# Distribution of shortest path lengths on trees of a given size in subcritical Erdős-Rényi networks

Barak Budnick, Ofer Biham, and Eytan Katzav
*Racah Institute of Physics, The Hebrew University, Jerusalem 9190401, Israel*

(Received 10 January 2023; accepted 2 October 2023; published 30 October 2023)

In the subcritical regime Erdős-Rényi (ER) networks consist of finite tree components, which are nonextensive in the network size. The distribution of shortest path lengths (DSPL) of subcritical ER networks was recently calculated using a topological expansion [E. Katzav, O. Biham, and A. K. Hartmann, Phys. Rev. E **98**, 012301 (2018)]. The DSPL, which accounts for the distance $\ell$ between any pair of nodes that reside on the same finite tree component, was found to follow a geometric distribution of the form $P(L=\ell|L<\infty)=(1-c)c^{\ell-1}$, where $0<c<1$ is the mean degree of the network. This result includes the contributions of trees of all possible sizes and topologies. Here we calculate the distribution of shortest path lengths $P(L=\ell|S=s)$ between random pairs of nodes that reside on the same tree component of a given size $s$. It is found that $P(L=\ell|S=s)=\frac{\ell+1}{s^\ell}\frac{(s-2)!}{(s-\ell-1)!}$. Surprisingly, this distribution does not depend on the mean degree $c$ of the network from which the tree components were extracted. This is due to the fact that the ensemble of tree components of a given size $s$ in subcritical ER networks is sampled uniformly from the set of labeled trees of size $s$ and thus does not depend on $c$. The moments of the DSPL are also calculated. It is found that the mean distance between random pairs of nodes on tree components of size $s$ satisfies $\mathbb{E}[L|S=s]\sim\sqrt{s}$, unlike small-world networks in which the mean distance scales logarithmically with $s$.

DOI: 10.1103/PhysRevE.108.044310

## I. INTRODUCTION

Random networks provide a useful framework for the analysis of a large variety of systems that consist of interacting objects [1–4]. One can distinguish between two major types of random networks: supercritical networks and subcritical networks. Supercritical networks form a giant component that encompasses a macroscopic fraction of all the nodes. The giant component may provide a useful description of networks in which the connectivity is essential, such as the world-wide-web, social networks, and infrastructure networks. The giant component is a small-world network, namely, the mean distance between pairs of nodes on the giant component scales logarithmically with its size. It includes a large number of cycles with a broad spectrum of cycle lengths [5–7]. These cycles provide redundancy in the connectivity between pairs of nodes via multiple paths. The redundancy helps to maintain the integrity of the giant component upon deletion of nodes or edges due to failures or attacks. The combination of the small-world property and the redundancy gives rise to highly efficient channels of transport and communication and to the robustness of the network. In contrast, subcritical networks consist of finite tree components that do not scale with the overall network size. In a tree topology, each pair of nodes is connected by a single path. Therefore, in subcritical networks, the shortest path between any pair of nodes that reside on the same tree component is, in fact, the only path between them. As a result, in subcritical networks each node of degree $k\geqslant 2$ is an articulation point, namely, its deletion would break the tree component on which it resides into at least two disconnected parts [8,9]. Moreover, each edge is a bredge (bridge edge), namely, its deletion would break the tree component on which it resides into two disconnected parts [10]. The subcritical tree components may describe the fragmented structure of secure compartmentalized networks, such as the communication networks of commercial enterprises, government agencies, and illicit organizations [11]. The structure of such networks may be determined by the trade-off between efficiency and security. When security considerations outweigh efficiency considerations, the number of communication lines may need to be reduced to a minimum, which is achieved in the case of tree structures. Other examples of fragmented networks include networks that suffered multiple failures, large-scale attacks, or epidemics, in which the remaining functional or uninfected nodes form small, isolated components [12,13]. In spite of their importance, the structural and statistical properties of subcritical networks have not attracted nearly as much attention as those of supercritical networks.

Random networks of the Erdős-Rényi (ER) type [14–16] are the simplest class of random networks and are used as a benchmark for the study of structure and dynamics in complex networks [17]. The ER network ensemble is a maximum-entropy ensemble, under the condition that the mean degree $\langle K\rangle=c$ is fixed. It is a special case of a broader class of random uncorrelated networks, referred to as configuration model networks [18–21]. In an ER network of $N$ nodes, each pair of nodes is independently connected with probability $p$, such that the mean degree is $c=(N-1)p$. It was recently shown that the ER graph structure is an asymptotic structure for networks that contract due to node deletion processes, which may result from failures, attacks or epidemics [22,23].

The degree distribution of ER networks follows a Poisson distribution of the form

$$P(K=k)=\frac{e^{-c}c^k}{k!}. \tag{1}$$

 

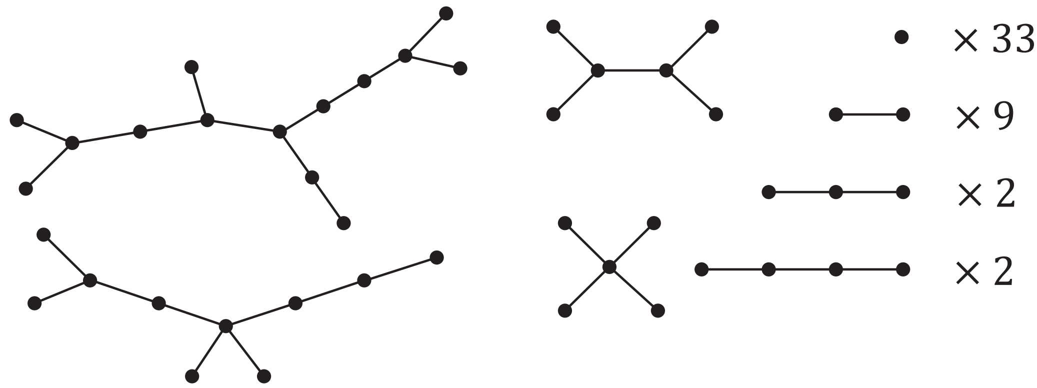


FIG. 1. The structure of a single instance of a subcritical ER network of $N = 100$ nodes with mean degree $c = 0.9$. It consists of 33 isolated nodes, nine dimers, two chains of three nodes, two chains of four nodes, and trees of 5, 6, 10, and 14 nodes.

ER networks exhibit a percolation transition at $c = 1$ such that for $c > 1$ (supercritical regime) there is a giant component [24], while for $0 < c < 1$ (subcritical regime) the network consists of small, isolated tree components [17,25]. In the special case of $c = 0$ the network consists of $N$ isolated nodes and the degree distribution degenerates into $P(K = k) = \delta_{k,0}$.

In Fig. 1 we present the structure of a single instance of a subcritical ER network of size $N = 100$ with mean degree $c = 0.9$. It consists of 33 isolated nodes, nine dimers, two chains of three nodes, two chains of four nodes and trees of 5, 6, 10, and 14 nodes.

In the asymptotic limit, ER networks exhibit duality with respect to the percolation threshold [17]. In a supercritical ER network of $N$ nodes the fraction of nodes that belong to the giant component is denoted by $0 < g \leqslant 1$, while the fraction of nodes that belong to the finite components is $1 - g$. Thus, the subcritical network that consists of the finite components is of size $N(1 - g)$. This network is in itself an ER network whose mean degree is $c' = c(1 - g)$, where $c' < 1$.

The distribution of tree sizes in subcritical ER networks with mean degree $0 < c < 1$ is given by [17,26,27]

$$P(S = s) = \frac{2s^{s-2}c^{s-1}e^{-cs}}{(2 - c)s!}. \tag{2}$$

In the special case of $c = 0$ this distribution degenerates into $P(S = s) = \delta_{s,1}$.

The mean tree size is given by [27]

$$\langle S \rangle = \frac{2}{2 - c}. \tag{3}$$

The expected number of trees in a network instance consisting of $N$ nodes is thus given by

$$N_T = \frac{N}{\langle S \rangle} = N\left(1 - \frac{c}{2}\right). \tag{4}$$

The variance of $P(S = s)$ is given by [27]

$$\mathrm{Var}(S) = \frac{2c}{(1 - c)(2 - c)^2}. \tag{5}$$

Note that $\mathrm{Var}(S)$ diverges as $c \to 1^-$, which implies that, near the percolation transition, some of the trees are very large.

Trees of a given size $s$ may exhibit different structures, where the number of distinct structures increases with $s$. An important distinction in this context is between labeled trees, in which nodes are distinguishable and carry labels, and unlabeled trees in which the nodes are indistinguishable. The number $T_s$ of distinct labeled tree configurations of size $s$ is given by the Cayley formula [28]

$$T_s = s^{s-2}. \tag{6}$$

Each one of these labeled tree configurations can be encoded by a unique sequence, refereed to as the Prüfer sequence [29]. The Prüfer sequence of a labeled tree of $s$ nodes is a string of $s - 2$ integers, taking values in the range of $1, 2, \ldots, s$. The Prüfer code provides a very powerful tool for the random sampling of labeled trees of a given size.

When the labels are removed, the number of distinct configurations is reduced since each unlabeled configuration corresponds to several labeled configurations. In the case of unlabeled trees, the number of nonisomorphic tree topologies, $n(s)$, which can be assembled from $s$ nodes quickly increases as a function of $s$. For example, the values of $n(s)$ for $s = 1, 2, \ldots, 13$ are 1, 1, 1, 2, 3, 6, 11, 23, 47, 106, 235, 551, and 1301, respectively [30]. An efficient algorithm for generating all the tree topologies that can be assembled from $s$ nodes, is presented in Refs. [31,32]. A list of all possible tree topologies up to $s = 13$ is presented in Ref. [30].

In Fig. 2 we present the tree topologies that consist of $s$ nodes for $s = 1, 2, \ldots, 7$. For $s \leqslant 3$ the linear chain topology is the only possible topology while for $s \geqslant 4$ more complex topologies appear and their number quickly increases. The number of labeled configuration associated with each one of the tree topologies is also shown. Note that the total number of labeled trees that consist of $s$ nodes add up to $s^{s-2}$, which is consistent with the Cayley formula (6).

While the local structure of a network is well characterized by the degree distribution, the distribution of shortest path lengths (DSPL), denoted by $P(L = \ell)$, provides a useful characterization of its large-scale structure. When two nodes $i$ and $j$ reside on the same connected component, the distance $\ell_{ij}$ between them is given by the length of the shortest path that connects them. When nodes $i$ and $j$ reside on different network components, there is no path connecting them and the distance between them is $\ell_{ij} = \infty$. The probability that two randomly selected nodes reside on the same component and thus are at a finite distance from each other, is

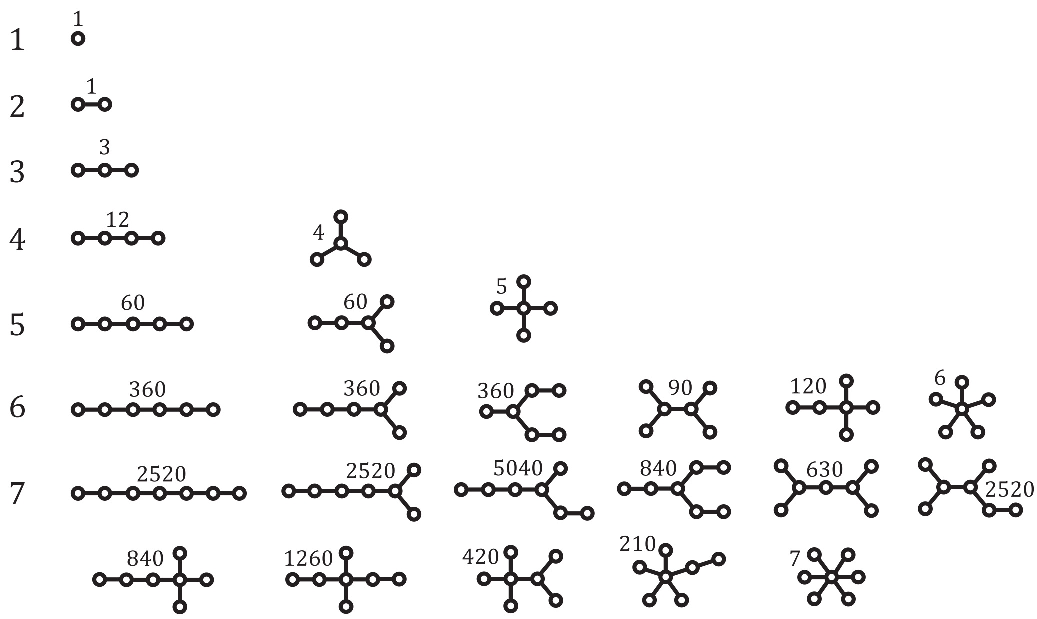


FIG. 2. The tree topologies that consist of $s$ nodes for $s = 1, 2, \dots, 7$. For $s \leqslant 3$ the linear chain topology is the only topology, while for $s \geqslant 4$ more complex topologies appear and their number quickly increases. The number of labeled configurations associated with each one of the tree topologies is also shown. Note that the total number of labeled trees that consist of $s$ nodes add up to $s^{s-2}$, which is consistent with the Cayley formula (6).

denoted $P(L < \infty) = 1 - P(L = \infty)$. The conditional DSPL between pairs of nodes that reside on the same component is denoted $P(L = \ell | L < \infty)$, where $\ell = 1, 2, \dots, N - 1$. The conditional DSPL satisfies

$$P(L = \ell | L < \infty) = \frac{P(L = \ell)}{P(L < \infty)}. \quad (7)$$

Note that $P(L = \ell | L < \infty)$ is well defined only for $c > 0$. This is due to the fact that $P(L < \infty) = 0$ for $c = 0$. Thus, the analysis presented below is focused on $0 < c < 1$.

The DSPL provides a natural platform for the study of dynamical processes on networks, such as diffusive processes, epidemic spreading, critical phenomena, synchronization, information propagation, and communication. For supercritical networks the DSPL was calculated using various theoretical approaches, which include recursion equations, generating functions, master equations, and branching processes [7,12,13,21,24,33–45]. In the special case of random regular graphs with $c \geqslant 3$, the giant component encompasses the whole network. In this case there is a closed-form analytical expression for $P(L = \ell)$ [34,40,44], which follows a discrete Gompertz distribution [46].

It was shown that the mean distance $\mathbb{E}[L|L < \infty] = \sum_{\ell=1}^{\infty} \ell P(L = \ell | L < \infty)$ scales like $\mathbb{E}[L|L < \infty] \sim \ln N / \ln c$, in agreement with rigorous results, showing that supercritical random networks are small-world networks [47–50]. It was also shown that the variance of the DSPL of supercritical random networks does not scale with $N$, and satisfies $\mathrm{Var}(L) \sim O(1)$ [40]. The statistical properties of distances in scale-free networks, which typically consist of a single connected component, were studied in Refs. [36,37,51]. Using an analytical argument it was shown that scale free networks with degree distributions of the form $P(k) \sim k^{-\gamma}$ are ultrasmall, namely, they exhibit a mean distance which scales like $\mathbb{E}[L] \sim \ln \ln N$ for $2 < \gamma < 3$. For $\gamma = 3$ it was shown that the mean distance scales like $\mathbb{E}[L] \sim \ln N / \ln \ln N$, while for $\gamma > 3$ it coincides with the common scaling of small world networks, namely, $\mathbb{E}[L] \sim \ln N$.

The DSPL of subcritical ER networks was recently studied using a topological expansion [27]. This analysis employs the fact that, in the subcritical regime, in the large-network limit, the network consists of finite tree components with no cycles [17,25]. It was found that for $0 < c < 1$ the DSPL between pairs of nodes that reside on the same tree component is given by [27]

$$P(L = \ell | L < \infty) = (1 - c)c^{\ell - 1}, \quad (8)$$

and that the probability that two random nodes reside on the same tree component is [27]

$$P(L < \infty) = \frac{c}{(1 - c)N}. \quad (9)$$

The corresponding tail distribution is given by

$$P(L > \ell | L < \infty) = c^{\ell}. \quad (10)$$

The mean distance between pairs of nodes that reside on the same tree component is

$$\mathbb{E}[L|L < \infty] = \frac{1}{1 - c}, \quad (11)$$

while the variance of the DSPL is given by

$$\mathrm{Var}(L|L < \infty) = \frac{c}{(1 - c)^2}. \quad (12)$$

While subcritical ER networks consist of finite tree components, in supercritical ER networks there is a coexistence between the giant component and the finite tree components. As a result, the DSPL of supercritical ER networks combines the contributions of the giant and finite components. Using the duality relations discussed above, the DSPL of the finite components of a supercritical ER network can be obtained from the analysis of its dual subcritical network [24,27].

In this paper we calculate the DSPL of finite tree components of size $s$, denoted by $P(L=\ell|S=s)$, in subcritical ER networks. This is done by expressing the overall distribution $P(L=\ell)$ as a linear combination of the corresponding conditional distributions $P(L=\ell|S=s)$, using the known distribution of tree sizes. Using an inverse transformation we extract the conditional distribution $P(L=\ell|S=s)$. Surprisingly, this distribution does not depend on the mean degree $c$ of the network from which the tree components were extracted. This is due to the fact that the ensemble of tree components of a given size $s$ in subcritical ER networks is sampled uniformly from the set of labeled trees of size $s$ and thus does not depend on $c$. This insight is corroborated by a direct combinatorial argument. We also calculate the DSPL over all tree components up to size $s$, denoted by $P(L=\ell|S\leqslant s)$ and examine its convergence towards the DSPL of the whole network, $P(L=\ell|L<\infty)$, as $s$ is increased. The moments of the DSPL are also calculated. It is found that the mean distance between random pairs of nodes on tree components of size $s$ satisfies $\mathbb{E}[L|S=s]\sim\sqrt{s}$, unlike small-world networks in which the mean distance scales logarithmically with $s$.

The paper is organized as follows: In Sec. II we consider the conditional DSPL on finite tree components. The moments of the DSPL are calculated in Sec. III. The results are discussed in Sec. IV and summarized in Sec. V.

## II. THE DISTRIBUTION OF SHORTEST PATH LENGTHS

Using the law of total probability, the DSPL of subcritical ER networks, given by Eq. (8), can be expressed in the form

$$P(L=\ell|L<\infty)=\sum_{s=2}^{\infty}P(L=\ell|S=s)\widehat{P}(S=s), \tag{13}$$

where $P(L=\ell|S=s)$ is the DSPL on tree components that consist of $s$ nodes and $\widehat{P}(S=s)$ is the distribution of tree sizes on which a pair of random nodes resides (given that they reside on the same tree component). In the analysis below we extract a closed-form expression for $P(L=\ell|S=s)$ by inverting the infinite system of linear equations given by Eq. (13). Unlike commonly used methods for the calculation of such distributions, which are based on combinatorial considerations, this approach is purely algebraic. It is essentially a top-down approach, in which the conditional distribution $P(L=\ell|S=s)$ is obtained from the overall distribution $P(L=\ell|L<\infty)$ via the distribution of tree sizes $P(S=s)$. This approach is advantageous over the complementary bottom-up approach, which would require a detailed knowledge of all the tree configurations of size $s$, their weights, and the DSPL over each and every one of them.

The distribution $\widehat{P}(S=s)$ is given by

$$\widehat{P}(S=s)=\frac{\binom{s}{2}}{\left\langle\binom{S}{2}\right\rangle}P(S=s), \tag{14}$$

where

$$\left\langle\binom{S}{2}\right\rangle=\sum_{s=2}^{\infty}\binom{s}{2}P(S=s) \tag{15}$$

is the mean number of pairs of nodes in a randomly selected tree component, and $P(S=s)$ is given by Eq. (2). This is due to the fact that the number of pairs of nodes on a tree component of size $s$ is given by the binomial coefficient $\binom{s}{2}$. The evaluation of $\left\langle\binom{S}{2}\right\rangle$ is presented in the Appendix. It yields

$$\left\langle\binom{S}{2}\right\rangle=\frac{c}{(1-c)(2-c)}, \tag{16}$$

where $0<c<1$. Inserting $P(S=s)$ from Eq. (2) and $\left\langle\binom{s}{2}\right\rangle$ from Eq. (16) into Eq. (14), we obtain

$$\widehat{P}(S=s)=(1-c)\frac{s^{s-2}c^{s-2}e^{-cs}}{(s-2)!}. \tag{17}$$

Inserting $\widehat{P}(S=s)$ from Eq. (17) and $P(L=\ell|L<\infty)$ from Eq. (8) into Eq. (13), we obtain

$$\sum_{s=2}^{\infty}\frac{s^{s-2}c^{s-1}e^{-cs}}{(s-2)!}P(L=\ell|S=s)=c^{\ell}. \tag{18}$$

This equation can be rewritten in the form

$$\sum_{s=2}^{\infty}\frac{s^{s-2}(ce^{-c})^{s}}{(s-2)!}P(L=\ell|S=s)=c^{\ell+1}. \tag{19}$$

The distribution $P(L=\ell|S=s)$ is obtained by inverting Eq. (19). In the inversion process we assume that $P(L=\ell|S=s)$ does not depend on the mean degree $c$. The results presented below show that such a solution indeed exists and is justified by a combinatorial argument. The resulting expression for $P(L=\ell|S=s)$ is verified by computer simulations.

Defining

$$x=ce^{-c} \tag{20}$$

enables us to express the left-hand side of Eq. (19) as a power series in $x$. For the analysis below, it will be useful to also express the right-hand side in terms of $x$ rather than $c$. To this end, we invert Eq. (20) and obtain

$$c=-W(-x), \tag{21}$$

where $W(x)$ is the Lambert $W$ function [52]. Equation (19) can now be written in the form

$$\sum_{s=2}^{\infty}\frac{s^{s-2}}{(s-2)!}P(L=\ell|S=s)x^{s}=[-W(-x)]^{\ell+1}. \tag{22}$$

From equation (3.2.2) in Ref. [53], which results from the Lagrange inversion formula, we obtain the identity

$$[W(x)]^{r}=(-r)\sum_{s=r}^{\infty}\frac{(-s)^{s-r-1}}{(s-r)!}x^{s}. \tag{23}$$

Using Eq. (23) we now express the right-hand side of Eq. (22) as a power series in $x$. Comparing the coefficients of $x^{s}$ on both sides of Eq. (22), we obtain the DSPL of tree components that consist of $s$ nodes in subcritical ER networks with $0 < c < 1$. It is given by

$$P(L = \ell | S = s) = \frac{(\ell + 1)}{s^{\ell}} \frac{(s-2)!}{(s-\ell-1)!}, \tag{24}$$

where $s \geqslant 2$ and $1 \leqslant \ell \leqslant s - 1$. This is the central result of the paper. Clearly, this distribution does not depend on the mean degree $c$ of the subcritical network from which the trees of size $s$ were extracted.

Unlike the DSPL of the whole network, which is a monotonically decreasing geometric distribution, $P(L = \ell | S = s)$ exhibits a peak. The location of the peak is referred to as the mode of the distribution and is denoted $\ell_{\rm mode}$. Since $P(L = \ell | S = s)$ exhibits a single peak, $\ell_{\rm mode}$ is the lowest integer for which $P(L = \ell + 1 | S = s) < P(L = \ell | S = s)$. Using Eq. (24), this inequality can be expressed in the form

$$\left( \frac{\ell + 2}{\ell + 1} \right) \left( \frac{s - \ell - 1}{s} \right) < 1. \tag{25}$$

The solution of this inequality (assuming positive $\ell$) is

$$\ell > \frac{\sqrt{4s+1} - 3}{2}. \tag{26}$$

The mode $\ell_{\rm mode}$ is the lowest integer that satisfies Eq. (26), namely,

$$\ell_{\rm mode} = \left\lceil \frac{\sqrt{4s+1} - 3}{2} \right\rceil, \tag{27}$$

where $\lceil x \rceil$ is the lowest integer that is larger than $x$, also known as the ceiling function. In the limit of large trees, the mode scales like $\ell_{\rm mode} \sim \sqrt{s}$.

It turns out that the DSPL given by Eq. (24) coincides with the DSPL of the ensemble obtained by uniformly random sampling over all the labeled tree configurations of size $s$ [54,55]. The DSPL over all the labeled tree configurations of size $s$ can be obtained from direct combinatorial considerations. To this end we pick a random pair of nodes $i$ and $j$ on a tree of size $s$. We count the number of possible configurations of labeled trees of size $s$, in which the distance between a given pair of nodes $i$ and $j$ is $\ell$. The fact that the distance between $i$ and $j$ is $\ell$ implies that there is a single path of length $\ell$ between them. This path consists of $\ell - 1$ intermediate nodes. The number of ways to select these $\ell - 1$ nodes from the $s - 2$ nodes (not including $i$ and $j$), where the order is important, is given by

$$\frac{(s-2)!}{(s-\ell-1)!} = \binom{s-2}{\ell-1} (\ell-1)!. \tag{28}$$

The path joining $i$ and $j$, which consists of $\ell + 1$ nodes (including $i$ and $j$), can be considered as the backbone of the tree. Each node on the backbone may be the root of a tree branch such that each one of the remaining $s - \ell - 1$ nodes belongs to one of these tree branches. This enables us to use the generalized Cayley formula [28,56,57], which provides the number of labeled tree configurations that consist of $\ell + 1$ nonempty disjoint tree components (also known as forests) with a total of $s$ nodes, namely,

$$T_{s,\ell+1} = (\ell + 1) s^{s-\ell-2}. \tag{29}$$

Note that Cayley formula of Eq. (6) is a special case of the generalized Cayley formula (29), namely $T_s = T_{s,1}$. The probability $P(L = \ell | S = s)$ is obtained by dividing the number of possible configurations of labeled trees of size $s$, in which the distance between a given pair of nodes $i$ and $j$ is $\ell$ by the total number $T_s$ of configurations of labeled trees of size $s$. It yields

$$P(L = \ell | S = s) = \frac{T_{s,\ell+1} \binom{s-2}{\ell-1} (\ell-1)!}{T_s}, \tag{30}$$

which is equivalent to Eq. (24). This equivalence suggests that the ensemble of trees of a given size $s$ in subcritical ER networks is equivalent to a uniformly random sampling among all the $T_s$ labeled tree configurations of size $s$. This is consistent with the fact that the DSPL given by Eqs. (24) and (30) does not depend on the mean degree $c$ of the network from which these trees were extracted. The equivalence between the two ensembles can be justified using the following argument: Given a finite connected component consisting of $s$ nodes in a subcritical ER network it is almost surely to exhibit a tree topology containing $s - 1$ edges [17]. For a set of $s$ nodes, the probability that these nodes will form a connected tree component of a given labeled configuration, which is isolated from the rest of the network, is given by

$$p^{s-1} (1-p)^{\binom{s}{2} - (s-1)} (1-p)^{s(N-s)}, \tag{31}$$

where the first term accounts for the $s - 1$ edges of the tree, the second term accounts for the probability that there are no additional edges between the nodes in the tree component, and the third term accounts for the probability that the tree is isolated from the rest of the network. In an ER network, in which the connectivity between different pairs of nodes is independent, this probability is the same for all possible configurations of labeled trees of size $s$.

Summing up the right-hand side of Eq. (24) from $\ell + 1$ to infinity, we obtain the tail distribution, which is given by

$$P(L > \ell | S = s) = \frac{(s-2)!}{s^{s-2}} \frac{s^{s-\ell-2}}{(s-\ell-2)!}, \tag{32}$$

where $\ell = 0, 1, 2, \ldots, s - 2$. It is a monotonically decreasing function that satisfies $P(L > 0 | S = s) = 1$ and $P(L > s - 2 | S = s) = (s-2)!/s^{s-2}$.

In Fig. 3 we present analytical results (solid lines) for the DSPL on trees of size $s$, denoted by $P(L = \ell | S = s)$, for $s = 10$, 20, 30 and 40, obtained from Eq. (24). The analytical results are in very good agreement with the results obtained from computer simulations carried out for $c = 0.5$ ($\times$) and $c = 0.8$ ($\circ$), which coincide with each other. These results confirm the validity of Eq. (24) as well as the fact that the ensemble of finite trees of a given size $s$ extracted from subcritical ER networks of mean degree $c$ does not depend on $c$.

In the simulations we generated subcritical ER networks of size $N = 10^4$ with mean degree $c = 0.5$ and $c = 0.8$. From these networks we picked tree components of the desired sizes, such as $s = 10$, 20, 30, and 40. The expected number

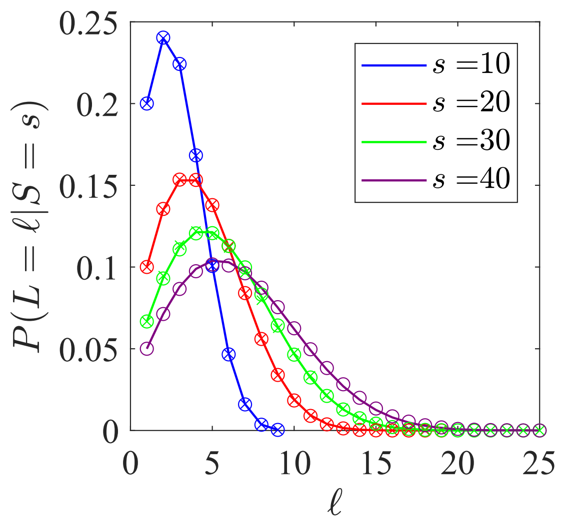


FIG. 3. Analytical results (solid lines) for the DSPL on trees of size $s$, denoted $P(L=\ell|S=s)$, for $s=10$, 20, 30, and 40 (left to right), obtained from Eq. (24). The analytical results are in very good agreement with the results obtained from computer simulations carried out for networks of size $N=10^4$, $c=0.5$ ($\times$), and $c=0.8$ ($\circ$), which coincide with each other. These results confirm the validity of Eq. (24) as well as the fact that the ensemble of finite trees of a given size $s$ extracted from subcritical ER networks of mean degree $c$ does not depend on $c$. Note that the simulation results for $c=0.5$ are shown only for $s=10$, 20, and 30, because trees of size $s=40$ are extremely rare in this case.

of trees of size $s$ in a network instance of size $N$ is given by

$$N_T(s)=N_T P(S=s). \tag{33}$$

Inserting $N_T$ from Eq. (4) and $P(S=s)$ from Eq. (2) into Eq. (33), we obtain

$$N_T(s)=N\frac{s^{s-2}c^{s-1}e^{-cs}}{s!}. \tag{34}$$

This result can be used in order to estimate the number of network instances which is required in order to obtain the desired number of trees of size $s$ that are needed for the statistical analysis. The distribution $P(S=s)$ is a quickly decreasing function of $s$. Thus, trees of size $s$ become less abundant as $s$ is increased. As a result, one needs a large number of network instances in order to obtain sufficient data for statistical analysis of large tree components. The results presented in Fig. 3 are based on 1500 instances of subcritical ER networks of size $N=10^4$ for each value of $c$. For $c=0.8$ these network instances yield 12 454 trees of size 10, 1823 trees of size 20, 500 trees of size 30, and 183 trees of size 40. For $c=0.5$ these network instances yield 3617 trees of size 10, 91 trees of size 20, 8 trees of size 30, and no trees of size 40. Therefore, In Fig. 3 the analytical results for $s=40$ are compared only with the simulation results for $c=0.8$ ($\circ$).

Another interesting distribution is the DSPL between pairs of nodes that reside on all tree components of size $s'\leqslant s$. It can be obtained from

$$P(L=\ell|S\leqslant s)=\frac{\sum_{s'=2}^{s}\widehat{P}(S=s')P(L=\ell|S=s')}{\sum_{s'=2}^{s}\widehat{P}(S=s')}. \tag{35}$$

Taking the limit of large $s$, $P(L=\ell|S\leqslant s)$ converges towards $P(L=\ell|L<\infty)$, as in Eq. (13). To explore this convergence it is convenient to replace the sums $\sum_{s'=2}^{s}$ in Eq. (35) by the difference $\sum_{s'=2}^{\infty}-\sum_{s'=s+1}^{\infty}$. Carrying out the first summations in the numerator and in the denominator, we obtain

$$P(L=\ell|S\leqslant s)=\frac{(1-c)c^{\ell-1}-\sum_{s'=s+1}^{\infty}\widehat{P}(S=s')P(L=\ell|S=s')}{1-\sum_{s'=s+1}^{\infty}\widehat{P}(S=s')}. \tag{36}$$

In Fig. 4 we present analytical results (solid lines) for the distribution $P(L=\ell|S\leqslant s)$ of shortest path lengths on all tree components of size smaller or equal to $s$, in subcritical ER networks with mean degree $c=0.8$. The analytical results obtained from Eq. (36), are presented for tree sizes of $s=10$, 20, 30, and 40 (top to bottom on the left-hand side). The analytical results are in very good agreement with the results obtained from computer simulations carried out for $c=0.8$ ($\circ$). As $s$ is increased, the distribution $P(L=\ell|S\leqslant s)$ converges towards the overall DSPL $P(L=\ell|L<\infty)$ of the subcritical ER network (dashed line).

## III. THE MEAN AND VARIANCE OF THE DISTRIBUTION OF SHORTEST PATH LENGTHS

To calculate the moments of the DSPL, we define the moment-generating function

$$M(x)=\sum_{\ell=1}^{s-1}e^{x\ell}P(L=\ell|S=s). \tag{37}$$

Inserting the probability $P(L=\ell|S=s)$ from Eq. (24) into Eq. (37) and carrying out the summation, we obtain

$$M(x)=\frac{s}{s-1}\left[\left(e^{-2x}-\frac{1}{s}\right)+\frac{e^{-x}}{s^s}(1-e^{-x})e^{s(x+e^{-x})}\Gamma(s+1,se^{-x})\right]. \tag{38}$$

where $\Gamma(a,z)$ is the incomplete Gamma function [52]. The $n$th moment of $P(L=\ell|S=s)$ is obtained by differentiating $M(x)$, with respect to $x$, $n$ times, namely,

$$\mathbb{E}[L^n|S=s]=\left.\frac{\partial^n M}{\partial x^n}\right|_{x=0}. \tag{39}$$

Inserting $n=1$ in Eq. (39), we obtain the mean distance between random pairs of nodes that reside on a tree component of size $s$. It is given by

$$\mathbb{E}[L|S=s]=\frac{s[e^s s^{-s}\Gamma(s+1,s)-2]}{s-1}. \tag{40}$$

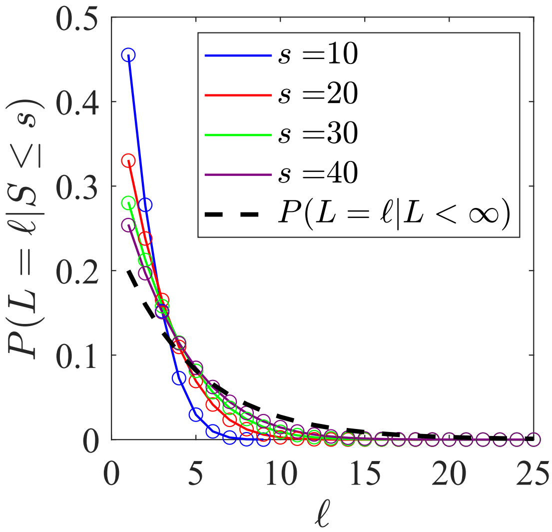


FIG. 4. Analytical results (solid lines) for $P(L=\ell|S\leqslant s)$ on tree components of size smaller or equal to $s$ in subcritical ER network with mean degree $c=0.8$, for $s=10$, 20, 30, and 40 (top to bottom on the left-hand side), obtained from Eq. (36). As $s$ is increased, $P(L=\ell|S\leqslant s)$ converges towards the overall DSPL, $P(L=\ell|L<\infty)$, of the subcritical ER network (dashed line). The analytical results are in very good agreement with the results obtained from computer simulations (○).

Inserting $n=2$ in Eq. (39), we obtain the second moment, which is given by

$$\mathbb{E}[L^2|S=s]=\frac{s[4+2s-3e^s s^{-s}\Gamma(s+1,s)]}{s-1}. \quad (41)$$

The variance of $P(L=\ell|S=s)$ is given by

$$\mathrm{Var}(L|S=s)=\mathbb{E}[L^2|S=s]-(\mathbb{E}[L|S=s])^2, \quad (42)$$

where $\mathbb{E}[L^2|S=s]$ is given by Eq. (41) and $\mathbb{E}[L|S=s]$ is given by Eq. (40).

For sufficiently large values of $s$ one can obtain simplified asymptotic expressions for the moments of the DSPL. To achieve this we use the double-asymptotic expansion of $\Gamma(s,s)$, given by equation 8.11.12 in Ref. [52], namely,

$$\Gamma(s,s)=s^{s-1}e^{-s}\left[\sqrt{\frac{\pi}{2}}\sqrt{s}-\frac{1}{3}+O\left(\frac{1}{\sqrt{s}}\right)\right]. \quad (43)$$

To evaluate the moments, we need a closed-form expression for $\Gamma(s+1,s)$. Using equation 8.8.2 in Ref. [52], we obtain

$$\Gamma(s+1,s)=s\Gamma(s,s)+s^s e^{-s}, \quad (44)$$

where $\Gamma(s,s)$ is given by Eq. (43). Equipped with these expressions, we can now obtain asymptotic expansions for the moments in the limit of large $s$. More specifically, the mean distance on a random tree of size $s$ is given by

$$\mathbb{E}[L|S=s]=\sqrt{\frac{\pi}{2}}\sqrt{s}-\frac{4}{3}+O\left(\frac{1}{\sqrt{s}}\right). \quad (45)$$

It is found that the mean distance between random pairs of nodes that reside on a tree component of size $s$ scales like square root of $s$. Comparing the right-hand sides of Eqs. (27) and (45), which show the mode $\ell_{\mathrm{mode}}$ and the mean distance

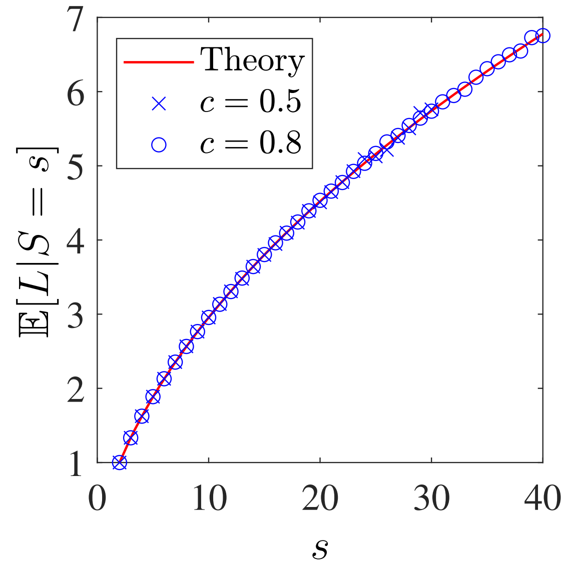


FIG. 5. Analytical results (solid line) for the mean distance $\mathbb{E}[L|S=s]$ between pairs of nodes that reside on the same tree component of size $s$, in a subcritical ER network, as a function of $s$. The analytical results are in very good agreement with the results obtained from computer simulations for subcritical ER networks of size $N=10^4$ and $c=0.5$ (×) and $c=0.8$ (○), which coincide with each other. Note that the simulation results for $c=0.5$ are shown only up to $s=30$, because in this case larger trees are rare.

$\mathbb{E}[L|S=s]$, respectively, it is found that while both of them scale like $\sqrt{s}$, the prefactor of the mean distance is larger than the prefactor of the mode. This implies that the distribution $P(L=\ell|S=s)$ is positively skewed. Interestingly, the scaling

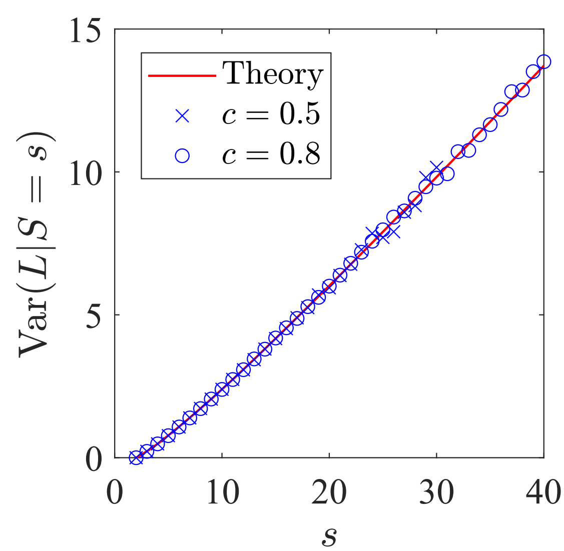


FIG. 6. The variance $\mathrm{Var}(L|S=s)$ of the DSPL between pairs of nodes that reside on the same tree component of size $s$, in a subcritical ER network, as a function of $s$. The analytical results are in very good agreement with the results obtained from computer simulations for subcritical ER networks of size $N=10^4$ and mean degree $c=0.5$ (×) and $c=0.8$ (○), which coincide with each other. Note that the simulation results for $c=0.5$ are shown only up to $s=30$, because in this case larger trees are rare.

of the mean distance, implied by Eq. (45), resembles the scaling of distances on two-dimensional lattices. It is in contrast with small-world random networks in which the mean distance scales like $\ln s$. This means that the tree components in subcritical ER networks are not small-world networks.

In Fig. 5 we present analytical results (solid line) for the mean distance $\mathbb{E}[L|S=s]$ between pairs of nodes that reside on the same tree component of size $s$, in a subcritical ER network, as a function of $s$. The analytical results are in very good agreement with the results obtained from computer simulations for subcritical ER networks of size $N=10^4$ with $c=0.5$ ($\times$) and $c=0.8$ ($\circ$), which coincide with each other. Note that the simulation results for $c=0.5$ are shown only up to $s=30$, because in this case larger trees are rare.

The second moment of the DSPL can be expressed by

$$\mathbb{E}[L^2|S=s] = 2s - 3\sqrt{\frac{\pi}{2}}\sqrt{s} + 2 + O\left(\frac{1}{\sqrt{s}}\right). \tag{46}$$

Combining the results presented above for the first and second moments, we obtain an asymptotic expression for the variance. It is given by

$$\mathrm{Var}(L|S=s) = \frac{4-\pi}{2}s - \sqrt{\frac{\pi}{18}}\sqrt{s} + O(1). \tag{47}$$

Thus, the standard deviation of the DSPL on trees of size $s$ scales like $\sqrt{s}$, namely, it scales like the mean distance $\mathbb{E}[L|S=s]$. Interestingly, the same qualitative relation is found in the DSPL of the whole subcritical ER network. This implies that $P(L=\ell|S=s)$ is relatively broad distribution, in contrast with the typical results for the DSPL of supercritical configuration model networks [39,40,44].

In Fig. 6 we present analytical results (solid line) for the variance $\mathrm{Var}(L|S=s)$ of the distribution of shortest path lengths between pairs of nodes that reside on the same tree component of size $s$, in a subcritical ER network, as a function of $s$. The analytical results are in very good agreement with the results obtained from computer simulations for subcritical ER networks of size $N=10^4$ and mean degree $c=0.5$ ($\times$) and $c=0.8$ ($\circ$), which coincide with each other. Note that the simulation results for $c=0.5$ are shown only up to $s=30$, because in this case larger trees are rare.

The cumulative mean distance between pairs of nodes that reside on a tree of size smaller or equal to $s$ is given by

$$\mathbb{E}[L|S\leqslant s] = \frac{\sum_{s'=2}^{s}\widehat{P}(S=s')\mathbb{E}[L|S=s']}{\sum_{s'=2}^{s}\widehat{P}(S=s')}. \tag{48}$$

To evaluate the right-hand side of Eq. (48), it is convenient to express the numerator and the denominator as differences between two infinite sums, namely,

$$\mathbb{E}[L|S\leqslant s] = \frac{\sum_{s'=2}^{\infty}\widehat{P}(S=s')\mathbb{E}[L|S=s'] - \sum_{s'=s+1}^{\infty}\widehat{P}(S=s')\mathbb{E}[L|S=s']}{\sum_{s'=2}^{\infty}\widehat{P}(S=s') - \sum_{s'=s+1}^{\infty}\widehat{P}(S=s')}. \tag{49}$$

The first term in the numerator amounts to $\mathbb{E}[L|L<\infty]$, which is given by Eq. (11), while the first term in the denominator is equal to 1 [due to the normalization of $\widehat{P}(S=s)$]. Equation (49) can thus be simplified to

$$\mathbb{E}[L|S\leqslant s] = \left(\frac{1}{1-c}\right)\frac{1-(1-c)\sum_{s'=s+1}^{\infty}\widehat{P}(S=s')\mathbb{E}[L|S=s']}{1-\sum_{s'=s+1}^{\infty}\widehat{P}(S=s')}. \tag{50}$$

Inserting $\widehat{P}(S=s)$ from Eq. (17) and $\mathbb{E}[L|S=s]$ from Eq. (45), which is accurate for sufficiently large $s$, into Eq. (50) and carrying out the summations, we obtain

$$\mathbb{E}[L|S\leqslant s] \simeq \left(\frac{1}{1-c}\right)\frac{1-\frac{(1-c)^2}{\sqrt{2\pi}c^2}(ce^{1-c})^{s+1}\left[\sqrt{\frac{\pi}{2}}\frac{1}{1-ce^{1-c}} - \frac{4}{3}\Phi\left(ce^{1-c},\frac{1}{2},s+1\right)\right]}{1-\frac{1-c}{\sqrt{2\pi}c^2}(ce^{1-c})^{s+1}\left[\Phi\left(ce^{1-c},\frac{1}{2},s+1\right)-\Phi\left(ce^{1-c},\frac{3}{2},s+1\right)\right]}, \tag{51}$$

where

$$\Phi(z,s,a) = \sum_{n=0}^{\infty}\frac{z^n}{(a+n)^s} \tag{52}$$

is the Lerch Phi transcendent [52]. Equation (51) is expected to be valid for large values of $s$.

In Fig. 7 we present analytical results (solid lines) for the mean distance $\mathbb{E}[L|S\leqslant s]$ between pairs of nodes that reside on the same tree component, for all tree components of size smaller or equal to $s$, in subcritical networks, as a function of the mean degree $c$. The results are presented for $s=10$, 20, 40, and 80 (from bottom to top). The analytical results, obtained from Eq. (51), are in very good agreement with the results obtained from computer simulations ($\circ$). As $s$ is increased, the mean distance $\mathbb{E}[L|S\leqslant s]$ converges towards the mean distance over the whole network, $\mathbb{E}[L|L<\infty]$ (dashed line), given by Eq. (11).

## IV. DISCUSSION

The ensemble of trees that appear in subcritical ER networks belong to the class of equilibrium trees [4]. These are trees that are formed by equilibrium processes. Their statistical properties can be analyzed using methods of equilibrium statistical mechanics. In this paper we calculated the DSPL of trees of a given size $s$ in subcritical ER networks. It was found that $P(L=\ell|S=s)$ is independent of the mean degree $c$ of the subcritical network from which these trees were extracted. It was also found that the mean distance on the ensemble of trees of size $s$ scales like $\mathbb{E}[L|S=s]\sim\sqrt{s}$. This scaling implies that the Hausdorff dimension of the trees is $D_H=2$, in agreement with earlier results obtained for other

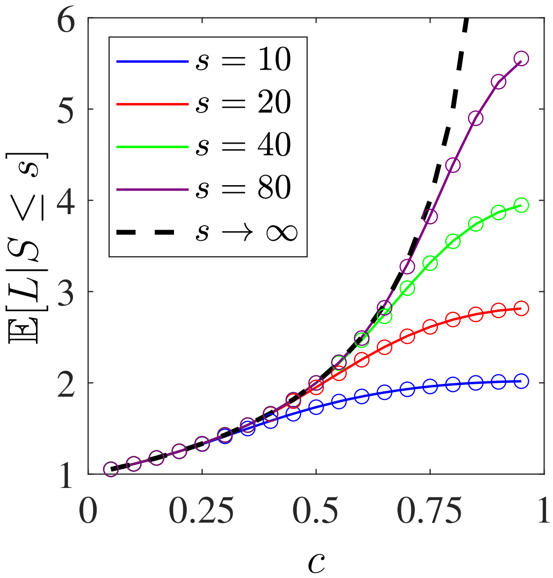


FIG. 7. Analytical results (solid lines), obtained from Eq. (51), for the mean distance $\mathbb{E}[L|S \leqslant s]$ between all pairs of nodes that reside on the same tree component, for all tree components of size smaller or equal to $s$, in subcritical networks, as a function of the mean degree $c$. The results are presented for $s = 10$, 20, 40, and 80 (from bottom to top). The analytical results are in very good agreement with the results obtained from computer simulations (○). As $s$ is increased, the mean distance $\mathbb{E}[L|S \leqslant s]$ converges towards the mean distance over the whole network, $\mathbb{E}[L|L < \infty]$ (dashed line), given by Eq. (11).

equilibrium trees [4]. It is in contrast with the scaling obtained in supercritical ER networks and other configuration model networks. In these networks the mean distance $\mathbb{E}[L]$ scales logarithmically with the network size $N$ and they are thus referred to as small-world networks.

Another important ensemble of trees consists of random recursive trees, which belong to the class of nonequilibrium trees. These trees grow via a kinetic process of node addition. The simplest model of random tree growth is the random attachment model. In this model, starting from a small seed network, at each time step a new node is added and is connected to one of the existing nodes uniformly at random. For simplicity we consider the case in which the seed network consists of a single node. Interestingly, the ensembles of equilibrium and nonequilibrium trees of size $s$ include the same set of tree configurations. However, their statistical properties are different due to the different weights assigned to each one of the possible configurations. In growing trees, the order in which the nodes are added is important. In particular, nodes that appeared early in the growth process are likely to gain more links than nodes that appeared at later stages [4].

The DSPL of the ensemble of random attachment trees of size $s$ was found to follow a Poisson distribution whose mean is given by $\mathbb{E}[L|S = s] = 2 \ln s$ [42]. This implies that the random attachment trees belong to the class of small-world networks, in which the mean distance scales logarithmically with the network size. These trees tend to form compact structures dominated by the nodes that appeared early in the growth process. This is in sharp contrast to the results obtained for the subcritical ER trees in which the mean distance scales like $\sqrt{s}$.

The methodology presented in this paper can be applied to the calculation of the distribution $P(L = \ell|S = s)$ in configuration model networks with various degree distributions $P(K = k)$, such as the exponential distribution and the power-law distribution. To this end, one needs to obtain the distribution $P(S = s)$ of tree sizes in the subcritical configuration model network under study and the DSPL of the whole network, $P(L = \ell|L < \infty)$ and to insert them into Eq. (13). The distribution $P(S = s)$ can be calculated using the generating function approach presented in Ref. [26]. The inversion of Eq. (13) to extract $P(L = \ell|S = s)$ is possible probably in those cases in which $P(L = \ell|S = s)$ is independent of the mean degree $c$. The validity of this condition will need to be tested on a case-by-case basis.

Apart from the DSPL there are other metric properties that characterize the large-scale structure of finite trees in subcritical configuration model networks. These include the distributions of eccentricities and diameters of trees of size $s$. The eccentricity is a property of a single node $i$ and it is equal to the largest distance between the given node $i$ and any other node in the tree. The diameter is a property of the whole tree and it is equal to the largest distance between any pair of nodes in the tree. The distribution of the largest diameter among all the trees in a subcritical ER network was recently studied [58,59]. It was found that this distribution follows a Gumbel distribution [60], which is one of the three distributions encountered in extreme-value theory.

The resistance distance between two nodes in a network is a measure of how difficult it is for electricity (or some other form of flow) to pass between these two nodes. In an unweighted network, the resistance distance is defined as the resistance between the two nodes, where the resistance of each edge is equal to 1 Ohm. The resistance distance can be thought of as a generalization of the concept of distance to networks, where the "distance" between two nodes is determined by the flow resistance between them rather than their physical separation. A more formal definition is given in Refs. [61,62], where it is also shown that it is a proper metric, satisfying for example the triangle inequality. In general, the resistance distance between two nodes will be smaller if there are more paths between the two nodes with lower resistance, and larger if there are fewer paths or if the paths have higher resistance. Random networks of resistors have been studied, mainly in two dimensions [63], and recently calculated for supercritical ER networks [64,65]. Interestingly, on tree graphs the shortest path between a pair of nodes $i$ and $j$ is in fact the only path between them. As a result, the resistance distance between $i$ and $j$ is equal to the shortest path length between them. This means that the results presented in this paper provide also the distribution of resistance distances in ER networks in the subcritical regime.

## V. SUMMARY

We calculated the distribution of shortest path lengths $P(L = \ell|S = s)$ between random pairs of nodes that reside on finite tree components of a given size $s$ in subcritical ER networks. It was found that $P(L = \ell|S = s) = \frac{\ell+1}{s^{\ell}} \frac{(s-2)!}{(s-\ell-1)!}$.

Surprisingly, this probability does not depend on the mean degree $c$ of the network from which these tree components were extracted. This is due to the fact that the ensemble of tree components of a given size $s$ in ER networks is sampled uniformly from the set of labeled trees of size $s$. The moments of the DSPL were also calculated. It was found that the mean distance between random pairs of nodes on tree components of size $s$ satisfies $\mathbb{E}[L|S=s] \sim \sqrt{s}$, unlike small-world networks in which the mean distance scales logarithmically with $s$.

### ACKNOWLEDGMENTS

This work was supported by the Israel Science Foundation Grant No. 1682/18.

### APPENDIX: THE GENERATING FUNCTION OF $P(S = s)$

The generating function of $P(S = s)$ is given by

$$H(u) = \sum_{s=1}^{\infty} u^s P(S = s). \tag{A1}$$

Inserting $P(S = s)$ from Eq. (2) into Eq. (A1), we obtain

$$H(u) = \frac{2}{2-c} \sum_{s=1}^{\infty} \frac{s^{s-2} c^{s-1} e^{-cs}}{s!} u^s. \tag{A2}$$

Rearranging terms on the right-hand side of Eq. (A2), we obtain

$$H(u) = -\frac{2}{c(2-c)} \sum_{s=1}^{\infty} \frac{1}{s} \frac{(-s)^{s-1}}{s!} (-uce^{-c})^s. \tag{A3}$$

Replacing the term $1/s$ on the right-hand side of Eq. (A3) by the integral expression

$$\frac{1}{s} = \int_0^{\infty} e^{-s\tau} d\tau, \tag{A4}$$

yields

$$H(u) = -\frac{2}{c(2-c)} \sum_{s=1}^{\infty} \int_0^{\infty} e^{-s\tau} d\tau \frac{(-s)^{s-1}}{s!} (-uce^{-c})^s. \tag{A5}$$

Exchanging the order of the sum and the integral on the right-hand side of Eq. (A5), we obtain

$$H(u) = -\frac{2}{c(2-c)} \int_0^{\infty} d\tau \sum_{s=1}^{\infty} \frac{(-s)^{s-1}}{s!} (-uce^{-c}e^{-\tau})^s. \tag{A6}$$

Using the series expansion of the Lambert $W$ function, which is given by

$$W(x) = \sum_{s=1}^{\infty} \frac{(-s)^{s-1} x^s}{s!}, \tag{A7}$$

we obtain

$$H(u) = -\frac{2}{c(2-c)} \int_0^{\infty} d\tau W(-uce^{-c}e^{-\tau}). \tag{A8}$$

Changing the integration variable from $\tau$ to $x = -uce^{-c}e^{-\tau}$, we obtain

$$H(u) = \frac{2}{c(2-c)} \int_{-uce^{-c}}^{0} W(x) \frac{dx}{x}. \tag{A9}$$

Changing the integration variable again, from $x$ to $y = W(x)$, which from the definition of the Lambert function implies that $x = ye^y$, we obtain

$$H(u) = \frac{2}{c(2-c)} \int_{W(-uce^{-c})}^{0} y\left(1 + \frac{1}{y}\right) dy. \tag{A10}$$

Carrying out the integration on the right-hand side of Eq. (A10), we obtain

$$H(u) = -\frac{1}{c(2-c)} \{[W(-uce^{-c})]^2 + 2W(-uce^{-c})\}. \tag{A11}$$

The moments of $P(S = s)$ can be obtained by taking suitable derivatives of $H(u)$. In particular, the mean tree size is

$$\langle S \rangle = \left. \frac{dH(u)}{du} \right|_{u=1} = \frac{2}{2-c}. \tag{A12}$$

and the second factorial moment is given by

$$\langle S(S-1) \rangle = \left. \frac{d^2 H(u)}{du^2} \right|_{u=1} = \frac{2c}{(1-c)(2-c)}. \tag{A13}$$

Using these results, it is found that the second moment of $P(S = s)$ is given by

$$\langle S^2 \rangle = \frac{2}{(1-c)(2-c)}, \tag{A14}$$

and the variance is given by

$$\mathrm{Var}(S) = \frac{2c}{(1-c)(2-c)^2}. \tag{A15}$$

It is also found that

$$\left\langle \binom{S}{2} \right\rangle = \frac{c}{(1-c)(2-c)}. \tag{A16}$$


[1] S. Havlin and R. Cohen, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, New York, 2010).

[2] E. Estrada, *The Structure of Complex Networks: Theory and Applications* (Oxford University Press, Oxford, 2011).

[3] M. E. J. Newman, *Networks*, 2nd ed. (Oxford University Press, Oxford, 2018).

[4] S. N. Dorogovtsev and J. F. F. Mendes, *The Nature of Complex Networks* (Oxford University Press, Oxford, 2022).

[5] E. Marinari and R. Monasson, Circuits in random graphs: From local trees to global loops, J. Stat. Mech.: Theory Exp. (2004) P09004.

[6] E. Marinari and G. Semerjian, On the number of circuits in random graphs, J. Stat. Mech.: Theory Exp. (2006) P06019.

[7] H. Bonneau, A. Hassid, O. Biham, R. Kühn, and E. Katzav, Distribution of shortest cycle lengths in random networks, Phys. Rev. E **96**, 062307 (2017).
[8] L. Tian, A. Bashan, D.-N. Shi, and Y.-Y. Liu, Articulation points in complex networks, Nat. Commun. **8**, 14223 (2017).
[9] I. Tishby, O. Biham, R. Kühn, and E. Katzav, Statistical analysis of articulation points in configuration model networks, Phys. Rev. E **98**, 062301 (2018).
[10] H. Bonneau, O. Biham, R. Kühn, and E. Katzav, Statistical analysis of edges and bredges in configuration model networks, Phys. Rev. E **102**, 012314 (2020).
[11] P. A. C. Duijn, V. Kashirin, and P. M. A. Sloot, The relative ineffectiveness of criminal network disruption, Sci. Rep. **4**, 4238 (2014).
[12] J. Shao, S. V. Buldyrev, R. Cohen, M. Kitsak, S. Havlin, and H. E. Stanley, Fractal boundaries of complex networks, Europhys. Lett. **84**, 48004 (2008).
[13] J. Shao, S. V. Buldyrev, L. A. Braunstein, S. Havlin, and H. E. Stanley, Structure of shells in complex networks, Phys. Rev. E **80**, 036105 (2009).
[14] P. Erdős and A. Rényi, On random graphs I, Math. Debrecen **6**, 290 (1959).
[15] P. Erdős and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).
[16] P. Erdős and A. Rényi, On the evolution of random graphs II, Bull. Inst. Int. Stat. **38**, 343 (1961).
[17] B. Bollobás, *Random Graphs*, 2nd ed. (Academic Press, London, 2001).
[18] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labeled regular graphs, Eur. J. Comb. **1**, 311 (1980).
[19] M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence, Random Struct. Algor. **6**, 161 (1995).
[20] M. Molloy and B. Reed, The size of the giant component of a random graph with a given degree sequence, Comb. Probab. Comput. **7**, 295 (1998).
[21] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distributions and their applications, Phys. Rev. E **64**, 026118 (2001).
[22] I. Tishby, O. Biham, and E. Katzav, Convergence towards an Erds-Rényi graph structure in network contraction processes, Phys. Rev. E **100**, 032314 (2019).
[23] I. Tishby, O. Biham, and E. Katzav, Analysis of the convergence of the degree distribution of contracting random networks towards a Poisson distribution using the relative entropy, Phys. Rev. E **101**, 062308 (2020).
[24] I. Tishby, O. Biham, E. Katzav, and R. Kühn, Revealing the micro-structure of the giant component in random graph ensembles, Phys. Rev. E **97**, 042318 (2018).
[25] R. Durrett, *Random Graph Dynamics* (Cambridge University Press, Cambridge, 2007).
[26] M. E. J. Newman, Component sizes in networks with arbitrary degree distributions, Phys. Rev. E **76**, 045101(R) (2007).
[27] E. Katzav, O. Biham, and A. K. Hartmann, Distribution of shortest path lengths in subcritical Erds-Rényi networks, Phys. Rev. E **98**, 012301 (2018).
[28] A. Cayley, A theorem on trees, Quart. J. Pure Appl. Math. **23**, 376 (1889).
[29] H. Prüfer, Neuer beweis eines satzes über permutationen, Arch. Math. Phys. **27**, 742 (1918).
[30] P. Steinbach, *Field Guide to Simple Graphs, Volume 3: The Book of Trees* (Design Lab, Albuquerque, 1990).
[31] R. A. Wright, B. Richmond, A. Odlyzko, and B. D. McKay, Constant time generation of free trees, SIAM J. Comput. **15**, 540 (1986).
[32] T. Beyer and S. M. Hedetniemi, Constant time generation of rooted trees, SIAM J. Comput. **9**, 706 (1980).
[33] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Metric structure of random networks, Nucl. Phys. B **653**, 307 (2003).
[34] R. van der Hofstad, G. Hooghiemstra, and P. van Mieghem, Distances in random graphs with finite variance degrees, Random Struct. Algor. **27**, 76 (2005).
[35] V. D. Blondel, J.-L. Guillaume, J. M. Hendrickx, and R. M. Jungers, Distance distribution in random graphs and application to network exploration, Phys. Rev. E **76**, 066101 (2007).
[36] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski, Distances in random graphs with finite mean and infinite variance degrees, Elect. J. Prob. **12**, 703 (2007).
[37] R. van der Hofstad and G. Hooghiemstra, Universality for distances in power-law random graphs, J. Math. Phys. **49**, 125209 (2008).
[38] H. van der Esker, R. van der Hofstad, and G. Hooghiemstra, Universality for the distance in finite variance random graphs, J. Stat. Phys. **133**, 169 (2008).
[39] E. Katzav, M. Nitzan, D. Ben-Avraham, P. L. Krapivsky, R. Kühn, N. Ross, and O. Biham, Analytical results for the distribution of shortest path lengths in random networks, Europhys. Lett. **111**, 26006 (2015).
[40] M. Nitzan, E. Katzav, R. Kühn, and O. Biham, Distance distribution in configuration-model networks, Phys. Rev. E **93**, 062309 (2016).
[41] S. Melnik and J. P. Gleeson, Simple and accurate analytical calculation of shortest path lengths, arXiv:1604.05521.
[42] C. Steinbock, O. Biham, and E. Katzav, Distribution of shortest path lengths in a class of node duplication network models, Phys. Rev. E **96**, 032301 (2017).
[43] C. Steinbock, O. Biham, and E. Katzav, Analytical results for the distribution of shortest path lengths in directed random networks that grow by node duplication, Eur. Phys. J. B **92**, 130 (2019).
[44] I. Tishby, O. Biham, R. Kühn, and E. Katzav, The mean and variance of the distribution of shortest path lengths of random regular graphs, J. Phys. A: Math. Theor. **55**, 265005 (2022).
[45] A. D. Jackson and S. P. Patil, Phases of small worlds: A mean field formulation, J. Stat. Phys. **189**, 40 (2022).
[46] B. Gompertz, 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, Philos. Trans. R. Soc. **115**, 513 (1825).
[47] F. Chung and L. Lu, The average distances in random graphs with given expected degrees, Proc. Natl. Acad. Sci. USA **99**, 15879 (2002).
[48] F. Chung and L. Lu, The average distance in a random graph with given expected degrees, Internet Math. **1**, 91 (2004).
[49] A. Fronczak, P. Fronczak, and J. A. Holyst, Average path length in random networks, Phys. Rev. E **70**, 056110 (2004).
[50] B. Bollobás, S. Janson, and O. Riordan, The phase transition in inhomogeneous random graphs, Random Struct. Algor. **31**, 3 (2007).

[51] R. Cohen and S. Havlin, Scale-free networks are ultrasmall, Phys. Rev. Lett. **90**, 058701 (2003).

[52] F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, *NIST Handbook of Mathematical Functions* (Cambridge University Press, Cambridge, 2010).

[53] I. M. Gessel, Lagrange inversion, J. Combin. Theory Ser. A **144**, 212 (2016).

[54] A. Meir and J. W. Moon, The distance between points in random trees, J. Comb. Theory **8**, 99 (1970).

[55] J. W. Moon, *Counting Labeled Trees* (Canadian Mathematical Congress, Ottawa, 1970).

[56] L. Takacs, On Cayley's formula for counting forests, J. Combin. Theory Ser. A **53**, 321 (1990).

[57] P. W. Shor, A new proof of Cayley's formula for counting labeled trees, J. Combin. Theory Ser. A **71**, 154 (1995).

[58] T. Łuczak, Random trees and random graphs, Random Struct. Algor. **13**, 485 (1998).

[59] A. K. Hartmann and M. Mézard, Distribution of diameters for Erds-Rényi random graphs, Phys. Rev. E **97**, 032128 (2018).

[60] E. J. Gumbel, Les valeurs extrêmes des distributions statistiques, Ann. Inst. Henri Poincaré **5**, 115 (1935).

[61] M. M. Deza and E. Deza, *Encyclopedia of Distances*, 4th ed. (Springer, Berlin, 2016).

[62] R. B. Bapat, *Graphs and Matrices*, 2nd ed. (Springer, London, 2014).

[63] B. Derrida and J. Vannimenus, A transfer-matrix approach to random resistor networks, J. Phys. A: Math. Gen. **15**, L557 (1982).

[64] P. Akara-pipattana, T. Chotibut, and O. Evnin, Resistance distance distribution in large sparse random graphs, J. Stat. Mech.: Theory Exp. (2022) 033404.

[65] P. Akara-pipattana and O. Evnin, Random matrices with row constraints and eigenvalue distributions of graph Laplacians, J. Phys. A: Math. Theor. **56**, 295001 (2023).