On the mean distance in scale free graphs

G. Hooghiemstra and P. Van Mieghem*

August 5, 2003

Abstract

We consider the degree graph, where all nodes have a pre-described degree distribution F, and where nodes are randomly connected in accordance to their degree. Based on a recent result [R. van der Hofstad *et al.*, http://ssor.twi.tudelft.nl/ gerardh/], we improve the approximation of the mean distance between arbitrarily chosen nodes given by [M.E.J. Newman *et al.*, Phys. Rev. E **64**, 026118, 2001]. Our new expression for the average distance involves the mean of the logarithm of the limit of a super-critical branching process. We compare simulations of the average distance with the results of Newman *e*t al. and with our new approach.

PACS number(s): 02.10.Ox, 02.50.Cw

1 Introduction

In several disciplines it was observed that graph models for real-world phenomena often possess a specific degree distribution. In electrical engineering and computer science, Faloutsos *et al.* [5] have shown that the degree distribution of the Internet graph follows a power law. Since their paper, a large amount of additional measurements to infer properties of the graph of the Internet have been published (see e.g. [12], [13]). In physics, Newman *et al.* [10, Introduction] have studied additional real-world networks described by various degree distributions. A general characteristic is that most of these 'real-world' degree distributions often differ considerably from the Poisson distribution that asymptotically (for a large number of nodes N) describes the degree distribution of the random graph $G_p(N)$. Barabasi [3] showed that preferential attachment of nodes gives rise to a class of graphs called 'scaled free networks'. Scale free graphs have a power law degree distribution which contrasts with small world networks and with random graphs. Preferential attachment implies that in dynamic models the nodes with the larger degree are the more likely candidates for attachment of new nodes.

The interest in the understanding of real-world graphs has led to the introduction of graph models that are specified by a certain degree sequence. In a series of papers ([10], [11] and [7]) the following graph model was studied. For a given discrete probability distribution f_j, $j = 0, 1, \ldots$, and an integer N we define the i.i.d. sequence $D_1, D_2, \ldots, D_N$ with (common) distribution

$$\mathbb{P}[D = j] = f_j, \quad j = 0, 1, 2, \ldots. \tag{1}$$

The construction of the involved graph model starts with N nodes, where node j has D_j (random!) outgoing links (initially called stubs). These stubs are connected by matching the stubs randomly. Each matched pair of stubs form together one edge or link. If the total number of stubs $D_1 + D_2 + \ldots + D_N$ in the random matching construction is odd an additional node with exactly one stub is introduced to remedy this minor difficulty. We refer to [7, Introduction] for more details. Let $F(x)$ be the cumulative distribution function of the discrete distribution f_j, i.e., $F(x) = \sum_{j=0}^{\lfloor x \rfloor} f_j$, where $\lfloor x \rfloor$ is the largest integer smaller than x. As in [7] we will assume that for some $\tau > 3$ and some positive c,

$$1 - F(x) \le c x^{-\tau+1}, \quad x > 0. \tag{2}$$

This graph model that includes heavy tail degrees with finite variance is called a *degree graph*, because the degree of each node has distribution F. Furthermore, we define

$$\mu = \mathbb{E}[D], \qquad \nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}. \tag{3}$$

The importance of the parameter ν will be clarified below.

We now turn to the main result of this paper. The distance (number of links or hops) between two arbitrary nodes is a characterizing property of a graph.

*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

Moreover, the computational complexity of a routing algorithm is nicely related to the distance distribution in a graph and in particular to the average distance. For $\tau > 3$, the average distance (or hopcount) H_N between node 1 and node 2 (or more generally between two arbitrary nodes) in the degree graph was shown heuristically in [10] to satisfy approximately

$$\mathbb{E}[H_N] \approx \frac{\log(N/\mu)}{\log \nu} + 1. \tag{4}$$

Concerning the asymptotic distribution of H_N, we have proved the following result [7].

Theorem 1.1 *Assume that $\tau > 3$ and that the expected value ν defined in (3) satisfies $\nu > 1$. For $n \geq 1$ define $a_n = \lfloor \log_\nu n \rfloor - \log_\nu n \in (-1, 0]$. Conditionally on node 1 and node 2 being connected, there exist random variables $\{R_a\}_{a\in(-1,0]}$ such that for each $\alpha > 0$ as $N \to \infty$,*

$$\begin{aligned} &\mathbb{P}\left[H_N - \lfloor \log_\nu N \rfloor = k\right] \\ &\quad = \mathbb{P}\left[R_{a_N} = k\right] + O((\log N)^{-\alpha}), \quad k \in \mathbb{Z}. \end{aligned} \tag{5}$$

Furthermore, if $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ are two independent copies of the random variable introduced below, then for fixed $a \in (-1, 0]$, and $k \in \mathbb{Z}$,

$$\begin{aligned} &\mathbb{P}\left[R_a > k\right] \\ &\quad = \mathbb{E}\left[e^{-\mu(\nu-1)^{-1}\nu^{a+k}\mathcal{W}^{(1)}\mathcal{W}^{(2)}} \,\middle|\, \mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0\right]. \end{aligned} \tag{6}$$

The random variables $\mathcal{W}^{(1)}$ and $\mathcal{W}^{(2)}$ are independent copies of the almost sure limit $\mathcal{W}$ of the normalized (delayed) branching process

$$\mathcal{W}_n = \mathcal{Z}_n/\mu\nu^{n-1}, \tag{7}$$

where $\mathcal{Z}_0 = 1$, having offspring distribution in the first generation given by (2), whereas in all further generations the offspring distribution is given by

$$g_j = \frac{(j+1)f_{j+1}}{\mu}, \quad j = 0, 1, \ldots. \tag{8}$$

The mean value of the discrete distribution $\{g_j\}$ is denoted by $\nu = \sum j g_j$ and satisfies (3). The involvement of $\{g_j\}$ becomes more transparent if one realizes that the stubs are matching according to size. Since the number of stubs at each node follows distribution $\{f_j\}$, sampling in accordance with size means sampling proportional to jf_j. We sample in fact proportional to $(j+1)f_{j+1}$, because one stub is needed for attachment, see also [10, (Section II.a, Reference (9)].

Theorem 1.1 suggests that the average distance H_N satisfies

$$\mathbb{E}[H_N] = \lfloor \log_\nu N \rfloor + \mathbb{E}[R_{a_N}] + o(1), \tag{9}$$

and that the asymptotic variance satisfies

$$\mathrm{Var}\,[H_N] = \mathrm{Var}\,[R_{a_N}] + o(1). \tag{10}$$

In this paper we will focus on the first moment $\mathbb{E}[R_a]$. Since we include an expression for the generating function $\varphi_{R_a}(z) = \mathbb{E}[z^{R_a}]$ in the appendix, extensions to higher moments can also be obtained. We show that for $\nu \in (1, 5]$, and with overwhelming accuracy:

$$\begin{aligned} \mathbb{E}[R_a] \;\approx\; & \frac{1}{2} - a - \left(\frac{\gamma + \log\mu - \log(\nu-1)}{\log\nu}\right) \\ & -2\frac{\mathbb{E}[\log\mathcal{W}|\mathcal{W} > 0]}{\log\nu}, \end{aligned} \tag{11}$$

where $\mathcal{W}$ is the limit of the delayed branching process introduced in (7), the parameters μ and ν are defined in (3) and where γ denotes Euler's constant ([1, 6.1.3]). Combining (9) and (11) forms the main result of this paper:

$$\begin{aligned} \mathbb{E}[H_N] \;\approx\; & \frac{\log N}{\log\nu} + \frac{1}{2} - \left(\frac{\gamma + \log\mu - \log(\nu-1)}{\log\nu}\right) \\ & -2\frac{\mathbb{E}[\log\mathcal{W}|\mathcal{W} > 0]}{\log\nu}. \end{aligned} \tag{12}$$

The paper is organized as follows. Section 2 discusses the computation of $\mathbb{E}[\log\mathcal{W}|\mathcal{W} > 0]$. Section 3 compares the value of (12) with simulations of the mean hopcount in two types of degree graphs. The important conclusion drawn from the presented computations in Sec. 2 and Sec. 3 is that the expression (12) for the average hopcount between two arbitrary nodes in a degree is surprisingly accurate, even for a small number of nodes N around $N = 500$. Mathematical derivations of the approximation (11) and details on the computation of $\mathbb{E}[\log\mathcal{W}|\mathcal{W} > 0]$ are deferred to the Appendix.

2 Computation of $\mathbb{E}[\log\mathcal{W}|\mathcal{W} > 0]$

Formula (12) requires the expected value $\mathbb{E}[\log\mathcal{W}|\mathcal{W} > 0]$. Unfortunately, little is known about the distribution of the limit random variable $\mathcal{W}$. The functional equation that determines the generating function of $\mathcal{W}$ seems difficult to solve. For the classical super-critical Galton-Watson branching process (our delayed branching process is a simple extension of this model) the distribution of $\mathcal{W}$ is

only explicitly known for geometrically distributed offspring (see [6, p. 92]). In this case $\mathcal{W}$ possesses an atom at 0 and conditionally on the event $\{\mathcal{W} > 0\}$, the distribution of $\mathcal{W}$ is exponential.

Here we use the Fast Fourier Transform (FFT) to approximate the distribution of $\mathcal{W}_n$ and from that distribution the value of $\mathbb{E}[\log \mathcal{W}_n | \mathcal{W}_n > 0]$. This approximation is based on the fact that the generating function of the size of generation n in a simple Galton-Watson process is the functional iterate of the generating function of the offspring distribution. We show in Appendix B that indeed $\mathbb{E}[\log \mathcal{W}_n | \mathcal{W}_n > 0]$ converges to $\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0]$, however to obtain conclusions about the involved error in this approximation is beyond the scope of this paper. We therefore extrapolate our results numerically.

Since in a super-critical Galton-Watson process the size of generation n grows exponentially with n, the memory in current computers confines the accuracy of the computation of $\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0]$ especially if the mean of the offspring distribution is substantially greater than 1. We expect that this drawback will be less pronounced in future implementations of the FFT. The exact expression of $\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0]$ for geometric offspring is used, in Appendix C, as a benchmark to evaluate the accuracy of the FFT.

Since $\mathbb{E}[\mathcal{W} | \mathcal{W} > 0] = \mathbb{E}[\mathcal{W}]/(1-\pi)$, where $\pi = \mathbb{P}(\mathcal{W} = 0)$ is the extinction probability, Jensen's inequality for concave functions provides us with an upperbound,

$$\begin{aligned} \mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0] &\leq \log(\mathbb{E}[\mathcal{W} | \mathcal{W} > 0]) \\ &= \log(\mathbb{E}[\mathcal{W}]) - \log(1-\pi). \end{aligned}$$

Using $\mathbb{E}[\mathcal{W}] = 1$, the Jensen's inequality becomes

$$\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0] \leq -\log(1-\pi). \tag{13}$$

Hence, we obtain a lower bound for the average distance (or hopcount)

$$\begin{aligned} \mathbb{E}[H_N] \geq \log_\nu N + \frac{1}{2} - \left(\frac{\gamma + \log \mu - \log(\nu - 1)}{\log \nu} \right) \\ + \frac{2 \log(1-\pi)}{\log \nu}. \end{aligned} \tag{14}$$

An obvious lower bound for $\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0]$, derived in Appendix B, is

$$\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0] \geq -\int_0^1 \frac{\mathbb{P}(0 < \mathcal{W} < y)}{1-\pi} \frac{dy}{y}.$$

Since the distribution of $\mathcal{W}$ is unknown, this lower bound is not very useful. It would be of interest to have an attainable lower bound for $\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0]$, in turn this would imply an upper bound for the average distance.

In case $\pi = 0$, conditioning on $\mathcal{W} > 0$ boils down to conditioning on the sure event and the Jensen inequality gives:

$$\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0] \leq \log(\mathbb{E}[\mathcal{W}]) = 0.$$

Since in this case the sequence $\{\log \mathcal{W}_n | \mathcal{W}_n > 0\} = \{\log \mathcal{W}_n\}$ is a supermartingale, because $\mathcal{W}_n$ is a martingale and the function $x \mapsto \log x$ is concave, we conclude that for $\pi = 0$, the sequence $\mathbb{E}[\log \mathcal{W}_n | \mathcal{W}_n > 0]$ is monotone decreasing.

A. Examples of two types of degree graphs

We now turn to the delayed branching process described in the introduction and discuss the numerical computation of $\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0]$. In the first generation the offspring is chosen from distribution F satisfying (2), and in further generations the offspring distribution has probability mass function $\{g_j\}$ given in (8). We performed the numerical calculation, for the delayed branching process, with a power law offspring (specified by (15) below) and for a binomial offspring distribution. Observe that for the power law distribution (15), the delayed branching process has extinction probability $\pi = 0$.

First consider distribution

$$F(k) = 1 - \left(\frac{1}{k}\right)^{\tau - 1}, \quad k = 1, 2, 3, \ldots, \tag{15}$$

with $\tau > 3$. Introducing Riemann's zeta function $\zeta(s) = \sum_{j=1}^{\infty} j^{-s}$, $s > 1$, we find

$$\mu = \mathbb{E}[D] = 1 + \zeta(\tau - 1) \quad \text{and} \quad \nu = \frac{2\zeta(\tau - 2)}{1 + \zeta(\tau - 1)}. \tag{16}$$

In Table 1 we take $\tau = 4.5$ and find

$$\mu = 2.1267\ldots \qquad \nu = 1.2615\ldots,$$

We only display the results in the last 4 generations, where the difference between numerical and theoretical mean is less than 0.1, which is our rule of thumb as explained in Appendix C. We will adopt $\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0] \approx -0.6$, as the correct value for $\tau = 4.5$ and will compute in the next section our estimate of $\mathbb{E}[H_N]$. Similarly, for $\tau = 5.5$, we have that

$$\mu = 2.0547\ldots \qquad \nu = 1.0967\ldots,$$

and the numerical results for the last 4 generations are given in Table 2. Extrapolating we adopt the value $\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0] \approx -0.45$.

Table 1: Results for $\tau = 4.5$

n	$\mu\nu^{n-1}$	FFT(approx)	$\mathbb{E}[\log \mathcal{W}_n \mid \mathcal{W}_n > 0]$
13	35.5570	34.5433	-0.5334
14	43.5952	43.5697	-0.5522
15	54.9972	54.9498	-0.5686
16	69.3813	69.2935	-0.5828

Table 2: Results for $\tau = 5.5$

n	$\mu\nu^{n-1}$	FFT(approx)	$\mathbb{E}[\log \mathcal{W}_n \mid \mathcal{W}_n > 0]$
43	99.3130	99.3086	-0.4263
44	108.9199	108.9138	-0.4283
45	119.4562	119.4475	-0.4301
46	131.0117	130.9993	-0.4319

Finally, we consider the binomial probabilities $f_j = \binom{N-1}{j} p^j (1-p)^{N-1-j}$, with mean $\mu = (N-1)p$, for which

$$g_j = \frac{(j+1)f_{j+1}}{Np} = \binom{N-2}{j} p^j (1-p)^{N-2-j},$$

so that $\{g_j\}$ has a binomial distribution with parameters $N-2$ and p. Hence $\nu = (N-2)p$. The last 4 relevant generations for the Poisson distribution with mean $\nu = 1.5$, which is the exact limit for large N, are given in Table 3.

Table 3: Results for the Poisson with $\nu = 1.5$

n	ν^n	FFT(approx)	$\mathbb{E}[\log \mathcal{W}_n \mid \mathcal{W} > 0]$
11	86.4976	86.4976	0.1048
12	129.7463	129.7463	0.1016
13	194.6195	194.6194	0.0992
14	291.9293	291.8817	0.0973

We approximate:

$$\mathbb{E}[\log \mathcal{W} \mid \mathcal{W} > 0] \approx 0.09.$$

For the binomial distribution, the degree distribution $\{f_j\}$ coincides with the degree distribution of the random graph $G_p(N)$; both distributions are binomial with parameters $N-1$ and p. In contrast to the random graph $G_p(N)$, the degree graph can have self-loops by construction. However, the hopcount in both graphs (the degree graph with binomial degrees with parameters $N-1$ and p, and the random graph $G_p(N)$) are asymptotically equal in distribution as follows from the coupling construction in [7]. This implies that we also found for the random graph $G_p(N)$ an asymptotic expression derived from (12) for the average distance (hopcount) between two arbitrary nodes.

3 Comparison of different formulas for $\mathbb{E}[H_N]$

For several values of N we compare a large sample confidence interval of the simulated hopcount, based on 1000 repetitions with the results (4) and (12).

For the power-law distribution (15) with $\tau = 4.5$ we have $\mu = 2.1267$, $\nu = 1.2615$ and $\mathbb{E}[\log \mathcal{W} \mid \mathcal{W} > 0] \approx -0.6$, we present the results in Table 4.

Table 4: Comparison for $\tau = 4.5$

N	simulation	$\mathbb{E}[H_N](4)$	$\mathbb{E}[H_N](12)$
500	(20.10, 21.34)	24.5	20.9
1000	(23.05, 24.22)	27.5	23.9
5000	(30.78, 32.01)	34.4	30.8
10,000	(33.51, 34.70)	37.4	33.8

For $\tau = 5.5$, we have $\mu = 2.0547$, $\nu = 1.0967$ and $\mathbb{E}[\log \mathcal{W} \mid \mathcal{W} > 0] \approx -0.45$, and present our findings in Table 5.

Table 5: Comparison for $\tau = 5.5$

N	simulation	$\mathbb{E}[H_N](4)$	$\mathbb{E}[H_N](12)$
500	(36.55, 39.19)	60.52	38.21
1000	(43.42, 45.96)	68.03	45.72
5000	(61.70, 64.32)	85.47	63.16
10,000	(68.98, 71.64)	92.98	70.66

For the degree graph with a binomial distribution with parameters $N-1$ and p, with $Np = 1.5$ for which $\mathbb{E}[\log \mathcal{W} \mid \mathcal{W} > 0] \approx 0.09$, we refer to Table 6.

In all cases the value (12) for $\mathbb{E}[H_N]$ is contained in the confidence interval obtained from the simulations, whereas the expression

$$\mathbb{E}[H_N] = \frac{\log N/\mu}{\log \nu} + 1,$$

overestimates the actual mean. These values demonstrate the accuracy of (12) for $\mathbb{E}[H_N]$ in degree graphs with a finite variance degree distribution.

Table 6: Comparison for binomial with $Np = 1.5$

N	simulation	$\mathbb{E}[H_N](4)$	$\mathbb{E}[H_N](12)$
500	(10.35, 11.34)	15.3	11.2
1000	(12.25, 13.49)	17.0	12.9
5000	(16.10, 17.34)	21.0	16.9
10,000	(18.25, 19.51)	22.7	18.6

4 Conclusions

We conclude that (12) for $\mathbb{E}[H_N]$ gives a genuine improvement over the Approximation (4). Inspection of both approximations shows that (4) overestimates the correct value of the mean $\mathbb{E}[H_N]$. The error is increasingly large when the Malthusian parameter ν is close to 1 because the factor $\log(N/\mu)$ is divided by $\log \nu$, which approaches zero as $\nu \to 1$. Apparently including the term

$$-\frac{1}{2} - \left(\frac{\gamma - \log(\nu - 1)}{\log \nu}\right) - 2\frac{\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0]}{\log \nu},$$

in the formula (4) corrects the overshoot.

We have shown how the more difficult quantity $\mathbb{E}[\log \mathcal{W} | \mathcal{W} > 0]$ can be computed and have presented a lower bound for $\mathbb{E}[H_N]$. At last, even for moderate values of the number of nodes N, Formula (12) for $\mathbb{E}[H_N]$ seems surprisingly accurate.

Acknowledgment. We thank Remco van der Hofstad for valuable discussions concerning branching processes.

References

[1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions,* Dover (1968).

[2] S. Asmussen, Some martingale methods in the limit theory of supercritical branching processes, in *Branching Processes*, edited by A. Joffe and P. Ney, Advances in Applied Probability **5**, Marcel Dekker, New York (1978).

[3] A.-L. Barabasi. *Linked, the New Science of Networks.* Perseus Publishing, Cambridge, Massachusetts (2002).

[4] M. S. Dubuc. La densité de la loi-limite d'un processus en cascade expansif, Z. Wahrscheinlichkeitstheorie Verw. Gebiete **19**, 281-290 (1971).

[5] C. Faloutsos, P. Faloutsos and M. Faloutsos, On power-law relationships of the Internet topology, Proceedings of the ACM SIGCOMM (1999).

[6] G.R. Grimmett and D.R. Stirzacker, *Probability and Random Processes,* Clarendon Press, Oxford (1982).

[7] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem, Random graphs with finite variance degrees, http://ssor.twi.tudelft.nl/ gerardh/ Preprint 2003 (unpublished).

[8] P. Jagers, *Branching Processes with Biological Applications,* John Wiley and Sons, London (1975).

[9] M. Loève, *Probability Theory,* second edition, Van Nostrand, New York (1960).

[10] 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, 2001; http://arxiv.org/pdf/cond-mat/0007235, 2000.

[11] H. Reittu and I. Norros, On large random graphs of the "Internet type", Cost279TD(01)19, Submitted to Performance Evaluation (2003).

[12] H Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, Network topology generators: Degree-based vs. structural, ACM Sigcomm, Pittsburgh, Pennsylvania, 19-23, USA (2002).

[13] A. Vázquez, R. Pastor-Satorras and A. Vespignani, Large-scale topological and dynamical properties of Internet, http://arxiv.org/pdf/cond-mat/0112400v1 (2001).

APPENDIX A: Generating function for R_a.

Denote by φ_{R_a} the generating function of the random variable R_a, introduced in Theorem 1.1. Then

$$\begin{aligned}
\varphi_{R_a}(z) &= \sum_{k=-\infty}^{\infty} z^k \mathbb{P}[R_A = k] \\
&= \sum_{k=-\infty}^{\infty} z^k \left(\mathbb{P}[R_A > k-1] - \mathbb{P}[R_A > k]\right) \\
&= (z-1) \sum_{k=-\infty}^{\infty} z^k \mathbb{P}[R_A > k].
\end{aligned}$$

Let $a \in (-1,0]$ be fixed. Denote by X a random variable having the same distribution as the random variable

$$\mu(\nu-1)^{-1}\nu^a \mathcal{W}^{(1)}\mathcal{W}^{(2)} | \mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0,$$

then from (6), using Fubini's theorem [9],

$$\varphi_{R_a}(z) = (z-1)\,\mathbb{E}\left[\sum_{k=-\infty}^{\infty} z^k \exp\{-\nu^k X\}\right].$$

Introducing $G(x,z) = G_1(x,z) + G_2(x,z)$, where

$$\begin{aligned} G_1(x,z) &= \sum_{k=0}^{\infty} z^k e^{-x\nu^k}, \\ G_2(x,z) &= \sum_{k=1}^{\infty} z^{-k} e^{-x\nu^{-k}} = \sum_{j=0}^{\infty} \frac{(-x)^j}{j!} \frac{1}{z\nu^j - 1}, \end{aligned}$$

yields

$$\begin{aligned} \varphi_{R_a}(z) &= (z-1)\,\mathbb{E}[G(X,z)] \qquad (17) \\ &= (z-1)\,\{\mathbb{E}[G_1(X,z) + G_2(X,z)]\}. \end{aligned}$$

Let us first consider $G_1(x,z)$. For $\sigma = Re(s) > 0$ and $k \geq 0$, we have [1, 6.1.1],

$$\frac{\Gamma(s)}{\nu^{ks}} = \int_0^{\infty} t^{s-1}\, e^{-\nu^k t}\, dt,$$

and

$$\Gamma(s) \sum_{k=0}^{\infty} \frac{z^k}{\nu^{ks}} = \int_0^{\infty} t^{s-1} \sum_{k=0}^{\infty} z^k e^{-\nu^k t}\, dt,$$

or, assuming $|z/\nu^s| < 1$,

$$\frac{\Gamma(s)\nu^s}{\nu^s - z} = \int_0^{\infty} t^{s-1} G_1(t,z)\, dt.$$

By the inverse Mellin transform, an alternative expression of $G_1(x,z)$ is, with $c > 0$,

$$G_1(x,z) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\Gamma(s)}{\nu^s - z} \left(\frac{\nu}{x}\right)^s ds.$$

By closing the contour over the half plane with negative real part, we encounter simple poles at $s = -k$ ($k \geq 0$) from $\Gamma(s)$ and simple poles at $s = \frac{\log z + 2k\pi i}{\log \nu}$ from $\frac{1}{\nu^s - z}$. By Cauchy's residue theorem,

$$\begin{aligned} &G_1(x,z) \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \frac{(-x)^k}{1 - z\nu^k} + \frac{1}{z \log \nu} \\ &\quad \times \sum_{k=-\infty}^{\infty} \Gamma\left(\frac{\log z + 2k\pi i}{\log \nu}\right) \left(\frac{\nu}{x}\right)^{\frac{\log z + 2k\pi i}{\log \nu}} \\ &= -G_2(x,z) + \frac{z^{-\log_\nu x}}{\log \nu} \\ &\quad \times \sum_{k=-\infty}^{\infty} \Gamma\left(\frac{\log z + 2k\pi i}{\log \nu}\right) e^{-2k\pi i \log_\nu x}. \end{aligned}$$

This implies the following representation for $G(x,z)$,

$$\begin{aligned} &G(x,z) \\ &= \frac{z^{-\log_\nu x}}{\log \nu} \sum_{k=-\infty}^{\infty} \Gamma\left(\frac{\log z + 2k\pi i}{\log \nu}\right) e^{-2k\pi i \log_\nu x}. \end{aligned}$$

Hence, from (17),

$$\begin{aligned} &\varphi_{R_a}(z) = \frac{(z-1)}{\log \nu} \\ &\quad \times \mathbb{E}\left[\sum_{k=-\infty}^{\infty} \Gamma\left(\frac{\log z + 2k\pi i}{\log \nu}\right) e^{-(\log z + 2k\pi i)\log_\nu X}\right], \end{aligned}$$

which shows the relation between R_a and $\log_\nu X$.

It is convenient to treat the term with index $k = 0$ separately,

$$\begin{aligned} \varphi_{R_a}(z) &= \mathbb{E}\left[\frac{z^{-\log_\nu X}(z-1)}{\log \nu} \Gamma\left(\frac{\log z}{\log \nu}\right)\right] \\ &\quad + \frac{(z-1)}{\log \nu} \sum_{k \neq 0} \Gamma\left(\frac{\log z + 2k\pi i}{\log \nu}\right) \\ &\quad \times \mathbb{E}\left[e^{-(\log z + 2k\pi i)\log_\nu X}\right]. \qquad (18) \end{aligned}$$

Since the last series is regular for all z (except for negative real values), the last term becomes arbitrary small for $z \to 1$ and we find

$$\begin{aligned} \varphi_{R_a}(1) &= \mathbb{E}\left[\lim_{z\to 1} \frac{z^{-\log_\nu X}(z-1)}{\log \nu} \Gamma\left(\frac{\log z}{\log \nu}\right)\right] \\ &= \lim_{z\to 1} \frac{(z-1)}{\log z} \Gamma\left(\frac{\log z}{\log \nu} + 1\right) \\ &= \lim_{z\to 1} \frac{(z-1)}{\log z} = 1. \end{aligned}$$

We proceed by investigating the term associated with the first term in (18),

$$\begin{aligned} &\frac{z^{-\log_\nu x}(z-1)}{\log \nu} \Gamma\left(\frac{\log z}{\log \nu}\right) \\ &= \frac{(1 - e^{-\log z})}{\log z} \Gamma\left(\frac{\log z}{\log \nu} + 1\right) e^{(1-\log_\nu x)\log z} \\ &= \frac{(e^{(1-\log_\nu x)\log z} - e^{-\log_\nu x \log z})}{\log z} \Gamma\left(\frac{\log z}{\log \nu} + 1\right). \end{aligned}$$

Expanding in a series yields

$$\begin{aligned} &\frac{z^{-\log_\nu x}(z-1)}{\log \nu} \Gamma\left(\frac{\log z}{\log \nu}\right) \\ &= \sum_{k=0}^{\infty} \frac{(1-\log_\nu x)^{k+1} - (-\log_\nu x)^{k+1}}{(k+1)!} \log^k z \\ &\quad \times \sum_{m=0}^{\infty} \frac{\Gamma^{(m)}(1)}{m! \log^m \nu} \log^m z. \end{aligned}$$

From this we find explicitly

$$\frac{d}{dz}\frac{z^{-\log_\nu x}(z-1)}{\log\nu}\Gamma\left(\frac{\log z}{\log\nu}\right)\Big|_{z=1} = \frac{1}{2} - \frac{\gamma}{\log\nu} - \frac{\log x}{\log\nu},$$

where $\gamma = .5772...$ denotes Euler's constant [1, 6.1.3]. Including the second term

$$\mathbb{E}[R_a] = \frac{1}{2} - \frac{\gamma}{\log\nu} - \mathbb{E}\left[\frac{\log X}{\log\nu}\right] + 2\mathbb{E}\left[\sum_{k=1}^{\infty}\frac{\cos\left[\frac{2k\pi}{\log\nu}\log\left(\frac{\nu}{X}\right) + \arg\Gamma\left(\frac{2k\pi i}{\log\nu}\right)\right]}{\sqrt{\log\nu}\sqrt{2k\sinh\frac{2k\pi^2}{\log\nu}}}\right].$$

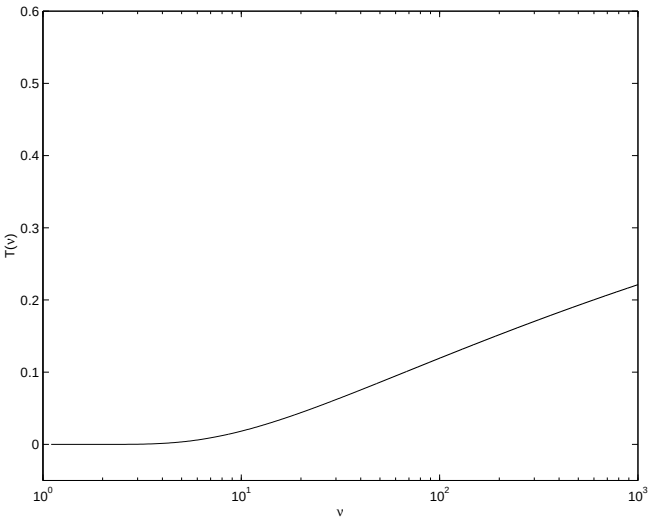

Figure 1: The function $T(\nu)$.

For $x > 0$, the rapidly converging series

$$\frac{2}{\sqrt{\log\nu}}\sum_{k=1}^{\infty}\frac{|\cos\left[\frac{2k\pi}{\log\nu}\log\left(\frac{\nu}{x}\right) + \arg\Gamma\left(\frac{2k\pi i}{\log\nu}\right)\right]|}{\sqrt{2k\sinh\frac{2k\pi^2}{\log\nu}}} \le \frac{2}{\sqrt{\log\nu}}\sum_{k=1}^{\infty}\frac{1}{\sqrt{2k\sinh\frac{2k\pi^2}{\log\nu}}} = T(\nu),$$

increases with ν, but is for $\nu \in (1, 5]$ about 1000 times smaller than the other terms. Hence

$$\frac{1}{2} - \frac{\gamma}{\log\nu} - \mathbb{E}\left[\frac{\log X}{\log\nu}\right] - T(\nu) \le \mathbb{E}[R_a] \le \frac{1}{2} - \frac{\gamma}{\log\nu} - \mathbb{E}\left[\frac{\log X}{\log\nu}\right] + T(\nu).$$

Since $X = \mu(\nu-1)^{-1}\nu^a\mathcal{W}^{(1)}\mathcal{W}^{(2)}|\mathcal{W}^{(1)}\mathcal{W}^{(2)} > 0$, and $\mathcal{W}^{(i)}$ for $i = 1, 2$ are independent, we end up with

$$\mathbb{E}[R_a] \approx \frac{1}{2} - a - \left(\frac{\gamma + \log\mu - \log(\nu-1)}{\log\nu}\right) - \frac{2\mathbb{E}[\log\mathcal{W}|\mathcal{W} > 0]}{\log\nu},$$

where the error in this approximation is bounded by the function $T(\nu)$ displayed in Figure 1.

Differentiating twice with respect to z, and omitting the tedious calculation which runs parallel to the case above, we obtain

$$\mathrm{Var}(R_a) \approx \frac{1}{12} + \frac{\pi^2}{6\log^2\nu} + 2\mathrm{Var}[\log_\nu\mathcal{W}|\mathcal{W} > 0].$$

APPENDIX B: On $\mathbb{E}[\log W|W > 0]$.

Consider a supercritical Galton-Watson process with $\mathcal{Z}_0 = 1$,

$$\mathcal{Z}_{n+1} = \xi_1 + \xi_2 + \ldots + \xi_{\mathcal{Z}_n}, \qquad n = 0, 1, \ldots,$$

where $\xi_1, \xi_2, \ldots$ is an i.i.d. sequence with mean $1 < \mathbb{E}[\xi] = \nu < \infty$ and generating function

$$g(s) = \sum_{k=0}^{\infty} s^k\mathbb{P}[\xi = k].$$

The martingale sequence is given by $\mathcal{W}_n = \mathcal{Z}_n/\nu^n$, $n = 1, 2, \ldots$ and $\mathcal{W}_n \to \mathcal{W}$, almost surely. The expectation $\mathbb{E}[\mathcal{W}] = 1$ if and only if [8],

$$\sum_{k=1}^{\infty}(k\log k)\mathbb{P}[\xi = k] < \infty.$$

Concerning the existence of $\mathbb{E}[\log\mathcal{W}|\mathcal{W} > 0]$, we write

$$\mathbb{E}[I_{(0,\infty)}(\mathcal{W})\log\mathcal{W}] = \mathbb{E}\left[I_{(0,\infty)}(\mathcal{W})\int_1^{\mathcal{W}}\frac{dy}{y}\right] = \mathbb{E}\left[-\int_0^1 I_{(0,y)}(\mathcal{W})\frac{dy}{y}\right] + \mathbb{E}\left[\int_1^{\infty} I_{(y,\infty)}(\mathcal{W})\frac{dy}{y}\right] = -\int_0^1\mathbb{P}[0 < \mathcal{W} < y]\frac{dy}{y} + \int_1^{\infty}\mathbb{P}[\mathcal{W} > y]\frac{dy}{y},$$

by Fubini's theorem [9]. Both integrals on the right-hand side are finite. The second because $\mathbb{E}[\mathcal{W}] = 1$, the first one because [4, Theorem 1] implies that $\mathbb{P}(0 < \mathcal{W} < y) \le Cy^\alpha$, $0 < y < 1$, for some positive α and some positive C. Similarly we have

$$\mathbb{E}[I_{(0,\infty)}(\mathcal{W}_n)\log\mathcal{W}_n] = -\int_0^1\mathbb{P}[0 < \mathcal{W}_n < y]\frac{dy}{y} + \int_1^{\infty}\mathbb{P}[\mathcal{W}_n > y]\frac{dy}{y}.$$

From estimates of the difference

$$\mathbb{P}[\mathcal{W}_n > y] - \mathbb{P}[\mathcal{W} > y],\ y > 0,$$

([2], and the references therein) and the form of the density of $\mathcal{W}|\mathcal{W} > 0$ [4, Theorem 1], one can conclude that $\mathbb{E}[\log \mathcal{W}_n|\mathcal{W}_n > 0]$ converges to $\mathbb{E}[\log \mathcal{W}|\mathcal{W} > 0]$.

Appendix C: Accuracy test.

We tested the speed of convergence of $\mathbb{E}[\log \mathcal{W}_n|\mathcal{W}_n > 0]$ to $\mathbb{E}[\log \mathcal{W}|\mathcal{W} > 0]$, using the FFT, for geometrically distributed offspring:

$$\mathbb{P}[\xi = k] = p(1-p)^k, \quad k \geq 0, \quad \nu = \mathbb{E}[\xi] = \frac{1-p}{p}.$$

In this case $\mathcal{W}|\mathcal{W} > 0$ has an exponential density on $(0, \infty)$ and

$$\mathbb{E}[\log \mathcal{W}|\mathcal{W} > 0] = -\gamma - \log(1 - \nu^{-1}),$$

where $\gamma = 0.5772\ldots$ is Euler's constant. Moreover, the extinction probability $\pi = \nu^{-1}$. We checked our numerical calculations for the case $\nu = 1.25$. As remarked in the introduction the maximum number of generations that can accurately be computed is limited since the support $\mathcal{Z}_n$ grows exponentially with n. The results with the `FFT` remain reliable only if the theoretical value of $\nu^n = \mathbb{E}[\mathcal{Z}_n]$ does not deviate too much from its numerical value using the `FFT`. As a rule of thumb we use that ν^n (second column of Table 7) does not differ more than 0.1 from its numerical value (third column of Table 7)

Table 7: Results for the geometric distribution with $\nu = 1.25$

n	ν^n	ν^n(num.)	$\mathbb{E}[\log \mathcal{W}|\mathcal{W} > 0]$
12	14.5519	14.5519	1.0087
13	18.1899	18.1899	1.0147
14	22.7374	22.7374	1.0192
15	28.4217	28.4217	1.0226
16	35.5271	35.5271	1.0251
17	44.4089	44.4089	1.0270
18	55.5112	55.5109	1.0285
19	69.3839	69.3836	1.0295
20	86.7362	86.6768	1.0301
21	108.4202	108.0140	1.0296
22	135.5253	133.6421	1.0256
23	169.4066	162.9979	1.0135
24	211.7582	194.4921	0.9869
25	264.6978	226.1251	0.9399
26	330.8722	255.9882	0.8690

An accurate estimate for $\mathbb{E}[\log \mathcal{W}|\mathcal{W} > 0]$ is the value 1.03 given in generation $n = 20$. We see that starting with generation $n = 21$ the numerical values of $\mathbb{E}[\log \mathcal{W}|\mathcal{W} > 0]$ are decreasing, whereas up to and including generation $n = 20$, the values are increasing. The decreasing values are most likely caused by the deviation between the numerically computed distribution and its theoretical counterpart (see the differences in the second and third column).

If we use extrapolation on the values $\mathbb{E}[\log \mathcal{W}_n|\mathcal{W}_n > 0]$ for the generations $n = 12, 13, \ldots, 20$, we obtain $\mathbb{E}[\log \mathcal{W}|\mathcal{W} > 0] = 1.032$. The theoretical value is $\mathbb{E}[\log \mathcal{W}|\mathcal{W} > 0] = -\gamma + \log(5) = 1.0322\ldots$.