OPEN ACCESS
IOP Publishing

Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **55** (2022) 265005 (23pp)

https://doi.org/10.1088/1751-8121/ac6f9a

The mean and variance of the distribution of shortest path lengths of random regular graphs

Ido Tishby[1,*], **Ofer Biham**[1,*], **Reimer Kühn**[2,*] and **Eytan Katzav**[1,*]

[1] Racah Institute of Physics, The Hebrew University, Jerusalem 9190401, Israel
[2] Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom

E-mail: ido.tishby@mail.huji.ac.il, biham@phys.huji.ac.il, reimer.kuehn@kcl.ac.uk and eytan.katzav@mail.huji.ac.il

Received 14 February 2022, revised 20 April 2022
Accepted for publication 13 May 2022
Published 7 June 2022

Abstract

The distribution of shortest path lengths (DSPL) of random networks provides useful information on their large scale structure. In the special case of random regular graphs (RRGs), which consist of N nodes of degree $c \geqslant 3$, the DSPL, denoted by $P(L = \ell)$, follows a discrete Gompertz distribution. Using the discrete Laplace transform we derive a closed-form (CF) expression for the moment generating function of the DSPL of RRGs. From the moment generating function we obtain CF expressions for the mean and variance of the DSPL. More specifically, we find that the mean distance between pairs of distinct nodes is given by $\langle L\rangle = \frac{\ln N}{\ln(c-1)} + \frac{1}{2} - \frac{\ln c - \ln(c-2) + \gamma}{\ln(c-1)} + \mathcal{O}\left(\frac{\ln N}{N}\right)$, where γ is the Euler–Mascheroni constant. While the leading term is known, this result includes a novel correction term, which yields very good agreement with the results obtained from direct numerical evaluation of $\langle L\rangle$ via the tail-sum formula and with the results obtained from computer simulations. However, it does not account for an oscillatory behavior of $\langle L\rangle$ as a function of c or N. These oscillations are negligible in sparse networks but detectable in dense networks. We also derive an expression for the variance $\mathrm{Var}(L)$ of the DSPL, which captures the overall dependence of the variance on c but does not account for the oscillations. The oscillations are due to the discrete nature of the shell structure around a random node. They reflect the profile of the filling of new shells

*Authors to whom any correspondence should be addressed.

as N is increased. The results for the mean and variance are compared to the corresponding results obtained in other types of random networks. The relation between the mean distance and the diameter is discussed.

Keywords: random network, random regular graph, distribution of shortest path lengths, moments, mean, variance

(Some figures may appear in colour only in the online journal)

1. Introduction

Random networks (or graphs) consist of a set of N nodes that are connected by edges in a way that is determined by some random process. They provide a useful conceptual framework for the study of a large variety of systems and processes in science, technology and society [1–5]. The local structure of a random network can be characterized by the degree distribution $P(k)$ and its moments $\langle K^n \rangle$, $n \geqslant 1$. In particular, the mean degree $\langle K \rangle$ provides the expected number of neighbors of a random node, while the variance $\mathrm{Var}(K) = \langle K^2 \rangle - \langle K \rangle^2$ accounts for the width of the degree distribution.

The large scale structure of a random network is captured by the distribution of shortest path lengths (DSPL), or the distance distribution, between pairs of distinct nodes. Properties of the DSPL, which is denoted by $P(L = \ell)$, have been studied in random networks with different degree distributions [6–26]. It was shown that in random networks whose degree distribution has a finite variance, the mean distance between pairs of distinct nodes scales like $\langle L \rangle \sim \ln N$ [6, 7, 12]. This implies that random networks are small-world networks [27–31]. Moreover, it was shown that scale-free networks, which exhibit a power-law degree distribution of the form $P(k) \sim k^{-\gamma}$, may be ultrasmall depending on the value of the exponent γ. In particular, for $2 < \gamma < 3$, where the variance of $P(k)$ diverges in the infinite system limit, the mean distance scales like $\langle L \rangle \sim \ln \ln N$ [7, 9, 32]. The variance $\mathrm{Var}(L)$ of the DSPL was also studied. It was shown that the DSPL of random networks is typically a narrow distribution, whose width does not grow as the network size is increased [7].

In the study of the DSPL it is convenient to use the tail distribution $P(L > \ell)$, which is the probability that the distance between a random pair of distinct nodes i and j is larger than ℓ. In networks that consist of more than one connected component, the distance between nodes that reside on different network components is $\ell = \infty$. In this case the DSPL is restricted to pairs of nodes that reside on the same connected component [15, 16, 23, 24]. In the analysis below we focus on networks that consist of a single connected component. In this case, the mean distance can be obtained from the tail-sum formula [33]

$$\langle L \rangle = \sum_{\ell=0}^{\infty} P(L > \ell). \tag{1}$$

In configuration model networks the degree of each node is drawn independently from a given degree distribution $P(k)$ and the connections are random and uncorrelated [6, 34, 35]. The configuration model generates maximum entropy ensembles in which the degree distribution $P(k)$ is fixed [6, 31, 34, 35]. Therefore, the configuration model provides a general and highly powerful platform for the analysis of statistical properties of networks. In configuration model networks, the DSPL is completely determined by the degree distribution $P(k)$ and the network size N. In particular, in the large network limit the mean distance can be approximated by [6, 29–31]

$$\langle L \rangle \simeq \frac{\ln N}{\ln \mu}, \tag{2}$$

where

$$\mu = \frac{\langle K^2 \rangle - \langle K \rangle}{\langle K \rangle} \tag{3}$$

is the mean of the excess degree distribution of nodes selected via a random edge. The excess degree distribution, given by [36]

$$P_{\text{excess}}(k) = \frac{(k+1)}{\langle K \rangle} P(k+1), \tag{4}$$

is obtained by selecting random edges and choosing randomly one of the two end-nodes of the selected edge. The excess degree of such end-node is obtained by extracting the edge that led to that node, reducing its degree by 1.

The random regular graph (RRG) is a special case of a configuration model network, in which the degree distribution is a degenerate distribution of the form $P(k) = \delta_{k,c}$, namely all the nodes are of the same degree c, where $c \geqslant 3$. In the special case of an RRG, the second moment of the degree distribution satisfies $\langle K^2 \rangle = c^2$ and the excess degree is $\mu = c - 1$. As a result, equation (2) is reduced to

$$\langle L \rangle \simeq \frac{\ln N}{\ln(c-1)}. \tag{5}$$

Recently, the DSPL of the Erdös–Rényi (ER) network [37–39] and other configuration model networks was calculated using an approach called the random path approach (RPA), which is based on recursion equations [15, 16, 20]. In general, the recursion equations are iterated step by step and the resulting distribution is evaluated numerically. The DSPL obtained from the recursion equations was found to be in very good agreement with the results obtained from computer simulations for configuration model networks with a broad range of degree distributions [16]. In the special case of RRGs of degree $c \geqslant 3$, the recursion equations yield a closed-form (CF) expression for the DSPL, whose tail distribution takes the form [16]

$$P(L > \ell) = \exp\left\{ -\frac{c}{(c-2)N} \left[(c-1)^{\ell} - 1 \right] \right\}. \tag{6}$$

The distribution shown in equation (6) is a discrete form of the Gompertz distribution [40, 41]. Equation (6) coincides with an earlier calculation of the DSPL of RRGs [8]. The existence of a CF expression for $P(L > \ell)$ opens the way to the derivation of compact formulae for the mean distance $\langle L \rangle$ and the variance $\mathrm{Var}(L)$ of the DSPL of RRGs. The mean distance $\langle L \rangle$ provides the typical length scale in the network, as well as its scaling with respect to the network size N and the degree c. The variance $\mathrm{Var}(L)$ provides the width of the peak in the probability mass function $P(L = \ell)$.

In this paper we derive a CF expression for the moment generating function of the DSPL of RRGs. The moment generating function is expressed in terms of a discrete Laplace transform. The discrete sum is then evaluated using the Euler–Maclaurin formula that applies for sufficiently large networks. Using the moment generating function we obtain CF expressions for the mean distance $\langle L \rangle$ and for the second moment $\langle L^2 \rangle$ of the DSPL. Using these results we also obtain, for the first time, a closed form expression for the variance $\mathrm{Var}(L)$. In previous studies [6, 8, 29, 42] it was found that in the large N limit the mean distance $\langle L\rangle$ can be approximated by equation (5). On top of this result, we obtain a novel correction term, which is of order 1, namely independent of N. Since the mean distance $\langle L\rangle$ is logarithmic in N, such correction of order 1 is often not negligible. Taking this correction into account yields a very good agreement with the results obtained from computer simulations. However, it does not account for an apparent oscillatory behavior of $\langle L\rangle$ as a function of c and as a function of N. These oscillations are negligible in the sparse-network limit but detectable in the dense-network limit. The CF expression obtained for $\mathrm{Var}(L)$ captures the overall dependence of the variance on the degree c. However, it does not account for the oscillatory behavior of $\mathrm{Var}(L)$ as a function of c, which becomes significant in the dense-network limit. The oscillations of $\langle L\rangle$ and $\mathrm{Var}(L)$ as a function of c and N are analyzed and discussed. In particular, it is shown that the oscillations become regular when plotted as a function of $\ln N/\ln(c-1)$, when c is kept fixed and N is varied. The results for the mean and variance of the DSPL are compared to the corresponding results obtained in other types of random networks. The relation to the diameter is also discussed.

The paper is organized as follows. In section 2 we describe the RRG. In section 3 we review the calculation of the DSPLs. In section 4 we derive a CF expression for the moment generating function of the DSPL. In section 5 we calculate the mean distance $\langle L\rangle$. In section 6 we calculate the variance of the DSPL. The results are discussed in section 7 and summarized in section 8.

2. The random regular graph

The RRG is a special case of a configuration model network, in which the degree distribution is a degenerate distribution of the form $P(k) = \delta_{k,c}$, namely all the nodes are of the same degree c. Here we focus on the case of $3 \leqslant c \leqslant N-1$, in which for a sufficiently large value of N the RRG consists of a single connected component [34, 35, 43]. RRGs of any finite size exhibit a local tree-like structure, while at larger scales there is a broad spectrum of cycle lengths. In that sense RRGs differ from Cayley trees, which maintain their tree structure by reducing the most peripheral nodes to leaf nodes of degree 1. They also differ from Bethe lattices which exhibit a tree structure of an infinite size.

The neighborhood of a given node i can be described by a shell structure, in which the first shell consists of the c neighbors of i and the second shell consists of the neighbors of the nodes of the first shell, which are at distance $\ell = 2$ from i. In general, the ℓth shell around i consists of all the nodes that are at a distance ℓ from i [13, 14]. In the infinite network limit, the ℓth shell around each node consists of $n(\ell) = c(c-1)^{\ell-1}$ nodes.

A convenient way to construct an RRG of size N and degree c (Nc must be an even number) is to prepare the N nodes such that each node is connected to c half edges or stubs [3, 44]. At each step of the construction, one connects a pair of random stubs that belong to two different nodes i and j that are not already connected, forming an edge between them. This procedure is repeated until all the stubs are exhausted. The process may get stuck before completion in case that all the remaining stubs belong to the same node or to pairs of nodes that are already connected. In such case one needs to perform some random reconnections in order to complete the construction.

3. The distribution of shortest path lengths

Consider an RRG consisting of N nodes of degree c. The distance ℓ_{ij} between a pair of nodes i and j is given by the length of the shortest path between i and j. The tail DSPL between pairs of random nodes is denoted by $P(L > \ell)$. In computer simulations this distribution is obtained by generating a large number of network instances from an ensemble of RRGs of a given size N and degree c. In each network instance one repeatedly selects pairs of random nodes, finds the shortest paths between them and measures their lengths. These results yield the distribution $P(L > \ell)$.

Another useful way to sample nodes is via a random edge. In this case one selects a random edge e and picks randomly one of the end-nodes $\tilde{i}$ of the selected edge. The edge e is then deleted, giving rise to a reduced network that includes all the nodes and edges, apart from e, also called the cavity graph [45]. The tail DSPL between pairs of distinct nodes consisting of a node $\tilde{i}$, selected via a random edge e, and a random node j, on the reduced network from which e is removed, is denoted by $\widetilde{P}(L > \ell)$.

Apart from the tail distributions $P(L > \ell)$ and $\widetilde{P}(L > \ell)$, it is also useful to consider various conditional distributions. Consider a pair of random nodes i and j, such that the distance between them is known to be larger than $\ell - 1$. The conditional probability $P(L > \ell | L > \ell - 1)$ provides the probability that the length of the shortest path between i and j is larger than ℓ, given that it is larger than $\ell - 1$. Similarly, for a pair of nodes consisting of a node $\tilde{i}$, which is selected via a random edge e, and a random node j, the conditional probability $\widetilde{P}(L > \ell | L > \ell - 1)$ is the probability that the distance between $\tilde{i}$ and j is larger than ℓ, given that it is larger than $\ell - 1$, in the reduced network from which e is removed.

Below we evaluate the conditional probability $P(L > \ell | L > \ell - 1)$, namely the probability that the shortest path length between a pair of random nodes i and j is larger than ℓ, given that it is larger than $\ell - 1$, using the RPA [15, 16, 20]. Since each node in the network (e.g. node i) has c neighbors, the boundary condition at $\ell = 1$ is given by

$$P(L > 1 | L > 0) = 1 - \frac{c}{N-1}. \tag{7}$$

For $\ell \geqslant 2$ the conditional probability can be expressed in the form

$$P(L > \ell | L > \ell - 1) = \widetilde{P}(L > \ell - 1 | L > \ell - 2)^c, \tag{8}$$

where $\widetilde{P}(L > \ell - 1 | L > \ell - 2)$ is the conditional DSPL between a node $\tilde{i}'$ selected via a random edge e and a random node j in the reduced network in which the edge e is removed. The conditional distribution $\widetilde{P}(L > \ell - 1 | L > \ell - 2)$ can be further expressed in the form

$$\widetilde{P}(L > \ell - 1 | L > \ell - 2) = \widetilde{P}(L > \ell - 2 | L > \ell - 3)^{c-1}, \tag{9}$$

where the power of $c - 1$ reflects the fact that in the reduced network the node $\tilde{i}'$ is of degree $c - 1$. In general, the conditional distribution $\widetilde{P}(L > \ell - m | L > \ell - m - 1)$ is given by

$$\widetilde{P}(L > \ell - m | L > \ell - m - 1) = \widetilde{P}(L > \ell - m - 1 | L > \ell - m - 2)^{c-1}, \tag{10}$$

where $m = 1, 2, \ldots, \ell - 2$. This cascade of recursion equations eventually leads to $\widetilde{P}(L > 1 | L > 0)$, which can be evaluated directly and is given by

$$\widetilde{P}(L > 1|L > 0) = 1 - \frac{c-1}{N-1}. \tag{11}$$

For sufficiently large N and for $c/N \ll 1$, equation (11) can be replaced by

$$\widetilde{P}(L > 1|L > 0) = \exp\left(-\frac{c-1}{N-1}\right). \tag{12}$$

Inserting $\widetilde{P}(L > \ell - 1|L > \ell - 2)$ from equation (9) into (8), we obtain

$$P(L > \ell|L > \ell - 1) = \widetilde{P}(L > \ell - 2|L > \ell - 3)^{c(c-1)}. \tag{13}$$

By repeatedly inserting $\widetilde{P}(L > \ell - m|L > \ell - m - 1)$ from equation (10) into the right-hand side of equation (13), we obtain

$$P(L > \ell|L > \ell - 1) = \widetilde{P}(L > \ell - m - 1|L > \ell - m - 2)^{c(c-1)^m}. \tag{14}$$

Finally, for $m = \ell - 2$ we obtain

$$P(L > \ell|L > \ell - 1) = \exp\left[-\frac{c(c-1)^{\ell-1}}{N-1}\right]. \tag{15}$$

Similarly, the conditional distribution $\widetilde{P}(L > \ell|L > \ell - 1)$ is given by

$$\widetilde{P}(L > \ell|L > \ell - 1) = \exp\left[-\frac{(c-1)^{\ell}}{N-1}\right]. \tag{16}$$

The tail distribution $P(L > \ell)$ is obtained from

$$P(L > \ell) = \prod_{\ell'=1}^{\ell} P(L > \ell'|L > \ell' - 1). \tag{17}$$

Inserting $P(L > \ell'|L > \ell' - 1)$ from equation (15) into (17) and replacing $N - 1$ by N, we obtain

$$P(L > \ell) = \exp\left[-\frac{c}{N}\sum_{\ell'=1}^{\ell}(c-1)^{\ell'-1}\right]. \tag{18}$$

Carrying out the summation, we obtain

$$P(L > \ell) = \begin{cases} \exp\left[-\eta\left(\mathrm{e}^{b\ell} - 1\right)\right] & \ell \geqslant 0 \\ 1 & \ell < 0, \end{cases} \tag{19}$$

where

$$\eta = \frac{c}{(c-2)N}, \tag{20}$$

and

$$b = \ln(c-1). \tag{21}$$

The tail distribution $P(L > \ell)$, given by equation (19), is a discrete version of the Gompertz distribution [40, 41]. It is in agreement with the DSPL of RRGs obtained in reference [8].

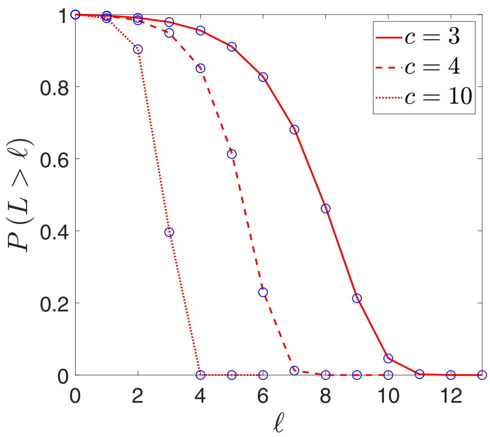

Figure 1. Analytical results for the tail distribution of shortest path lengths $P(L > \ell)$ for ensembles of RRGs of size $N = 1000$ and degrees $c = 3$ (solid line), $c = 4$ (dashed line) and $c = 10$ (dotted line), obtained from equations (19)–(21). The analytical results are in very good agreement with the results obtained from computer simulations (circles). As c is increased, the sigmoid function shifts to the left, which implies that distances in the network become shorter. It also becomes steeper, which implies that the DSPL becomes narrower.

In figure 1 we present analytical results for the tail distribution $P(L > \ell)$ of shortest path lengths for ensembles of RRGs of size $N = 1000$ and degrees $c = 3$ (solid line), $c = 4$ (dashed line) and $c = 10$ (dotted line). The analytical results, obtained from equations (19)–(21), are in very good agreement with the results obtained from computer simulations (circles). The tail distribution, which follows a discrete Gompertz distribution, exhibits a monotonically decreasing sigmoid-like shape, or a smoothed Heaviside step function. As c is increased, the sigmoid function shifts to the left, which implies that distances in the network become shorter. The sigmoid function also becomes steeper, which implies that the DSPL becomes narrower. This means that for sufficiently large values of c the majority of pairs of nodes are at equal distance from each other.

The probability mass function of the DSPL can be recovered from the tail distribution by

$$P(L = \ell) = P(L > \ell - 1) - P(L > \ell). \tag{22}$$

As mentioned in the introduction, it reflects the shell structure around a random node i, where the ℓth shell consists of the nodes that reside at a distance ℓ from i. The expected number of nodes in the ℓth shell is given by

$$n(\ell) = (N - 1)P(L = \ell). \tag{23}$$

The Gompertz distribution frequently appears in the analysis of life spans and survival. It describes situations in which the mortality rate grows exponentially with time. The shell structure around a random node can be interpreted in terms of survival probabilities. To explain this property we express equation (17) in the form

$$P(L > \ell) = P(L > \ell | L > \ell - 1)P(L > \ell - 1). \tag{24}$$

This implies that the probability of a random node to remain outside the first ℓ shells is given by the probability to be outside the first $\ell - 1$ shells times a survival probability, which is given by equation (15). The complementary probability $P(L = \ell|L > \ell - 1) = 1 - P(L > \ell|L > \ell - 1)$ is the probability that such random node ends up in the ℓth shell. In the context of survival analysis this probability corresponds to the hazard function or the mortality rate.

4. The moment generating function of the DSPL

In this section we apply a systematic method for the calculation of moments of the DSPL, which is based on the discrete Laplace transform of the tail distribution $P(L > \ell)$. The discrete Laplace transform of some function $f(\ell)$ is given by

$$\mathcal{L}\{f\}(s) = \sum_{\ell=0}^{\infty} \mathrm{e}^{-s\ell} f(\ell), \tag{25}$$

which is related to the one-sided Z-transform and to the starred transform [46]. Inserting $f(\ell) = P(L > \ell)$ we obtain

$$\mathcal{L}\{P(L > \ell)\}(s) = \sum_{\ell=0}^{N-2} \mathrm{e}^{-s\ell} P(L > \ell), \tag{26}$$

where the summation is truncated above $N - 2$. This reflects the fact that in a network that consists of N nodes, $P(L > \ell) = 0$ for $\ell \geqslant N - 1$. In spite of this fact, note that equation (19) yields non-zero results for the probabilities $P(L > \ell)$ even for $\ell \geqslant N - 1$. However, these probabilities are vanishingly small. Therefore, the upper limit of the sum in equation (26) can be changed from $N - 2$ to ∞ without affecting the results. We also note that the tail distribution $P(L > \ell)$ is defined as the distribution of distances between pairs of distinct nodes. Thus, it satisfies $P(L > 0) = 1$. Using these observations, equation (26) can be written in the form

$$\mathcal{L}\{P(L > \ell)\}(s) = 1 + \sum_{\ell=1}^{\infty} \mathrm{e}^{-s\ell} P(L > \ell). \tag{27}$$

The sum on the right-hand side of equation (27) can be approximated by the corresponding integral. However, a more accurate result can be obtained using the Euler–Maclaurin formula, in which the difference between the sum and the integral is systematically approximated in terms of derivatives of the integrand, which are evaluated at the end-points of the interval [47, 48]. Applying the Euler–Maclaurin formula (equation (2.10.1) in reference [49]), we obtain

$$\begin{aligned} \mathcal{L}\{P(L > \ell)\}(s) = 1 + \int_0^{\infty} \mathrm{e}^{-s\ell} P(L > \ell) d\ell - \frac{1}{2} \\ - \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!} \frac{\mathrm{d}^{2k-1}}{\mathrm{d}\ell^{2k-1}} \left[\mathrm{e}^{-s\ell} P(L > \ell)\right]\Big|_{\ell=0}, \end{aligned} \tag{28}$$

where B_{2k} is the Bernoulli number of order $2k$ [49]. Carrying out the integration, we obtain

$$\int_0^{\infty} \mathrm{e}^{-s\ell} P(L > \ell) \mathrm{d}\ell = \frac{\exp\left[\left(\frac{c}{c-2}\right)\frac{1}{N}\right]}{\ln(c-1)} \mathrm{E}_{1+\frac{s}{\ln(c-1)}} \left[\left(\frac{c}{c-2}\right)\frac{1}{N}\right], \tag{29}$$

where $\mathrm{E}_n(x)$ is the generalized exponential integral (equation (8.19.3) in reference [49]). In order to evaluate the sum in equation (28), we expand the summand up to first order in $1/N$.

Inserting $P(L>\ell)$ from equation (19), we obtain

$$\mathrm{e}^{-s\ell}P(L>\ell)=\mathrm{e}^{-s\ell}-\left(\frac{c}{c-2}\right)\frac{1}{N}\left[\mathrm{e}^{-(s-b)\ell}-\mathrm{e}^{-s\ell}\right]+\mathcal{O}\left(\frac{1}{N^2}\right), \tag{30}$$

where $b=\ln(c-1)$. Expanding equation (30) in powers of ℓ, we obtain

$$\begin{aligned}\mathrm{e}^{-s\ell}P(L>\ell)=1&+\left[1+\left(\frac{c}{c-2}\right)\frac{1}{N}\right]\sum_{r=1}^{\infty}\frac{(-s)^r}{r!}\ell^r\\&-\left(\frac{c}{c-2}\right)\frac{1}{N}\sum_{r=1}^{\infty}\frac{(b-s)^r}{r!}\ell^r+\mathcal{O}\left(\frac{1}{N^2}\right).\end{aligned} \tag{31}$$

Differentiating equation (31) $(2k-1)$ times with respect to ℓ and evaluating the result at $\ell=0$, we obtain

$$\begin{aligned}\frac{\mathrm{d}^{2k-1}}{\mathrm{d}\ell^{2k-1}}\left[\mathrm{e}^{-s\ell}P(L>\ell)\right]\bigg|_{\ell=0}&=-\left[1+\left(\frac{c}{c-2}\right)\frac{1}{N}\right]s^{2k-1}\\&\quad+\left(\frac{c}{c-2}\right)\frac{1}{N}(s-b)^{2k-1}+\mathcal{O}\left(\frac{1}{N^2}\right).\end{aligned} \tag{32}$$

Inserting the right-hand side of equation (32) into (28) and using the identity (equation (4.19.6) in reference [49])

$$\sum_{k=1}^{\infty}\frac{B_{2k}x^{2k-1}}{(2k)!}=\frac{1}{2}\coth\left(\frac{x}{2}\right)-\frac{1}{x}, \tag{33}$$

we obtain

$$\begin{aligned}\mathcal{L}\{P(L>\ell)\}(s)=&\frac{\exp\left[\left(\frac{c}{c-2}\right)\frac{1}{N}\right]}{\ln(c-1)}\mathrm{E}_{1+\frac{s}{\ln(c-1)}}\left[\left(\frac{c}{c-2}\right)\frac{1}{N}\right]+\frac{1}{2}\\&+\left[\frac{1}{2}\coth\left(\frac{s}{2}\right)-\frac{1}{s}\right]\left[1+\left(\frac{c}{c-2}\right)\frac{1}{N}\right]\\&-\left(\frac{c}{c-2}\right)\frac{1}{N}\left[\frac{1}{2}\coth\left(\frac{s-b}{2}\right)-\frac{1}{s-b}\right]\\&+\mathcal{O}\left(\frac{1}{N^2}\right).\end{aligned} \tag{34}$$

In the large network limit equation (34) can be reduced to

$$\begin{aligned}\mathcal{L}\{P(L>\ell)\}(s)=&\frac{1}{\ln(c-1)}\mathrm{E}_{1+\frac{s}{\ln(c-1)}}\left[\left(\frac{c}{c-2}\right)\frac{1}{N}\right]\\&+\frac{1}{2}+\left[\frac{1}{2}\coth\left(\frac{s}{2}\right)-\frac{1}{s}\right]+\mathcal{O}\left(\frac{1}{N}\right).\end{aligned} \tag{35}$$

The moment generating function of the DSPL is denoted by

$$M(s)=\sum_{\ell=0}^{\infty}\mathrm{e}^{s\ell}P(L=\ell). \tag{36}$$

Inserting $P(L=\ell)$ from equation (22) into (36), we can rewrite $M(s)$ in terms of the tail distribution in the form

$$M(s) = 1 + (\mathrm{e}^{s} - 1)\sum_{\ell=0}^{\infty} \mathrm{e}^{s\ell} P(L > \ell). \tag{37}$$

Using equation (26) we express the moment generating function in terms of the Laplace transform, namely

$$M(s) = 1 + (\mathrm{e}^{s} - 1)\mathcal{L}\{P(L > \ell)\}(-s). \tag{38}$$

In the sections below we use the moment generating function to obtain closed form expressions for the mean and variance of the DSPL.

5. The mean distance

The mean distance $\langle L\rangle$ between pairs of distinct nodes in an RRG can be calculated using the tail-sum formula [33]

$$\langle L\rangle = \sum_{\ell=0}^{N-2} P(L > \ell), \tag{39}$$

where $P(L > \ell)$ is given by equation (19). Since equation (19) is highly accurate for $c \ll N$, equation (39) is expected to yield accurate results for the mean distance except for the limit of very dense networks. However, this expression provides little insight on the behavior of $\langle L\rangle$ and its dependence on c and N. Since $P(L > \ell)$, given by equation (19), is vanishingly small for $\ell \geqslant N - 1$, equation (39) can be replaced by

$$\langle L\rangle = \sum_{\ell=0}^{\infty} P(L > \ell). \tag{40}$$

Below we use the moment generating function $M(s)$ to obtain a CF expression for $\langle L\rangle$, which is valid for sufficiently large networks. It is given by

$$\langle L\rangle = \left.\frac{\mathrm{d}M(s)}{\mathrm{d}s}\right|_{s=0}. \tag{41}$$

Inserting $M(s)$ from equations (34) and (38) into equation (41) and carrying out the differentiation, we obtain

$$\begin{aligned}\langle L\rangle = {} & \frac{\exp\left[\left(\frac{c}{c-2}\right)\frac{1}{N}\right]}{\ln(c-1)} \mathrm{E}_1\left[\left(\frac{c}{c-2}\right)\frac{1}{N}\right] + \frac{1}{2} \\ & + \frac{1}{2}\left(\frac{c}{c-2}\right)\frac{1}{N}\left\{\coth\left[\frac{\ln(c-1)}{2}\right] - \frac{2}{\ln(c-1)}\right\} + \mathcal{O}\left(\frac{1}{N^2}\right),\end{aligned} \tag{42}$$

where $\mathrm{E}_1(x)$ is the exponential integral, also denoted as $\mathrm{Ei}(x)$ [49]. Expanding the right-hand side of equation (42) in powers of $1/N$, we obtain

$$\langle L\rangle = \frac{\ln N}{\ln(c-1)} + \frac{1}{2} - \frac{\ln\left(\frac{c}{c-2}\right)+\gamma}{\ln(c-1)} + \frac{1}{\ln(c-1)}\left(\frac{c}{c-2}\right)\frac{\ln N}{N} + \frac{c}{c-2}\left\{\frac{\ln\left(\frac{c-2}{c}\right)-\gamma}{\ln(c-1)} + \frac{1}{2}\coth\left[\frac{\ln(c-1)}{2}\right]\right\}\frac{1}{N} + \mathcal{O}\left(\frac{\ln N}{N^2}\right), \tag{43}$$

where $\gamma = 0.577\ldots$ is the Euler–Mascheroni constant [49, 50]. For sufficiently large values of N, the corrections of orders $\ln N/N$ and $1/N$ can be neglected, leading to

$$\langle L\rangle = \frac{\ln N}{\ln(c-1)} + \frac{1}{2} - \frac{\ln\left(\frac{c}{c-2}\right)+\gamma}{\ln(c-1)} + \mathcal{O}\left(\frac{\ln N}{N}\right). \tag{44}$$

In figure 2 we present analytical results (solid line) for the mean distance $\langle L\rangle$ between pairs of nodes in an RRG of size $N = 1000$ as a function of the degree c, obtained from the CF expression of equation (44). We also present the results ($\times$ symbols) obtained from a direct numerical evaluation (DE) of the sum in equation (40), where $P(L > \ell)$ is taken from equation (19). Since the discrete Gompertz distribution provides highly accurate results for the DSPL, the results obtained from the DE of equation (40) are expected to be accurate. Indeed, they are found to be in very good agreement with the results obtained from computer simulations (circles). The analytical results obtained from equation (44) are in very good agreement with the results obtained from direct evaluation and computer simulations. The widely known result of $\langle L\rangle = \ln N/\ln(c-1)$, given by equation (5), is also shown ($+$ symbols), indicating that the subleading correction presented in equation (44) is important. However, close inspection of figure 2 reveals that on top of the overall trend, the mean distance $\langle L\rangle$ exhibits an oscillatory behavior as a function of c, which is not captured by equation (44). The amplitude of these oscillations is negligible in the sparse-network limit and becomes detectable for dense networks. Moreover, the period of the oscillations increases as c is increased.

Interestingly, in the dense-network limit the DE of the mean distance $\langle L\rangle$ can be done using an approximate form of the tail-sum formula of equation (40). It is given by

$$\langle L\rangle \simeq (r-1) + P(L > r-1) + P(L > r) + P(L > r+1), \tag{45}$$

where

$$r = \left\lfloor \frac{\ln N}{\ln(c-1)} \right\rfloor, \tag{46}$$

and $\lfloor x \rfloor$ is the integer part of x. The first term on the right-hand side of equation (45) accounts for the sum over the probabilities $P(L > \ell)$ for $\ell = 0, 1, \ldots, r-2$, which in the dense-network limit can be approximated by 1. The next three terms account for the range of distances in which the tail distribution decreases sharply. In the dense-network limit this range is narrow, while the probabilities $P(L > \ell)$ for $\ell \geqslant r+2$ are negligible.

To analyze the oscillations, it is convenient to consider the difference $\langle L\rangle_{\mathrm{DE}} - \langle L\rangle_{\mathrm{CF}}$, where $\langle L\rangle_{\mathrm{DE}}$ is the mean distance obtained from DE of the mean of the discrete Gompertz distribution by equation (40), while $\langle L\rangle_{\mathrm{CF}}$ is the mean distance obtained from the CF expression of equation (44).

In figure 3 we present ($\times$ symbols) the difference $\langle L\rangle_{\mathrm{DE}} - \langle L\rangle_{\mathrm{CF}}$ as a function of $\ln N/\ln(c-1)$ where the network size N is fixed at $N = 10^6$ and the mean degree c is varied. The mean distance $\langle L\rangle_{\mathrm{DE}}$ is obtained from equation (40) and $\langle L\rangle_{\mathrm{CF}}$ is obtained

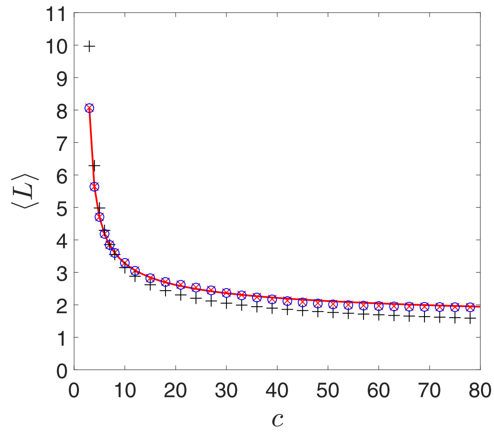

Figure 2. Analytical results (solid line) for the mean distance $\langle L \rangle$ between pairs of random nodes in RRGs of size $N = 1000$ as a function of the degree c, obtained from equation (44). The results obtained from DE of the sum in equation (40) are also shown ($\times$ symbols), where $P(L > \ell)$ is taken from equation (19). The direct evaluation results are found to be in very good agreement with the results obtained from computer simulations (circles). The analytical results are in very good agreement with the results obtained from direct evaluation and computer simulations, except for small oscillatory discrepancies discussed in the text. The widely known result of $\langle L \rangle = \ln N/\ln(c-1)$ is also shown ($+$ symbols), indicating that the subleading term presented in equation (44) is important. The subleading correction is found to be negative for $c \leqslant 7$ and positive for $c > 7$.

from equation (44). It is found that this difference exhibits oscillations as a function of $\ln N/\ln(c-1)$, whose wavelength is equal to 1. The amplitude of the oscillations decreases as $\ln N/\ln(c-1)$ is increased. The difference $\langle L \rangle_{\mathrm{DE}} - \langle L \rangle_{\mathrm{CF}}$ vanishes at integer and half-integer values of $\ln N/\ln(c-1)$. It is found that the maxima of the oscillations take place when the fractional part of $\ln N/\ln(c-1)$ is approximately 1/4, while the minima take place when the fractional part is approximately 3/4. We also present approximated results (solid line) in which $\langle L \rangle_{\mathrm{DE}}$ is evaluated using equation (45). The two curves are found to be in very good agreement except for the limit of sparse networks in which equation (45) is not expected to provide accurate results. It is important to note that the amplitude of the oscillations is very small compared to the mean distance $\langle L \rangle$. Thus, equation (44) provides a very good approximation for $\langle L \rangle$.

In figure 4 we present ($\times$ symbols) the difference $\langle L \rangle_{\mathrm{DE}} - \langle L \rangle_{\mathrm{CF}}$ as a function of $\ln N/\ln(c-1)$, where the degree c is fixed at $c = 30$ and the network size N is varied. The mean distance $\langle L \rangle_{\mathrm{DE}}$ is obtained from equation (40) and $\langle L \rangle_{\mathrm{CF}}$ is obtained from equation (44). It is found that this difference exhibits oscillations as a function of $\ln N/\ln(c-1)$, whose wavelength is equal to 1. For sufficiently large values of N the amplitude of the oscillations is a constant. The difference $\langle L \rangle_{\mathrm{DE}} - \langle L \rangle_{\mathrm{CF}}$ vanishes at integer and half-integer values of $\ln N/\ln(c-1)$. It is found that the maxima of the oscillations take place when the fractional part of $\ln N/\ln(c-1)$ is approximately 1/4, while the minima take place when the fractional part is approximately 3/4. We also present approximated results (solid line) in which $\langle L \rangle_{\mathrm{DE}}$ is evaluated using equation (45). The two curves are found to be in very good agreement.

The oscillatory behavior presented in figure 4 and specifically the fact that the wavelength is 1 and the amplitude is constant, implies that the difference $\langle L \rangle_{\mathrm{DE}} - \langle L \rangle_{\mathrm{CF}}$ depends on N only via the fractional part

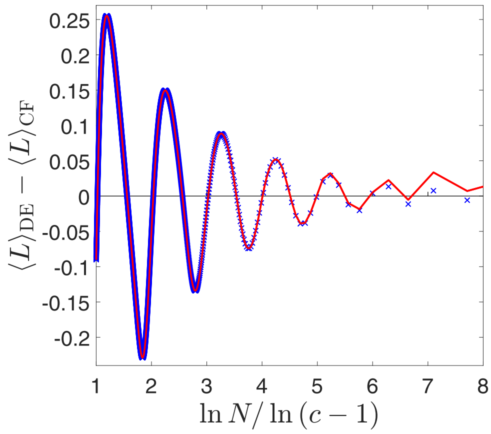

Figure 3. The difference $\langle L\rangle_{\mathrm{DE}} - \langle L\rangle_{\mathrm{CF}}$ ($\times$ symbols) as a function of $\ln N/\ln(c-1)$ where the network size is fixed at $N = 10^6$ and the mean degree c is varied. The mean distance $\langle L\rangle_{\mathrm{DE}}$ is obtained from equation (40) and $\langle L\rangle_{\mathrm{CF}}$ is obtained from equation (44). This difference exhibits oscillations as a function of $\ln N/\ln(c-1)$, whose wavelength is equal to 1. The amplitude of the oscillations decreases as $\ln N/\ln(c-1)$ is increased. This implies that the oscillations are negligible in the sparse-network limit and become detectable as the network becomes more dense. We also present approximated results (solid line) in which $\langle L\rangle_{\mathrm{DE}}$ is evaluated using equation (45). The two curves are found to be in very good agreement except for the limit of sparse networks in which equation (45) is not expected to apply. In general, the amplitude of the oscillations is very small compared to the mean distance $\langle L\rangle$. Thus, equation (44) provides a very good approximation for $\langle L\rangle$.

$$\phi = \frac{\ln N}{\ln(c-1)} - \left\lfloor \frac{\ln N}{\ln(c-1)} \right\rfloor, \tag{47}$$

which can be considered as a phase and takes values in the range $0 \leqslant \phi < 1$. Below we show that for sufficiently large N this can be rigorously justified. In the large N limit, where $\eta \ll 1$, equation (19) can be reduced to

$$P(L > \ell) = \exp\left(-\eta \mathrm{e}^{b\ell}\right), \tag{48}$$

which can also be written in the form

$$P(L > \ell) = \exp\left(-\mathrm{e}^{b\ell + \ln\,\eta}\right). \tag{49}$$

Using this expression, one can show that

$$P(L > \ell + r) = \begin{cases} \exp\left[-\left(\dfrac{c}{c-2}\right)\mathrm{e}^{\ln(c-1)(\ell-\phi)}\right] & \ell \geqslant 0 \\ 1 & \ell < 0. \end{cases} \tag{50}$$

This implies that the dependence of $P(L > \ell + r)$ on the network size N is only via the phase ϕ.

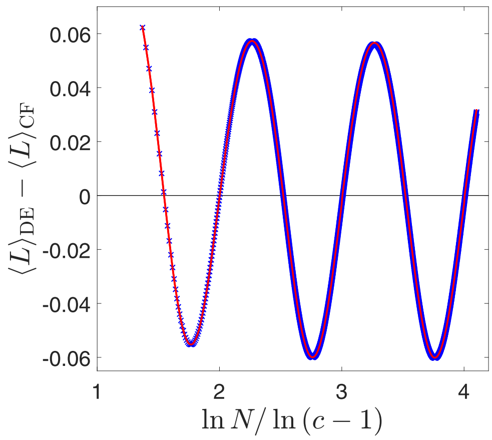

Figure 4. The difference $\langle L\rangle_{\rm DE} - \langle L\rangle_{\rm CF}$ ($\times$ symbols) as a function of $\ln N/\ln(c-1)$ where the mean degree c is fixed at $c = 30$ and the network size N is varied. The mean distance $\langle L\rangle_{\rm DE}$ is obtained from equation (40) and $\langle L\rangle_{\rm CF}$ is obtained from equation (44). This difference exhibits oscillations as a function of $\ln N/\ln(c-1)$, whose wavelength is equal to 1 and the amplitude is a constant. We also present approximated results (solid line) in which $\langle L\rangle_{\rm DE}$ is evaluated using equation (45). The two curves are found to be in very good agreement except for the limit of sparse networks in which equation (45) is not expected to apply. In general, the amplitude of the oscillations is very small compared to the mean distance $\langle L\rangle$. Thus, equation (44) provides a very good approximation for $\langle L\rangle$.

The mean distance can be expressed by

$$\langle L\rangle = \sum_{\ell=-\infty}^{\infty} \ell P(L=\ell), \tag{51}$$

which is justified because $P(L=\ell) = 0$ for $\ell \leqslant 0$. Shifting the summation variable by r to the left, we obtain the equivalent expression

$$\langle L\rangle = \sum_{\ell=-\infty}^{\infty} (\ell+r)P(L=\ell+r). \tag{52}$$

Summing up separately the two terms on the right-hand side, we obtain

$$\langle L\rangle = r + \sum_{\ell=-\infty}^{\infty} \ell P(L=\ell+r). \tag{53}$$

Expressing the probability $P(L=\ell+r)$ in terms of the tail distribution, we obtain

$$\langle L\rangle = r + \sum_{\ell=-\infty}^{\infty} \ell \left[P(L>\ell+r-1) - P(L>\ell+r)\right]. \tag{54}$$

Subtracting $\langle L\rangle_{\rm CF}$, given by equation (44) from (54), we obtain

$$\langle L\rangle_{\mathrm{DE}} - \langle L\rangle_{\mathrm{CF}} = -\phi - \frac{1}{2} + \frac{\ln\left(\frac{c}{c-2}\right) + \gamma}{\ln(c-1)} + \sum_{\ell=-\infty}^{\infty} \ell\left[P(L > \ell + r - 1) - P(L > \ell + r)\right]. \tag{55}$$

Inserting the probabilities $P(L > \ell + r)$ from equation (50) into (55), one finds that $\langle L\rangle_{\mathrm{DE}} - \langle L\rangle_{\mathrm{CF}}$ is only a function of the degree c and the phase ϕ. This implies that its dependence on the network size N is only via the phase ϕ. Note that the argument above holds when the network size N is sufficiently large, such that equation (50) is a good approximation for equation (19).

6. The variance of the DSPL

Using the tail-sum formula, the second moment of the DSPL can be expressed in the form [33]

$$\langle L^2\rangle = \sum_{\ell=0}^{N-2} (2\ell + 1)P(L > \ell). \tag{56}$$

Since the probability $P(L > \ell)$, given by equation (19), is vanishingly small for $\ell \geqslant N - 1$, the upper limit of the summation can be changed from $N - 2$ to ∞, without any noticeable change in the result. Therefore,

$$\langle L^2\rangle = 1 + \sum_{\ell=1}^{\infty} (2\ell + 1)P(L > \ell). \tag{57}$$

The variance of the DSPL is given by

$$\mathrm{Var}(L) = \langle L^2\rangle - \langle L\rangle^2. \tag{58}$$

Inserting $\langle L\rangle$ from equation (40) and inserting $\langle L^2\rangle$ from equation (57) into (58) yields highly accurate results for the variance of the DSPL. However, this expression provides little insight on the behavior of the variance and its dependence on c and N.

Below we use the moment generating function $M(s)$ to obtain a CF expression for $\langle L^2\rangle$, which is valid for large networks. It is given by

$$\langle L^2\rangle = \left.\frac{\mathrm{d}^2 M(s)}{\mathrm{d}s^2}\right|_{s=0}. \tag{59}$$

Inserting $M(s)$ from equations (34) and (38) into equation (59) and carrying out the differentiations, we obtain

$$\begin{aligned}\langle L^2\rangle = \frac{\exp\left[\frac{c}{(c-2)N}\right]}{6[\ln(c-1)]^2}\Bigg\{6\left(\ln\left[\frac{c}{(c-2)N}\right]\right)^2 + 12\gamma\,\ln\left[\frac{c}{(c-2)N}\right] \\ + \pi^2 + 6\gamma^2 - \frac{12c}{(c-2)N}\,{}_3F_3\left[\begin{matrix}1,1,1\\2,2,2\end{matrix}\middle| -\frac{c}{(c-2)N}\right]\Bigg\} \\ + \frac{\exp\left[\frac{c}{(c-2)N}\right]}{\ln(c-1)}\mathrm{E}_1\left[\left(\frac{c}{c-2}\right)\frac{1}{N}\right] + \frac{1}{3} + \mathcal{O}\left(\frac{1}{N^2}\right),\end{aligned} \tag{60}$$

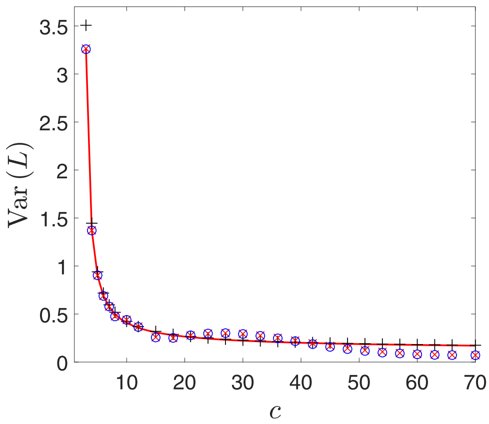

Figure 5. Closed form analytical results (solid line) for the variance Var(L) of the distribution of shortest path lengths for an RRG of size $N = 1000$ as a function of the degree c, obtained from equation (58), where $\langle L \rangle$ is given by equation (42) and $\langle L^2 \rangle$ is given by equation (60). The results obtained from DE of the sums in equations (40) and (57) for $\langle L \rangle$ and $\langle L^2 \rangle$, respectively are also shown ($\times$ symbols). The direct evaluation results are found to be in very good agreement with the results obtained from computer simulations (circles). In the regime of sparse networks the analytical results obtained from the closed form expressions of equations (42) and (60) are in very good agreement with the results obtained from direct evaluation and computer simulations. However, for dense networks the results obtained from direct evaluation and computer simulations exhibit oscillations that are not captured by the CF expression. Instead, the CF expression captures the overall trend of Var(L) vs c as if the oscillations are averaged out. We also present the results obtained from the simpler expression of equation (62) ($+$ symbols). These results are found to be in good agreement with the results obtained from the more complete expressions of equations (42) and (60), except for a small discrepancy for very small values of c.

where $\mathrm{E}_1(x)$ is the exponential integral and ${}_3F_3[\ ]$ is the generalized hypergeometric function [49]. Performing an asymptotic expansion for large N, we obtain

$$\langle L^2 \rangle = \frac{1}{[\ln(c-1)]^2}\left(\ln\left[\frac{(c-2)N}{c}\right]\right)^2 + \frac{\ln(c-1)-2\gamma}{[\ln(c-1)]^2}\ln\left[\frac{(c-2)N}{c}\right] + \frac{\pi^2+6\gamma^2-6\gamma\ln(c-1)}{6[\ln(c-1)]^2} + \frac{1}{3} + \mathcal{O}\left[\left(\frac{\ln N}{N}\right)^2\right]. \tag{61}$$

To obtain the variance of the DSPL we insert $\langle L^2 \rangle$ from equation (60) and $\langle L \rangle$ from equation (42) into (58). While this result is expected to be relatively precise the resulting expression is complicated. A simpler expression for the variance can be obtained by inserting $\langle L^2 \rangle$ from equation (61) and $\langle L \rangle$ from equation (44) into (58). In this case the resulting expression can be simplified to the form

$$\mathrm{Var}(L) = \frac{\pi^2}{6[\ln(c-1)]^2} + \frac{1}{12} + \mathcal{O}\left(\frac{\ln N}{N}\right). \tag{62}$$

This result implies that except for the limit of very small networks, the variance Var(L) does not depend on the network size N but only on the degree c. Moreover Var(L) is a

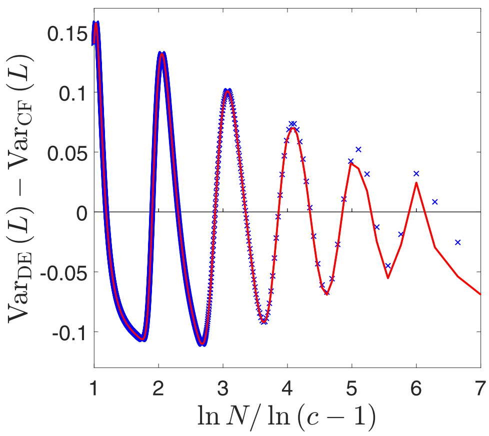

Figure 6. The difference $\mathrm{Var}_{\mathrm{DE}}(L) - \mathrm{Var}_{\mathrm{CF}}(L)$ ($\times$ symbols) between the variance obtained from DE of equations (40) and (57) and from the closed form expressions given by equations (42) and (60), as a function of $\ln N/\ln(c-1)$ for $N = 10^6$. This difference exhibits oscillations as a function of $\ln N/\ln(c-1)$, whose wavelength is equal to 1. The amplitude of the oscillations decreases as $\ln N/\ln(c-1)$ is increased. This implies that the oscillations are negligible in the sparse-network limit and become more pronounced as the network becomes more dense. We also present approximated results (solid line) in which $\mathrm{Var}_{\mathrm{DE}}(L)$ is evaluated using equation (64). The two curves are found to be in very good agreement except for the limit of sparse networks in which equation (64) is not expected to provide accurate results.

monotonically decreasing function of c, whose largest value, obtained at $c = 3$ is $\mathrm{Var}(L) \simeq 3.51$. We thus conclude that the DSPL of RRGs is a narrow distribution, whose width decreases as c is increased.

In figure 5 we present analytical results (solid line) for the variance $\mathrm{Var}(L)$ for RRGs of size $N = 1000$ as a function of the degree c obtained from equation (58), where $\langle L\rangle$ is given by equation (42) and $\langle L^2\rangle$ is given by equation (60). We also present the results obtained from a DE ($\times$ symbols) of the sums in equations (40) and (57) for $\langle L\rangle$ and $\langle L^2\rangle$, respectively. These results are found to be in very good agreement with the results obtained from computer simulations (circles). In the regime of sparse networks the analytical results obtained from the closed form expressions of equations (42) and (60) are in very good agreement with the results obtained from direct evaluation and computer simulations. For dense networks the results obtained from direct evaluation and computer simulations exhibit some oscillations that are not captured by the closed form expressions. Instead, the CF expressions capture the overall trend of $\mathrm{Var}(L)$ vs c as if the oscillations are averaged out. We also present the results obtained from the simpler expression of equation (62) ($+$ symbols). For very small values of c, equation (62) is found to slightly over-estimate the variance, while for larger values of c it is in very good agreement with the results obtained from equations (42) and (60).

In the dense-network limit, the tail sum formula (56) can be approximated by

$$\langle L^2\rangle \simeq (r-1)^2 + \sum_{\ell=r-1}^{r+1} (2\ell+1)P(L > \ell), \tag{63}$$

where r is given by equation (46). Inserting $\langle L^2\rangle$ from equation (63) and $\langle L\rangle$ from equation (45) into (58) and rearranging terms, we obtain

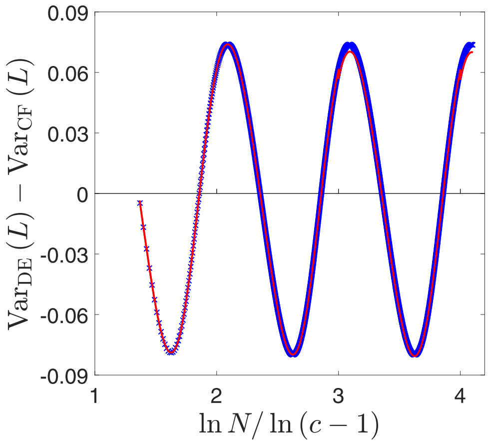

Figure 7. The difference $\mathrm{Var}_{\mathrm{DE}}(L) - \mathrm{Var}_{\mathrm{CF}}(L)$ ($\times$ symbols) between the variance obtained from DE of equations (40) and (57) and from the closed form expressions given by equations (42) and (60), as a function of $\ln N/\ln(c-1)$ where the degree is fixed at $c = 30$ and the network size N is varied. This difference exhibits oscillations as a function of $\ln N/\ln(c-1)$, whose wavelength is equal to 1 and the amplitude is a constant. We also present approximated results (solid line) in which $\mathrm{Var}_{\mathrm{DE}}(L)$ is evaluated using equation (64). The two curves are found to be in very good agreement except for the limit of sparse networks in which equation (64) is not expected to provide accurate results.

$$\begin{aligned}\mathrm{Var}(L) \simeq\ & P(L > r-1) + 3P(L > r) + 5P(L > r+1) \\ & - [P(L > r-1) + P(L > r) + P(L > r+1)]^2. \qquad (64)\end{aligned}$$

To analyze the discrepancy between the results obtained from the CF expressions and those obtained from the DE of the variance, it is convenient to consider the difference $\mathrm{Var}_{\mathrm{DE}}(L) - \mathrm{Var}_{\mathrm{CF}}(L)$, where $\mathrm{Var}_{\mathrm{DE}}(L)$ is the variance obtained from the DE of equations (40) and (57) and $\mathrm{Var}_{\mathrm{CF}}(L)$ is the variance obtained from the closed form expressions of equations (42) and (60). It turns out that this difference exhibits oscillations similar to those obtained for $\langle L\rangle_{\mathrm{DE}} - \langle L\rangle_{\mathrm{CF}}$, between positive and negative values as a function of c. Moreover, the period and the amplitude of these oscillations increase as c is increased.

In figure 6 we present ($\times$ symbols) the difference $\mathrm{Var}_{\mathrm{DE}}(L) - \mathrm{Var}_{\mathrm{CF}}(L)$ as a function of $\ln N/\ln(c-1)$ for $N = 10^6$. It is found that this difference exhibits oscillations as a function of $\ln N/\ln(c-1)$, whose wavelength is equal to 1. The amplitude of the oscillations decreases as $\ln N/\ln(c-1)$ is increased. It is found that the maxima of the oscillations take place at integer values of $\ln N/\ln(c-1)$, while the minima take place at half-integer values. This means that around integer values of $\ln N/\ln(c-1)$ the CF expression provides an under-estimated value for $\mathrm{Var}(L)$, while around half-integer values of $\ln N/\ln(c-1)$ the CF expressions provide an over-estimated value for $\mathrm{Var}(L)$. We also present approximated results (solid line) in which $\mathrm{Var}_{\mathrm{DE}}(L)$ is evaluated using equation (64). The two curves are found to be in very good agreement except for the limit of sparse networks in which equation (64) is not expected to provide accurate results.

In figure 7 we present ($\times$ symbols) the difference $\mathrm{Var}_{\mathrm{DE}}(L) - \mathrm{Var}_{\mathrm{CF}}(L)$ as a function of $\ln N/\ln(c-1)$ where the degree c is fixed at $c = 30$ and the network size N is varied. It is found that this difference exhibits oscillations as a function of $\ln N/\ln(c-1)$, whose wavelength is equal to 1 and the amplitude is a constant. It is found that the maxima of the oscillations take place at integer values of $\ln N/\ln(c-1)$, while the minima take place at half-integer values. This means that around integer values of $\ln N/\ln(c-1)$ the CF expression provides an under-estimated value for $\mathrm{Var}(L)$, while around half-integer values of $\ln N/\ln(c-1)$ the CF expressions provide an over-estimated value for $\mathrm{Var}(L)$. We also present approximated results (solid line) in which $\mathrm{Var}_{\mathrm{DE}}(L)$ is evaluated using equation (64). The two curves are found to be in very good agreement except for the limit of sparse networks in which equation (64) is not expected to provide accurate results. The oscillatory behavior presented in figure 7 implies that the difference $\mathrm{Var}_{\mathrm{DE}}(L) - \mathrm{Var}_{\mathrm{CF}}(L)$ depends on N only via the phase ϕ. This can be justified using an argument similar to the one presented for the mean distance at the end of section 5.

7. Discussion

In configuration model networks, in the large network limit, the mean distance can be approximated by equation (2). It is interesting to compare this result to the corresponding results in models of growing networks. Here we focus on a model of random networks that grow by node duplication (ND), introduced in reference [51]. In this model, at each time step a random (mother) node is duplicated. The daughter node is connected to the mother node and is also connected to each one of the neighbors of the mother node with probability p. This model exhibits a phase transition at $p = 1/2$ between the sparse-network regime at $p < 1/2$ and the dense-network regime at $p > 1/2$. In the sparse-network regime, the mean distance is given by [21]

$$\langle L_{\mathrm{ND}} \rangle = 2(1-\eta)\ln N + \mathcal{O}(1), \tag{65}$$

where $\eta = p + 2p^3 + \mathcal{O}(p^4)$. Thus, the mean distance of a ND network scales logarithmically with N as in the case of configuration model networks. However, it was found that for a given network size N, the mean distance $\langle L_{\mathrm{ND}} \rangle$ is significantly larger than the mean distance of a configuration model network with the same degree distribution [21]. The variance of the DSPL of ND networks is given by [21]

$$\mathrm{Var}(L_{\mathrm{ND}}) = 2(1-\eta)\ln N + \mathcal{O}(1). \tag{66}$$

Comparing equations (65) and (66), one observes that the mean and variance are the same, which is typical to Poisson-like distributions. This implies that the DSPL of ND networks is a broad distribution whose width scales like $(\ln N)^{1/2}$. This is in contrast to the case of RRGs, in which the variance is very small and does not scale with the network size.

In directed ND networks not all the pairs of nodes are connected by directed paths. Conditioning on pairs of nodes that are connected by directed paths, it was found that $\mathbb{E}[L|L < \infty] \sim \ln N$, which is similar to the result for undirected ND networks. However, the variance scales like $\mathrm{Var}(L) \sim (\ln N)^2$, which means that the distribution is much broader [22].

Another quantity which is closely related to the mean distance is the mean diameter $\langle D \rangle$ [52–54]. Unlike the mean distance $\langle L \rangle$ which is averaged over all pairs of nodes in each network instance as well as over the ensemble, the mean diameter is averaged only over the ensemble (each network instance provides a single value of the diameter). It was shown that the mean diameter of an ensemble of RRGs is given by [55]

$$\langle D \rangle = \frac{\ln N}{\ln(c-1)} + \frac{\ln \ln N}{\ln(c-1)} + \mathcal{O}(1). \tag{67}$$

It thus turns out that the expressions for the mean distance $\langle L\rangle$ and the mean diameter $\langle D\rangle$ share the same leading term. This is not an obvious result. It appears to reflect the fact that in the shell structure around a random node, most nodes are concentrated in the last few shells. Such networks, in which most of the distances are almost the same, are called idemetric networks [56]. Having said that, the subleading term for the mean diameter scales like $\ln\ln N$, unlike the case of $\langle L\rangle$ in which it does not depend on N. This reflects the fact that the diameter of an RRG of size N is the maximal distance between all pairs of nodes. As N is increased the number of such pairs increases, making it more probable to find a pair of nodes that are farther away from each other.

8. Summary

The DSPL of RRGs follows a discrete Gompertz distribution. Using the discrete Laplace transform and the Euler–Maclaurin formula we derived a closed-form expression for the moment generating function of the DSPL. From the moment generating function we obtained expressions for the mean distance $\langle L\rangle$ and for the variance $\mathrm{Var}(L)$ of the DSPL. More specifically, it was found that the mean distance is given by

$$\langle L\rangle \simeq \frac{\ln N}{\ln(c-1)} + \frac{1}{2} - \frac{\ln\left(\frac{c}{c-2}\right)+\gamma}{\ln(c-1)}, \tag{68}$$

and the variance of the DSPL is given by

$$\mathrm{Var}(L) \simeq \frac{\pi^2}{6[\ln(c-1)]^2} + \frac{1}{12}. \tag{69}$$

The result for $\langle L\rangle$ extends known results by adding a correction term, which yields very good agreement with the results obtained from DE of $\langle L\rangle$ via the tail-sum formula and with the results obtained from computer simulations. The expression obtained for the variance captures the overall dependence of the variance on the degree c. However, it turns out that on top of the overall trend, the mean distance $\langle L\rangle$ and the variance $\mathrm{Var}(L)$ also exhibit some oscillatory behavior, which is not captured by the CF expressions. The oscillations of the mean and variance are due to the discrete nature of the shell structure around a random node. They reflect the profile of the filling of new shells as N is increased, or as c is decreased. It was shown that these oscillations depend on N only via the phase ϕ, defined in equation (47). This implies regular oscillations with wavelength 1 as a function of $\ln N/\ln(c-1)$, when N is varied and c is kept fixed. The results for the mean and variance of the DSPL were compared to the corresponding results obtained in other types of random networks. The relation between the mean distance and the diameter was also discussed.

Acknowledgments

The authors wish to thank one of the anonymous referees for insightful comments and analysis that helped to clarify the nature of the oscillations of the mean and variance of the DSPL. This work was supported by the Israel Science Foundation Grant No. 1682/18.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Ofer Biham https://orcid.org/0000-0001-9251-812X
Reimer Kühn https://orcid.org/0000-0001-8030-1951
Eytan Katzav https://orcid.org/0000-0001-7555-7717

References

[1] Dorogovtsev S N and Mendes J F F 2003 *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford: Oxford University Press)
[2] Havlin S and Cohen R 2010 *Complex Networks: Structure, Robustness and Function* (Cambridge: Cambridge University Press)
[3] Newman M E J 2018 *Networks: An Introduction* 2nd edn (Oxford: Oxford University Press)
[4] Estrada E 2011 *The Structure of Complex Networks: Theory and Applications* (Oxford: Oxford University Press)
[5] Barrat A, Barthélemy M and Vespignani A 2012 *Dynamical Processes on Complex Networks* (Cambridge: Cambridge University Press)
[6] Newman M E J, Strogatz S H and Watts D J 2001 Random graphs with arbitrary degree distributions and their applications *Phys. Rev.* E **64** 026118
[7] Dorogovtsev S N, Mendes J F F and Samukhin A N 2003 Metric structure of random networks *Nucl. Phys.* B **653** 307
[8] van der Hofstad R, Hooghiemstra G and Van Mieghem P 2005 Distances in random graphs with finite variance degrees *Random Struct. Alg.* **27** 76
[9] van der Hofstad R, Hooghiemstra G and Znamenski D 2007 Distances in random graphs with finite mean and infinite variance degrees *Electron. J. Probab.* **12** 703
[10] Blondel V D, Guillaume J-L, Hendrickx J M and Jungers R M 2007 Distance distribution in random graphs and application to network exploration *Phys. Rev.* E **76** 066101
[11] van der Hofstad R and Hooghiemstra G 2008 Universality for distances in power-law random graphs *J. Math. Phys.* **49** 125209
[12] van der Esker H, van der Hofstad R and Hooghiemstra G 2008 Universality for the distance in finite variance random graphs *J. Stat. Phys.* **133** 169
[13] Shao J, Buldyrev S V, Cohen R, Kitsak M, Havlin S and Stanley H E 2008 Fractal boundaries of complex networks *Europhys. Lett.* **84** 48004
[14] Shao J, Buldyrev S V, Braunstein L A, Havlin S and Stanley H E 2009 Structure of shells in complex networks *Phys. Rev.* E **80** 036105
[15] Katzav E, Nitzan M, ben-Avraham D, Krapivsky P L, Kühn R, Ross N and Biham O 2015 Analytical results for the distribution of shortest path lengths in random networks *Europhys. Lett.* **111** 26006
[16] Nitzan M, Katzav E, Kühn R and Biham O 2016 Distance distribution in configuration-model networks *Phys. Rev.* E **93** 062309
[17] Melnik S and Gleeson J P 2016 Simple and accurate analytical calculation of shortest path lengths (arXiv:1604.05521)
[18] Goldental A, Vardi R, Sardi S, Sabo P and Kanter I 2015 Broadband macroscopic cortical oscillations emerge from intrinsic neuronal response failures *Front. Neural Circuits* **9** 65
[19] Goldental A, Uzan H, Sardi S and Kanter I 2017 Oscillations in networks of networks stem from adaptive nodes with memory *Sci. Rep.* **7** 2700
[20] Bonneau H, Hassid A, Biham O, Kühn R and Katzav E 2017 Distribution of shortest cycle lengths in random networks *Phys. Rev.* E **96** 062307

[21] Steinbock C, Biham O and Katzav E 2017 Distribution of shortest path lengths in a class of node duplication network models *Phys. Rev.* E **96** 032301
[22] Steinbock C, Biham O and Katzav E 2019 Analytical results for the distribution of shortest path lengths in directed random networks that grow by node duplication *Eur. Phys. J.* B **92** 130
[23] Tishby I, Biham O, Katzav E and Kühn R 2018 Revealing the microstructure of the giant component in random graph ensembles *Phys. Rev.* E **97** 042318
[24] Katzav E, Biham O and Hartmann A 2018 Metric properties of subcritical Erdös–Rényi networks *Phys. Rev.* E **98** 012301
[25] Asher E, Sanhedrai H, Panduranga N K, Cohen R and Havlin S 2020 Distance distribution in extreme modular networks *Phys. Rev.* E **101** 022313
[26] Jackson A D and Patil S P 2021 Phases of small worlds: a mean field formulation (arXiv:2103.14001)
[27] Travers J and Milgram S 1969 An experimental study of the small world problem *Sociometry* **32** 425
[28] Watts D J and Strogatz S H 1998 Collective dynamics of small-world networks *Nature* **393** 440
[29] Chung F and Lu L 2002 The average distances in random graphs with given expected degrees *Proc. Natl Acad. Sci. USA* **99** 15879
[30] Chung F and Lu L 2003 The average distance in a random graph with given expected degrees *Internet Math.* **1** 91
[31] Fronczak A, Fronczak P and Holyst J A 2004 Average path length in random networks *Phys. Rev.* E **70** 056110
[32] Cohen R and Havlin S 2003 Scale-free networks are ultrasmall *Phys. Rev. Lett.* **90** 058701
[33] Pitman J 1993 *Probability* (Berlin: Springer)
[34] Molloy M and Reed B 1995 A critical point for random graphs with a given degree sequence *Random Struct. Alg.* **6** 161
[35] Molloy M and Reed B 1998 The size of the giant component of a random graph with a given degree sequence *Combinator. Probab. Comput.* **7** 295
[36] Newman M E J and Park J 2003 Why social networks are different from other types of networks *Phys. Rev.* E **68** 036122
[37] Erdös P and Rényi A 1959 On random graphs I *Publ. Math.* **6** 290
[38] Erdös P and Rényi A 1960 On the evolution of random graphs *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17
[39] Erdös P and Rényi A 1961 On the evolution of random graphs II *Bull. Inst. Int. Stat.* **38** 343
[40] Gompertz B 1825 On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies *Phil. Trans. R. Soc.* A **115** 513
[41] Shklovskii B 2005 A simple derivation of the Gompertz law for human mortality *Theory Biosci.* **123** 431
[42] Shimizu N 2020 The average distance and the diameter of dense random regular graphs *Electron. J. Combinator.* **27** P3.62
[43] Bollobas B 2001 *Random Graphs* 2nd edn (New York: Academic)
[44] Coolen A C C, Annibale A and Roberts E 2017 *Generating Random Networks and Graphs* (Oxford: Oxford University Press)
[45] Mézard M and Parisi G 2003 The cavity method at zero temperature *J. Stat. Phys.* **111** 1
[46] Phillips C L, Nagle H T and Chakrabortty A 2015 *Digital Control System: Analysis and Design* 4th edn (Harlow: Pearson Education)
[47] Apostol T M 1999 An elementary view of Euler's summation formula *Am. Math. Mon.* **106** 409
[48] Bender C M and Orzag S A 1999 *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory* (Berlin: Springer)
[49] Olver F W J, Lozier D M, Boisvert R R and Clark C W 2010 *NIST Handbook of Mathematical Functions* (Cambridge: Cambridge University Press)
[50] Finch S R 2003 *Mathematical Constants* (Cambridge: Cambridge University Press)
[51] Lambiotte R, Krapivsky P L, Bhat U and Redner S 2016 Structural transitions in densifying networks *Phys. Rev. Lett.* **117** 218301
[52] Fernholz D and Ramachandran V 2007 The diameter of sparse random graphs *Random Struct. Alg.* **31** 482
[53] Riordan O and Wormald N 2010 The diameter of sparse random graphs *Combinator. Probab. Comput.* **19** 835

[54] Chung F and Lu L 2001 The diameter of sparse random graphs *Adv. Appl. Math.* **26** 257
[55] Bollobás B and Fernandez de la Vega W 1982 The diameter of random regular graphs *Combinatorica* **2** 125
[56] Barmpalias G, Huang N, Lewis-Pye A, Li A, Li X, Pan Y and Roughgarden T 2019 The idemetric property: when most distances are (almost) the same *Proc. R. Soc.* A **475** 20180283