Distribution of shortest cycle lengths in random networks

Haggai Bonneau,[1] Aviv Hassid,[1] Ofer Biham,[1] Reimer Kühn,[2] and Eytan Katzav[1]

[1]*Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel*
[2]*Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom*

We present analytical results for the distribution of shortest cycle lengths (DSCL) in random networks. The approach is based on the relation between the DSCL and the distribution of shortest path lengths (DSPL). We apply this approach to configuration model networks, for which analytical results for the DSPL were obtained before. We first calculate the fraction of nodes in the network which reside on at least one cycle. Conditioning on being on a cycle, we provide the DSCL over ensembles of configuration model networks with degree distributions which follow a Poisson distribution (Erdős-Rényi network), degenerate distribution (random regular graph), and a power-law distribution (scale-free network). The mean and variance of the DSCL are calculated. The analytical results are found to be in very good agreement with the results of computer simulations.

I. INTRODUCTION

Network models provide a useful conceptual framework for the study of a large variety of systems and processes in science, technology, and society [1–4]. These models consist of nodes and edges, where the nodes represent physical objects, while the edges represent the interactions between them. Unlike regular lattices in which all the nodes have the same coordination number, network models are characterized by a degree distribution $P(K = k)$, $k = 0, 1, 2, \ldots$, with a mean degree denoted by $\langle K \rangle$. An important distinction is between networks which exhibit a narrow degree distribution (such as the Poisson distribution) and those which exhibit a broad degree distribution, which is typically a power-law distribution of the form $P(K = k) \sim k^{-\gamma}$. The latter networks are called scale-free networks. They exhibit some highly connected nodes, called hubs, which are essential for the integrity of these networks and play a dominant role in dynamical processes.

While pairs of adjacent nodes exhibit direct connections, the interactions between most pairs of nodes are mediated by intermediate nodes and edges. A pair of nodes, i and j, may be connected by many paths of different lengths. However, the distance, ℓ_{ij}, between nodes i and j is given by the length of the shortest path between them. The mean distance between all pairs of nodes in a network is denoted by $\langle L \rangle$. A central feature of random networks is the small-world property, namely the fact that the mean distance scales like $\langle L \rangle \sim \ln N$ where N is the network size [5–8]. Moreover, it was shown that scale-free networks may be ultrasmall depending on the exponent γ. In particular, for $2 < \gamma < 3$, their mean distance scales like $\langle L \rangle \sim \ln \ln N$ [9].

The distribution of shortest path lengths (DSPL) between all pairs of nodes in a network is a fundamental property of the network structure. The DSPL regulates the temporal evolution of dynamical processes on networks, such as signal propagation [10], navigation [11–13], and epidemic spreading [14,15]. Properties of the DSPL have been studied in different types of networks [16–23]. However, in spite of its importance it has not attracted nearly as much attention as the degree distribution.

Recently, an analytical approach was developed for calculating the DSPL [24] in the Erdős-Rényi (ER) network,

which is the simplest mathematical model of a random network [25–27]. The study of the DSPL was later extended to other network models [28–30]. Using recursion equations, analytical results for the DSPL were obtained in different regimes, including sparse and dense networks of small as well as asymptotically large sizes. The resulting distributions were found to be in good agreement with the results of computer simulations.

ER networks are random graphs which exhibit a Poisson degree distribution, with no degree-degree correlations between pairs of adjacent nodes. In fact, ER networks can be considered as a maximum entropy ensemble, under the constraint that the mean degree is fixed. Moreover, the broader class of configuration model networks generates maximum entropy ensembles under conditions in which the entire degree distribution is constrained [2,16,31–33]. For any given degree distribution, one can produce an ensemble of configuration model networks and perform a statistical analysis of its properties. Therefore, the configuration model provides a powerful platform for the analysis of random networks. It is the ideal model to use as a null model when one tries to analyze an empirical network of which the degree distribution is known. To this end, one constructs configuration model networks of the same size and the same degree distribution as the empirical network. Properties of interest such as the DSPL [34], the betweenness centrality [35], and the abundance of network motifs [36–38] are compared between the two networks. The discrepancies provide a rigorous test of the systematic features of the empirical network versus the corresponding ensemble of random networks.

In addition to open paths between pairs of distinct nodes, networks may exhibit cycles, namely closed paths which return to their initial nodes. The length of a cycle is given by the number of edges (or nodes) which reside along the cycle. The shortest possible cycle is the triangle, of length $\ell = 3$. The longest possible cycle is a Hamiltonian cycle of length $\ell = N$. Some nodes in a network may not reside on any cycle. Other nodes may reside on one or more cycles. In the latter case, the shortest among these cycles is of particular importance. The shortest cycle on which a given node resides provides the shortest feedback loop for signals originated from that node and the strongest correlations between signals reaching

the node via different links. Therefore, the distribution of shortest cycle lengths (DSCL) provides useful information on chemical networks [39], biological networks [40], feedback processes [41], oscillations [42–44], and synchronization [4] in complex networks, as well as for ranking of nodes [45,46]. Moreover, the partition functions of statistical physics models on networks can be expressed in terms of the combinatorial properties of the cycles, using high temperature expansions and low temperature expansions [47].

An important class of networks consists of tree networks, in which any pair of nodes is connected by a single path. Thus, in tree networks the shortest path between any pair of nodes is the only path between them and there are no cycles. Tree structures appear in the dilute limit of random networks such as the ER network and the configuration model network, below the percolation transition. Above the percolation transition long cycles start to emerge in the giant cluster. As the network becomes more strongly connected, the size of the giant cluster increases and the cycles become more numerous and shorter.

In this paper we present analytical results for the DSCL in configuration model networks. We first calculate the probability that a random node resides on at least one cycle. We then calculate the DSCL for all the nodes which reside on at least one cycle. We apply this approach to networks with Poisson, degenerate, and power-law degree distributions. It is found that the analytical results are in very good agreement with numerical simulations. Using the tail-sum formula we calculate the mean and the variance of the DSCL for these networks.

The paper is organized as follows. In Sec. II we present the configuration model. In Sec. III we consider the percolation transition and the giant cluster in configuration model networks. In Sec. IV we consider properties of the DSPL to be used in the calculation of the DSCL. In Sec. V we present analytical results for the fraction of nodes which reside on at least one cycle. In Sec. VI we present analytical results for the DSCL of configuration model networks, expressed in terms of the degree distributions and the DSPL. In Sec. VII we apply these results to ER networks, regular graphs, and scale-free networks. The results are discussed in Sec. VIII and summarized in Sec. IX. In Appendix A we present the short-distance behavior of the DSPL between pairs of nodes of given degrees. In Appendices B, D, and E we summarize the properties of the giant clusters in ER networks, random regular graphs, and scale-free networks, respectively. In Appendix C we provide some explicit expressions for the probabilities that random nodes of given degrees reside on at least one cycle.

II. THE CONFIGURATION MODEL

The configuration model is a maximum entropy ensemble of networks under the condition that the degree distribution is imposed [2,16]. Here we focus on the case of undirected networks, in which all the edges are bidirectional. To construct such a network of N nodes, one can draw the degrees of all nodes from a desired degree distribution, $P(K = k)$, producing a degree sequence of the form $\{k_i\}_{i=1,\dots,N}$ (where $\sum k_i$ must be even). The mean degree over the ensemble of networks is $\langle K \rangle = \sum_k k P(K = k)$. For brevity, in the rest of the paper we use a more compact notation, in which $P(K = k)$

is replaced by $P(k)$, except for a few places in which the more detailed notation is needed for clarity.

A convenient way to construct a configuration model network is to prepare the N nodes such that each node, i, is connected to k_i half edges [2]. Pairs of half edges from different nodes are then chosen randomly and are connected to each other in order to form the network. The result is a network with the desired degree sequence but no correlations. Note that towards the end of the construction the process may get stuck. This may happen in the case in which the only remaining pairs of half edges belong to the same node or to pairs of nodes which are already connected to each other. In such cases one may perform some random reconnections in order to enable completion of the construction.

III. THE PERCOLATION TRANSITION AND THE GIANT CLUSTER

Configuration model networks generically consist of many connected components. In some cases the size of the largest component scales linearly with the network size, N. In such cases, the largest component is called a giant cluster. All the other components are nonextensive and are called finite or isolated components, and below are referred to as nongiant components. The size of the giant cluster is determined by the degree distribution, $P(k)$. Some families of degree distributions can be parametrized such that in a certain range of parameters there is no giant cluster, while in the complementary range there is a giant cluster. On the boundary between these two domains in the parameter space there is a phase transition, which is referred to as a percolation transition.

Consider a configuration model network of N nodes with a given degree distribution $P(k)$. In this paper we will employ two different sampling procedures. The degrees of nodes which are sampled randomly from the network follow the overall degree distribution $P(k)$. However, nodes which are sampled as random neighbors of random nodes follow a modified degree distribution, which takes the form

$$\widetilde{P}(k) = \frac{k}{\langle K \rangle} P(k). \tag{1}$$

This is due to the fact that such nodes are selected proportionally to their degrees. Each one of these degree distributions has a generating function associated with it. The generating function of $P(k)$ is

$$G_0(x) = \sum_{k=0}^{\infty} P(k) x^k, \tag{2}$$

while the generating function of $\widetilde{P}(k)$ is

$$G_1(x) = \sum_{k=0}^{\infty} \widetilde{P}(k) x^{k-1}. \tag{3}$$

From the definitions of $G_0(x)$ and $G_1(x)$ in Eqs. (2) and (3), respectively, we find that $G_0(1) = 1$ and $G_1(1) = 1$. In some networks there are no isolated nodes (of degree $k = 0$) and no leaf nodes (of degree $k = 1$). In such networks $P(k) > 0$ only for $k \geqslant 2$. For these networks we find that $G_0(0) = 0$ and $G_1(0) = 0$. This implies that in such networks both $x = 0$ and $x = 1$ are fixed points of both $G_0(x)$ and $G_1(x)$.

In what follows we review the well known analysis of the percolation probability in configuration model networks, following Refs. [1,2]. Our main motivation for doing so is that it allows us to highlight two lesser known facts about the problem, which we will need in our evaluation of the DSCL below. These concern the degree-dependent probabilities of randomly chosen nodes and randomly chosen neighbors of randomly chosen nodes to belong to the giant cluster. The probability that a random node resides on the giant cluster is denoted by g. In the case in which a giant cluster exists, $g > 0$, while in the case in which there is no giant cluster, $g = 0$. To obtain the probability g, one needs to first calculate the probability $\tilde{g}$ that a random neighbor of a random node, i, belongs to the giant cluster in the reduced network, which does not include the node i. In the thermodynamic limit, $N \to \infty$, the probability $\tilde{g}$ is given as a solution of the self-consistency equation [1]

$$1 - \tilde{g} = G_1(1 - \tilde{g}). \qquad (4)$$

The left hand side of this equation is the probability that a random neighbor of a random node does not reside on the giant cluster. The right hand side represents the same quantity in terms of its neighbors, namely as the probability that none of the neighbors of such node resides on the giant cluster. Once $\tilde{g}$ is known, the probability g can be obtained from

$$g = 1 - G_0(1 - \tilde{g}). \qquad (5)$$

This relation is based on the same consideration as Eq. (4), where the difference is that the reference node is a random node rather than a random neighbor of a random node.

Below we consider the more specific case of nodes of a given degree. The probability that a random node of a given degree, k, resides on the giant cluster is denoted by g_k. Using the degree distribution, $P(k)$, the probability, g, that a random node of an unspecified degree resides on the giant cluster can be expressed in terms of g_k by

$$g = \sum_{k=0}^{\infty} P(k) g_k. \qquad (6)$$

Such a node resides on the giant cluster if at least one of its k neighbors resides on the giant cluster. Therefore,

$$g_k = 1 - (1 - \tilde{g})^k. \qquad (7)$$

Thus, high degree nodes are more likely to reside on the giant cluster than low degree nodes. Similarly, the probability $\tilde{g}$ that a random neighbor of a random node resides on the giant cluster can be expressed in the form

$$\tilde{g} = \sum_{k=0}^{\infty} \widetilde{P}(k) \tilde{g}_k, \qquad (8)$$

where $\tilde{g}_k$ is the probability that a random neighbor of a random node resides on the giant cluster, under the condition that its degree is k. Using similar considerations, we find that the probability $\tilde{g}_k$ is given by

$$\tilde{g}_k = 1 - (1 - \tilde{g})^{k-1}. \qquad (9)$$

In Appendices B, D, and E we apply these considerations to the analysis of the giant clusters in ER networks, random regular graphs, and scale-free networks, respectively.

IV. THE DISTRIBUTION OF SHORTEST PATH LENGTHS

Consider a pair of random nodes, i and j, in a network of N nodes. Assuming that the two nodes reside on the same connected component, they may be connected to each other by a large number of paths. The distance between the two nodes is equal to the length of the shortest among these paths (possibly more than one). Below we briefly review the approach introduced in Ref. [28] for the calculation of the DSPL in configuration model networks of a given size, N, and a given degree distribution, $P(k)$. The DSPL can be expressed in the form of a tail distribution, where $P_{\mathrm{PL}}(L > \ell)$ is the probability that the shortest path length between a random pair of nodes is larger than ℓ. The tail distribution can be expressed as a product of the form

$$P_{\mathrm{PL}}(L > \ell) = \prod_{\ell'=1}^{\ell} P_{\mathrm{PL}}(L > \ell' | L > \ell' - 1), \qquad (10)$$

where $P_{\mathrm{PL}}(L > \ell | L > \ell - 1)$ is the conditional probability that the distance between a random pair of nodes is larger than ℓ conditioned on it being larger than $\ell - 1$. In the analysis below we use different types of tail distributions for the DSPL. In Table I we summarize these distributions and list the equations from which each one of them can be evaluated.

A path of length ℓ from node i to node j can be decomposed into a single edge connecting node i and node $r \in \partial_i$ (where ∂_i is the set of all nodes directly connected to i), and a shorter path of length $\ell - 1$ connecting r and j. Thus, the existence of a path of length ℓ between nodes i and j can be ruled out if there is no path of length $\ell - 1$ between any of the nodes $r \in \partial_i$ and j. For sufficiently large networks, the argument presented above translates into the recursion equation [28]

$$P_{\mathrm{PL}}(L > \ell | L > \ell - 1) = G_0[\widetilde{P}_{\mathrm{PL}}(L > \ell - 1 | L > \ell - 2)], \qquad (11)$$

where the generating function $G_0(x)$ is given by Eq. (2). Here we distinguish between the conditional probability $P_{\mathrm{PL}}(L > \ell | L > \ell - 1)$ between nodes i and j and the probability $\widetilde{P}_{\mathrm{PL}}(L > \ell | L > \ell - 1)$ between a node $r \in \partial_i$ and node j, on the reduced network from which node i was removed. The reason for this distinction is that the former probability involves two random nodes, while the latter probability involves a node, r, which is a random neighbor of a random node, and a random node, j. The conditional probability $\widetilde{P}_{\mathrm{PL}}(L > \ell | L > \ell - 1)$ satisfies the recursion equation

$$\widetilde{P}_{\mathrm{PL}}(L > \ell | L > \ell - 1) = G_1[\widetilde{P}_{\mathrm{PL}}(L > \ell - 1 | L > \ell - 2)], \qquad (12)$$

where $G_1(x)$ is given by Eq. (3), which is valid for $\ell \geqslant 2$.

The case of $\ell = 1$ deserves special attention. On a network of size N (sufficiently large), the probability that two random nodes are not connected is given by [28]

$$P_{\mathrm{PL}}(L > 1 | L > 0) \simeq 1 - \frac{\langle K \rangle}{N - 1} + O\left(\frac{1}{N^2}\right), \qquad (13)$$

TABLE I. The different tail distributions for the DSPL and DSCL and the equations from which each one of them can be evaluated

Distribution	Equation	Description
$P_{\mathrm{PL}}(L > \ell \| L > \ell - 1)$	Eq. (11)	Conditional DSPL between pairs of random nodes[a]
$\widetilde{P}_{\mathrm{PL}}(L > \ell \| L > \ell - 1)$	Eq. (12)	Conditional DSPL between random nodes and RNRNs[a,b]
$\widehat{P}_{\mathrm{PL}}(L > \ell \| L > \ell - 1)$	Eq. (18)	Conditional DSPL between pairs of RNRNs[a]
$P_{\mathrm{PL}}(L > \ell)$	Eq. (10)	DSPL between pairs of random nodes[a]
$\widetilde{P}_{\mathrm{PL}}(L > \ell)$	Eq. (15)	DSPL between random nodes and RNRNs[a]
$\widehat{P}_{\mathrm{PL}}(L > \ell)$	Eq. (19)	DSPL between pairs of RNRNs[a]
$Q_{\mathrm{PL}}(L > \ell)$	Eq. (20)	Overall DSPL between pairs of random nodes
$\widetilde{Q}_{\mathrm{PL}}(L > \ell)$	Eq. (21)	Overall DSPL between random nodes and RNRNs
$\widehat{Q}_{\mathrm{PL}}(L > \ell)$	Eq. (22)	Overall DSPL between pairs of RNRNs
$P_{\mathrm{PL}}(L > \ell \| k, k')$	Eq. (23)	DSPL between pairs of random nodes of degrees k and k'[a]
$\widetilde{P}_{\mathrm{PL}}(L > \ell \| k, k')$	Eq. (24)	DSPL between random nodes and RNRNs of degrees k and k'[a]
$\widehat{P}_{\mathrm{PL}}(L > \ell \| k, k')$	Eq. (25)	DSPL between pairs of RNRNs of degrees k and k'[a]
$Q_{\mathrm{PL}}(L > \ell \| k, k')$	Eq. (26)	Overall DSPL between pairs of random nodes of degrees k and k'
$\widetilde{Q}_{\mathrm{PL}}(L > \ell \| k, k')$	Eq. (27)	Overall DSPL between random nodes and RNRNs of degrees k and k'
$\widehat{Q}_{\mathrm{PL}}(L > \ell \| k, k')$	Eq. (28)	Overall DSPL between pairs of RNRNs of degrees k and k'
$P_{\mathrm{CL}}(L > \ell \| k)$	Eq. (35)	DSCL of nodes of degree k
$P_{\mathrm{CL}}(L > \ell)$	Eq. (36)	DSCL

[a]For pairs of nodes which reside on the same connected component.
[b]RNRNs: random neighbors of random nodes.

while the probability that a random neighbor of a random node and a random node are not connected is given by

$$\widetilde{P}_{\mathrm{PL}}(L > 1 | L > 0) \simeq 1 - \frac{\langle K^2 \rangle - \langle K \rangle}{\langle K \rangle (N - 1)} + O\left(\frac{1}{N^2}\right). \quad (14)$$

The difference between Eqs. (13) and (14) is due to the fact that the degree distribution $\widetilde{P}(k)$ of random neighbors of random nodes, given by Eq. (1), is generically distinct from the degree distribution $P(k)$ of random nodes.

Actually, there are two other types of DSPLs in random networks, which are needed for the analysis of shortest cycles. One of them is the DSPL between a random node and a random neighbor of a random node, denoted by $\widetilde{P}_{\mathrm{PL}}(L > \ell)$. The other one is the DSPL between two random neighbors of random nodes, denoted by $\widehat{P}_{\mathrm{PL}}(L > \ell)$. The DSPL between a random node and a random neighbor of a random node is expressed as a product of the form

$$\widetilde{P}_{\mathrm{PL}}(L > \ell) = \prod_{\ell'=1}^{\ell} \widetilde{P}_{\mathrm{PL}}(L > \ell' | L > \ell' - 1), \quad (15)$$

where $\widetilde{P}_{\mathrm{PL}}(L > \ell' | L > \ell' - 1)$ is obtained by iterating Eq. (12), using Eq. (14) as an initial condition.

The DSPL between two random neighbors of random nodes, $\widehat{P}_{\mathrm{PL}}(L > \ell)$, requires careful attention. The initial condition in this case, namely the probability that two such nodes are not connected on a network of size N, is

$$\widehat{P}_{\mathrm{PL}}(L > 1 | L > 0)$$
$$= \sum_{k=0}^{\infty} \widetilde{P}(k) \sum_{k'=0}^{\infty} \widetilde{P}(k') \left[1 - \frac{k' - 1}{(N - 1)\langle K \rangle} \right]^{k-1}. \quad (16)$$

Using a binomial approximation and performing the summations, we obtain

$$\widehat{P}_{\mathrm{PL}}(L > 1 | L > 0)$$
$$= 1 - \frac{\langle K \rangle}{N - 1} \left(\frac{\langle K^2 \rangle - \langle K \rangle}{\langle K \rangle^2} \right)^2 + O\left(\frac{1}{N^2}\right). \quad (17)$$

This initial condition is fed into the recursion equation

$$\widehat{P}_{\mathrm{PL}}(L > \ell | L > \ell - 1) = G_1[\widehat{P}_{\mathrm{PL}}(L > \ell - 1 | L > \ell - 2)]. \quad (18)$$

The DSPL between random neighbors of random nodes is then obtained as a product of the conditional probabilities:

$$\widehat{P}_{\mathrm{PL}}(L > \ell) = \prod_{\ell'=1}^{\ell} \widehat{P}_{\mathrm{PL}}(L > \ell' | L > \ell' - 1). \quad (19)$$

In the analysis above, we considered only pairs of nodes which reside on the same cluster. Since not all pairs of random nodes reside on the same cluster, the DSPL needs to be adjusted. Taking a random pair of nodes, i and j, the probability that they reside on the same cluster is negligible, unless they both reside on the giant cluster. The probability that both nodes reside on the giant cluster is g^2. Therefore, the probability that the distance between them is infinite is $P_{\mathrm{PL}}(L = \infty) = 1 - g^2$. This implies that the DSPL between all pairs of nodes in the network (without assuming that they reside on the same cluster) is

$$Q_{\mathrm{PL}}(L > \ell) = g^2 P_{\mathrm{PL}}(L > \ell) + (1 - g^2). \quad (20)$$

Using a similar argument for the DSPL between a random node and a random neighbor of a random node, we find that the DSPL between all such pairs is given by

$$\widetilde{Q}_{\mathrm{PL}}(L > \ell) = g \tilde{g} \widetilde{P}_{\mathrm{PL}}(L > \ell) + (1 - g \tilde{g}). \quad (21)$$

Similarly, the DSPL between all pairs of random neighbors of random nodes is given by

$$\widehat{Q}_{\mathrm{PL}}(L > \ell) = \tilde{g}^2 \widehat{P}_{\mathrm{PL}}(L > \ell) + (1 - \tilde{g}^2). \qquad (22)$$

In cases where $g < 1$, the overall DSPLs, $Q(L > \ell)$, $\widetilde{Q}(L > \ell)$, and $\widehat{Q}(L > \ell)$, approach a nonzero asymptotic value at large ℓ, unlike the original DSPLs, $P(L > \ell)$, $\widetilde{P}(L > \ell)$, and $\widehat{P}(L > \ell)$, which decay to zero.

To obtain the DSPL between random pairs of nodes of known degrees, consider two random nodes, i and j, of degrees k and k', respectively, which do not share any neighbors and thus the distance between them satisfies $\ell > 2$. Since node i has k neighbors and node j has k' neighbors, the probability that the distance between them is longer than ℓ is equal to the probability that the distance between any neighbor of i to any neighbor of j is longer than $\ell - 2$. Therefore,

$$P_{\mathrm{PL}}(L > \ell|k,k') = [\widehat{P}_{\mathrm{PL}}(L > \ell - 2)]^{kk'}, \qquad (23)$$

where $\widehat{P}_{\mathrm{PL}}(L > \ell)$ is the DSPL between two random neighbors of random nodes, given by Eq. (19). Similarly, the DSPL between a random node, of degree k, and a random neighbor of a random node, of degree k', is given by

$$\widetilde{P}_{\mathrm{PL}}(L > \ell|k,k') = [\widehat{P}_{\mathrm{PL}}(L > \ell - 2)]^{k(k'-1)}. \qquad (24)$$

The DSPL between pairs of random neighbors of random nodes, under the condition that their degrees are k and k', is given by

$$\widehat{P}_{\mathrm{PL}}(L > \ell|k,k') = [\widehat{P}_{\mathrm{PL}}(L > \ell - 2)]^{(k-1)(k'-1)}. \qquad (25)$$

It is important to note that Eqs. (23)–(25) are valid for $\ell > 2$. The corresponding equations for the conditional probabilities $P_{\mathrm{PL}}(L > \ell|k,k')$, $\widetilde{P}_{\mathrm{PL}}(L > \ell|k,k')$, and $\widehat{P}_{\mathrm{PL}}(L > \ell|k,k')$ with $\ell = 1,2$ are presented in Appendix A.

Using the results presented above we now provide the overall DSPLs, between random pairs of nodes of known degrees. Considering two random nodes of degrees k and k', we obtain

$$Q_{\mathrm{PL}}(L > \ell|k,k') = g_k g_{k'} P_{\mathrm{PL}}(L > \ell|k,k') + (1 - g_k g_{k'}), \qquad (26)$$

where g_k is given by Eq. (7). Similarly, the DSPL between a random node of degree k and a random neighbor of a random node, of degree k', is given by

$$\widetilde{Q}_{\mathrm{PL}}(L > \ell|k,k') = g_k \tilde{g}_{k'} \widetilde{P}_{\mathrm{PL}}(L > \ell|k,k') + (1 - g_k \tilde{g}_{k'}), \qquad (27)$$

where $\tilde{g}_{k'}$ is given by Eq. (9). Lastly, the DSPL between pairs of random neighbors of random nodes, conditioned on their degrees, k and k', is given by

$$\widehat{Q}_{\mathrm{PL}}(L > \ell|k,k') = \tilde{g}_k \tilde{g}_{k'} \widehat{P}_{\mathrm{PL}}(L > \ell|k,k') + (1 - \tilde{g}_k \tilde{g}_{k'}). \qquad (28)$$

The moments of $P_{\mathrm{PL}}(L > \ell)$ provide useful information about the network. The nth moment, $\langle L^n \rangle_{\mathrm{PL}}$, can be obtained using the tail-sum formula [48]

$$\langle L^n \rangle_{\mathrm{PL}} = \sum_{\ell=0}^{N-2} [(\ell + 1)^n - \ell^n] P_{\mathrm{PL}}(L > \ell). \qquad (29)$$

Note that the sum in Eq. (29) does not extend to ∞ because the longest possible shortest path in a network of size N is $N - 1$. The mean distance in configuration model networks has been studied extensively [7,8,16,19,21]. It was found that

$$\langle L \rangle_{\mathrm{PL}} \simeq \frac{\ln N}{\ln \left(\frac{\langle K^2 \rangle - \langle K \rangle}{\langle K \rangle} \right)} + O(1). \qquad (30)$$

The width of the distribution can be characterized by the variance $\sigma_{\mathrm{PL}}^2 = \langle L^2 \rangle_{\mathrm{PL}} - \langle L \rangle_{\mathrm{PL}}^2$.

V. THE FRACTION OF NODES WHICH RESIDE ON AT LEAST ONE CYCLE

In this section we calculate the probability $P(i \in \mathrm{cycle})$ that a random node, i, resides on at least one cycle. To do so, we first calculate the conditional probability, $P(i \in \mathrm{cycle}|k)$, that a node of a given degree, k, resides on at least one cycle. Actually, this probability can be expressed by $P(i \in \mathrm{cycle}|k) = 1 - P(i \notin \mathrm{cycle}|k)$. Clearly, nodes of degree $k = 0$ or 1 cannot reside on any cycle and thus

$$P(i \notin \mathrm{cycle}|0) = P(i \notin \mathrm{cycle}|1) = 1. \qquad (31)$$

For a node of degree $k \geqslant 2$ to reside on a cycle, two of its neighbors must be connected to each other on the reduced network from which i is removed. The probability that a neighbor of a random node i, on the reduced network from which i is removed, is part of the giant cluster of the reduced network is equal to $\tilde{g}$. The probability that a given pair of neighbors will reside on the giant cluster of the reduced network is $\tilde{g}^2$. This pair of neighbors will reside on the same component only if this component is the giant cluster (up to negligible probability). Hence, the probability that a given pair of neighbors of i is not connected is $1 - \tilde{g}^2$. Since there are $\binom{k}{2}$ pairs of neighbors of node i, the probability that none of these pairs are connected on the reduced network from which node i is removed is

$$P(i \notin \mathrm{cycle}|k) = (1 - \tilde{g}^2)^{\binom{k}{2}}. \qquad (32)$$

Note that this result is based on the assumption that the paths between different pairs of neighbors of i are independent. This assumption is expected to hold in ensembles of uncorrelated networks, such as the configuration model or any other network model in which the clustering coefficient is small.

Using the arguments discussed above we find that the probability that a random node of unspecified degree resides on at least one cycle is given by

$$P(i \in \mathrm{cycle}) = 1 - \left[P(K = 0) + P(K = 1) \right.$$
$$\left. + \sum_{k=2}^{\infty} P(k) P(i \notin \mathrm{cycle}|k) \right], \qquad (33)$$

where $P(i \notin \mathrm{cycle}|k)$ is given by Eq. (32). Note that in the case of $\tilde{g} = 0$ one can show, using Eq. (5), that $g = 0$ as well, meaning that there is no giant cluster. Equation (33) shows that under these conditions $P(i \in \mathrm{cycle}) = 0$. This reflects the fact that for a network below the percolation threshold, in the thermodynamic limit, the number of cycles does not scale with N [49,50]. Thus, essentially all the components are trees.

One should point out that Eq. (32) does not take into account certain correlations between pairs of neighbors of node i. To demonstrate this point, consider the k neighbors of node i. We will denote their degrees by $k_1, k_2, \ldots, k_k$. These degrees are independent of each other and are all drawn from the same distribution, $\widetilde{P}(k)$. However, the probability that a node, r_m, resides on the giant cluster depends on its degree, k_m, and is given by g_{k_m} [Eq. (7)]. Since each neighbor of i participates in $k-1$ such pairs, the probabilities that different pairs reside on the giant cluster are not independent. Each one of these nodes may connect to each of the other $k-1$ neighbors, with a probability which depends on its degree. Therefore, these $k-1$ probabilities are not independent, unlike the assumption made in Eq. (32). To account for these correlations we express $P(i \notin \text{cycle}|k)$ in the form

$$P(i \notin \text{cycle}|k) = \sum_{k_1,k_2,\ldots,k_k} \prod_{r=1}^{k} \widetilde{P}(k_r) \prod_{m<n} \left(1 - \tilde{g}_{k_m}\tilde{g}_{k_n}\right), \quad (34)$$

where the product runs over all pairs of neighbors of node i.

In summary, we have presented two approaches to the calculation of $P(i \notin \text{cycle})$. The simpler approach of Eq. (32) provides a good approximation in most cases. For highly heterogeneous networks one may need the more detailed approach of Eq. (34), which is much more elaborate to implement. More specifically, it requires summation over all possible degree sequences of length k, which becomes prohibitive when k is large.

VI. THE DISTRIBUTION OF SHORTEST CYCLE LENGTHS

Consider a random node, i, in a configuration model network of size N with degree distribution $P(k)$. A node of degree $K \geqslant 2$ may reside on one or more cycles. Here we focus on the shortest among these cycles. More specifically, we calculate the distribution of lengths of the shortest cycles on which a random node of degree k resides. We denote the neighbors of node i by $r_1, r_2, \ldots, r_k$. A cycle of length ℓ on which i resides consists of the edges connecting i to two of its neighbors, r_m and r_n, and a path of length $\ell - 2$ connecting r_m and r_n. The number of possible shortest cycles is $\binom{k}{2}$, namely the number of pairs of neighbors of i. In Fig. 1 we present an illustration of the cycles on which a random reference node (black filled circle) resides. This node has $k = 4$ neighbors (empty circles). The edges between the reference node and its neighbors are shown by dashed lines. The paths connecting pairs of neighbors are shown by solid lines. The shortest among these paths is shown by a thick solid line (blue) of length 2; thus the shortest cycle on which the reference node resides is of length $\ell = 4$. The other paths, of lengths 3 and 4, are shown by thin solid lines (red).

The tail distribution of the lengths of shortest cycles on which random nodes of degree k reside is denoted by $P_{\text{CL}}(L > \ell|K = k)$. In order that the shortest cycle will be longer than ℓ, the distances between all pairs of neighbors must satisfy $L > \ell - 2$. Therefore

$$P_{\text{CL}}(L > \ell|k) = \widehat{Q}_{\text{PL}}(L > \ell - 2)^{\binom{k}{2}}, \quad (35)$$
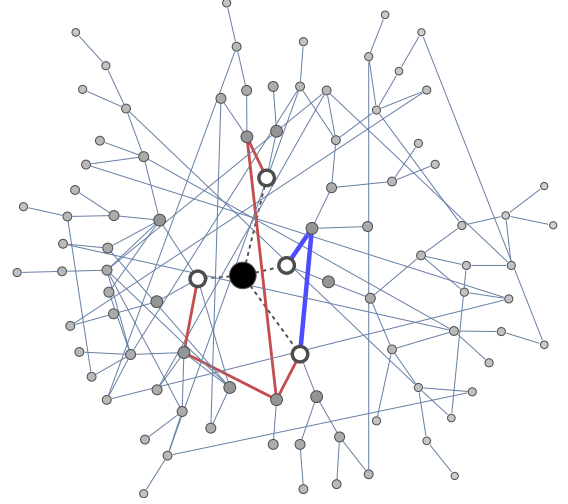

FIG. 1. Illustration of the cycles on which a random reference node (black filled circle) resides. The reference node has 4 neighbors (empty circles). The edges connecting the reference node to its neighbors are shown by dashed lines. The paths connecting pairs of neighbors are shown by solid lines. The shortest path, shown by a thick solid line (blue) is of length 2; thus the shortest cycle on which the reference node resides is of length $\ell = 4$. The other paths between neighbors of the reference node, which are of lengths 3 and 4, are shown by narrower solid lines (red). They form cycles of lengths 5 and 6, respectively.

where $\widehat{Q}_{\text{PL}}(L > \ell)$ is given by Eq. (22). This equation is based on the assumption that the distances between all pairs of neighbors of node i are independent of each other. This assumption is expected to be satisfied in configuration model networks. Note that nodes of degrees $k = 0$ and 1 do not reside on any cycle, and thus $P_{\text{CL}}(L > \ell|K = 0) = P_{\text{CL}}(L > \ell|K = 1) = 1$ for any value of ℓ.

For a random node, i, of unknown degree, the DSCL is obtained by averaging over all possible degrees according to

$$P_{\text{CL}}(L > \ell) = \sum_{k=0}^{\infty} P(k) P_{\text{CL}}(L > \ell|k). \quad (36)$$

Writing this equation in a more explicit form, we obtain

$$P_{\text{CL}}(L > \ell) = P(K = 0) + P(K = 1)$$
$$+ \sum_{k=2}^{\infty} P(k) \widehat{Q}_{\text{PL}}(L > \ell - 2)^{\binom{k}{2}}. \quad (37)$$

This equation is expected to provide an accurate description of the DSCL of configuration model networks, when the degree distribution is not too broad. The corresponding probability distribution function, $P_{\text{CL}}(L = \ell)$, can be easily obtained by

$$P_{\text{CL}}(L = \ell) = P_{\text{CL}}(L > \ell - 1) - P_{\text{CL}}(L > \ell). \quad (38)$$

Similarly to the discussion of $P(i \in \text{cycle})$ in the previous section, to obtain more accurate results for $P_{\text{CL}}(L > \ell)$ in a network which exhibits a broad degree distribution, one needs to take into account the heterogeneity of the network. Consider the first shell around the random node, i, which consists of the nodes $r_1, r_2, \ldots, r_k$, of degrees $k_1, k_2, \ldots, k_k$. The distribution

of shortest path lengths between a pair of neighbors, r_m and r_n, depends on their degrees, k_m and k_n. Therefore, in this analysis one should use the conditional probabilities $\widehat{Q}_{\mathrm{PL}}(L > \ell - 2|k_m,k_n)$. The shortest cycle on which i resides consists of the shortest path among all the paths connecting the $\binom{k}{2}$ pairs of neighbors of i. Since each neighbor, such as r_m, of degree k_m, participates in $k - 1$ such pairs, these conditional distributions are not independent. Thus, one should properly condition on the degrees of pairs of neighbors. Implementing these considerations, one can replace Eq. (35) by

$$P_{\mathrm{CL}}(L > \ell|k)$$
$$= \sum_{k_1,k_2,\dots,k_k} \prod_{r=1}^{k} \widetilde{P}(k_r) \prod_{m<n} \widehat{Q}_{\mathrm{PL}}(L > \ell - 2|k_m,k_n), \quad (39)$$

where $\widehat{Q}_{\mathrm{PL}}(L > \ell - 2|k_m,k_n)$ is given by Eq. (28). Actually, for $\ell \geqslant N$ this equation coincides with Eq. (34). This is due to the fact that the maximal length of a cycle is $\ell = N$. Hence, the probability that the length of the shortest cycle is larger than N is equivalent to the probability that there is no cycle. Plugging Eq. (39) into Eq. (36), we obtain a more accurate expression for the DSCL.

In practice, for networks with broad degree distributions, the summation over the whole range of values of k and $k_1,k_2,\dots,k_k$ may be impractical. In such cases, one can evaluate Eq. (39) using Monte Carlo methods [51]. The simplest approach is to draw the degree k from the distribution $P(k)$ and then draw the k degree $k_1,k_2,\dots,k_k$ from the distribution $kP(k)/\langle K \rangle$. One then calculates $\widehat{Q}_{\mathrm{PL}}(L > \ell - 2|k_m,k_n)$ for all the $\binom{k}{2}$ combinations of degrees, k_m and k_n, and multiplies them to obtain one data point for $P_{\mathrm{CL}}(L > \ell|k)$. In Fig. 2 we present flow charts illustrating the sequence of intermediate steps in the calculation of the DSCL. The simpler approach of Eq. (35) is shown in Fig. 2(a) and the more detailed approach of Eq. (39) is shown in Fig. 2(b).

The mean of the DSCL is given by the first moment

$$\langle L \rangle_{\mathrm{CL}} = \sum_{\ell=2}^{N-1} P_{\mathrm{CL}}(L > \ell). \quad (40)$$

The variance of the DSCL is given by

$$\sigma_{\mathrm{CL}}^2 = \langle L^2 \rangle_{\mathrm{CL}} - \langle L \rangle_{\mathrm{CL}}^2, \quad (41)$$

where

$$\langle L^2 \rangle_{\mathrm{CL}} = \sum_{\ell=2}^{N-1} (2\ell + 1) P_{\mathrm{CL}}(L > \ell). \quad (42)$$

Similarly, higher order moments can be obtained using the tail-sum formula, as in Eq. (29).

VII. APPLICATIONS TO SPECIFIC NETWORK MODELS

Here we apply the approach presented above for the calculation of the DSCL in three examples of configuration model networks, namely ER networks, random regular graphs, and scale-free networks.
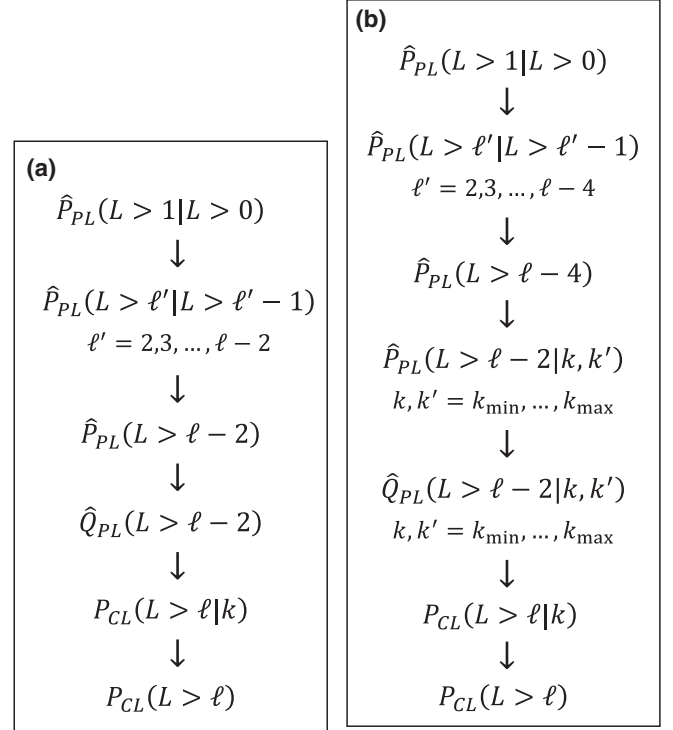

FIG. 2. Flow charts illustrating the sequence of intermediate steps in the calculation of the distribution of shortest cycle lengths, $P_{\mathrm{CL}}(L > \ell)$: (a) in the simpler approach of Eq. (35); (b) in the more detailed approach of Eq. (39).

A. Erdős-Rényi networks

The Erdős-Rényi (ER) network is the simplest kind of a random network, and a special case of the configuration model, in which only the mean degree, $\langle K \rangle = c$, is constrained. ER networks can be constructed by independently connecting each pair of nodes with probability $p = c/(N - 1)$. In the thermodynamic limit the resulting degree distribution follows a Poisson distribution of the form

$$P(k) = \frac{e^{-c}c^k}{k!}. \quad (43)$$

In Appendix B we briefly summarize the properties of the giant cluster of the ER network and present a closed form expression for g as a function of c. More generally, in ER networks there is no distinction between the statistical properties of a random node and a random neighbor of a random node. As a result, $\tilde{g} = g$ and the different DSPLs are identical, namely $P_{\mathrm{PL}}(L = \ell) = \widetilde{P}_{\mathrm{PL}}(L = \ell) = \widehat{P}_{\mathrm{PL}}(L = \ell)$. Similarly, for the overall DSPLs we obtain $Q_{\mathrm{PL}}(L = \ell) = \widetilde{Q}_{\mathrm{PL}}(L = \ell) = \widehat{Q}_{\mathrm{PL}}(L = \ell)$. Inserting the degree distribution of Eq. (43) into the generating functions $G_0(x)$ and $G_1(x)$ in Eqs. (11) and (12), respectively, one obtains the conditional probabilities $P_{\mathrm{PL}}(L > \ell|L > \ell - 1)$. Inserting them into Eq. (10), one obtains the tail DSPL between pairs of nodes which reside on the same cluster, denoted by $P_{\mathrm{PL}}(L > \ell)$. This DSPL essentially accounts only for pairs of nodes which both reside on the giant cluster, because for a pair of nodes on the nongiant components it is extremely unlikely that they reside on the same nongiant component. In order to obtain
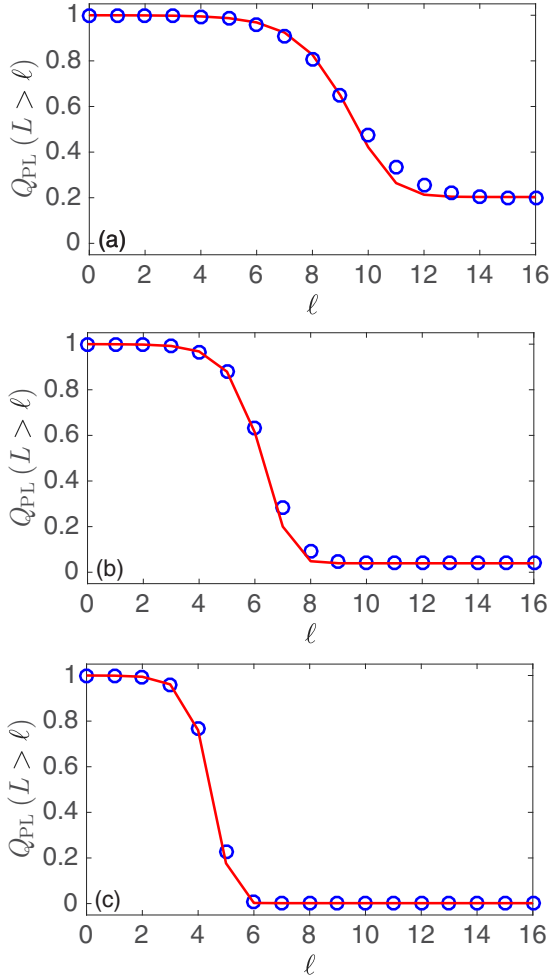
FIG. 3. The tail distribution of shortest path lengths, $Q_{\mathrm{PL}}(L > \ell)$, for ER networks of $N = 10^4$ nodes, and mean degree $c = 2.5$ (a), $c = 4$ (b), and $c = 7$ (c). The analytical results (solid lines), obtained from Eq. (20), are found to be in very good agreement with the results of computer simulations (circles), which were averaged over 10 instances of the network. The tail distributions exhibit the characteristic shape of a monotonically decreasing sigmoid function, with a nonzero asymptotic value at large distances. The asymptotic value of $Q_{\mathrm{PL}}(L > \ell)$ at large distances accounts for the probability that two randomly selected nodes do not reside on the same cluster. As c is increased, the inflection point, which corresponds to the peak of the DSPL, shifts to the left; namely distances in the network become shorter, in agreement with the prediction of small-world theory. Concurrently, the asymptotic tail moves down, reflecting the increasing size of the giant cluster.

the overall DSPL between all pairs of nodes, one needs to adjust the results for the fraction of pairs of nodes in which both of them reside on the giant cluster, which is given by g^2. Inserting the probability $P_{\mathrm{PL}}(L > \ell)$ into Eq. (20), one obtains the overall DSPL, $Q_{\mathrm{PL}}(L > \ell)$.

In Fig. 3 we present the tail distributions $Q_{\mathrm{PL}}(L > \ell)$ for ER networks of $N = 10^4$ nodes with mean degree $c = 2.5$ (a), $c = 4$ (b), and $c = 7$ (c). The analytical results (solid lines), obtained from Eq. (20), are found to be in very good agreement with the results of computer simulations (circles). The tail distributions exhibit the characteristic shape
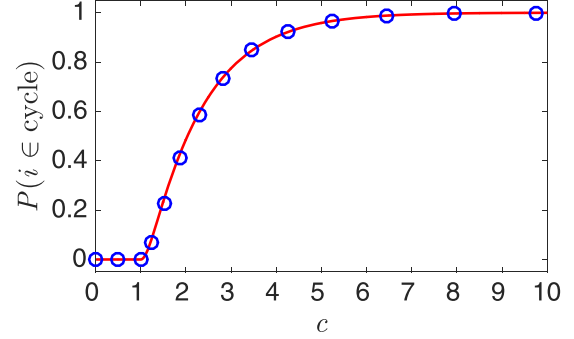

FIG. 4. The probability $P(i \in \mathrm{cycle})$ that a random node, i, resides on at least one cycle, versus the mean degree, c, in an ER network of $N = 10^4$ nodes. The analytical results (solid line), obtained from Eq. (44), are found to be in excellent agreement with the results of computer simulations (circles).

of a monotonically decreasing sigmoid function between two plateaus. Their inflection points coincide with the peaks of the corresponding probability distribution functions. The tail distributions $Q_{\mathrm{PL}}(L > \ell)$ exhibit nonzero asymptotic values at large distances, which account for the probability that two randomly selected nodes do not reside on the same cluster, and thus the distance between them is $\ell = \infty$. As c is increased, the inflection point shifts to the left, which means that distances in the network become shorter. This can be understood in the framework of small-world theory, where the mean distance is given by $\langle L \rangle \simeq \ln N / \ln c$. Concurrently, the asymptotic value of $Q_{\mathrm{PL}}(L > \ell)$ decreases, due to the increasing size of the giant cluster.

Using Eqs. (32) and (33), the probability that a random node in an ER network resides on at least one cycle can be expressed in the form

$$P(i \in \mathrm{cycle}) = 1 - \sum_{k=0}^{\infty} \frac{e^{-c} c^k}{k!} (1 - g^2)^{k(k-1)/2}, \qquad (44)$$

where g is given by Eq. (B1). In Fig. 4 we present the probability $P(i \in \mathrm{cycle})$ as a function of the mean degree, c, for ER networks of $N = 10^4$ nodes. The analytical results (solid lines), obtained from Eq. (44), are found to be in very good agreement with the results of computer simulation (circles). It is found that for $c < 1$ there are no cycles and thus $P(i \in \mathrm{cycle}) = 0$. As c is increased above 1, the probability $P(i \in \mathrm{cycle})$ increases sharply.

To obtain more accurate results, we consider a random node i of a given degree, k, and express the probability that it resides on at least one cycle in the form

$$P(i \in \mathrm{cycle}|k) = 1 - \sum_{k_1, k_2, \ldots, k_k} \prod_{r=1}^{k} \widetilde{P}(k_r) \prod_{m < n} \left(1 - \tilde{g}_{k_m} \tilde{g}_{k_n}\right).$$

$$(45)$$

In the case of an ER network, where $P(k)$ is a Poisson distribution, $\tilde{g}_k = g_k$ and $k_r P(k_r)/\langle K \rangle = P(k_r - 1)$, where $k_r - 1$ is the degree of the rth neighbor of node i on the reduced network from which i was removed. Therefore, in the

case of an ER network

$$P(i \in \text{cycle}|k) = 1 - \sum_{k_1,k_2,\dots,k_k} \prod_{r=1}^{k} P(k_r) \prod_{m<n} \left(1 - g_{k_m} g_{k_n}\right). \tag{46}$$

The evaluation of this product requires moments of g_k, which can be expressed in a closed form as

$$\langle g_k^n \rangle = \sum_{r=0}^{n} \binom{n}{r} (-1)^r e^{-c[1-(1-g)^r]}. \tag{47}$$

The two lowest order moments are

$$\langle g_k \rangle = 1 - e^{-cg} = g \tag{48}$$

and

$$\langle g_k^2 \rangle = 1 - 2e^{-cg} + e^{-cg(2-g)} = g^2 + (1-g)^2 \left(e^{cg^2} - 1\right). \tag{49}$$

Inserting these moments into Eq. (46) we find that the probability that a node of degree $k = 2$ resides on at least one cycle is

$$P(i \in \text{cycle}|K = 2) = g^2. \tag{50}$$

Incidentally, this result coincides with the simpler form which comes from Eq. (32). For nodes of degree $k = 3$

$$P(i \in \text{cycle}|K = 3) = 3g^2 - 3g^2\langle g_k^2 \rangle + \langle g_k^2 \rangle^3. \tag{51}$$

At this order the result already deviates from those obtained from the simpler approach of Eq. (32). Analytical expressions for $P(i \in \text{cycle}|K = k)$ with $k = 4$ and 5 are presented in Appendix C.

In Fig. 5 we present the conditional probability $P(i \in \text{cycle}|K = k)$ that a random node of degree k resides on at least one cycle as a function of the mean degree c for $k = 2$ (a), $k = 3$ (b), and $k = 5$ (c). The analytical results obtained from the simpler approach of Eq. (32) are shown in dashed lines. The analytical results obtained from the more detailed approach of Eq. (34) are given explicitly in Eqs. (50), (51), (C1), and (C2). These results are shown in solid lines. Incidentally, the two analytical curves coincide for $k = 2$, while for $k = 3$ and 5, the more detailed theory is found to be in a better agreement with the results of computer simulations (circles).

The DSCL of an ER network is given by

$$P_{\text{CL}}(L > \ell) = (1+c)e^{-c} + \sum_{k=2}^{\infty} \frac{e^{-c}c^k}{k!} Q_{\text{PL}}(L > \ell - 2)^{\binom{k}{2}}. \tag{52}$$

In Fig. 6 we present the tail distributions $P_{\text{CL}}(L > \ell)$ for ER networks of $N = 10^4$ nodes, where $c = 2.5$ (a), $c = 4$ (b), and $c = 7$ (c). The analytical results (solid lines), obtained from Eq. (52), are in good agreement with computer simulations (circles). The tail distribution exhibits a monotonically decreasing sigmoid shape from the $P_{\text{CL}}(L > \ell) = 1$ plateau on the left to $P_{\text{CL}}(L > \ell) = P(i \notin \text{cycle})$ on the right, since the height of the second plateau represents the fraction of nodes which do not reside on any cycle. This fraction decreases as the mean degree, c, is increased, namely the probability
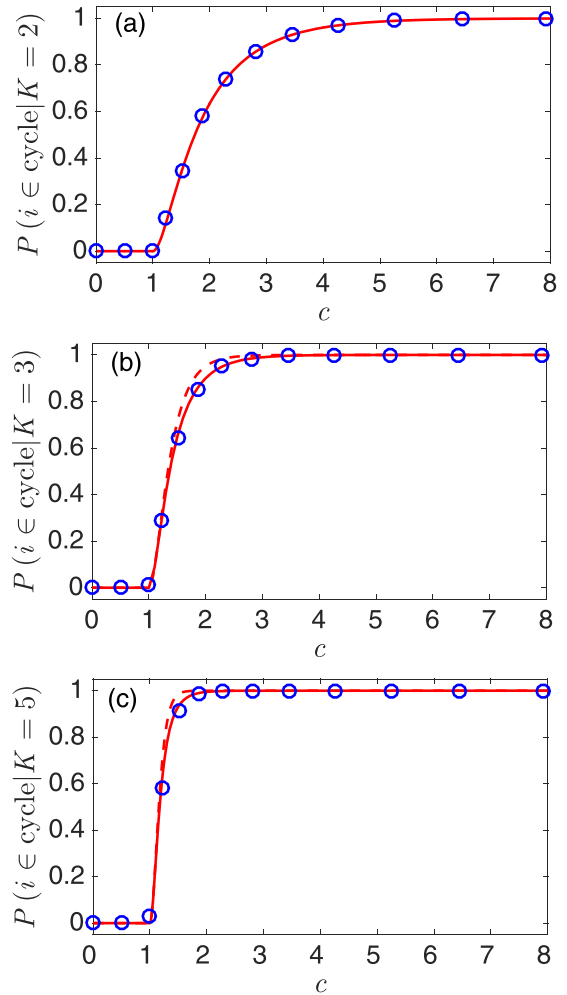

FIG. 5. The conditional probability $P(i \in \text{cycle}|K = k)$ that a random node i of degree $k = 2$ (a), $k = 3$ (b), and $k = 5$ (c) resides on at least one cycle, as a function of the mean degree c in an ER network of $N = 10^4$ nodes. For $0 < c < 1$ there are no cycles and therefore $P(i \in \text{cycle}|K = k) = 0$. For $c > 1$ the probability that a random node, i, of a given degree, k, resides on at least one cycle increases monotonically with c. This is due to the fact that as c is increased the degrees of its neighbors increase and they are thus more likely to be connected to each other on the reduced network which does not include the node i. The analytical results obtained from the simpler approach, described by Eq. (32), are shown in dashed lines, while the analytical results obtained from the more detailed approach, described by Eq. (34), are shown in solid lines. Incidentally, the two analytical curves coincide for $k = 2$, while for $k = 3$ and 5, the more detailed theory is found to be in better agreement with the results of computer simulations (circles).

that a random node resides on at least one cycle increases as c is increased. The peak of the corresponding probability distribution function, $P_{\text{CL}}(L = \ell)$, shifts to the left as c is increased. These results imply that as the network becomes more strongly connected the shortest cycles become more numerous and shorter.

In Fig. 7 we present the conditional tail distribution $P_{\text{CL}}(L > \ell|K = k)$ for an ER network of $N = 10^4$ nodes and $c = 2.5$, where $k = 2$ (a), $k = 3$ (b), and $k = 5$ (c). The analytical results obtained from the simpler approach
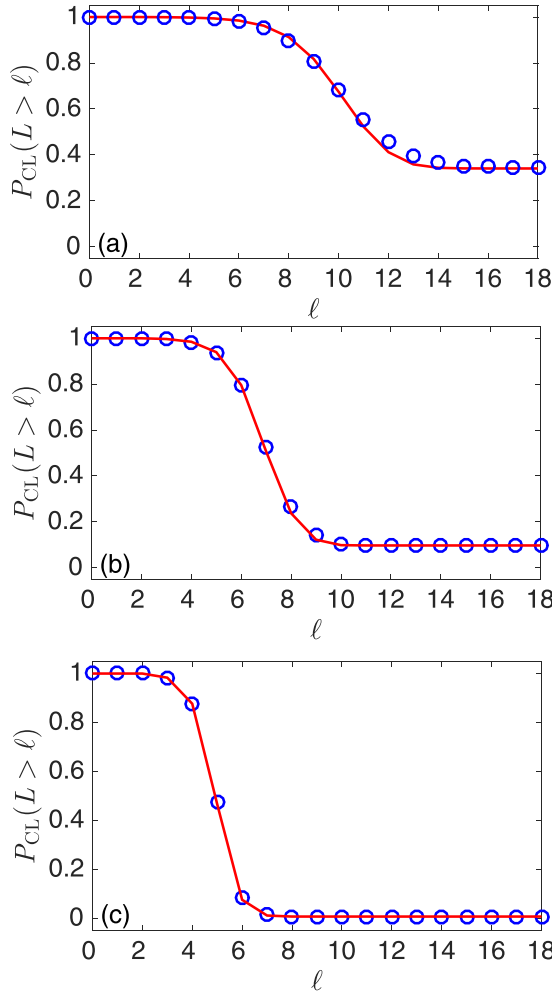
FIG. 6. The tail distribution of shortest cycle lengths, $P_{\mathrm{CL}}(L > \ell)$, for ER networks of $N = 10^4$ nodes, where $c = 2.5$ (a), $c = 4$ (b), and $c = 7$ (c). The analytical results (solid lines), obtained from Eq. (52), are found to be in very good agreement with the results of computer simulations (circles).
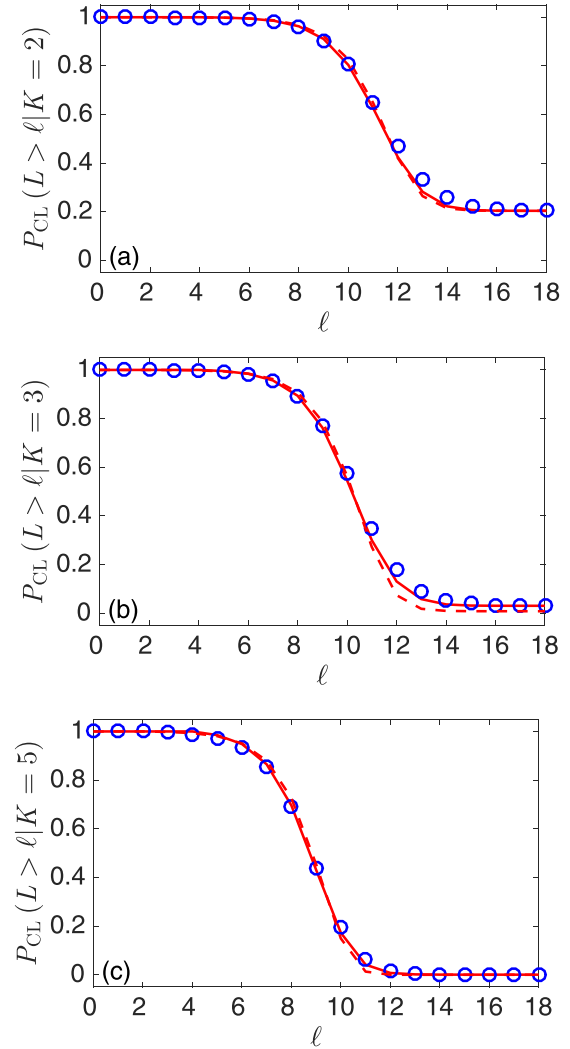

FIG. 7. The conditional distribution of shortest cycle lengths, $P_{\mathrm{CL}}(L > \ell | K = k)$, for ER networks of $N = 10^4$ nodes, where $c = 2.5$ and $k = 2$ (a), $k = 3$ (b), and $k = 5$ (c). The analytical results obtained from the simpler approach of Eq. (35) are shown in dashed lines, while the analytical results obtained from the more detailed approach of Eq. (39) are shown in solid lines. The more detailed theory is found to be in better agreement with the results of computer simulations (circles).

of Eq. (35) are shown in dashed lines, while the analytical results obtained from the more detailed approach of Eq. (39) are shown in solid lines. The two analytical curves are almost indistinguishable for $k = 2$, and are both in very good agreement with the results of computer simulations (circles). For $k = 3$ and 5, the more detailed theory provides a better agreement with the results of computer simulations (circles).

The conditional tail distribution retains the qualitative features of the sigmoid shape. The asymptotic value at large ℓ is $P_{\mathrm{CL}}(L > \ell) = P(i \notin \mathrm{cycle})$, which decreases as k is increased, which means that the probability that a random node of degree k resides on at least one cycle increases as k is increased. The peak of the corresponding probability distribution function, $P_{\mathrm{CL}}(L = \ell | K = k)$, shifts to the left as k is increased, which means that for node of higher degree the shortest cycle is shorter.

The probability that a random node, i, of degree k resides on at least one cycle is a monotonically increasing function of k. The length ℓ of the shortest cycle tends to decrease as a function of k. This is due to the fact that the length of the

shortest cycle is determined by the shortest path among all the paths connecting neighbors of i, and the number of such pairs increases quadratically with k.

In Fig. 8 we present analytical results for the mean, $\langle L \rangle_{\mathrm{CL}}$, of the DSCL as a function of the mean degree, c, for ER networks of $N = 10^3$ nodes (solid line). The results are in very good agreement with computer simulations (circles). The mean, $\langle L \rangle_{\mathrm{CL}}$, is a monotonically decreasing function of c. It exhibits a sharp decrease in the dilute network limit, which becomes more moderate as the network becomes more dense. For comparison, we also present analytical (dashed line) and numerical ($\times$) results for the mean, $\langle L \rangle_{\mathrm{PL}}$, of the DSPL as a function of c (dashed line). It is found that for the entire range of values of c, the mean of the DSCL is slightly larger than the mean of the corresponding DSPL. This can be understood as
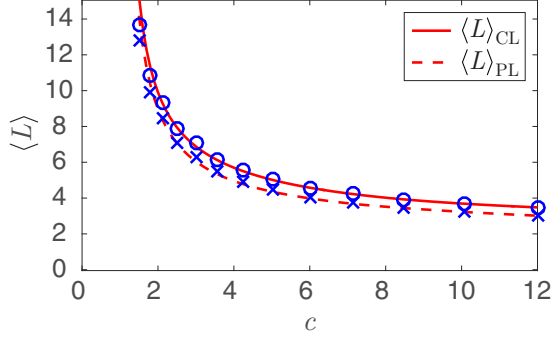
FIG. 8. Analytical results for the mean, $\langle L \rangle_{\mathrm{CL}}$, of the distribution of shortest cycle lengths (solid line), and for the mean, $\langle L \rangle_{\mathrm{PL}}$, of the distribution of shortest path lengths (dashed line), as a function of the mean degree, c, for ER networks of $N = 10^3$ nodes. Both curves are found to be in very good agreement with the corresponding results obtained from computer simulations (circles for $\langle L \rangle_{\mathrm{CL}}$ and $\times$ for $\langle L \rangle_{\mathrm{PL}}$). It is found that $\langle L \rangle_{\mathrm{CL}}$ is slightly larger than $\langle L \rangle_{\mathrm{PL}}$ for all values of c.

follows. The length of the shortest cycle on which a random node, i, resides consists of the shortest path between a pair of its neighbors, plus 2 for the two edges connecting i to these neighbors. This suggests that $\langle L \rangle_{\mathrm{CL}}$ should be longer by about two units than $\langle L \rangle_{\mathrm{PL}}$. However, the shortest path between neighbors of i which is incorporated in the shortest cycle is the shortest among the shortest paths connecting *all* pairs of neighbors of i. Thus, it tends to be shorter than the path between two random nodes. As a result, the difference $\Delta = \langle L \rangle_{\mathrm{CL}} - \langle L \rangle_{\mathrm{PL}}$ is smaller than 2.

In Fig. 9 we present the standard deviation of the DSCL, σ_{CL}, as a function of the mean degree, c, for ER networks of $N = 10^3$ nodes. For small values of c, the analytical results (solid line) underestimate the standard deviation, as can be seen from the comparison with the results of computer simulations (circles). We also show the analytical (dashed line) and numerical ($\times$) results for the standard deviation of the
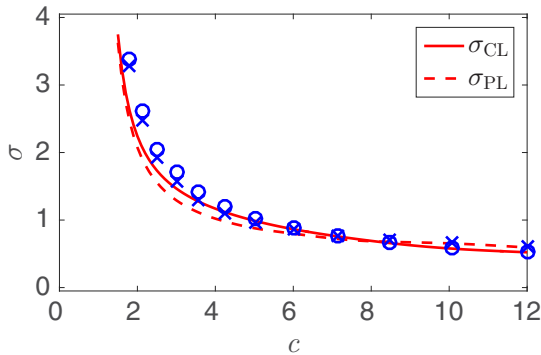

FIG. 9. Analytical results for the standard deviations, σ_{CL}, of the distribution of shortest cycle lengths (solid line), and σ_{PL} of the distribution of shortest path lengths (dashed line), as a function of the mean degree, c, for ER networks of $N = 10^3$ nodes. For small values of c, the analytical results appear to underestimate the standard deviations with respect to the results of computer simulations for σ_{CL} (circles) and σ_{PL} ($\times$).

DSPL, σ_{PL}, for the same networks, which exhibits the same qualitative features.

B. Random regular graphs

In a random regular graph with $c \geqslant 3$ the giant cluster encompasses the whole network. Therefore, $g = \widetilde{g} = 1$ (for more details see Appendix D). Moreover, in this case the DSPLs and the overall DSPLs are identical since all pairs of nodes reside on the giant cluster. The generating functions for the random regular graph are given by Eqs. (D1) and (D2). The DSCL is given by

$$P_{\mathrm{CL}}(L > \ell) = [\widehat{P}_{\mathrm{PL}}(L > \ell - 2)]^{\binom{c}{2}}. \tag{53}$$

In order to proceed we shall first calculate the conditional probabilities $\widehat{P}_{\mathrm{PL}}(L > \ell | L > \ell - 1)$ using the recursion equation (18) and the initial condition (17). This yields

$$\widehat{P}_{\mathrm{PL}}(L > \ell | L > \ell - 1) = \left[1 - \frac{(c-1)^2}{(N-1)c} \right]^{(c-1)^{\ell-1}}. \tag{54}$$

Assuming that the size of the network is large $N \gg 1$, we can approximate the above to

$$\ln \widehat{P}_{\mathrm{PL}}(L > \ell | L > \ell - 1) \simeq -\frac{(c-1)^{\ell+1}}{cN} + O\left(\frac{1}{N^2} \right). \tag{55}$$

By inserting the conditional distribution into Eq. (19) we can obtain the tail distribution

$$\widehat{P}_{\mathrm{PL}}(L > \ell) \simeq \exp\left[-\frac{(c-1)^2}{cN} \frac{(c-1)^\ell - 1}{c-2} \right]. \tag{56}$$

We can use this DSPL inside Eq. (53) to get

$$P_{\mathrm{CL}}(L > \ell) \simeq \exp\left[-\frac{(c-1)^3}{2N} \frac{(c-1)^{\ell-2} - 1}{c-2} \right]. \tag{57}$$

In Fig. 10 we present the DSCL for random regular graphs of $N = 10^3$ nodes with $c = 3$ (a), $c = 5$ (b), and $c = 7$ (c). The analytical results (solid lines), obtained from Eq. (57), are found to be in excellent agreement with the results of computer simulations (circles). Since Eq. (57) is based on exact results for the DSPL, we conjecture that it is an exact result for the DSCL of the random regular graph.

C. Scale-free networks

Consider a configuration model network with a power-law degree distribution, $P(k)$, given by

$$P(k) = \frac{k^{-\gamma}}{\zeta(\gamma, k_{\min}) - \zeta(\gamma, k_{\max} + 1)}, \tag{58}$$

where $k_{\min} \leqslant k \leqslant k_{\max}$ and $\zeta(s,a)$ is the Hurwitz zeta function [52]. Here we focus on the case in which $\gamma > 2$, in which the mean degree $\langle K \rangle$ is bounded even for $k_{\max} \to \infty$. We further restrict our analysis to the case in which $k_{\min} \geqslant 2$, namely the network does not include isolated nodes and leaf nodes. Under these conditions $g = \widetilde{g} = 1$, namely the giant cluster encompasses the entire network (for more details see Appendix E). As a result, $\widehat{Q}_{\mathrm{PL}}(L > \ell) = \widehat{P}_{\mathrm{PL}}(L > \ell)$. Thus,
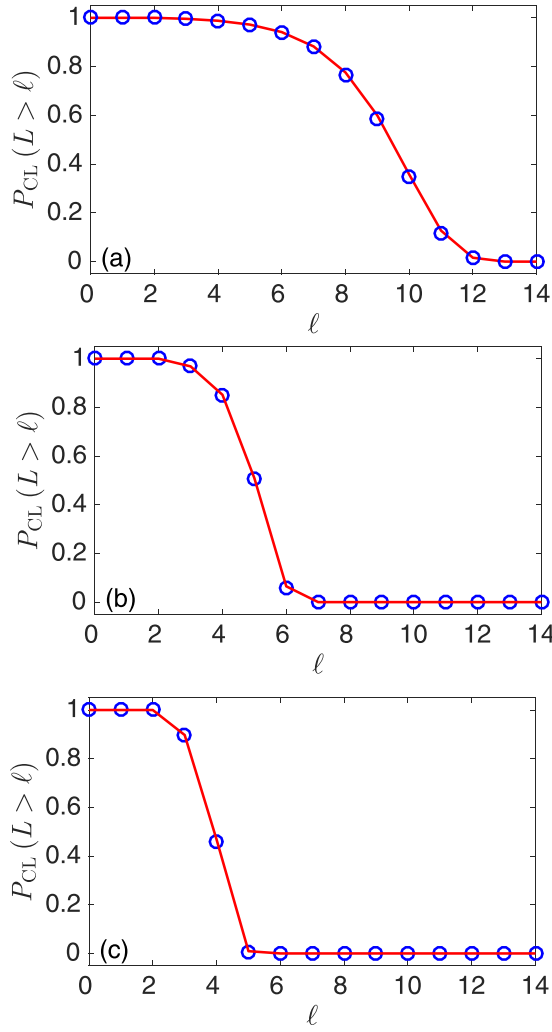
FIG. 10. The tail distributions of shortest cycle lengths, $P_{CL}(L > \ell)$, for random regular graphs of $N = 10^3$ nodes and $c = 3$ (a), $c = 5$ (b), and $c = 7$ (c). The analytical results (solid line), obtained from Eq. (57), are found to be in excellent agreement with the results of computer simulations (circles).

the DSCL can be expressed in the form

$$P_{CL}(L > \ell) = \sum_{k=2}^{\infty} P(k) \widehat{P}_{PL}(L > \ell - 2)^{\binom{k}{2}}, \quad (59)$$

where $\widehat{P}_{PL}(L > \ell - 2)$ is calculated using Eqs. (17)–(19), where $\langle K \rangle$ and $\langle K^2 \rangle$ are given by Eqs. (E1) and (E2), respectively.

In Fig. 11 we present the tail distribution, $P_{CL}(L > \ell)$, for a configuration model network of $N = 10^3$ nodes and a power-law degree distribution with $\gamma = 2.5$ and $k_{min} = 3$ (a), 5 (b), and 8 (c). The analytical results obtained from the simpler approach of Eq. (35) are shown in dashed lines, while the analytical results obtained from the more detailed approach of Eq. (39) are shown in solid lines. The results of the more detailed approach were obtained from 10^4 Monte Carlo samplings of the degrees $k, k_1, k_2, \ldots, k_k$. Both results are found to be in very good agreement with the results of comp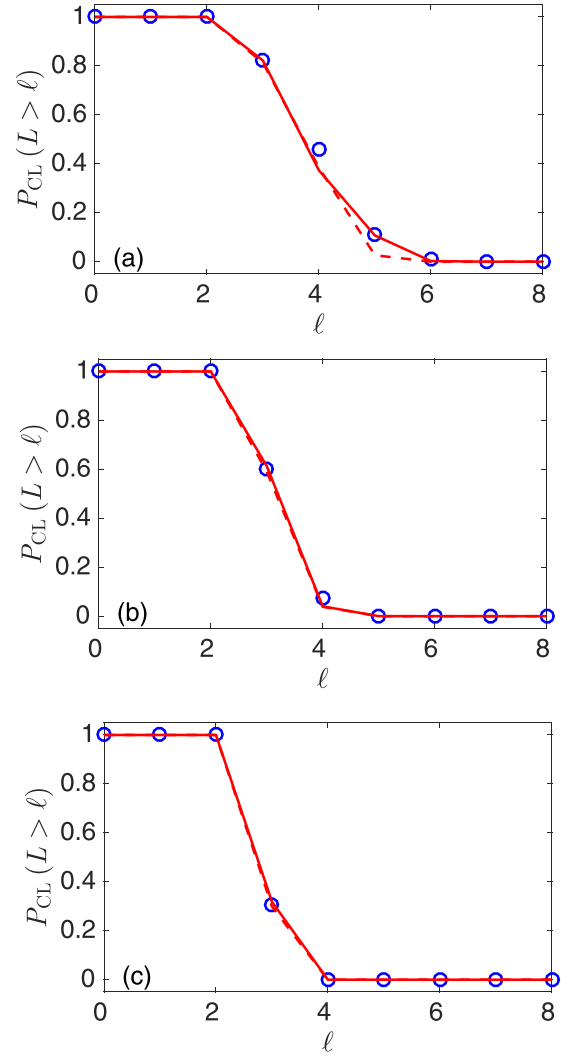uter simulations (circles), except for one data point of the simpler approach, for $k_{min} = 3$, at $\ell = 5$, which is significantly

FIG. 11. The tail distribution of shortest cycle lengths, $P_{CL}(L > \ell)$, for scale free networks of $N = 10^3$ nodes and a power-law degree distribution with $\gamma = 2.5$ and $k_{min} = 3$ (a), 5 (b), and 8 (c). The analytical results obtained from the simpler approach of Eq. (35) are shown in dashed lines, while the analytical results obtained from the more detailed approach of Eq. (39) are shown in solid lines. Both results are in very good agreement with the results of computer simulations (circles), except for some deviation of the simpler approach for $k_{min} = 3$.

lower than the simulation result. It is observed that as k_{min} is increased, the distances in the network become shorter.

VIII. DISCUSSION

An important distinction in network theory is between networks which exhibit a tree structure and networks which include cycles. In network growth models, the existence of cycles is determined by the growth rules of the network. For example, in the Barabási-Albert model [53,54], the existence of cycles depends on the number of nodes, m, which are added at each time step. In the case in which $m = 1$, the model gives

rise to a stochastic tree structure [55,56], while for $m \geqslant 2$ it forms cycles.

In equilibrium networks such as configuration model networks, one can distinguish between three situations, which are determined by the degree distribution $P(k)$. In the subpercolation regime of dilute networks, the network consists of finite tree components, whose size does not scale with N. In this regime, the number of cycles does not scale with N. Above percolation, the network consists of a giant cluster, which includes cycles, in addition to many finite components. As the network becomes denser, the number of cycles increases and their typical length becomes shorter. In the regime of dense networks, the giant cluster encompasses the entire network and there are many short cycles.

The degree distribution plays a crucial role in shaping the properties of cycles in a network. In particular, isolated nodes (of degree $k = 0$) and leaf nodes (of degree $k = 1$) cannot reside on any cycle. Only nodes of degrees $k \geqslant 2$ may reside on a cycle. Still, some nodes of degrees $k \geqslant 2$ do not reside on any cycle. Instead, they reside on a tree component which can be either isolated or connected to the giant cluster.

There are interesting connections between the DSCL and the DSPL of a configuration model network. For a random node, i, the cycles on which it resides consist of paths between pairs of neighbors of i and two edges from i to these neighbors. The shortest cycle length is thus given by the shortest path between all pairs of neighbors of i plus 2. A naive expectation would thus be that the shortest cycles are longer than the shortest paths by 2 units. From Fig. 8 we observe that the mean cycle length is longer than the mean path length by about one unit over a broad range of values of c in the ER network. To understand this point, we recall that the shortest cycle on which a random node i of degree k resides is composed of the shortest path among all the $\binom{k}{2}$ paths connecting pairs of neighbors of i. Another issue is the fact that the degrees of the neighbors of i are not uniformly sampled from $P(k)$ but from $\widetilde{P}(k)$. The mean path length between pairs of neighbors of i is given by $\langle \widehat{L} \rangle$, while the mean path length between pairs of random nodes is given by $\langle L \rangle$. Clearly, the path lengths between nodes of higher degrees are shorter than between nodes of lower degrees, as can be seen from Eqs. (23)–(25). It is thus interesting to compare the mean degrees of $P(k)$ and $\widetilde{P}(k)$. The former is given by $\langle K \rangle$ while the latter is $\langle K^2 \rangle / \langle K \rangle$. In our context, the effective degree of a neighbor of a random node i is given by the connective constant $\mu = \langle K^2 \rangle / \langle K \rangle - 1$, where the edge connecting i and its neighbor is removed. It turns out that μ may be larger than, equal to, or smaller than $\langle K \rangle$ in different network ensembles. In the case of the ER ensemble, a special symmetry gives rise to $\mu = \langle K \rangle$. In the random regular graph, it turns out that $\mu = c - 1$ and thus $\mu < \langle K \rangle$. In configuration models with a power-law degree distribution and $2 < \gamma \leqslant 3$, the moment $\langle K^2 \rangle$ diverges and thus $\mu > \langle K \rangle$. In those cases in which $\mu > \langle K \rangle$, the mean distance between neighbors of i is smaller than the mean distance between random nodes, and vice versa. Therefore, the difference between the mean of the DSCL and the mean of the DSPL is determined by a combination of these conflicting effects.

The results presented above have implications for the stability of configuration model networks to node deletion

processes due to failures or attacks. In particular, if a node of degree $k \geqslant 2$, which does not reside on any cycle, is removed, the network breaks down to k separate components. Thus, nodes of degree $k \geqslant 2$ which do not reside on any cycle are articulation points [57].

In this paper we have studied configuration model networks in which the DSCL is completely determined by the degree distribution $P(k)$. Recently, other network ensembles were introduced, which include many short cycles, where the cycle lengths are controlled by various constraints [58,59]. It would be interesting to generalize the calculation of the DSCL to such networks.

Knowing the properties of cycles is important for the study of many dynamical processes on networks. For example, shortest cycles provide the fastest feedback paths in the network and introduce correlations between the signals arriving at a given node via different links. It was found that in neural circuits the lengths of the shortest cycles determine the frequencies of broadband spontaneous macroscopic neural oscillations [42–44]. In a broader context, feedback processes are affected by the entire spectrum of cycle lengths, up to the longest possible length of the Hamiltonian cycles. The number of cycles of a given length was studied extensively in Refs. [38,60–64].

In the context of network control theory, it was shown that dynamical processes on complex networks can be identified and controlled by a small set of "determining nodes," which can be identified from the network structure alone, regardless of the specific properties of the dynamical process. Moreover, this set must include at least one node from each one of the feedback loops in the network [65,66]. This approach was recently applied [41] to the analysis of real biological, technological, and social networks, providing predictions for the set of nodes whose control can push the network dynamics towards any desired asymptotic state (fixed point, cycle, or limit cycle).

Analytical techniques for treating spin models on networks are mostly exact on tree structures. Utilizing the local tree structure of random networks, they provide accurate results for short range properties. However, in order to obtain insight about collective and long range correlations, one needs to take into account the large scale structure, which notably involves the statistics of loops as done recently in Refs. [67,68].

IX. SUMMARY

We presented an analytical approach for the calculation of the distribution of shortest cycle lengths in configuration model networks. This approach is based on a fundamental relation between the distribution of shortest cycle lengths and the distribution of shortest path lengths in such networks. It employs an analytical approach for the calculation of the distribution of shortest path lengths, presented in Ref. [28]. We use this approach for the calculation of the DSCL in Erdős-Rényi networks, random regular graphs, and scale-free configuration model networks, and obtain very good agreement with the results of computer simulations. The mean and standard deviation of the DSCL in these networks are also calculated. We also obtain a closed form expression for the

fraction of nodes which do not reside on any cycle. While in this paper we have focused on the case of undirected networks, cycles are known to be important also in directed networks, in contexts such as gene regulation networks, neural networks, and food webs [69]. It would thus be interesting to study the DSCL on directed networks. Another interesting direction is the study of properties of long cycles [70]. In this context, an open question is the distribution of longest cycle lengths on random networks.

APPENDIX A: THE CONDITIONAL DSPL FOR SHORT DISTANCES

The conditional DSPLs presented in Eqs. (23)–(25) apply for the case in which $\ell > 2$. Here we provide the expressions for the conditional DSPLs for the special cases of $\ell = 1$ and 2. Starting from $\ell = 1$, the probability that two random nodes of degrees k and k' are not connected to each other is given by

$$P_{\mathrm{PL}}(L > 1|k,k') = 1 - \frac{kk'}{(N-1)\langle K \rangle}. \quad (A1)$$

Similarly, when the node of degree k is selected as a random neighbor of a random node, while the node of degree k' is a random node, one obtains

$$\widetilde{P}_{\mathrm{PL}}(L > 1|k,k') = 1 - \frac{(k-1)k'}{(N-2)\langle K \rangle}. \quad (A2)$$

Finally, for the case in which both nodes are selected as random neighbors of random nodes, one obtains

$$\widehat{P}_{\mathrm{PL}}(L > 1|k,k') = 1 - \frac{(k-1)(k'-1)}{(N-3)\langle K \rangle}. \quad (A3)$$

Proceeding to $\ell = 2$, one first evaluates the conditional probability $P_{\mathrm{PL}}(L > 2|L > 1; k,k')$, which is given by

$$P_{\mathrm{PL}}(L > 2|L > 1; k,k')$$
$$= \left[1 - \sum_{k''=0}^{\infty} P(k'') \frac{kk'k''(k''-1)}{N(N-1)\langle K \rangle^2} \right]^{N-2}. \quad (A4)$$

Carrying out the summation and multiplying by $P_{\mathrm{PL}}(L > 1|k,k')$, we obtain

$$P_{\mathrm{PL}}(L > 2|k,k') = 1 - \frac{\langle K^2 \rangle kk'}{N\langle K \rangle^2} + O\left(\frac{1}{N^2}\right), \quad (A5)$$

which is valid under the assumption that $\langle K^2 \rangle$ is finite. Using similar considerations, one can show that

$$\widetilde{P}_{\mathrm{PL}}(L > 2|k,k') = 1 - \frac{\langle K^2 \rangle(k-1)k'}{N\langle K \rangle^2} + O\left(\frac{1}{N^2}\right) \quad (A6)$$

and

$$\widehat{P}_{\mathrm{PL}}(L > 2|k,k') = 1 - \frac{\langle K^2 \rangle(k-1)(k'-1)}{N\langle K \rangle^2} + O\left(\frac{1}{N^2}\right). \quad (A7)$$

APPENDIX B: THE GIANT CLUSTER IN ERDŐS-RÉNYI NETWORKS

In the asymptotic limit the ER network exhibits a percolation transition at $c = 1$, such that for $c < 1$ the network consists only of finite components while for $c > 1$ there is a giant cluster. At a higher value of the connectivity, namely at $c = \ln N$, there is a second transition, above which the giant cluster encompasses the entire network and there are no nongiant components. We denote the probability that a randomly selected node belongs to the giant cluster by $g = g(c)$. Clearly, $g(c) = 0$ for $c \leqslant 1$ and $g(c) = 1$ for $c > \ln N$. For intermediate values of c, in the range of $1 < c < \ln N$, the probability that a random node belongs to the giant cluster is given by $1 - g = \exp(-cg)$ [50]. Solving for g, one obtains

$$g(c) = 1 + \frac{W(-ce^{-c})}{c}, \quad (B1)$$

where $W(x)$ is the Lambert W function [52]. ER networks exhibit a special property, resulting from the Poisson degree distribution, Eq. (43), which satisfies $\tilde{P}(k) = P(k-1)$, where $\widetilde{P}(k)$ is given by Eq. (1). This implies that for the Poisson distribution, the two generating functions defined in the main text are identical, namely $G_0(x) = G_1(x) = e^{-c(1-x)}$. As a consequence of Eqs. (4) and (5), in ER networks $\tilde{g} = g$. This means that in ER networks there is no distinction between the statistical properties of a random node and a random neighbor of a random node.

APPENDIX C: THE PROBABILITY $P(i \in \text{cycle}|k)$ IN ERDŐS-RÉNYI NETWORKS FOR $k = 4$ AND 5

Here we present analytical results for $P(i \in \text{cycle})|k)$ for $k = 4$ and 5 in an ER network, obtained from Eq. (46). It is found that

$$P(i \in \text{cycle}|K = 4) = 6g^2 - 3g^4 - 12g^2\langle g_k^2 \rangle + 12g^2\langle g_k^2 \rangle^2 + 4g^3\langle g_k^3 \rangle + 4\langle g_k^2 \rangle^3 - 12g\langle g_k^2 \rangle^2\langle g_k^3 \rangle - 3\langle g_k^2 \rangle^4 + 6\langle g_k^2 \rangle^2\langle g_k^3 \rangle^2 - \langle g_k^3 \rangle^4 \quad (C1)$$

and

$$\begin{aligned} P(i \in \text{cycle}|K = 5) = {}& 30g^4\langle g_k^2 \rangle - 5g^4\langle g_k^4 \rangle - 15g^4 - 60g^3\langle g_k^3 \rangle\langle g_k^2 \rangle + 20g^3\langle g_k^3 \rangle - 70g^2\langle g_k^2 \rangle^3 + 60g^2\langle g_k^2 \rangle^2 + 30g^2\langle g_k^4 \rangle\langle g_k^2 \rangle^2 \\ & - 30g^2\langle g_k^2 \rangle + 60g^2\langle g_k^3 \rangle^2\langle g_k^2 \rangle + 10g^2 + 12\langle g_k^2 \rangle^5 - 15\langle g_k^4 \rangle\langle g_k^2 \rangle^4 - 15\langle g_k^2 \rangle^4 - 70\langle g_k^3 \rangle^2\langle g_k^2 \rangle^3 \\ & + 10\langle g_k^4 \rangle^2\langle g_k^2 \rangle^3 + 120g\langle g_k^3 \rangle\langle g_k^2 \rangle^3 + 10\langle g_k^2 \rangle^3 + 30\langle g_k^3 \rangle^2\langle g_k^2 \rangle^2 - 60g\langle g_k^3 \rangle\langle g_k^2 \rangle^2 + 60\langle g_k^3 \rangle^2\langle g_k^4 \rangle\langle g_k^2 \rangle^2 \\ & - 60g\langle g_k^3 \rangle\langle g_k^4 \rangle\langle g_k^2 \rangle^2 + 30\langle g_k^3 \rangle^4\langle g_k^2 \rangle - 60g\langle g_k^3 \rangle^3\langle g_k^2 \rangle - 30\langle g_k^3 \rangle^2\langle g_k^4 \rangle^2\langle g_k^2 \rangle - \langle g_k^4 \rangle^5 - 5\langle g_k^3 \rangle^4 \\ & + 10\langle g_k^3 \rangle^2\langle g_k^4 \rangle^3 - 15\langle g_k^3 \rangle^4\langle g_k^4 \rangle + 20g\langle g_k^3 \rangle^3\langle g_k^4 \rangle. \end{aligned} \quad (C2)$$

These conditional probabilities can be evaluated explicitly, as a function of c, by inserting the moments

$$\langle g_k^3 \rangle = 1 - 3e^{-cg} + 3e^{-cg(2-g)} - e^{-cg(3-3g+g^2)} \quad \text{(C3)}$$

and

$$\langle g_k^4 \rangle = 1 - 4e^{-cg} + 6e^{-cg(2-g)} - 4e^{-cg(3-3g+g^2)}$$
$$+ e^{-c[1-(1-g)^4]}. \quad \text{(C4)}$$

Higher order moments can be obtained from Eq. (47).

APPENDIX D: THE GIANT CLUSTER IN RANDOM REGULAR GRAPHS

In this appendix we show that in random regular networks the giant cluster encompasses the entire network, namely $g = 1$. In the random regular graph, the degree distribution is $P(k) = \delta_{k,c}$, where $\delta_{k,c}$ is the Kronecker symbol and c is an integer. We will focus on the case of $c \geqslant 3$. In this case

$$G_1(x) = x^{c-1} \quad \text{(D1)}$$

and

$$G_0(x) = x^c. \quad \text{(D2)}$$

For $c = 1$ the random regular graph consists of dimers, while for $c = 2$ it consists of loops of various lengths. A fully developed network is obtained only for $c \geqslant 3$, and this will be the case of interest in the present work. To obtain the size of the giant cluster we look for solutions of Eq. (4). Inserting $G_1(x)$ from Eq. (D1) into Eq. (4) we obtain $1 - \tilde{g} = (1 - \tilde{g})^{c-1}$. It is easy to see that $\tilde{g} = 0$ and $\tilde{g} = 1$ are solutions of this equation. Moreover, for $c \geqslant 3$ the expression on the right hand side is smaller than the expression on the left hand side for any $0 < \tilde{g} < 1$. Thus, for $c \geqslant 3$ there are no other solutions for Eq. (4). This proves that $\tilde{g}$ may be either 0 or 1. Inserting these solutions into Eq. (5) we find that in both cases g is equal to $\tilde{g}$, namely $g = 0$ or 1. The solution $g = 0$ stands for the case in which there is no giant cluster, while the solution $g = 1$ implies that the giant cluster encompasses the entire network. In order to determine which of these possible solutions is the relevant one, we use the criterion of Molloy and Reed for the existence of a giant cluster [32,33]. It states that if $\langle K^2 \rangle > 2\langle K \rangle$ then there is a giant cluster, namely $g > 0$. In the case of a random regular graph $\langle K \rangle = c$ and $\langle K^2 \rangle = c^2$. Thus, for $c \geqslant 3$ the

Molloy and Reed criterion is satisfied and $g > 0$. Hence, the only possible solution is $g = 1$, namely the giant cluster of a random regular graph with $c \geqslant 3$ encompasses the entire network.

APPENDIX E: THE GIANT CLUSTER IN SCALE-FREE NETWORKS

In this appendix we consider a configuration model with a power-law degree distribution of the form $P(k) \sim k^{-\gamma}$, where the degrees are bounded in the range $k_{\min} \leqslant k \leqslant k_{\max}$. The normalized degree distribution is given by Eq. (58). The mean degree is

$$\langle K \rangle = \frac{\zeta(\gamma - 1, k_{\min}) - \zeta(\gamma - 1, k_{\max} + 1)}{\zeta(\gamma, k_{\min}) - \zeta(\gamma, k_{\max} + 1)}, \quad \text{(E1)}$$

while the second moment of the degree distribution is

$$\langle K^2 \rangle = \frac{\zeta(\gamma - 2, k_{\min}) - \zeta(\gamma - 2, k_{\max} + 1)}{\zeta(\gamma, k_{\min}) - \zeta(\gamma, k_{\max} + 1)}. \quad \text{(E2)}$$

For $\gamma \leqslant 2$ the mean degree diverges when $k_{\max} \rightarrow \infty$. For $2 < \gamma \leqslant 3$ the mean degree is bounded while the second moment, $\langle K^2 \rangle$, diverges. For $\gamma > 3$ both moments are bounded. It can be shown that for $\gamma > 2$ and $k_{\min} \geqslant 2$ (where nodes of degrees 0 and 1 do not exist), $\langle K^2 \rangle > 2\langle K \rangle$, namely the Molloy and Reed criterion is satisfied and the network exhibits a giant cluster [32,33]. Below we show that under these conditions the giant cluster encompasses the entire network. Inserting a power-law degree distribution with $k_{\max} \rightarrow \infty$ into Eq. (4) we obtain

$$G_1(x) = \frac{\Phi(x, \gamma - 1, k_{\min})}{\zeta(\gamma - 1, k_{\min})} x^{k_{\min} - 1}, \quad \text{(E3)}$$

where $\Phi(x, \gamma, k)$ is the Lerch transcendent [71]. It can be shown that for any $0 < x < 1$ the relation $\Phi(x, s, k) < \Phi(1, s, k) = \zeta(s, k)$ is satisfied, provided that $k > 0$. Therefore, $G_1(x) < x^{k_{\min} - 1}$ for any value of x in the range $0 < x < 1$. In the case in which $k_{\min} \geqslant 2$ the inequality $x^{k_{\min} - 1} < x$ is also satisfied for $0 < x < 1$. Thus, $x = 0$ and $x = 1$ are the only fixed points of the generating function $G_1(x)$. Inserting $1 - \tilde{g}$ instead of x and using the criterion of Molloy and Reed [32,33], we find that the only possible value of $\tilde{g}$ is $\tilde{g} = 1$. Inserting this value in Eq. (5) one finds that $g = 1$, namely that in scale-free configuration model networks with $\gamma > 2$ and $k_{\min} \geqslant 2$, the giant cluster encompasses the entire network.

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

[2] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).

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

[4] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2012).

[5] J. Travers and S. Milgram, Socimetry **32**, 425 (1969).

[6] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[7] F. Chung and L. Lu, Proc. Natl. Acad. Sci. U.S.A. **99**, 15879 (2002)

[8] F. Chung and L. Lu, Internet Mathematics **1**, 91 (2003).

[9] R. Cohen and S. Havlin, Phys. Rev. Lett. **90**, 058701 (2003).

[10] A. Maáyan, S. L. Jenkins, S. Neves, A. Hasseldine, E. Grace, B. Dubin-Thaler, N. J. Eungdamrong, G. Weng, P. T. Ram, J. J. Rice, A. Kershenbaum, G. A. Stolovitzky, R. D. Blitzer, and R. Iyengar, Science **309**, 1078 (2005).

[11] E. W. Dijkstra, Numer. Math. **l**, 269 (1959).

[12] D. Delling, P. Sanders, D. Schultes, and D. Wagner, Engineering route planning algorithms, in *Algorithmics of Large and Complex Networks: Design, Analysis, and Simulation*, edited by J. Lerner, D. Wagner, and K. A. Zweig (Springer, Heidelberg, 2009), p. 117.

[13] I. Abraham, D. Delling, A. V. Goldberg, and R. F. Werneck, J. Exp. Algorithmics **18**, 1.3 (2013).

[14] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[15] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Rev. Mod. Phys. **87**, 925 (2015).

[16] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).

[17] V. D. Blondel, J.-L. Guillaume, J. M. Hendrickx, and R. M. Jungers, Phys. Rev. E **76**, 066101 (2007).

[18] S. N. Dorogotsev, J. F. F. Mendes, and A. N. Samukhin, Nucl. Phys. B **653**, 307 (2003).

[19] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski, Electron. J. Probab. **12**, 703 (2007).

[20] R. van der Hofstad and G. Hooghiemstra, J. Math. Phys. **49**, 125209 (2008).

[21] H. van der Esker, R. van der Hofstad, and G. Hooghiemstra, J. Stat. Phys. **133**, 169 (2008).

[22] J. Shao, S. V. Buldyrev, R. Cohen, M. Kitsak, S. Havlin, and H. E. Stanley, Europhys. Lett. **84**, 48004 (2008).

[23] J. Shao, S. V. Buldyrev, L. A. Braunstein, S. Havlin, and H. E. Stanley, Phys. Rev. E **80**, 036105 (2009).

[24] E. Katzav, M. Nitzan, D. ben-Avraham, P. L. Krapivsky, R. Kühn, N. Ross, and O. Biham, Europhys. Lett. **111**, 26006 (2015).

[25] P. Erdős and A. Rényi, Publicationes Mathematicae Debrecen **6**, 290 (1959).

[26] P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).

[27] P. Erdős and A. Rényi, Bulletin of the International Statistical Institute **38**, 343 (1961).

[28] M. Nitzan, E. Katzav, R. Kühn, and O. Biham, Phys. Rev. E **93**, 062309 (2016).

[29] S. Melnik and J. P. Gleeson, arXiv:1604.05521.

[30] C. Steinbock, O. Biham, and E. Katzav, Phys. Rev. E **96**, 032301 (2017).

[31] A. Fronczak, P. Fronczak, and J. A. Holyst, Phys. Rev. E **70**, 056110 (2004).

[32] B. Molloy and A. Reed, Rand. Struct. Alg. **6**, 161 (1995).

[33] B. Molloy and A. Reed, Comb. Probab. Comput. **7**, 295 (1998).

[34] L. Giot *et al.*, Science **302**, 1727 (2003).

[35] K.-I. Goh, E. Oh, B. Kahng, and D. Kim, Phys. Rev. E **67**, 017101 (2003).

[36] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).

[37] G. Caldarelli, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **38**, 183 (2004).

[38] K. Klemm and P. F. Stadler, Phys. Rev. E **73**, 025101(R) (2006).

[39] P. M. Gleiss, P. F. Stadler, A. Wagner, and D. A. Fell, Adv. Complex Syst. **4**, 207 (2001).

[40] S. Klamt and A. von Kamp, BMC Bioinf. **10**, 181 (2009).

[41] J. G. T. Zañudo, G. Yang, and R. Albert, Proc. Natl. Acad. Sci. U.S.A. **114**, 7234 (2017).

[42] N. Vladimirov, Y. Tu, and R. D. Traub, Front. Comput. Neurosci. **6**, 17 (2012).

[43] A. Goldental, R. Vardi, S. Sardi, P. Sabo, and I. Kanter, Front. Neural Circuits **9**, 65 (2015).

[44] A. Goldental, H. Uzan, S. Sardi, and I. Kanter, Sci. Rep. **7**, 2700 (2017).

[45] V. Van Kerrebroeck and E. Marinari, Phys. Rev. Lett. **101**, 098701 (2008).

[46] P.-L. Giscard and R. C. Wilson, *Proceedings of Complex Networks, 2017* (The Sixth International Conference on Complex Networks and Their Applications), pp 14–28.

[47] J. M. Yeomans, *Statistical Mechanics of Phase Transitions* (Clarendon Press, Oxford, 1992).

[48] J. Pitman, *Probability* (Springer, New York, 1993).

[49] L. Takács, J. Appl. Probab. **25**, 359 (1988).

[50] B. Bollobás, *Random Graphs* (Cambridge University Press, Cambridge, 2001).

[51] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999).

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

[53] A.-L. Barabasi and R. Albert, Science **286**, 509 (1999).

[54] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[55] M. Drmota and B. Gittenberger, Random Struct. Alg. **10**, 421 (1997).

[56] M. Drmota and H.-K. Hwang, Adv. Appl. Probab. **37**, 321 (2005).

[57] L. Tian, A. Bashan, D.-N. Shi, and Y.-Y. Liu, Nat. Commun. **8**, 14223 (2017).

[58] E. S. Roberts and A. C. C. Coolen, ESAIM: Proc. **47**, 97 (2014).

[59] A. C. C. Coolen, J. Phys.: Conf. Ser. **699**, 012022 (2016).

[60] G. Bianconi and M. Marsili, J. Stat. Mech. (2005) P06005.

[61] E. Marinari and R. Monasson, J. Stat. Mech. (2004) P09004.

[62] E. Marinari, R. Monasson, and G. Semerjian, Europhys. Lett. **73**, 8 (2006).

[63] E. Marinari and G. Semerjian, J. Stat. Mech. (2006) P06019.

[64] J. D. Noh, Eur. Phys. J. B **66**, 251 (2008).

[65] B. Fiedler, A. Mochizuki, G. Kurosawa, and D. Saito, J. Dynam. Differ. Equat. **25**, 563 (2013).

[66] A. Mochizuki, B. Fiedler, G. Kurosawa, and D. Saito, J. Theor. Biol. **335**, 130 (2013).

[67] A. Montanari and T. Rizzo, J. Stat. Mech. (2005) P10011.

[68] A. Altieri *et al.*, arXiv:1707.08499.

[69] S. Johnson and N. S. Jones, Proc. Natl. Acad. Sci. U.S.A. **114**, 5618 (2017).

[70] E. Marinari, G. Semerjian, and V. Van Kerrebroeck, Phys. Rev. E **75**, 066708 (2007).

[71] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, 6th ed. (Academic Press, San Diego, 2000).