o(1)), \tag{5.16}$$

so that the largest connected component has size $qN(1+o(1))$ with probability $1+o(1)$, as claimed.

To see that all other connected components in G have size at most $\gamma \log N$, we note that in G' the removed edges are all connected to nodes with degree $\lceil N^{\frac{1}{4}-\epsilon}\rceil$. We first show that with overwhelming probability these nodes are already in the largest connected component in G'. Since in G' only the largest connected component has at least N^δ nodes for any $\delta > 0$ and since $\gamma \log N = o(N^\delta)$, it suffices to check that nodes in G' with degree $\lceil N^{\frac{1}{4}-\epsilon}\rceil$ are connected to at least N^δ other nodes. Since the probability of picking a node different from the ones already connected to the node under observation is bounded from below by $1 - N^{2(\frac{1}{4}-\epsilon)-1}$ (since all degrees in G' are bounded above by $\lceil N^{\frac{1}{4}-\epsilon}\rceil$), the probability that at most N^δ different nodes are chosen is bounded by the probability that a binomial random variable, with parameters $p = 1 - N^{2(\frac{1}{4}-\epsilon)-1}$ and $n = \lceil N^{\frac{1}{4}-\epsilon}\rceil$, is bounded from above by N^δ. By (5.5), this probability is negligible whenever $\delta < \frac{1}{4} - \epsilon$. Thus, we may assume that all nodes with degree $\lceil N^{\frac{1}{4}-\epsilon}\rceil$ are in the largest connected component in G'. Therefore, we obtain that the nodes that must be added to G' to form G are attached to the largest connected component of G'. Thus, the size of the second largest connected component of G is bounded from above by the size of the second largest connected component of G', which is bounded from above by $\gamma \log N$. $\square$

A Appendix.

A.1 Proof of Proposition 3.4

In this part of the appendix, we prove Proposition 3.4, which we restate here for convenience as Proposition A.1.1. At the end of this section, we restate and prove Proposition 3.5.

Proposition A.1.1 *There exist* $\alpha_2, \beta_2 > 0$ *such that*

$$\mathbb{P}\Big(\sum_{n=0}^{\infty}(n+1)|g_n^{(N)} - g_n| \geq N^{-\alpha_2}\Big) \leq N^{-\beta_2}. \tag{A.1.1}$$

In the proof, we need the following lemma.

Lemma A.1.2 *Fix* $\tau > 1$. *For each non-negative integer* s, *there exists a constant* $C > 0$, *such that*

$$\sum_{j=m}^{n}(j+1)^s f_{j+1} \quad \leq \quad Cm^{-(\tau-1-s)} + Ch(n). \tag{A.1.2}$$

where

$$h(n) = \begin{cases} 0, & s < \tau - 1, \\ \log(n+1), & s = \tau - 1, \\ (n+1)^{s-\tau+1}, & s > \tau - 1. \end{cases}$$

We defer the proof of Lemma A.1.2 to the end of this section.
Proof of Proposition A.1.1. Fix $a, b, \alpha > 0$. Define

$$\begin{aligned} F \quad = \quad & \Big\{|\frac{L_N}{\mu N} - 1| \leq N^{-\alpha}\Big\} \cap \Big\{\frac{1}{N}\sum_{i=1}^{N}(D_i+1)^2 I[D_i \geq N^a] \leq N^{-b}\Big\} \\ & \cap \Big\{\frac{1}{N}\sum_{n=0}^{N^a}(n+1)^2\Big|\sum_{i=1}^{N}\big(I[D_i = n+1] - f_{n+1}\big)\Big| \leq N^{-b}\Big\}. \end{aligned} \tag{A.1.3}$$

The constants a, b and α will be chosen appropriately in the proof. The strategy of the proof is as follows. We will prove that

$$\mathbb{P}(F^c) \leq N^{-\beta_2}, \tag{A.1.4}$$

for some $\beta_2 > 0$, and that on F,

$$\sum_{n=0}^{\infty}(n+1)|g_n^{(N)} - g_n| \leq N^{-\alpha_2}, \tag{A.1.5}$$

for some α_2. This proves Proposition A.1.1. We start by showing (A.1.5).

We bound

$$\sum_{n=0}^{\infty}(n+1)|g_n^{(N)} - g_n| \leq \sum_{n=0}^{\infty}(n+1)|g_n^{(N)} - \frac{N\mu}{L_N}g_n| + (\nu+1)\Big|\frac{N\mu}{L_N} - 1\Big|. \tag{A.1.6}$$

The second term is bounded by $(\nu+1)N^{-\alpha}$ by the first event in F. The first term in (A.1.6) can be bounded, for N sufficiently large, as, again using the first event in F,

$$\begin{aligned} \sum_{n=0}^{\infty}(n+1)|g_n^{(N)} - \frac{N\mu}{L_N}g_n| \quad &= \quad \frac{1}{L_N}\sum_{n=0}^{\infty}(n+1)^2\Big|\sum_{i=1}^{N}\big(I[D_i = n+1] - f_{n+1}\big)\Big| \\ &\leq \quad \frac{2}{\mu N}\sum_{n=0}^{\infty}(n+1)^2\Big|\sum_{i=1}^{N}\big(I[D_i = n+1] - f_{n+1}\big)\Big|. \end{aligned} \tag{A.1.7}$$

We next split the sum over n into $n > N^a$ and $n \leq N^a$ for some appropriately chosen $a \in (0, 1]$. On F, the contribution from $n \leq N^a$ is at most $\frac{1}{\mu}N^{-b}$, whereas we can bound the contribution from $n > N^a$ by

$$\frac{2}{\mu N}\sum_{n=N^a}^{\infty}(n+1)^2\sum_{i=1}^{N}(I[D_i = n+1] + f_{n+1}) = \frac{2}{\mu N}\sum_{i=1}^{N}(D_i+1)^2 I[D_i \geq N^a] + \frac{2}{\mu}\sum_{n=N^a}^{\infty}(n+1)^2 f_{n+1}.$$

For $\tau > 3$, the second term is bounded by $CN^{-a(\tau-3)}$ by Lemma A.1.2. The first term is bounded by $\frac{\mu}{2}N^{-b}$ by the second event in F. Thus, we obtain (A.1.5) with $\alpha_2 = \min\{b, a(\tau-3)\}$.

We now prove (A.1.4). For this, we use that F is an intersection of three events which we will write as F_1, F_2 and F_3, so that

$$\mathbb{P}(F^c) \leq \mathbb{P}(F_1^c) + \mathbb{P}(F_2^c) + \mathbb{P}(F_3^c). \tag{A.1.8}$$

The first probability is bounded by $\mathbb{P}(F_1^c) \leq c \cdot N^{2\alpha-1}$, by Lemma 3.1. For $\mathbb{P}(F_2^c)$, we use the Markov inequality, to obtain that

$$\mathbb{P}(F_2^c) \leq N^b\mathbb{E}\big[(D_1+1)^2 I[D_1 \geq N^a]\big] \leq N^{b-a(\tau-3)}, \tag{A.1.9}$$

by Lemma A.1.2. For $\mathbb{P}(F_3^c)$, we use in turn the Markov inequality, Cauchy-Schwarz in the form $\sum_{n=0}^{N^a} b_n \leq (\sum_{n=0}^{N^a} 1^2 \sum_{n=0}^{N^a} b_n^2)^{\frac{1}{2}}$, and the Jensen inequality applied to $x \mapsto \sqrt{x}$ (a concave function), to obtain

$$\begin{aligned}
\mathbb{P}(F_3^c) &\leq N^{b-1}\mathbb{E}\big[\sum_{n=0}^{N^a}(n+1)^2\big|\sum_{i=1}^{N}\big(I[D_i = n+1] - f_{n+1}\big)\big|\big] \qquad \text{(A.1.10)}\\
&\leq N^{b-1}(N^a+1)^{\frac{1}{2}}\mathbb{E}\Big(\sum_{n=0}^{N^a}(n+1)^4\big(\sum_{i=1}^{N}\big(I[D_i = n+1] - f_{n+1}\big)\big)^2\Big)^{1/2}\\
&\leq 2N^{b+a/2-1}\sum_{n=0}^{N^a}(n+1)^4\mathbb{E}\big(\sum_{i=1}^{N}\big(I[D_i = n+1] - f_{n+1}\big)\big)^2\big]\Big)^{1/2}\\
&\leq 2N^{b+a/2-1}\Big(\sum_{n=0}^{N^a}(n+1)^4 N f_{n+1}\Big)^{1/2} \leq 2N^{b+a/2-1/2}N^{a\max\{0,5-\tau\}/2},
\end{aligned}$$

where in the last inequalities, we have used Lemma A.1.2 and

$$\mathbb{E}\big[\big(\sum_{i=1}^{N}[I[D_i = n+1] - f_{n+1}]\big)^2\big] = \mathrm{Var}\left(\sum_{i=1}^{N} I[D_i = n+1]\right) = Nf_{n+1}(1-f_{n+1}) \leq Nf_{n+1}.$$

Thus, we obtain the statement in Proposition A.1.1 with

$$\beta_2 = \min\{1/2 - b - a\max\{1, 6-\tau\}/2, a(\tau-3) - b, (2\alpha-1)\}.$$

By picking first b small, and then a small, we see that $\alpha_2, \beta_2 > 0$. □

Remark A.1.3 *When (1.2) holds for some $\tau > 2$ (rather than $\tau > 3$), then the above proof can be repeated to show the weaker result that*

$$\mathbb{P}\big(\sum_{n=0}^{\infty}|g_n^{(N)} - g_n| \geq N^{-\alpha_2}\big) \leq N^{-\beta_2}. \tag{A.1.11}$$

Indeed, in the definition of the event F in (A.1.3), we can replace $(D_i+1)^2$ by (D_i+1) in the second event, and $(n+1)^2$ by $(n+1)$ in the third event. Then, by adapting the above argument, the event F implies that $\sum_{n=0}^{\infty}|g_n^{(N)} - g_n| \leq N^{-\alpha_2}$. The proof that $\mathbb{P}(F^c) \leq N^{-\beta_2}$ can be adapted accordingly.

Proof of Lemma A.1.2. Define a density $f(x) = \sum_{j=0}^{\infty} f_j I[j \le x < j+1]$, and the corresponding distribution function $\tilde{F}(x) = \int_0^x f(u)\, du$. Then for integer-valued $j > 0$,

$$\tilde{F}(j) = f_0 + \ldots + f_{j-1} = F(j-1), \quad F(j-1) \le \tilde{F}(x) \le F(j), \quad x \in (j, j+1).$$

Moreover

$$\sum_{j=m}^{n} (j+1)^s f_{j+1} \le \int_{m+1}^{n+2} x^s f(x)\, dx = -\int_{m+1}^{n+2} x^s\, d(1-\tilde{F}(x)).$$

Using partial integration and the upper bound

$$1 - \tilde{F}(x) \le 1 - F(j-1) \le c(j-1)^{1-\tau},$$

for $x \in (j, j+1)$, we conclude that

$$\begin{aligned}\sum_{j=m}^{n} (j+1)^s f_{j+1} &\le (m+1)^s(1-\tilde{F}(m+1)) - (n+2)^s(1-\tilde{F}(n+2)) + \int_{m+1}^{n+2} (1-\tilde{F}(x))\, dx^s \\ &\le c\left[m^{1+s-\tau} + \int_m^{n+1} y^{s-\tau}\, dy\right].\end{aligned}$$

This yields the upper bound. □

We finally prove Proposition 3.5, which we restate as Proposition A.1.4.

Proposition A.1.4 *There exist $\varepsilon, \beta, \eta > 0$ such that for all $j \le (1+2\eta)\log_\nu N$, as $N \to \infty$,*

$$\mathbb{P}\Big(\frac{1}{N}\Big|\sum_{i=1}^{j} \mathcal{Z}^{(1)}_{\lceil i/2\rceil}\mathcal{Z}^{(2)}_{\lfloor i/2\rfloor} - \sum_{i=1}^{j} \hat{Z}^{(1)}_{\lceil i/2\rceil}\hat{Z}^{(2)}_{\lfloor i/2\rfloor}\Big| > N^{-\varepsilon}\Big) = O(N^{-\beta}). \tag{A.1.12}$$

Proof. Let

$$F_N = \{\sum_{n=0}^{\infty} n|g_n^{(N)} - g_n| < N^{-\alpha_2}\}, \tag{A.1.13}$$

then according to Proposition A.1.1 we have $\mathbb{P}(F_N^c) \le N^{-\beta_2}$. We claim that for all $i \ge 1$,

$$\mathbb{E}_N|\mathcal{Z}_i^{(1)} - \hat{Z}_i^{(1)}| \le \max\{\nu - \alpha_N, \nu_N - \alpha_N\}\sum_{m=1}^{i} \mathbb{E}_N[\hat{Z}_m^{(1)}](\max\{\nu, \nu_N\})^{i-m}, \tag{A.1.14}$$

where

$$\alpha_N = \sum_{n=0}^{\infty} n \min\{g_n, g_n^{(N)}\}. \tag{A.1.15}$$

We first prove (A.1.14). For $\mathcal{Z}_i^{(1)} \ne \hat{Z}_i^{(1)}$, the coupling is not successful in at least one of the generations m, $1 \le m \le i$. Let m be the first generation for which the coupling is unsuccessful. There are at most $\hat{Z}_m^{(1)}$ nodes for which the coupling can fail. If the coupling fails for a node, the expected difference between the offspring of that node is bounded above by (compare the paragraph around (3.5)),

$$\begin{aligned}&\sum_{n=1}^{\infty} n\mathbb{P}(X^{(N)} = n, Y = \infty) - \sum_{n=1}^{\infty} n\mathbb{P}(X = n, Y^{(N)} = \infty) \\ &= \sum_{n=1}^{\infty} n\big(g_n^{(N)} - \min\{g_n, g_n^{(N)}\}\big) - \sum_{n=1}^{\infty} n\big(g_n - \min\{g_n, g_n^{(N)}\}\big) \\ &= (\nu_N - \alpha_N) - (\nu - \alpha_N) \le \max\{\nu - \alpha_N, \nu_N - \alpha_N\}.\end{aligned}$$

From generation $m+1$ on, we again have two BP's with laws g and $g^{(N)}$, so that the expected offspring is bounded by $(\max\{\nu,\nu_N\})^{i-m}$. This demonstrates the claim (A.1.14).

Furthermore, since $\mathbb{E}_N[\hat{Z}^{(1)}_m] = D_1\nu_N^{m-1}$, we end up with

$$\mathbb{E}_N|\mathcal{Z}^{(1)}_i - \hat{Z}^{(1)}_i| \le \max\{\nu-\alpha_N, \nu_N-\alpha_N\} i D_1 (\max\{\nu,\nu_N\})^{i-1}. \tag{A.1.16}$$

By (A.1.13), on F_N we have that

$$\max\{\nu-\alpha_N, \nu_N-\alpha_N\} \le \sum_n n|g_n - g_n^{(N)}| < N^{-\alpha_2},$$

$$\frac{\max\{\nu,\nu_N\}}{\nu} = 1+\nu^{-1}\max\{0, \sum_n n(g_n-g_n^{(N)})\} = 1+O(N^{-\alpha_2}).$$

We bound the left hand side of (A.1.12) by

$$\mathbb{P}\left(\Big|\frac{\sum_{i=1}^j \mathcal{Z}^{(1)}_{\lceil i/2\rceil}\mathcal{Z}^{(2)}_{\lfloor i/2\rfloor}}{N} - \frac{\sum_{i=1}^j \hat{Z}^{(1)}_{\lceil i/2\rceil}\hat{Z}^{(2)}_{\lfloor i/2\rfloor}}{N}\Big| > N^{-\varepsilon}\right) \tag{A.1.17}$$

$$\le \mathbb{P}\left(\left|\frac{\sum_{i=1}^j \mathcal{Z}^{(2)}_{\lfloor i/2\rfloor}(\mathcal{Z}^{(1)}_{\lceil i/2\rceil} - \hat{Z}^{(1)}_{\lceil i/2\rceil})}{\sqrt{N}\sqrt{N}}\right| > \frac{1}{2}N^{-\varepsilon}\right) + \mathbb{P}\left(\left|\frac{\sum_{i=1}^j \hat{Z}^{(1)}_{\lceil i/2\rceil}(\mathcal{Z}^{(2)}_{\lfloor i/2\rfloor} - \hat{Z}^{(2)}_{\lfloor i/2\rfloor})}{\sqrt{N}\sqrt{N}}\right| > \frac{1}{2}N^{-\varepsilon}\right).$$

Both terms on the right hand side of (A.1.17) can be treated similarly and we will only do the first one. We have uniformly in $i \le (\frac{1}{2}+\eta)\log_\nu N$,

$$N^{-1/2}\max\left\{\mathbb{E}[\mathcal{Z}^{(2)}_i], \mathbb{E}[\hat{Z}^{(1)}_i]\right) = \max\{N^\eta, N^\eta\cdot(1+O(N^{-\alpha_2}))^{(\frac{1}{2}+\eta)\log_\nu N}\}, \tag{A.1.18}$$

on F_N. For $j \le (1+2\eta)\log_\nu N$, using the abbreviation

$$T_N = \frac{1}{N}\Big|\sum_{i=1}^j \mathcal{Z}^{(2)}_{\lfloor i/2\rfloor}(\mathcal{Z}^{(1)}_{\lceil i/2\rceil} - \hat{Z}^{(1)}_{\lceil i/2\rceil})\Big|,$$

we have

$$\begin{aligned}\mathbb{P}\Big(T_N > N^{-\varepsilon}\Big) &\le \mathbb{P}(F_N^c) + \mathbb{P}(T_N > N^{-\varepsilon}, F_N) \le N^{-\beta_2} + \mathbb{E}\left[\mathbb{P}_N\left(T_N I_{F_N} > N^{-\varepsilon}\right)\right] \\ &\le N^{-\beta_2} + \mathbb{E}\left[N^\varepsilon \mathbb{E}_N\left[T_N I_{F_N}\right]\right].\end{aligned}$$

From (A.1.16), the uniform bound in (A.1.18), and the estimates on F_N, we obtain

$$\begin{aligned}\mathbb{E}\left[N^\varepsilon\mathbb{E}_N\left[T_N I_{F_N}\right]\right] &\le \mathbb{E}\left[N^{\varepsilon-1}\mathbb{E}_N\left[\sum_{i=1}^j \Big|\mathcal{Z}^{(2)}_{\lfloor i/2\rfloor}(\mathcal{Z}^{(1)}_{\lceil i/2\rceil} - \hat{Z}^{(1)}_{\lceil i/2\rceil})\Big|\cdot I_{F_N}\right]\right] \\ &\le \mathbb{E}\left[N^{\eta+\varepsilon-\frac{1}{2}}\sum_{i=1}^j \mathbb{E}_N\Big|\mathcal{Z}^{(1)}_{\lceil i/2\rceil} - \hat{Z}^{(1)}_{\lceil i/2\rceil})\Big|\cdot I_{F_N}\right] \\ &\le 2N^{\eta+\varepsilon-\frac{1}{2}}\mathbb{E}\left[\sum_{i=1}^{\lceil j/2\rceil}\mathbb{E}_N\Big|\mathcal{Z}^{(1)}_i - \hat{Z}^{(1)}_i)\Big|\cdot I_{F_N}\right] \\ &\le 2N^{\eta+\varepsilon-\frac{1}{2}}\mathbb{E}[D_1\sum_{i=1}^{\lceil j/2\rceil} N^{-\alpha_2} i(1+O(N^{-\alpha_2}))^{i-1}] \\ &\le \mu N^{\varepsilon+\eta-\alpha_2}\sum_{i=1}^{\lceil(\frac{1}{2}+\eta)\log_\nu N\rceil} i(1+O(N^{-\alpha_2}))^{i-1} \\ &\le 2\mu N^{\varepsilon+\eta-\alpha_2}\cdot(\log_\nu N)^2\cdot N^{(\frac{1}{2}+\eta)\log_\nu(1+O(N^{-\alpha_2}))},\end{aligned}$$

using that for $x = 1+O(N^{-\alpha_2}) > 1$, we have $\sum_{i=1}^n ix^{i-1} \le n^2x^n$. This proves the proposition since $(\log_\nu N)^2\cdot N^{(\frac{1}{2}+\eta)\log_\nu(1+O(N^{-\alpha_2}))}$ can be bounded by any small power of N, and ε and η can both be taken arbitrarily small, whereas $\alpha_2 > 0$. □

A.2 Proof of Proposition 3.2

In this second part of the appendix, we restate our main result on the coupling between the SPG and the BP with offspring distribution $\{g_n^{(N)}\}$ once more and give a full proof.

Proposition A.2.1 *There exist $\eta, \beta > 0$, $\alpha > \frac{1}{2} + \eta$ and a constant C, such that for all $j \leq (\frac{1}{2} + \eta) \log_\nu N$,*

$$\mathbb{P}\Big((1 - N^{-\alpha}\nu^j)\hat{Z}_j^{(1)} \leq Z_j^{(1)} \leq (1 + N^{-\alpha}\nu^j)\hat{Z}_j^{(1)}\Big) \geq 1 - CjN^{-\beta}. \tag{A.2.1}$$

This proof is divided into several lemmas. It is rather involved, and we may think of Proposition A.2.1 as one of the key estimates of the paper. We start with an explanation of the different steps in this proof.

The proof of Proposition A.2.1 proceeds by induction with respect to j. Note that for all $j \leq (\frac{1}{2} + \eta) \log_\nu N$, we have $N^{-\alpha}\nu^j \leq N^{(\frac{1}{2}+\eta)-\alpha} \to 0$, as $N \to \infty$ and when $\alpha > \eta$. When at level $j - 1$, the event in the statement of the proposition holds, we have

$$|\hat{Z}_{j-1}^{(1)} - Z_{j-1}^{(1)}| \leq \frac{\nu^{j-1}}{N^\alpha} \hat{Z}_{j-1}^{(1)},$$

so that we control the difference between the number of stubs $Z_{j-1}^{(1)}$ and the number of children $\hat{Z}_{j-1}^{(1)}$. The absolute value of this difference is bounded by $\hat{Z}_{j-1}^{(1)}$ times a fraction that converges to 0. For generation j we have to control the difference $\hat{Z}_j^{(1)} - Z_j^{(1)}$. Differences in generation j arise from differences in generation $j-1$ and from drawing stubs with label 2 or label 3. If a label 2 stub is chosen, then the SPG will contain a loop or cycle and hence no free stubs in level j are created, whereas in the BP a non-negative number of offspring is attached. If a label 3 stub is chosen, then the corresponding node with described number of children is attached in the BP, whereas for the SPG we have to resample until we draw a stub labeled 1 or 2. Hence, if $Z_j^{(1)} \geq \hat{Z}_j^{(1)}$, so that the number of free stubs attached to nodes at distance $j-1$ of the SPG exceeds the number of children in generation j of the BP, then this overshoot can only be caused by drawing label 3 stubs. The number of stubs with label 3 is bounded by the total number drawn in the SPG, i.e., by

$$\sum_{i=1}^{j-1} Z_i^{(1)} \leq \sum_{i=1}^{j-1} (1 + N^{-\alpha}\nu^i)\hat{Z}_i^{(1)} \leq 2\sum_{i=1}^{j-1} \hat{Z}_i^{(1)}.$$

For $Z_j^{(1)} \leq \hat{Z}_j^{(1)}$, the number of stubs with level 2 or 3 both matter and their total amount is bounded by

$$\sum_{i=1}^{j} Z_i^{(1)} = \sum_{i=1}^{j-1} Z_i^{(1)} + Z_j^{(1)} \leq \sum_{i=1}^{j-1} (1 + N^{-\alpha}\nu^i)\hat{Z}_i^{(1)} + \hat{Z}_j^{(1)} \leq 2\sum_{i=1}^{j} \hat{Z}_i^{(1)}.$$

In both cases the probability of drawing a label 2 or 3 stub is bounded by

$$\frac{2\sum_{i=1}^{j} \hat{Z}_i^{(1)}}{L_N} \leq \frac{2N^{\frac{1}{2}+\delta}}{L_N}, \tag{A.2.2}$$

on the event where $\sum_{i=1}^{j} \hat{Z}_i^{(1)} \leq N^{\frac{1}{2}+\delta}$. Using that L_N is of order $\mathbb{E}[L_N] = \mu N$ (see Lemma 3.1), this probability is sufficiently small to allow us to use Chebychev's inequality.

The main lemmas in this section are Lemma A.2.7 and Lemma A.2.9. Together, they prove the induction step described above. Lemmas A.2.2 up to A.2.6 are preparations, the most important one being Lemma A.2.6. This lemma shows that if the *total* progeny up to and including generation j of $\{\hat{Z}_i^{(1)}\}$ is larger than $N^{\frac{1}{2}-\delta}$, for some $\delta > 0$, then with overwhelming probability also each of the sizes of the last two generations, i.e., $\hat{Z}_{j-1}^{(1)}$ and $\hat{Z}_j^{(1)}$, exceed $N^{\frac{1}{2}-2\delta}$.

As before, we will abbreviate the conditional probability and expectation given $D_1, \ldots, D_N$ by $\mathbb{P}_N$ and $\mathbb{E}_N$.

Lemma A.2.2 *For $0<\eta<\frac{1}{2}$ and all $j\geq 1$,*

$$\mathbb{P}_N\big(Z_j^{(1)}\neq \hat{Z}_j^{(1)}, \sum_{i=1}^{j}\hat{Z}_i^{(1)}\leq N^{\frac{1}{2}-\eta}\big)\leq \frac{N^{-2\eta}}{(L_N/N)}, \quad a.s. \tag{A.2.3}$$

Lemma A.2.2 together with Lemma 3.1 prove Proposition A.2.1 for all j such that the total size of the BP is at most $N^{\frac{1}{2}-\eta}$.

Proof. We denote by l the first stub which is grown differently in the SPG and in the BP. Assume that this l^{th} stub is in the j^{th} generation or earlier.

Before the growth of the l^{th} stub, the BP and the SPG are identical. Thus, we must have that $l\leq \sum_{i=1}^{j}\hat{Z}_i^{(1)}$. Hence, as we reach to the l^{th} stub, the number of stubs having either label 2 or 3 is bounded above by $\sum_{i=1}^{j}\hat{Z}_i^{(1)}\leq N^{\frac{1}{2}-\eta}$. A difference in the SPG and the BP can only arise when we draw a stub for the BP having label 2 or 3. Thus, the probability that the l^{th} stub is the first to create a difference between the SPG and the BP is bounded above by $N^{\frac{1}{2}-\eta}/L_N$. Therefore,

$$\mathbb{P}_N\big(Z_j^{(1)}\neq \hat{Z}_j^{(1)}, \sum_{i=1}^{j}\hat{Z}_i^{(1)}\leq N^{\frac{1}{2}-\eta}\big)\leq \sum_{l=1}^{N^{\frac{1}{2}-\eta}}\frac{N^{\frac{1}{2}-\eta}}{L_N}=\frac{N^{-2\eta}}{(L_N/N)}.$$

□

Recall that $\nu_N=\sum_{n=0}^{\infty}ng_n^{(N)}$ is the expected offspring of the BP $\{\hat{Z}^{(1)}\}_j$ under $\mathbb{P}_N$. Note from Proposition A.1.1 that ν_N is close to ν with probability close to one. In the statement of the next lemma, we write

$$D_N^{(N)}=\max_{1\leq i\leq N}D_i. \tag{A.2.4}$$

Lemma A.2.3 *For every $\gamma>0$,*

$$\mathbb{P}\big(D_N^{(N)}\geq N^{\gamma}\big)\leq cN^{1-(\tau-1)\gamma}. \tag{A.2.5}$$

Proof. We use Boole's inequality to obtain from (1.2) that

$$\mathbb{P}\big(D_N^{(N)}\geq N^{\gamma}\big)\leq \sum_{i=1}^{N}\mathbb{P}(D_i\geq N^{\gamma})\leq cN^{1-(\tau-1)\gamma}. \tag{A.2.6}$$

□

Lemma A.2.4 *For $\eta,\delta\in(-\frac{1}{2},\frac{1}{2})$, and all $j\leq(\frac{1}{2}+\eta)\log_\nu N$, there exists $\beta_2>0$ such that*

$$\mathbb{P}\big(\sum_{i=1}^{j}\hat{Z}_i^{(1)}\geq N^{\frac{1}{2}+\delta}\big)\leq CN^{\eta-\delta}+N^{-\beta_2}. \tag{A.2.7}$$

Proof. By Proposition 3.4, we can include the indicator that $|\nu_N-\nu|\leq N^{-\alpha_2}$; this explains the additional error term $N^{-\beta_2}$. By the Markov inequality, we obtain for $j\leq(\frac{1}{2}+\eta)\log_\nu N$,

$$\mathbb{P}\big(\sum_{i=1}^{j}\hat{Z}_i^{(1)}\geq N^{\frac{1}{2}+\delta}, |\nu_N-\nu|\leq N^{-\alpha_2}\big)\leq N^{-\frac{1}{2}-\delta}\mathbb{E}\big(\sum_{i=1}^{j}\hat{Z}_i^{(1)}I[|\nu_N-\nu|\leq N^{-\alpha_2}]\big).$$

The expectation on the right-hand side can be computed by conditioning:

$$\begin{aligned}\mathbb{E}\big[\hat{Z}_i^{(1)}I[|\nu_N-\nu|\leq N^{-\alpha_2}]\big]&=\mathbb{E}[\mathbb{E}_N\big[\hat{Z}_i^{(1)}I[|\nu_N-\nu|\leq N^{-\alpha_2}]\big]\\&=\mathbb{E}[I[|\nu_N-\nu|\leq N^{-\alpha_2}]\cdot D_1\nu_N^{i-1}]\leq(\nu+N^{-\alpha_2})^{i-1}\mathbb{E}[D_1].\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{P}\big(\sum_{i=1}^{j}\hat{Z}_i^{(1)} \geq N^{\frac12+\delta}\big) &\leq N^{-\beta_2}+\mu N^{-\frac12-\delta}\sum_{i=1}^{j}(\nu+N^{-\alpha_2})^{i-1}\\ &\leq N^{-\beta_2}+\mu N^{-\frac12-\delta}\frac{(\nu+N^{-\alpha_2})^j-1}{(\nu+N^{-\alpha_2})-1}\leq N^{-\beta_2}+CN^{\eta-\delta}.\end{aligned}$$

□

In the lemma below, we write d for a random variable with discrete distribution $\{g_n^{(N)}\}$ given in (3.1), and $\mathrm{Var}_N(d)$ for the variance of d under $\mathbb{P}_N$. Furthermore, we let, for any $0 < a < \frac12$,

$$A_N = A_N(a,\gamma,\alpha_2) = \left\{\left|\frac{L_N}{\mu N}-1\right| \leq N^{-a}\right\} \cap \{D_N^{(N)} \leq N^\gamma\} \cap \{|\nu_N-\nu| \leq N^{-\alpha_2}\},$$

then, according to Proposition 3.4, Lemmas 3.1 and A.2.3, we have

$$\mathbb{P}(A_N^c) = O(N^{-\epsilon}), \tag{A.2.8}$$

where $\epsilon = b \wedge ((\tau-1)\gamma-1) \wedge \beta_2 > 0$ whenever $\gamma > 1/(\tau-1)$. On A_N, we have

$$\frac{1}{\mu(1+N^{-a})} \leq \frac{N}{L_N} \leq \frac{1}{\mu(1-N^{-a})}. \tag{A.2.9}$$

This will be used in the following lemma.

Lemma A.2.5 *For every $\gamma > 0$,*

$$\mathbb{E}\big(\mathrm{Var}_N(d) I[A_N]\big) \leq CN^{(4-\tau)^+\gamma}, \tag{A.2.10}$$

where $x^+ = \max(0,x)$.

Proof. Since the variance of a random variable is bounded by its second moment,

$$\mathrm{Var}_N(d) \leq \sum_{n=0}^{\infty} n^2 g_n^{(N)} = \sum_{n=0}^{\infty}\sum_{j=1}^{N}\frac{n^2(n+1)}{L_N}I[D_j=n+1] \leq \frac{1}{L_N}\sum_{j=1}^{N} D_j^3,$$

and so, for $\tau \in (3,4]$,

$$\mathbb{E}\big(\mathrm{Var}_N(d)I[A_N]\big) \leq \sum_{j=1}^{N}\mathbb{E}\big[\frac{1}{L_N}D_j^3 I[A_N]\big] \leq \frac{N}{\mu N}\mathbb{E}\big[D^3 I[D \leq N^\gamma]\big] \leq C\sum_{i=1}^{\lceil N^\gamma\rceil} i^3 f_i \leq N^{\gamma(4-\tau)},$$

by Lemma A.1.2. For $\tau > 4$, the third moment of D is finite, and the result is also true even without the indicator $I[D_N^{(N)} \leq N^\gamma]$. □

Lemma A.2.6 *For all $(\frac12-2\eta)\log_\nu N \leq j \leq (\frac12+2\eta)\log_\nu N$, there exists $\delta,\beta > 0$ such that*

$$\mathbb{P}\big(\sum_{i=1}^{j}\hat{Z}_i^{(1)} \geq N^{\frac12-\delta}, \hat{Z}_{j-1}^{(1)} \leq N^{\frac12-2\delta}\big) \leq CN^{-\beta}, \tag{A.2.11}$$

$$\mathbb{P}\big(\sum_{i=1}^{j}\hat{Z}_i^{(1)} \geq N^{\frac12-\delta}, \hat{Z}_{j}^{(1)} \leq N^{\frac12-2\delta}\big) \leq CN^{-\beta}. \tag{A.2.12}$$

Remark: The statements of the lemma are almost identical, the difference being that the index of $\hat{Z}^{(1)}_{j-1}$ in the first statement is replaced by the index j in the second statement. We will be satisfied with a proof for the first statement only, the proof with index j is a straightforward extension.

Proof. Since $\sum_{i=1}^{j} \hat{Z}^{(1)}_i \geq N^{\frac12-\delta}$, there must be an $i \leq j \leq (\frac12+2\eta)\log_\nu N$ such that for N large enough

$$\hat{Z}^{(1)}_i \geq N^{\frac12-\delta}/j \geq \frac{N^{\frac12-\delta}}{(\frac12+2\eta)\log_\nu N} \geq N^{\frac12-\frac32\delta}.$$

We write I for the first $i \leq j$ such that $\hat{Z}^{(1)}_i \geq N^{\frac12-\frac32\delta}$. It suffices to bound

$$\sum_{i=1}^{j} \mathbb{P}\big(\sum_{k=1}^{j} \hat{Z}^{(1)}_k \geq N^{\frac12-\delta}, I=i, \hat{Z}^{(1)}_{j-1} \leq N^{\frac12-2\delta}\big). \tag{A.2.13}$$

The contribution from $I=j-1$ is 0. When $I=j$, then $\hat{Z}^{(1)}_j \geq N^{\frac12-\frac32\delta}$, but $\hat{Z}^{(1)}_{j-1} \leq N^{\frac12-2\delta}$ so that from the Markov inequality

$$\begin{aligned}
&\mathbb{P}\big(\sum_{k=1}^{j} \hat{Z}^{(1)}_k \geq N^{\frac12-\delta}, I=j, \hat{Z}^{(1)}_{j-1} \leq N^{\frac12-2\delta}\big) \qquad \text{(A.2.14)}\\
&\quad\leq \mathbb{E}\Big[I[\hat{Z}^{(1)}_{j-1} \leq N^{\frac12-2\delta}]\mathbb{P}_N\big(\hat{Z}^{(1)}_j \geq N^{\frac12-\frac32\delta}\big|\hat{Z}^{(1)}_{j-1}\big)\Big]\\
&\quad\leq N^{-\frac12+\frac32\delta}\mathbb{E}\Big[I[\hat{Z}^{(1)}_{j-1} \leq N^{\frac12-2\delta}]\mathbb{E}_N\big[\hat{Z}^{(1)}_j\big|\hat{Z}^{(1)}_{j-1}\big]\Big]\\
&\quad= N^{-\frac12+\frac32\delta}\mathbb{E}\Big[I[\hat{Z}^{(1)}_{j-1} \leq N^{\frac12-2\delta}]\nu_N\hat{Z}^{(1)}_{j-1}\Big] \leq CN^{-\frac12+\frac32\delta}N^{\frac12-2\delta} = CN^{-\delta/2}.
\end{aligned}$$

Thus, we are left to deal with the cases where $I < j-1$. Then, there exists an $i < j-1$ such that $\hat{Z}^{(1)}_i \geq N^{\frac12-\frac32\delta}$, but $\hat{Z}^{(1)}_{j-1} \leq N^{\frac12-2\delta}$. Thus, there must be a first $s \geq i$ such that $\hat{Z}^{(1)}_{s+1} \leq \hat{Z}^{(1)}_s$. Consequently, $\hat{Z}^{(1)}_s \geq \hat{Z}^{(1)}_i \geq N^{\frac12-\frac32\delta}$. We will bound, uniformly in s,

$$\mathbb{P}(\hat{Z}^{(1)}_{s+1} \leq \hat{Z}^{(1)}_s, \hat{Z}^{(1)}_s \geq N^{\frac12-\frac32\delta}) \leq N^{-\beta}, \tag{A.2.15}$$

for some $\beta > 0$. This proves (A.2.11), since the total number of possible i and s with $i \leq s \leq j$ is bounded by $(\log_\nu N)^2$.

We use Lemma A.2.3 to see that we may include the indicator on A_N for any $\gamma > 1/(\tau-1)$. We will use the Chebychev inequality and Lemma A.2.5 to obtain that

$$\begin{aligned}
&\mathbb{P}(\hat{Z}^{(1)}_{s+1} \leq \hat{Z}^{(1)}_s, \hat{Z}^{(1)}_s \geq N^{\frac12-\frac32\delta}, A_N) \qquad \text{(A.2.16)}\\
&\quad= \mathbb{E}\Big[I[\hat{Z}^{(1)}_s \geq N^{\frac12-\frac32\delta}, A_N]\mathbb{P}_N\big(\hat{Z}^{(1)}_{s+1} \leq \hat{Z}^{(1)}_s\big|\hat{Z}^{(1)}_s\big)\Big]\\
&\quad\leq \mathbb{E}\Big[I[\hat{Z}^{(1)}_s \geq N^{\frac12-\frac32\delta}, A_N]\mathbb{P}_N\big(\big|\hat{Z}^{(1)}_{s+1} - \nu_N\hat{Z}^{(1)}_s\big| \geq (\nu_N-1)\hat{Z}^{(1)}_s\big|\hat{Z}^{(1)}_s\big)\Big]\\
&\quad\leq \mathbb{E}\Big[I[\hat{Z}^{(1)}_s \geq N^{\frac12-\frac32\delta}, A_N](\nu_N-1)^{-2}\frac{\mathrm{Var}_N(d_1)}{\hat{Z}^{(1)}_s}\Big]\\
&\quad\leq CN^{(4-\tau)^+\gamma-\frac12+\frac32\delta} \leq N^{-\beta},
\end{aligned}$$

with $C = 2(\nu-1)^{-2}$, and since $(4-\tau)^+\gamma < 1/2$ and $\delta > 0$ can be taken arbitrarily small. □

We are now ready to give the proof of Proposition A.2.1.

Proof of Proposition A.2.1.

We first set the stage for the proof by induction in j. Fix $\eta < \delta < 2\eta$, and $\alpha > \frac12+\eta$, and define

$$E_j = \big\{\forall i \leq j : (1-N^{-\alpha}\nu^i)\hat{Z}^{(1)}_i \leq Z^{(1)}_i \leq (1+N^{-\alpha}\nu^i)\hat{Z}^{(1)}_i\big\}. \tag{A.2.17}$$

We will prove by induction that for all $j \leq (\frac{1}{2}+\eta)\log_\nu N$,

$$\mathbb{P}\big(E_j^c\big) \leq CjN^{-\beta}, \tag{A.2.18}$$

which implies Proposition A.2.1 by taking the complementary event. First, by Lemma A.2.2 and A.2.4 and since $\eta < \delta$ we see that it is sufficient to prove for $j \leq (\frac{1}{2}+\eta)\log_\nu N$,

$$\mathbb{P}\big(E_j^c, N^{\frac{1}{2}-\delta} \leq \sum_{i=1}^{j} \hat{Z}_i^{(1)} \leq N^{\frac{1}{2}+\delta}\big) \leq CjN^{-\beta}.$$

For $j < (\frac{1}{2}-2\eta)\log_\nu N$, we bound

$$\mathbb{P}\Big(E_j^c, N^{\frac{1}{2}-\delta} \leq \sum_{i=1}^{j} \hat{Z}_i^{(1)} \leq N^{\frac{1}{2}+\delta}\Big) \leq \mathbb{P}\Big(\sum_{i=1}^{j} \hat{Z}_i^{(1)} \geq N^{\frac{1}{2}-\delta}\Big) \leq N^{-\beta} + CN^{-2\eta+\delta},$$

by the Markov inequality and using Proposition 3.4 in a similar way as in Lemma A.2.4. Hence, the statement in (A.2.18) follows for $j < (\frac{1}{2}-2\eta)\log_\nu N$. This initializes the induction in j.

To advance the induction, we bound

$$\begin{aligned}\mathbb{P}\big(E_j^c, N^{\frac{1}{2}-\delta} \leq \sum_{i=1}^{j} \hat{Z}_i^{(1)} \leq N^{\frac{1}{2}+\delta}\big) &\leq \mathbb{P}(E_{j-1}^c) + \mathbb{P}\big(E_j^c \cap E_{j-1}, N^{\frac{1}{2}-\delta} \leq \sum_{i=1}^{j} \hat{Z}_i^{(1)} \leq N^{\frac{1}{2}+\delta}\big) \\ &\leq C(j-1)N^{-\beta} + \mathbb{P}\big(E_j^c \cap E_{j-1}, N^{\frac{1}{2}-\delta} \leq \sum_{i=1}^{j} \hat{Z}_i^{(1)} \leq N^{\frac{1}{2}+\delta}\big),\end{aligned}$$

where the last inequality follows by the induction hypothesis. Thus, it suffices to prove that

$$\mathbb{P}\big(E_j^c \cap E_{j-1}'\big) \leq CN^{-\beta}, \tag{A.2.19}$$

where

$$E_{j-1}' = E_{j-1} \cap \{N^{\frac{1}{2}-\delta} \leq \sum_{i=1}^{j} \hat{Z}_i^{(1)} \leq N^{\frac{1}{2}+\delta}\}.$$

Note that

$$E_j^c \cap E_{j-1}' = \Big(\{Z_j^{(1)} < (1-N^{-\alpha}\nu^j)\hat{Z}_j^{(1)}\} \cap E_{j-1}'\Big) \bigcup \Big(\{Z_j^{(1)} > (1+N^{-\alpha}\nu^j)\hat{Z}_j^{(1)}\} \cap E_{j-1}'\Big). \tag{A.2.20}$$

We write the disjoint events on the right-hand side of (A.2.20) as $E_{j,<}^c$ and $E_{j,>}^c$ and bound the probability of these events separately. We will start with $E_{j,<}^c$. This result is stated in the following lemma:

Lemma A.2.7 *There exists $\beta > 0$ such that for all $(\frac{1}{2}-2\eta)\log_\nu N < j \leq (\frac{1}{2}+\eta)\log_\nu N$,*

$$\mathbb{P}(E_{j,<}^c) \leq CN^{-\beta}. \tag{A.2.21}$$

Proof. We note that on $E_{j,<}^c$, we have that

$$\sum_{i=1}^{j} Z_i^{(1)} \leq \sum_{i=1}^{j}(1+\nu^i N^{-\alpha})\hat{Z}_i^{(1)} \leq (1+N^{\frac{1}{2}+\eta}N^{-\alpha})\sum_{i=1}^{j} \hat{Z}_i^{(1)} \leq 2N^{\frac{1}{2}+\delta},$$

because $\alpha > \frac{1}{2}+\eta$. Thus, for every stub which is grown simultaneously for the BP and the SPG, there is a probability bounded from above by $2N^{\frac{1}{2}+\delta}/L_N$ that a difference is created between the BP and the SPG (such a difference is called a miscoupling). Denote by U the number of stubs where such a difference occurs. Then, U is bounded from above by a binomial random variable with $n = N^{\frac{1}{2}+\delta}$ and $p = 2N^{\frac{1}{2}+\delta}/L_N$. Thus, by the Markov inequality, we have,

$$\mathbb{P}_N(U \geq N^a) \leq \frac{2N^{-a+1+2\delta}}{L_N}.$$

Using (A.2.9), we obtain, for $2\delta < a$,

$$\mathbb{P}(U \geq N^a) \leq CN^{-a+2\delta} + N^{-b} \leq N^{-\beta}. \tag{A.2.22}$$

Observe that differences between $Z_j^{(1)}$ and $\hat{Z}_j^{(1)}$ can only arise through (i) different numbers of stubs in the $(j-1)^{\text{st}}$ generation, and (ii) differences created in the j^{th} generation which we previously called miscouplings. In the first case, the difference in the number of stubs is bounded from below by an independent draw from $g^{(N)}$. A miscoupling occurs if we draw a stub with label 2 or 3. Hence,

$$Z_j^{(1)} - \hat{Z}_j^{(1)} \geq -\sum_{i=1}^{\left(\hat{Z}_{j-1}^{(1)} - Z_{j-1}^{(1)}\right)^+} d_i - \sum_{i=1}^{U} \tilde{d}_i, \tag{A.2.23}$$

where $\{d_i\}_{i\geq 1}$ are independent draws from $g^{(N)}$ and $\{\tilde{d}_i\}_{i\geq 1}$ are draws from $g^{(N)}$, conditionally on drawing a stub labeled 2 or 3. On E'_{j-1}, we have that

$$\left(\hat{Z}_{j-1}^{(1)} - Z_{j-1}^{(1)}\right)^+ \leq N^{-\alpha}\nu^{j-1}\hat{Z}_{j-1}^{(1)}, \tag{A.2.24}$$

so that on $E_{j<}^c$, introducing the notation $\alpha_{N,j} = N^{-\alpha}\nu^{j-1}\hat{Z}_{j-1}^{(1)}$,

$$\sum_{i=1}^{\alpha_{N,j}} d_i + \sum_{i=1}^{U} \tilde{d}_i \geq \sum_{i=1}^{\left(\hat{Z}_{j-1}^{(1)} - Z_{j-1}^{(1)}\right)^+} d_i + \sum_{i=1}^{U} \tilde{d}_i > N^{-\alpha}\nu^j\hat{Z}_j^{(1)}. \tag{A.2.25}$$

Combining this with (A.2.22) and using the definition of $\alpha_{N,j}$, we see that in order to prove (A.2.21) it suffices to show that

$$\mathbb{P}\Big(\Big\{\sum_{i=1}^{\alpha_{N,j}} d_i + \sum_{i=1}^{\lceil N^a\rceil} \tilde{d}_i > N^{-\alpha}\nu^j\hat{Z}_j^{(1)}\Big\} \cap E'_{j-1}\Big) \leq CN^{-\beta}. \tag{A.2.26}$$

We will first show that on E'_{j-1} the term $\sum_{i=1}^{\lceil N^a\rceil} \tilde{d}_i$ is small compared to $N^{-\alpha}\nu^j\hat{Z}_j^{(1)}$, if we choose a sufficiently small. On E'_{j-1}, we have $\sum_{i=1}^{j} Z_i^{(1)} \geq N^{\frac{1}{2}-\delta}$, and so, with probability larger than $1 - CN^{-\beta}$, according to Lemma A.2.6, we have that also $Z_j^{(1)} \geq N^{\frac{1}{2}-2\delta}$. Hence, using that for all i, $\tilde{d}_i \leq \max_{1\leq i\leq N} D_i = D_N^{(N)}$, and the inequality of Lemma A.2.3, we get

$$\begin{aligned}
\mathbb{P}\Big(\Big\{\sum_{i=1}^{\lceil N^a\rceil} \tilde{d}_i > \frac{1}{2}N^{-\alpha}\nu^j\hat{Z}_j^{(1)}\Big\} \cap E'_{j-1}\Big) &\leq CN^{-\beta} + \mathbb{P}\Big(\sum_{i=1}^{\lceil N^a\rceil} \tilde{d}_i > \frac{1}{2}N^{-\alpha}\nu^j N^{\frac{1}{2}-2\delta}\Big) \\
&\leq CN^{-\beta} + \mathbb{P}\Big(\sum_{i=1}^{\lceil N^a\rceil} \tilde{d}_i > \frac{1}{2}N^{\frac{1}{2}-2\eta-\alpha} N^{\frac{1}{2}-2\delta}\Big) \\
&\leq CN^{-\beta} + \mathbb{P}\Big(N^a D_N^{(N)} > \frac{1}{2}N^{1-2\eta-\alpha-2\delta}\Big) \\
&\leq CN^{-\beta} + cN^{1-(\tau-1)\gamma},
\end{aligned}$$

where $\gamma = 1 - 2\eta - \alpha - 2\delta - a < \frac{1}{2}$, but can be taken arbitrary close to $\frac{1}{2}$. Since $\tau > 3$, we then have that $cN^{1-(\tau-1)\gamma} < N^{-\beta}$.

Hence it suffices to prove the statement in (A.2.26) without the term $\sum_{i=1}^{\lceil N^a\rceil}\tilde d_i$, that is, it suffices to prove

$$\mathbb{P}\Big(\sum_{i=1}^{\alpha_{N,j}} d_i > \frac{1}{2}N^{-\alpha}\nu^j\hat Z_j^{(1)}, E'_{j-1}\Big) \le CN^{-\beta}. \tag{A.2.27}$$

Since we can write $\hat Z_j^{(1)} = \sum_{i=1}^{\hat Z_{j-1}^{(1)}} d_i$, and, using again Lemma A.2.6, we have that E'_{j-1} implies $\hat Z_{j-1}^{(1)} \ge N^{\frac12-2\delta}$, with probability larger than $1-CN^{-\beta}$, it is sufficient to prove that

$$\mathbb{P}\Big((1-N^{-\alpha}\nu^j)\sum_{i=1}^{\alpha_{N,j}} d_i > \frac{1}{2}N^{-\alpha}\nu^j\sum_{i=\alpha_{N,j}+1}^{\hat Z_{j-1}^{(1)}} d_i, Z_{j-1}^{(1)} \ge N^{\frac12-2\delta}\Big) \le N^{-\beta}. \tag{A.2.28}$$

Now $\mathbb{E}_N[d]=\nu_N$ and, given $\hat Z_{j-1}^{(1)}$, the variance of $\sum_{i=1}^{\hat Z_{j-1}^{(1)}}(d_i-\nu_N)$ equals $\hat Z_{j-1}^{(1)}\mathrm{Var}_N(d)$. Therefore, by the Chebychev inequality,

$$\begin{aligned}
&\mathbb{P}_N\Big((1-N^{-\alpha}\nu^j)\sum_{i=1}^{\alpha_{N,j}} d_i > \frac{1}{2}N^{-\alpha}\nu^j\sum_{i=\alpha_{N,j}+1}^{\hat Z_{j-1}^{(1)}} d_i\Big|\hat Z_{j-1}^{(1)}\Big)\\
&\quad\le \mathbb{P}_N\Big((1-N^{-\alpha}\nu^j)\sum_{i=1}^{\alpha_{N,j}}(d_i-\nu_N) - N^{-\alpha}\nu^j\sum_{i=\alpha_{N,j}+1}^{\hat Z_{j-1}^{(1)}}(d_i-\nu_N) > \frac{1}{2}\nu_N\alpha_{N,j}(\nu-1)\Big|\hat Z_{j-1}^{(1)}\Big)\\
&\quad\le \frac{4\hat Z_{j-1}^{(1)}\mathrm{Var}_N(d_1)}{(\nu_N\alpha_{N,j}(\nu-1))^2} = \frac{4\mathrm{Var}_N(d_1)}{\hat Z_{j-1}^{(1)}N^{-2\alpha}\nu^{2j}(1-\nu^{-1})^2\nu_N^2}.
\end{aligned}$$

We use Lemma A.2.5. Hence, by intersecting with the event $I[A_N]$ and its complement, and using (A.2.8), we obtain for $j \ge (\frac12-2\eta)\log_\nu N$,

$$\begin{aligned}
&\mathbb{P}\Big((1-N^{-\alpha}\nu^j)\sum_{i=1}^{\alpha_{N,j}} d_i > N^{-\alpha}\nu^j\sum_{i=\alpha_{N,j}+1}^{\hat Z_{j-1}^{(1)}} d_i, Z_{j-1}^{(1)} \ge N^{\frac12-2\delta}\Big)\\
&\quad\le c_1N^{-\epsilon} + c_2\frac{\mathbb{E}\Big[\mathrm{Var}_N(d_1)I[A_N]\Big]}{N^{\frac12-2\delta}N^{-2\alpha}\nu^{2j}}\\
&\quad\le c_1N^{-\epsilon} + c_2N^{2\alpha+2\delta-\frac12-1+4\eta}N^{(4-\tau)^+\gamma} \le c_1N^{-\epsilon}+c_2N^{-\beta} \le CN^{-\beta},
\end{aligned}$$

by fixing $\alpha > \frac12+\eta$ so that the exponent is negative (using that $\gamma<\frac12$ and $(4-\tau)^+\le 1$), and writing $\beta = \frac32-2\alpha-2\delta-4\eta-(4-\tau)^+\gamma > 0$. This proves (A.2.28) and completes the proof of Lemma A.2.7. □

Before turning to the proof of the bound on $\mathbb{P}(E_{j,>}^c)$ in Lemma A.2.9 below, we start with a preparatory lemma and some definitions. Suppose we have L objects divided into N groups of sizes $d_1,\ldots,d_N$, so that $L=\sum_{i=1}^N d_i$. Suppose we draw an object at random, and we define a random variable by d_I-1 when the object is taken from the I^{th} group. This gives a distribution $g^{(\vec d)}$, i.e.,

$$g_n^{(\vec d)} = \frac{1}{L}\sum_{i=1}^N d_iI[d_i=n+1]. \tag{A.2.29}$$

Clearly, $g^{(N)} = g^{(\vec D)}$, where $\vec D=(D_1,\ldots,D_N)$.

We next label M of the L objects, and suppose that the distribution $g^{(\vec d)}(M)$ is obtained in a similar way from drawing conditionally on drawing an unlabelled object. More precisely, we remove the labelled objects from all objects thus creating new $d'_1, \ldots, d'_N$, $\sum d'_i = L - M$, and we let $g^{(\vec{d})}(M) = g^{(\vec{d}')}$. Even though this is not indicated, the law $g^{(\vec{d})}(M)$ depends on what objects have been labelled.

Lemma A.2.8 below shows that the law $g^{(\vec{d})}(M)$ can be bounded above and below by two specific ways of labeling the M objects. Before we can state the lemma, we need to describe those specific labellings.

For a vector $\vec{d}$, we let $d_{(1)}, \ldots, d_{(N)}$ be the ordered vector, so that $d_{(1)} = \min_{i=1,\ldots,N} d_i$ and $d_{(N)} = \max_{i=1,\ldots,N} d_i$. Then the laws $f^{(\vec{d})}(M)$ and $h^{(\vec{d})}(M)$, respectively, are defined by successively decreasing $d_{(N)}$ and $d_{(1)}$ respectively, by one. Thus,

$$f_n^{(\vec{d})}(1) = \frac{1}{L-1}\sum_{i=1}^{N-1} d_{(i)} I[d_{(i)} = n+1] + \frac{(d_{(N)}-1)I[d_{(N)}-1 = n+1]}{L-1} \tag{A.2.30}$$

$$h_n^{(\vec{d})}(1) = \frac{1}{L-1}\sum_{i=2}^{N} d_{(i)} I[d_{(i)} = n+1] + \frac{(d_{(1)}-1)I[d_{(1)}-1 = n+1]}{L-1}. \tag{A.2.31}$$

For $f^{(\vec{d})}(M)$ and $h^{(\vec{d})}(M)$, respectively, we repeat the above change M times. Here we note that when $d_{(1)} = 1$, and for $h^{(\vec{d})}(1)$ we decrease it by one, that we only keep the $d_i \geq 1$. Thus, in this case, the number of groups of objects is decreased by 1.

Finally, we write that $f \preceq g$ when the distribution f is stochastically dominated by g, i.e., when $\sum_{i=0}^n f_i \geq \sum_{i=0}^n g_i$ for all $n \geq 0$. Similarly, we write that $X \preceq Y$ when for the probability mass functions f_X, f_Y we have that $f_X \preceq f_Y$.

We next prove stochastic bounds on the distribution $g^{(\vec{d})}(M)$ that are uniform in the choice of the M labelled objects.

Lemma A.2.8 *For all choices of M labelled objects*

$$f^{(\vec{d})}(M) \preceq g^{(\vec{d})}(M) \preceq h^{(\vec{d})}(M). \tag{A.2.32}$$

Thus, the expectation and variance of the random variable $X(M)$ with probability mass function $g^{(\vec{d})}(M)$ are bounded by

$$\mathbb{E}[X(M)] \leq \mathbb{E}[\overline{X}(M)], \qquad \mathrm{Var}[X(M)] \leq \mathbb{E}[\overline{X}(M)^2], \tag{A.2.33}$$

where $\overline{X}(M)$ has probability mass function $h^{(\vec{d})}(M)$.

Moreover, when $X_1, \ldots, X_l$ are draws from $g^{(\vec{d})}(M_1), \ldots, g^{(\vec{d})}(M_l)$, where the only dependence between the X_i resides in the labelled objects, then

$$\sum_{i=1}^{l} \underline{X}_i \preceq \sum_{i=1}^{l} X_i \preceq \sum_{i=1}^{l} \overline{X}_i, \tag{A.2.34}$$

where $\{\underline{X}_i\}_{i=1}^l$ and $\{\overline{X}_i\}_{i=1}^l$, respectively, are i.i.d. copies of $\underline{X}$ and $\overline{X}$ with laws $f^{(\vec{d})}(M)$ and $h^{(\vec{d})}(M)$ for $M = \max_{i=1}^l M_i$, respectively.

In the proof of Proposition A.2.1, we will only use the upper bounds in Lemma A.2.8.

Proof. In order to prove (A.2.32), we will use induction in M. We note that $f^{(\vec{d})}(0) = g^{(\vec{d})}(0) = h^{(\vec{d})}(0) = g^{(\vec{d})}$, and this initializes the induction. To advance the induction, we note that we need to investigate the effect of labelling one extra object. For $f^{(\vec{d})}(M)$, we need to maximize the cumulative distribution function, whereas for $h^{(\vec{d})}(M)$, we need to minimize it. Clearly, (A.2.30-A.2.31) are optimal. This advances the induction. The statement in (A.2.33) follows from (A.2.32)

To prove (A.2.34), we see that for every j, conditionally on the 'past' $(X_1, \ldots, X_{j-1})$, the random variable X_j is stochastically bounded by $\underline{X}_j$ and $\overline{X}_j$, respectively. This completes the proof of Lemma A.2.8. □

Lemma A.2.9 *There exists $\beta > 0$ such that for all $j \leq (\frac{1}{2} + \eta)\log_\nu N$,*

$$\mathbb{P}(E_{j,>}^c) \leq CN^{-\beta}. \tag{A.2.35}$$

Proof. The proof of Lemma A.2.9 follows the proof of Lemma A.2.7, and we focus on the differences only.

Let V denote the number of stubs out of the $\hat{Z}_{j-1}^{(1)}$ stubs that are attached to stubs with label 3 in the BP. Since for each stub in the $(j-1)^{\text{st}}$ generation, on E'_{j-1}, we have that there are at most $2\sum_{i=1}^{j-1} Z_i^{(1)} \leq 2N^{\frac{1}{2}+\delta}$ stubs with label 3, we have that V is bounded from above by a binomial random variable with $n = N^{\frac{1}{2}+\delta}$ and $p = 2N^{\frac{1}{2}+\delta}/L_N$. Thus, by the Markov inequality, we have that for any $a > 2\delta$,

$$\mathbb{P}(V \geq N^a) \leq CN^{-\beta}, \qquad \text{with} \qquad \beta = a - 2\delta > 0, \tag{A.2.36}$$

where we can take a arbitrarily small by choosing $\delta > 0$ small.

We thus assume that $V \leq N^a$. We next proceed by investigating $\mathbb{P}(E_{j,>}^c)$. Now, on $E_{j,>}^c \cap E_{j-1}$, we have that

$$Z_j^{(1)} > (1 + N^{-\alpha}\nu^j)\hat{Z}_j^{(1)}. \tag{A.2.37}$$

Thus, $Z_j^{(1)}$ is larger than $\hat{Z}_j^{(1)}$. We note that $Z_j^{(1)}$ can only become larger than $\hat{Z}_j^{(1)}$ from (a) a redraw and the redraw exceeds the original draw from $g^{(N)}$; and (b) stubs in $Z_{j-1}^{(1)}$ that are not in $\hat{Z}_{j-1}^{(1)}$ which give rise to new stubs. On E_{j-1}, we thus have that (recalling that $\alpha_{N,j} = N^{-\alpha}\nu^{j-1}\hat{Z}_{j-1}^{(1)}$)

$$Z_j^{(1)} - \hat{Z}_j^{(1)} \leq \sum_{i=1}^{\alpha_{N,j}} d'_i + \sum_{i=1}^{V} d''_i, \tag{A.2.38}$$

where d'_i, d''_i are drawn from the appropriate conditional distributions given that we pick a stub with label unequal to 3.

We note that each of the d'_i, d''_i is obtained by drawing from stubs conditionally on labels not being 3. Since the total number of stubs labeled 3 is throughout the growth process bounded above by $2\sum_{i=1}^{j-1} Z_i^{(1)} \leq 2N^{\frac{1}{2}+\delta}$, on $V \leq N^a$, we obtain that by Lemma A.2.8, $\{d'_i\}_{i=1}^{\alpha_{N,j}}$ and $\{d''_i\}_{i=1}^{V}$ are bounded above by $\alpha_{N,j} + \lceil N^a \rceil$ independent copies of $\overline{X}_i(2N^{\frac{1}{2}+\delta})$, where for any M, $\overline{X}_i(M)$ has probability distribution $h^{(\vec{D})}(M)$.

We note that by (A.2.33) and Proposition 3.4, the expectation of $\overline{X}_i(2N^{\frac{1}{2}+\delta})$ is bounded above by $\nu + N^{-\alpha_2}$ for some $\alpha_2 > 0$, and the variance of $\overline{X}_i(2N^{\frac{1}{2}+\delta})$ obeys the same bound as $\mathrm{Var}_N(d)$ in Lemma A.2.5. Thus, we can copy the remaining part of the proof from the proof of Lemma A.2.7. □

A.3 Proof of Proposition 3.3

In this section, we prove Proposition 3.3. In fact, we will prove a slightly different result, as formulated in the next proposition. This proposition summarizes the coupling results, and will be instrumental both in this paper, as well as in [25], in which we investigate the case where $\tau \in (2,3)$.

Proposition A.3.1 *Fix $\tau > 2$, and assume that (1.2) holds. For any m such that, for any $\eta > 0$ small enough,*

$$\mathbb{P}(\sum_{j=1}^{m} \hat{Z}_j^{(i)} \geq N^\eta) = o(1), \tag{A.3.1}$$

there exist independent branching processes $\mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}$, such that

$$\lim_{N\to\infty} \mathbb{P}(Z_m^{(i)} = \mathcal{Z}_m^{(i)}) = 1. \tag{A.3.2}$$

Remark: For fixed m, by the Markov inequality, (A.3.1) indeed holds. Therefore, Proposition 3.3 follows from (A.3.2). We are left to prove Proposition A.3.1.
Proof. By (A.3.1), it suffices to show that $\mathbb{P}(Z_m^{(i)} = \mathcal{Z}_m^{(i)}, \sum_{j=1}^m \hat{Z}_j^{(i)} < N^\eta) = 1 + o(1)$. For this, we use Lemma A.2.2 to conclude that, for $\eta < 1/2$,

$$\mathbb{P}(Z_m^{(i)} = \mathcal{Z}_m^{(i)}, \sum_{j=1}^m \hat{Z}_j^{(i)} < N^\eta) = \mathbb{P}(\hat{Z}_m^{(i)} = \mathcal{Z}_m^{(i)}, \sum_{j=1}^m \hat{Z}_j^{(i)} < N^\eta) + o(1). \tag{A.3.3}$$

By the coupling between $\hat{Z}_m^{(i)}$ and $\mathcal{Z}_m^{(i)}$, a miscoupling occurs with probability equal to p_N defined in (3.6). Therefore, by Remark A.1.3, the probability of a miscoupling for the offspring of a given individual is bounded from above by $N^{-\alpha_2}$ with probability $1 + O(N^{-\beta_2})$. On the event that $\sum_{j=1}^m \hat{Z}_j^{(i)} < N^\eta$, the number of individuals that need to be coupled is bounded from above by N^η. We thus obtain that for any $\eta < \alpha_2$,

$$\mathbb{P}(\hat{Z}_m^{(i)} \neq \mathcal{Z}_m^{(i)}, \sum_{j=1}^m \hat{Z}_j^{(i)} < N^\eta, p_N \leq N^{-\alpha_2}) \leq N^\eta N^{-\alpha_2} = o(1), \tag{A.3.4}$$

which completes the proof. □

A.4 Limits of the form $\lim_{n\to\infty} \sum_{j=1}^n a_j \nu^{-n}$

Suppose that $\nu > 1$ and that $a_1, a_2, \ldots$ is a sequence of real numbers with the property that $\lim_{n\to\infty} a_n \nu^{-n} = a$. We claim that

$$\lim_{n\to\infty} \sum_{j=1}^n \frac{a_j}{\nu^n} = \frac{\nu a}{\nu - 1}. \tag{A.4.1}$$

Indeed, we can write:

$$\sum_{j=1}^n \frac{a_j}{\nu^n} = \frac{a_n}{\nu^n}\left(1 + \frac{a_{n-1}}{a_n} + \ldots + \frac{a_{j-1}}{a_n} + \ldots + \frac{a_1}{a_n}\right).$$

For fixed j as $n \to \infty$:

$$\frac{a_n}{\nu^n}\left(\frac{a_{j-1}}{a_n} + \ldots + \frac{a_1}{a_n}\right) \to 0,$$

since $\nu > 1$. On the other hand for $k \geq j$,

$$\frac{1-\varepsilon}{\nu^{n-k}} \leq \frac{a_k}{a_n} = \frac{1}{\nu^{n-k}} \frac{a_k/\nu^k}{a_n/\nu^n} \leq \frac{1+\varepsilon}{\nu^{n-k}}.$$

Hence for j sufficiently large,

$$-\varepsilon + \sum_{k=j}^n \frac{1-\varepsilon}{\nu^{n-k}} \leq \left(1 + \frac{a_{n-1}}{a_n} + \ldots + \frac{a_{j-1}}{a_n} + \ldots + \frac{a_1}{a_n}\right) \leq \varepsilon + \sum_{k=j}^n \frac{1+\varepsilon}{\nu^{n-k}}.$$

Since for $n \to \infty$ (and $\nu > 1$),

$$\sum_{k=j}^n \frac{1}{\nu^{n-k}} = \sum_{k=0}^{n-j} \frac{1}{\nu^k} = \frac{1 - (1/\nu)^{n-j+1}}{1 - 1/\nu} \to \frac{\nu}{\nu - 1},$$

we obtain (A.4.1). Suppose now that we have two sequences $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ with the property that $\lim_{n\to\infty} a_n \nu^{-n} = a$ and $\lim_{n\to\infty} b_n \nu^{-n} = b$. We claim that

$$\lim_{n\to\infty} \sum_{j=1}^n \frac{a_j b_j}{\nu^{2n}} = \frac{\nu^2 ab}{\nu^2 - 1}. \tag{A.4.2}$$

Indeed the sequence $a_n\nu^{-n} \cdot b_n\nu^{-n}$ converges to ab and so we obtain (A.4.2) from (A.4.1).

For the proof of (4.20) we have to distinguish between the cases where the upper bound of the summation in (4.20) is either even or odd. We start with the odd case $\sigma_N + k + 1 = 2n + 1$:

$$\sum_{i=2}^{2n+1} \frac{a_{\lceil i/2\rceil} b_{\lfloor i/2\rfloor}}{\nu^{2n}} = \left(\frac{a_1b_1 + a_2b_2 + \ldots + a_nb_n}{\nu^{2n}}\right) + \left(\frac{a_2b_1 + a_3b_2 + \ldots + a_{n+1}b_n}{\nu^{2n}}\right).$$

According to (A.4.2), the first series on the right hand side converges to $\frac{\nu^2 ab}{\nu^2-1}$. Because we miss one additional factor ν in the denominator of the second series on the right hand side we conclude from again (A.4.2) that this second series converges to: $\nu \cdot \frac{\nu^2 ab}{\nu^2-1}$. Hence

$$\lim_{n\to\infty} \sum_{i=2}^{2n+1} \frac{a_{\lceil i/2\rceil} b_{\lfloor i/2\rfloor}}{\nu^{2n}} = \frac{\nu^2 ab}{\nu^2-1} + \nu\frac{\nu^2 ab}{\nu^2-1} = \frac{\nu^2 ab}{\nu-1}. \tag{A.4.3}$$

Applying this to our almost sure converging sequence with

$$a_j = \mathcal{Z}_j^{(1)}, \quad b_j = \mathcal{Z}_j^{(2)}, \quad a = \mu\mathcal{W}^{(1)}/\nu \quad b = \mu\mathcal{W}^{(2)}/\nu, \text{ and } \sigma_N + k = 2n,$$

we obtain,

$$\lim_{N\to\infty} \frac{\sum_{i=2}^{\sigma_N+k+1} \mathcal{Z}^{(1)}_{\lceil i/2\rceil}\mathcal{Z}^{(2)}_{\lfloor i/2\rfloor}}{\mu^2\nu^{\sigma_N+k}} = \lim_{N\to\infty} \frac{\sum_{i=2}^{2n+1} a_{\lceil i/2\rceil} b_{\lfloor i/2\rfloor}}{\mu^2\nu^{2n}} = \frac{1}{\mu^2}\frac{\nu^2 ab}{\nu-1} = \frac{\mathcal{W}^{(1)}\mathcal{W}^{(2)}}{\nu-1}.$$

For $\sigma_N + k + 1 = 2n$, the even case:

$$\sum_{i=2}^{2n} \frac{a_{\lceil i/2\rceil} b_{\lfloor i/2\rfloor}}{\nu^{2n-1}} = \left(\frac{a_1b_1 + a_2b_2 + \ldots + a_nb_n}{\nu^{2n-1}}\right) + \left(\frac{a_2b_1 + a_3b_2 + \ldots + a_nb_{n-1}}{\nu^{2n-1}}\right).$$

According to (A.4.2), the first series on the right hand side converges to $\nu\frac{\nu^2 ab}{\nu^2-1}$, and the second to: $\frac{\nu^2 ab}{\nu^2-1}$. The even and odd case together yields the limit (4.20) of step 4.

Acknowledgement. The work of RvdH was supported in part by Netherlands Organisation for Scientific Research (NWO). We thank Dmitri Znamenski for the Figures 1 and 2 and for useful comments on a previous version. We thank the two referees for many suggestions that improved on the readability of the paper.

References

[1] W. Aiello, F. Chung and L. Lu. A random graph model for power law graphs. *Experiment. Math.* **10**, no. 1, 53–66, 2001.

[2] W. Aiello, F. Chung and L. Lu. Random evolution of massive graphs. In Handbook of Massive Data Sets, J. Abello, P.M. Pardalos and M.G.C. Resende, eds., Kluwer Academic, Dordrecht, 97–122, 2002.

[3] R. Albert and A.-L. Barabási. Emergence of Scaling in Random Networks. *Science* **286**: 509-512, 1999.

[4] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.,* **74**, 47-97, 2002.

[5] N. Alon and J. Spencer. *The Probabilistic Method,* 2nd Edition. John Wiley and Sons, New York (2000).

[6] S. Asmussen. Some martingale methods in the limit theory of supercritical branching processes, *Branching processes, A. Joffe and P. Ney (Editors), Marcel Dekker, New York and Basel*, pp. 1-26, 1978.

[7] K.B. Athreya and P.E. Ney. *Branching Processes,* Springer, Berlin 1972.

[8] A.-L. Barabási. *Linked: The New Science of Networks,* Perseus Publishing, Cambridge, Massachusetts, 2002.

[9] B. Bollobás. *Random Graphs,* 2nd edition. Academic Press, New York, 2001.

[10] B. Bollobás, C. Borgs, J.T. Chayes and O. Riordan. Directed scale-free graphs. Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003), 132–139, ACM, New York, 2003.

[11] B. Bollobás and O. Riordan. The diameter of scale-free random graphs. Preprint Dept. of Math. Sciences, University of Memphis, 2003.

[12] B. Bollobás and O. Riordan. Mathematical results on scale-free random graphs. *Handbook of graphs and networks*, 1–34, Wiley-VCH, Weinheim, 2003.

[13] B. Bollobás and O. Riordan. Coupling scale-free and classical random graphs. Preprint, 2003.

[14] B. Bollobás, O. Riordan, J. Spencer and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, **18**, 279–290, 2001.

[15] B. Bollobás and F. de la Vega. The diameter of random regular graphs. *Combinatorica*, **2** 125-134, 1982.

[16] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener Graph Structure in the Web. *Computer Networks*, **33**, 309-320, 2000.

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

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

[19] C. Cooper and A.A. Frieze. A general model of web graphs. *Random Structures Algorithms*, **22** (3), 311–335, 2003.

[20] S.N. Dorogovtsev, A.V. Goltsev and J.F.F. Mendes. Pseudofractal scale-free web, *Phys. Rev. E* **65**, 066122, 2002.

[21] S.N. Dorogovtsev, J.F.F. Mendes and A.N. Samuhkin. Metric structure of random networks, *Nucl. Phys. B.* 653, 307, 2003.

[22] C. Faloutsos, P. Faloutsos and M. Faloutsos. On power-law relationships of the internet topology, *Computer Communications Rev.*, **29**, 251-262, 1999.

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

[24] R. van der Hofstad, G. Hooghiemstra and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees. In preparation.

[25] R. van der Hofstad, G. Hooghiemstra and D. Znamenski. Distances in random graphs with infinite mean degrees. Preprint 2004.

[26] R. van der Hofstad, G. Hooghiemstra and D. Znamenski. Connected components in random graphs with i.i.d. degrees. In preparation.

[27] G. Hooghiemstra and P. Van Mieghem. On the mean value of the logarithm of a martingale limit in branching processes, Preprint 2003. [Online available from `http://ssor.twi.tudelft.nl/ gerardh/`].

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

[29] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Trawling the Web for emerging cyber communities. *Computer Networks,* **31**,(11-16), 1481–1493, 1999.

[30] R. Kumar, P. Raghavan, S. Rajagopalan D. Sivakumar, A. Tomkins and E. Upfal. Stochastic Models for the Web Graph. *42st Annual IEEE Symposium on Foundations of Computer Science*, 57–65, 2000.

[31] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence, *Random Structures and Algorithms*, **6**, 161-179, 1995.

[32] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence, *Combin. Probab. Comput.*, **7**, 295-305, 1998.

[33] M.E.J. Newman. The structure and function of complex networks. *SIAM Rev.* **45**, no. 2, 167–256, 2003.

[34] V. Paxson. End-to-end routing behaviour in the Internet. *IEEE Transac. Networking*, **5**(5), 601–615, 1997.

[35] M.E.J. Newman, S.H. Strogatz, and D.J. Watts. Random graphs with arbitrary degree distribution and their application, *Phys. Rev. E*, **64**, 026118, 1-17.

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

[37] S. H. Strogatz. Exploring complex networks. *Nature*, 410(8), 268–276, March 2001.

[38] H Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topology generators: Degree-based vs. structural. *ACM Sigcomm'02, Pittsburgh, Pennsylvania, pp. 19-23, USA*, 2002.

[39] H. Thorisson. *Coupling, Stationarity and Regeneration*. Springer, New York 2000.

[40] P. van Mieghem, G. Hooghiemstra and R. van der Hofstad. A scaling law for the hopcount. Report 2000125, Delft University of Technology, Delft, The Netherlands. [Online Available on `http://www.nas.its.tudelft.nl/people/Piet/papers/hopcount.pdf`].

[41] D. J. Watts. *Small Worlds, The Dynamics of Networks between Order and Randomness.* Princeton University Press, Princeton, New Jersey, 1999.

[42] D. Williams. *Diffusions, Markov Processes and Martingales,* John Wiley and Sons, 1979.