General formulation of long-range degree correlations in complex networks

Yuka Fujiki, Taro Takaguchi, and Kousuke Yakubo

Phys. Rev. E **97**, 062308 — Published 11 June 2018

DOI: 10.1103/PhysRevE.97.062308

General formulation of long-range degree correlations in complex networks

Yuka Fujiki*
Department of Applied Physics, Hokkaido University, Sapporo 060-8628, Japan

Taro Takaguchi†
National Institute of Information and Communications Technology, Tokyo 184-8795, Japan

Kousuke Yakubo‡
Department of Applied Physics, Hokkaido University, Sapporo 060-8628, Japan
(Dated: May 22, 2018)

We provide a general framework for analyzing degree correlations between nodes separated by more than one step (i.e., beyond nearest neighbors) in complex networks. One joint and four conditional probability distributions are introduced to fully describe long-range degree correlations with respect to degrees k and k' of two nodes and shortest path length l between them. We present general relations among these probability distributions and clarify the relevance to nearest-neighbor degree correlations. Unlike nearest-neighbor correlations, some of these probability distributions are meaningful only in finite-size networks. Furthermore, as a baseline to determine the existence of intrinsic long-range degree correlations in a network other than inevitable correlations caused by the finite-size effect, the functional forms of these probability distributions for random networks are analytically evaluated within a mean-field approximation. The utility of our argument is demonstrated by applying it to real-world networks.

PACS numbers: 89.75.Hc, 89.75.Fb, 02.70.Rr

I. INTRODUCTION

In many networks describing complex real systems, the number of edges from a node, namely degree, widely fluctuates from node to node, and degree distributions often exhibit power-law behavior [1]. For such networks, significant interest now concentrates on the issue of correlations between degrees of two nodes. In particular, degree correlations between adjacent nodes have been extensively studied so far [2–10]. Nearest neighbor degree correlations (NNDCs) in complex networks are related to their fundamental structural properties, such as clustering [11–14], community structures [15], the average path length [16], and fractality [17–19]. In addition, NNDCs influence various dynamics on networks, such as epidemic spreading [20–24], synchronization phenomena [25–30], strategic games [31–34], and resilience to failures [35–39].

It has, however, been pointed out recently that NNDCs are not enough to characterize structural properties of complex networks. For example, scale-free fractal networks are known to exhibit negative NNDCs (namely, disassortative mixing) [17]. Thus, hub nodes in such a network are almost never connected directly by an edge. In actual fractal networks, like the World Wide Web or synthetic graphs [18, 40], however, hub nodes are not only nonadjacent to, but also repulsive over a long-range distance to each other [41]. As another example, Orsini *et al.* [42] found that many local and even global structural features of real-world complex networks are closely reproduced by random graphs with the same degree sequences, clustering, and NNDCs as those for the real networks. However, some sort of global properties, such as the shortest path length distributions, betweenness distributions, and community structures, cannot be explained by these local characteristics. This implies that intrinsic non-local degree correlations in these networks cannot be described by NNDCs as a local characteristic. Furthermore, it has been demonstrated that the shortest path length between hub nodes influences functions or dynamical properties of networks [43–46]. For understanding non-local structural properties, it is important and useful to provide a framework to describe degree correlations between nodes beyond nearest neighbors, namely, long-range degree correlations (LRDCs).

There have been several proposals for formulating LRDCs in complex networks. Rybski *et al.* [47] describe LRDCs by fluctuations of the degree along shortest paths between two nodes. This is an analogy to fluctuation analysis used in correlated time series. Mayo *et al.* [48] defined the long-range assortativity and the average lth neighbor degree to quantify LRDCs (the same definition of the long-range assortativity was independently employed in [49]). The long-range assortativity r_l is the Pearson correlation coefficient between degrees of pairs of nodes separated by the shortest path length l from each other. The average lth neighbor degree $k_l(k)$ is the average degree of nodes separated by l from a node of degree k. They found that social networks exhibit disassortative degree correlations on long-range scales, while nonsocial networks do not indicate such a tendency. The

* y-fujiki@eng.hokudai.ac.jp
† taro.takaguchi.cp@gmail.com
‡ yakubo@eng.hokudai.ac.jp

two-walks degree assortativity proposed by Allen-Perkins *et al.* [50] is another type of assortativity measure beyond nearest neighbors. This quantity is defined as the Pearson correlation coefficient of the sum of the nearest-neighbor degrees of adjacent nodes, which reflects second neighbor degree correlations. These quantities enable us to pick up some specific aspects of LRDCs. However, if we perform a global and multilateral analysis of LRDCs, a more general framework is required to obtain various information of LRDCs.

In this work, we provide a general framework for analyzing LRDCs in complex networks of either finite or infinite size. In order to fully describe correlations between degrees k and k' of two nodes separated by a shortest path length l, one joint probability distribution and four conditional probability distributions are introduced as functions of k, k', and l. NNDCs can be described by these probability distributions as a special case of $l = 1$. These five distributions are not independent of each other, and we present general relations among them. In addition, the functional forms of these probability distributions for random networks are analytically evaluated within a mean-field approximation. By comparing the distributions for a given network with those for the corresponding random network with the same degree distribution, one can judge whether the network possesses intrinsic LRDCs other than inevitable correlations caused by the finite-size effect, and obtain detailed information about degree correlations. Finally, we demonstrate the usefulness of our argument by applying it to real-world networks.

The rest of this paper is organized as follows. In Sec. II, we introduce the probability distributions characterizing LRDCs and present general relations between them. In Sec. III, the functional forms of the probability distributions for random networks are analytically evaluated. In Sec. IV, the utility of our argument is tested by calculating the distributions for real-world networks. Section V is devoted to the summary and remarks.

II. JOINT AND CONDITIONAL PROBABILITY DISTRIBUTIONS

Degree correlations between nearest-neighbor nodes (namely, NNDCs) are completely described by the joint probability distribution $P_{\rm nn}(k, k')$ that two end nodes of a randomly chosen edge have the degrees k and k'. We can define the conditional probability from $P_{\rm nn}(k, k')$ by $P_{\rm nn}(k'|k) = P_{\rm nn}(k, k')/\sum_{k'} P_{\rm nn}(k, k')$, which is the probability that a node adjacent to a randomly chosen node of degree k has the degree k'. If the degree distribution function $P(k)$ is given, the probability $P_{\rm nn}(k'|k)$ also identifies NNDCs. We extend this idea to LRDCs. All information pertaining to correlations between degrees k and k' of two nodes separated by a shortest path length l (namely, LRDCs) is included in the joint probability distribution $P(k, k', l)$ that randomly chosen two nodes have the degrees k and k' and the shortest path length between them is l. From this joint distribution, four conditional probability distributions can be constructed as follows,

$$P(l|k,k') = \frac{P(k,k',l)}{\sum_l P(k,k',l)}, \tag{1a}$$

$$P(k'|k,l) = \frac{P(k,k',l)}{\sum_{k'} P(k,k',l)}, \tag{1b}$$

$$P(k,k'|l) = \frac{P(k,k',l)}{\sum_{k,k'} P(k,k',l)}, \tag{1c}$$

$$P(k',l|k) = \frac{P(k,k',l)}{\sum_{k',l} P(k,k',l)}. \tag{1d}$$

The meanings of these probability distributions, as well as the joint distribution, are listed in Table I. These conditional distributions also describe LRDCs. The probability distributions in Table I are normalized as $\sum_{k,k',l} P(k,k',l) = \sum_l P(l|k,k') = \sum_{k'} P(k'|k,l) = \sum_{k,k'} P(k,k'|l) = \sum_{k',l} P(k',l|k) = 1$. Here, we note that the sum over l includes the distance (l_∞) between disconnected node pair. The following should also be emphasized. We are interested in sparse networks from the correspondence of real-world networks. The probability distributions $P(k,k',l)$, $P(l|k,k')$, and $P(k',l|k)$ are meaningless for such sparse networks with infinitely large components, because the average shortest path length $\langle l \rangle$ diverges and values of these distribution functions become always zero for finite l. In contrast, $P(k'|k,l)$ and $P(k,k'|l)$ can be properly defined even for infinite networks.

TABLE I. Meanings of one joint and four conditional probability distributions characterizing LRDCs in networks.

Probability distribution	Meaning
$P(k,k',l)$	Probability that randomly chosen two nodes have the degrees k and k' and the distance between them is l
$P(l\|k,k')$	Probability that randomly chosen two nodes of degrees k and k' are separated by l, namely, the shortest path length distribution between nodes of degrees k and k'
$P(k'\|k,l)$	Probability that a node separated by l from a randomly chosen node of degree k has the degree k', namely, the degree distribution of a node separated by l from a node of degree k
$P(k,k'\|l)$	Probability that randomly chosen two nodes separated by l from each other have the degrees k and k'
$P(k',l\|k)$	Probability that a randomly chosen node has the degree k' and is separated by l from a node of degree k

Using the joint distribution $P(k,k',l)$, the degree distribution $P(k)$ and the shortest path length distribution $R(l)$ are presented by

$$P(k) = \sum_{k',l} P(k,k',l), \quad (2)$$

and

$$R(l) = \sum_{k,k'} P(k,k',l), \quad (3)$$

respectively. It is convenient to introduce the probability distribution $Q(k|l)$ defined by

$$Q(k|l) = \sum_{k'} P(k,k'|l), \quad (4)$$

which is the probability that one end node of a randomly chosen l-chain has the degree k, where l-chain is a shortest path between two nodes separated by l. This is an extension of the probability $Q_{\rm nn}(k)$ that one end node of an *edge* has the degree k to a long-range node pair in the sense of $Q_{\rm nn}(k) = Q(k|l=1)$. With the aid of $Q(k|l)$, we have

$$\sum_{k'} P(k,k',l) = Q(k|l)R(l). \quad (5)$$

Equations (2), (3), and (5), as well as the obvious relation

$$\sum_{l} P(k,k',l) = P(k)P(k'), \quad (6)$$

form sum rules of the joint distribution $P(k,k',l)$. Considering these sum rules, Eq. (1) yields several general relations between the conditional distributions, $P(k)$, and $R(l)$, such as,

$$P(k',l|k) = P(k')P(l|k,k'), \quad (7)$$
$$P(k,k'|l) = Q(k|l)P(k'|k,l), \quad (8)$$
$$P(k)P(k',l|k) = Q(k'|l)R(l)P(k|k',l), \quad (9)$$

and

$$P(k)P(k')P(l|k,k') = R(l)P(k,k'|l). \quad (10)$$

Equations (9) and (10) can be considered as direct consequences of the Bayes' theorem that relates $P(A|B)$ and $P(B|A)$ for events A and B.

The joint distribution $P_{\rm nn}(k,k')$ and the conditional distributions $P_{\rm nn}(k'|k)$ describing NNDCs are included in the above long-range probability distributions as a special case of $l=1$. In fact, we have

$$P_{\rm nn}(k,k') = P(k,k'|l=1), \quad (11)$$

and

$$P_{\rm nn}(k'|k) = P(k'|k,l=1). \quad (12)$$

Similarly, the degree distribution $Q_{\rm nn}(k)$ of an end node of a randomly chosen edge is given by

$$Q_{\rm nn}(k) = Q(k|l=1) = \frac{kP(k)}{\langle k\rangle}, \quad (13)$$

where $\langle k\rangle = \sum_k kP(k)$ is the average degree. Then, Eq. (8) with $l=1$ is reduced to the well-known relation $P_{\rm nn}(k,k') = kP(k)P_{\rm nn}(k'|k)/\langle k\rangle$ [21].

Considering the above correspondence, we can easily extend indices characterizing NNDCs to those for LRDCs. For example, the long-range assortativity r_l can be defined as

$$\begin{aligned} r_l &= \frac{4\langle kk'\rangle_l - \langle k+k'\rangle_l^2}{2\langle k^2+k'^2\rangle_l - \langle k+k'\rangle_l^2}, \\ &= \frac{\langle kk'\rangle_l - \langle k\rangle_l^2}{\langle k^2\rangle_l - \langle k\rangle_l^2}, \end{aligned} \quad (14)$$

where $\langle f(k,k')\rangle_l = \sum_{k,k'} f(k,k')P(k,k'|l)$. This quantity is the Pearson correlation coefficient between degrees of nodes separated by l from each other. For $l=1$, r_l is reduced to the conventional nearest-neighbor assortativity[4]. Another example is the average degree of lth neighbor nodes, which is given by [51]

$$k_l(k) = \sum_{k'} k'P(k'|k,l). \quad (15)$$

This is an extension of the average degree, $k_{\rm nn}(k) = \sum_{k'} k'P_{\rm nn}(k'|k)$, of nearest neighbors of a node of degree k to that for lth neighbors. The quantities r_l and $k_l(k)$ are equivalent to those proposed by Ref. [48]. Besides extensions of existing indices for NNDCs, it is also possible to introduce completely new measures characterizing LRDCs, such as the strength of long-range repulsive correlations between hubs, by using the probability distributions listed in Table I.

III. BASELINE FOR COMPARISON

In the previous section, we introduced five fundamental probability distribution functions describing LRDCs in complex networks. However, even if we know these distributions for a given network, we cannot judge whether the network possesses LRDCs or not. This is due to the lack of a baseline for comparison, i.e., it has not yet been clarified how these distribution functions behave for a network in which the degrees of two nodes separated by an arbitrary distance are not correlated. In this section, we discuss such a baseline for comparison to analyze what kind of LRDCs a given network possesses.

A. General remarks

A nearest-neighbor uncorrelated network (NNUN) is a network in which the degree of one end node of an edge is independent of the degree of another end node. Thus, the joint probability $P_{\rm nn}(k,k')$ in an NNUN is given by the product $Q_{\rm nn}(k)Q_{\rm nn}(k')$. Extending this idea, a long-range uncorrelated network (LRUN) in which the degrees of any two nodes are not correlated regardless of their inter-distance is considered to be a network satisfying the relation

$$P(k,k'|l) = Q(k|l)Q(k'|l), \tag{16}$$

for all l. This equation implies that the degrees k and k' of two nodes separated by l are independent of each other.

While $P_{\rm nn}(k,k')$ for an NNUN has the simple functional form as $P_{\rm nn}(k,k') = kk'P(k)P(k')/\langle k\rangle^2$, it is difficult to obtain an exact expression of $P(k,k'|l)$ for an LRUN because $Q(k|l)$ itself depends on LRDCs. Nevertheless, we can generally conclude that $P(k'|k,l)$ for an LRUN does not depend on k. This comes immediately from the relation for LRUNs

$$P(k'|k,l) = Q(k'|l), \tag{17}$$

which is obtained by substituting Eq. (16) into Eq. (8). Equation (17) with $l=1$ leads to the well-known relation $P_{\rm nn}(k'|k) = k'P(k')/\langle k\rangle$ for NNUNs [21].

It is meaningful to consider whether LRUNs satisfying Eq. (16) actually exist or not. A prime candidate for LRUNs is random networks with a given degree distribution function, namely, networks generated by the configuration model [52]. In finite random networks, however, large degree nodes cannot be far away from each other as would be expected from Eq. (16) due to the finite-size effect. In fact, as shown in Appendix, we have numerically confirmed that the relation between $P(k,k'|l)$ and $Q(k|l)$ for random networks deviates from Eq. (16) for $l \gg \langle l\rangle$ while Eq. (16) holds for $l \ll \langle l\rangle$, where $\langle l\rangle$ is the average shortest path length. This implies that even random networks are not rigorously long-range uncorrelated in the sense of Eq. (16). The fact that Eq. (16) holds for $l \ll \langle l\rangle$, however, shows that LRDCs in random networks are caused only by the finite-size effect. Although the existence or non-existence of rigorous LRUNs may be a nontrivial mathematical problem, it is rather important from a practical viewpoint to investigate LRDCs of a given network by comparing the probability distributions with those for corresponding random networks with the same degree distribution. Comparison with random networks as the baseline enables us to evaluate intrinsic LRDCs other than inevitable degree correlations arising from the finite-size effect.

B. Mean-field approximation

In this subsection, we calculate the probability distribution functions for random networks with a given degree distribution as the baseline for comparison. Hereafter, we denote the probability distributions for random networks by adding the subscript "0". Once one of the joint and conditional probability distributions is obtained, other distributions can be calculated by using Eq. (1). We then focus on $P_0(l|k,k')$ at first, which is the shortest path length distribution between nodes of degrees k and k'. Such a length distribution for random networks has been calculated recently by Melnik and Gleeson [53]. Therefore, we can utilize their result for calculating $P_0(l|k,k')$. The outline of their argument is shown below, in which some necessary modifications are implemented[54].

Let us introduce the probability $\rho(l|k,k')$ that the distance between randomly chosen two nodes of degrees k and k' is equal to or less than l. The probability $P_0(l|k,k')$ is then presented by

$$P_0(l|k,k') = \begin{cases} \rho(0|k,k') & (l=0), \\ \rho(l|k,k') - \rho(l-1|k,k') & (l \geq 1), \\ 1 - \lim_{l\to\infty} \rho(l|k,k') & (l = l_\infty). \end{cases} \tag{18}$$

The last expression for $l = l_\infty$ is for disconnected node pairs in a network composed of multiple components. The normalization condition of $P(l|k,k')$ is thus written as $\sum_{l=0}^{\infty} P(l|k,k') + P(l_\infty|k,k') = 1$.

Under the local tree assumption and the mean-field approximation, $\rho(l|k,k')$ for a random network is given by [53]

$$\rho(l|k,k') = 1 - [1-\rho(0|k,k')]\,[1-\bar{q}(l-1|k,k')]^k\,, \tag{19}$$

where $\bar{q}(l|k,k')$ is the probability that an adjacent node of a randomly chosen node i_k of degree k lies within the distance l from a node $j_{k'}$ of degree k' under the condition that i_k is separated by more than l from $j_{k'}$. The first factor $1-\rho(0|k,k')$ of the second term in the right-hand side represents the probability that the node i_k is not the node $j_{k'}$ itself. The second factor $[1-\bar{q}(l-1|k,k')]^k$ means the probability that all adjacent nodes of i_k are separated by more than $l-1$ from $j_{k'}$ under the condition that i_k is separated by more than $l-1$ from $j_{k'}$. Thus, the rough meaning of Eq. (19) is that the probability that the node i_k lies within the distance l from the node $j_{k'}$ is equal to the probability that at least one of k adjacent nodes of i_k lies within the distance $l-1$ from $j_{k'}$. Furthermore, let us introduce the probability $q(l|k,k')$ that a randomly chosen node i_k of degree k with at least one neighboring node, say h, separated by more than l from a node $j_{k'}$ of degree k' lies within the distance l. Then, we have the following relation between $q(l|k,k')$ and $\bar{q}(l-1|k,k')$ similar to Eq. (19),

$$q(l|k,k') = 1 - [1-\rho(0|k,k')]\,[1-\bar{q}(l-1|k,k')]^{k-1}\,. \tag{20}$$

The right-hand side of this equation implies the probability that at least one node of $k-1$ adjacent nodes of i_k other than h lies within the distance $l-1$ from $j_{k'}$.

Using $q(l|k,k')$, the probability $\bar{q}(l|k,k')$ is expressed by

$$\begin{aligned}\bar{q}(l|k,k') &= \sum_{k''} P_{\mathrm{nn}}(k''|k) q(l|k'',k'), \\ &= \frac{1}{\langle k\rangle} \sum_{k''} k'' P(k'') q(l|k'',k'). \end{aligned} \tag{21}$$

Since $\bar{q}(l|k,k')$ is actually independent of k, we denote it simply by $\bar{q}(l|k')$. Multiplying $kP(k)/\langle k\rangle$ on both sides of Eq. (20), summing over k, and using Eq. (21), we have the recursion equation for $\bar{q}(l|k')$,

$$\begin{aligned}\bar{q}(l|k') = 1 &- G_1\left[1-\bar{q}(l-1|k')\right] \\ &+ \frac{k'}{N\langle k\rangle}\left[1-\bar{q}(l-1|k')\right]^{k'-1}, \end{aligned} \tag{22}$$

where N is the number of nodes in the network and $G_1(x)$ is the generating function defined by $G_1(x) = \sum_k x^{k-1} kP(k)/\langle k\rangle$. Here, we used the obvious relation,

$$\rho(0|k,k') = \frac{\delta_{kk'}}{NP(k)}. \tag{23}$$

Equation (22) can be solved iteratively with the initial condition [53],

$$\bar{q}(0|k') = \frac{k'}{N\langle k\rangle}. \tag{24}$$

Using the solution of $\bar{q}(l|k')$ and Eqs. (18) and (19), we can calculate $P_0(l|k,k')$. The joint probability distribution $P_0(k,k',l)$ is computed by $P(k)P(k')P_0(l|k,k')$ from Eqs. (1a) and (6), and other conditional probability distributions listed in Table I are determined from $P_0(k,k',l)$ by using Eq. (1).

We should remark the accuracy of the mean-field approximation in the above calculation. The probability $\rho(l|k,k')$ must be equal to $\rho(l|k',k)$ from the definition. However, $\rho(l|k,k')$ calculated from Eq. (19) is actually not symmetric with respect to k and k'. In fact, $\rho(l|k,k')$ for $l=1$ and $k \neq k'$, calculated as

$$\begin{aligned}\rho(1|k,k') &= 1 - \left(1 - \frac{k'}{N\langle k\rangle}\right)^{k} \\ &= \frac{kk'}{N\langle k\rangle} - \frac{1}{2}\frac{k(k-1)k'^2}{(N\langle k\rangle)^2} + \cdots, \end{aligned} \tag{25}$$

is asymmetric in the order of N^{-2}. This is due to the difference in accuracy of the mean-field treatment for nearest neighbors of the nodes of degrees k and k'. The mean-field approximation for neighboring nodes of a large degree node is more accurate than that of a small degree node. Since $\bar{q}(l|k')$ is iteratively calculated for the distance l from the source node of degree k' according to Eq. (22), $\rho(l|k,k')$ with $k<k'$ is more accurate than $\rho(l|k',k)$. Therefore, we first calculate $\rho(l|k,k')$ for $k<k'$ by Eq. (19), then transfer it to $\rho(l|k',k)$ in actual computations. Another remark on the mean-field approximation is related to the component-size distribution. We assume that $\rho(l|k,k')$ does not depend on the size of the component that the source node of degree k' belongs to. This implies that the distribution function of the component size is assumed to be relatively narrow. If a random network with a given degree distribution $P(k)$ is very close to its percolation transition point, however, the component-size distribution becomes wide, and then the mean-field calculations have poor accuracy.

C. Infinite tree-like networks

For infinitely large sparse networks, only $P(k'|k,l)$ and $P(k,k'|l)$ are meaningful among five distributions, as mentioned in Sec. II. It is easy to calculate these conditional distributions for infinite random networks with tree-like structures. Let us consider $P_0(k'|k,l)$ at first. Since this is the probability that a node separated by l from a node of degree k has the degree k', $P_0(k'|k,l)$ must satisfy the relation,

$$P_0(k'|k,l) = \sum_{k''} P_{\mathrm{nn}}(k'|k'') P_0(k''|k,l-1), \tag{26}$$

where the nearest-neighbor degree distribution function $P_{\mathrm{nn}}(k'|k'')$ is given by $k'P(k')/\langle k\rangle$ for random networks. Using the obvious relation $P_0(k'|k,0) = \delta_{kk'}$, we can solve the above equation as,

$$P_0(k'|k,l) = \begin{cases} \delta_{kk'} & (l=0), \\ \dfrac{k'P(k')}{\langle k\rangle} & (l \geq 1). \end{cases} \tag{27}$$

Thus, we have immediately, from Eq. (17),

$$Q_0(k|l) = \frac{kP(k)}{\langle k\rangle} \quad (l \geq 1). \tag{28}$$

The probability $P_0(k,k'|l)$ for $l \geq 1$ is then calculated from Eq. (8) as

$$P_0(k,k'|l) = \frac{kk'P(k)P(k')}{\langle k\rangle^2}. \tag{29}$$

We should note that $P_0(k,k'|l)$ and $P_0(k'|k,l)$ for infinite tree-like random networks are equivalent to $P_{\mathrm{nn}}(k,k')$ and $P_{\mathrm{nn}}(k'|k)$, respectively, independently of l.

It is reasonable to consider that the above expressions of $P_0(k'|k,l)$, $Q_0(k|l)$, and $P_0(k,k'|l)$ for infinitely large networks hold approximately for $l \ll \langle l\rangle$ even in finite random networks. While we have shown in Sec. III A that $P_0(k'|k,l)$, in general, does not depend on k, our result here indicates that this probability distribution is independent of l too if $l \ll \langle l\rangle$.

D. Numerical confirmation

In order to confirm the validity of our analytical evaluation of the probability distribution functions for random networks, we compare the distributions $P_0(l|k,k')$,

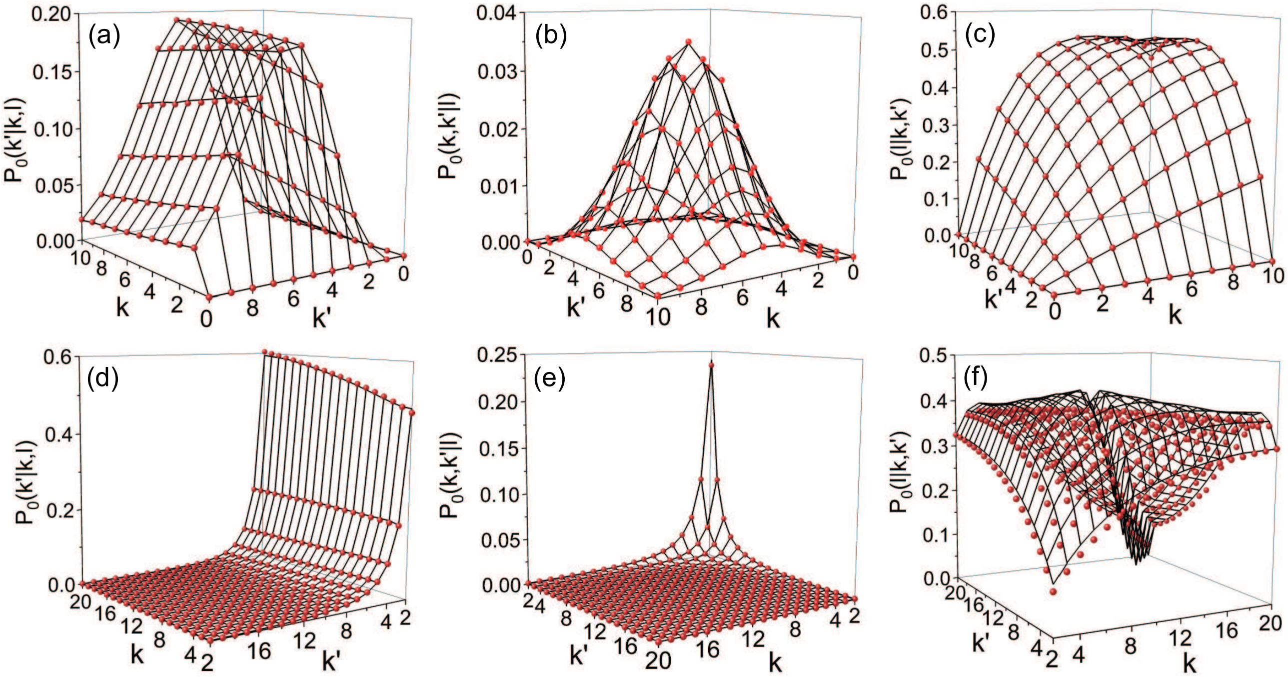

FIG. 1. (Color online) Probability distributions $P_0(k'|k,l)$, $P_0(k,k'|l)$, and $P_0(l|k,k')$ as functions of k and k' for $l=4$ and $N=1,000$. Wireframes indicate the distributions calculated analytically by the method explained in Sec. III B, while dots represent those measured for numerically realized random networks. Upper three panels [(a)-(c)] are the results for Erdős-Rényi random graphs with $\langle k\rangle = 5.0$, and lower panels [(d)-(f)] for scale-free random networks with the degree distribution $P(k) \propto k^{-3}$.

$P_0(k'|k,l)$, and $P_0(k,k'|l)$ obtained by the method explained in Sec. III B with those measured for synthetic random networks. Figure 1 shows the dependence of these distributions on k and k' for $l=4$. The wireframe in each panel indicates the analytically calculated distributions, while dots represent numerical results. The upper three panels give the results for Erdős-Rényi random graphs with $\langle k\rangle = 5.0$ and $N=1,000$. We have dared to employ relatively small networks to check the validity of the method for finite sizes. Numerical results are obtained by averaging over 100 realizations of Erdős-Rényi random graphs. The average path length of these networks is $\langle l\rangle = 4.5$. The lower three panels present the results for scale-free random networks with $N=1,000$ and the degree distribution function of $P(k) \propto k^{-3}$ for $2 \le k \le 50$ and $P(k)=0$ otherwise. Numerical results show the averages over $10,000$ realizations generated by the configuration model. The average degree and the average path length are $\langle k\rangle = 3.1$ and $\langle l\rangle = 5.4$, respectively. These plots demonstrate that the analytical treatment based on the mean-field approximation well reproduces numerical results even for finite networks.

We also verified the argument in Sec. III C by calculating $Q_0(k|l)$ for Erdős-Rényi random graphs. Figure 2 compares this probability distribution calculated by Eq. (28) with $Q_0(k|l)$ measured numerically. The Erdős-Rényi random graphs are the same as in Fig. 1. Thus, the average path length is $\langle l\rangle = 4.5$ for these random graphs. As shown by the numerical results for $l=1$, 2, and 3, $Q_0(k|l)$ measured numerically is almost independent of l and is well described by Eq. (28), if l is sufficiently smaller than $\langle l\rangle$. On the contrary, if l becomes close to or larger than $\langle l\rangle$, numerically computed $Q_0(k|l)$ deviates from Eq. (28), as shown by the results for $l=4$ and 5.

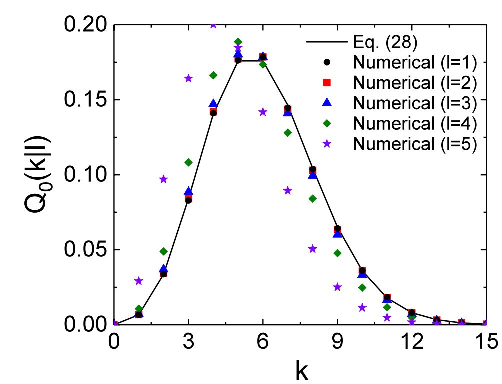

FIG. 2. (Color online) Probability distribution $Q_0(k|l)$ for Erdős-Rényi random graphs with $N=1,000$ and $\langle k\rangle = 5.0$. Solid line indicates $Q_0(k|l)$ given by Eq. (28), while symbols represent the numerical results for $l=1$ (circles), 2 (boxes),3 (triangles), 4 (diamonds), and 5 (stars) averaged over 100 network realizations.

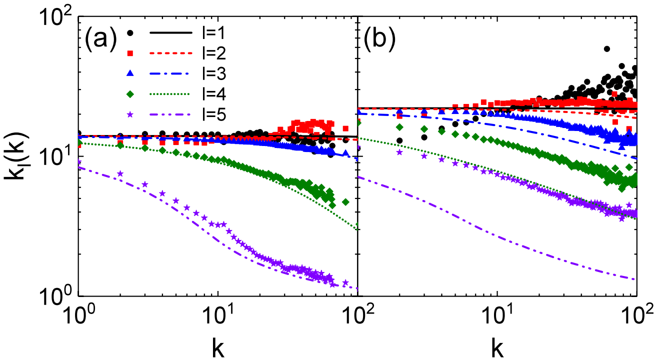

FIG. 3. (Color online) Average lth neighbor degree $k_l(k)$ of a typical node of degree k for (a) the Gnutella network [55] and (b) the coauthorship network [56]. Symbols represent $k_l(k)$ of the real-world networks as a function of k at fixed values of l, and curves indicate those of corresponding random networks with the same degree sequences as the real networks.

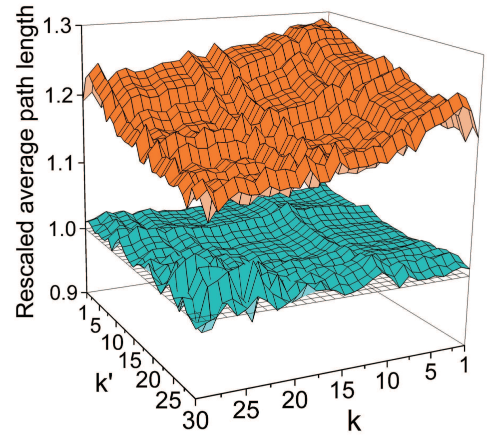

FIG. 4. (Color online) Rescaled average shortest path length $\langle l(k,k')\rangle_{\rm res}$ between nodes of degree k and k'. The lower and upper surfaces represent the results for the Gnutella network and the coauthorship network, respectively. The bottom flat mesh indicates $\langle l(k,k')\rangle_{\rm res} = 1$.

These results prove that Eq. (28) holds for $l \ll \langle l \rangle$ even in finite networks.

IV. REAL-WORLD NETWORKS

Finally, we investigate LRDCs in two real-world complex networks by using the probability distributions listed in Table I. One is the Gnutella peer-to-peer network [55] and the other is the coauthorship network [56]. The Gnutella network has $N = 10,876$ nodes and $E = 39,994$ edges. Thus, the average degree is $\langle k \rangle = 7.4$. This network consists of a single connected component. The average path length $\langle l \rangle$ and the maximum shortest path length (diameter) $l_{\rm max}$ are 4.6 and 10, respectively. The Spearman's degree-rank correlation coefficient ρ [10, 57] characterizing the NNDC is measured as $\rho = 0.0$, which implies no NNDC in the Gnutella network. The coauthorship network possesses $N = 23,133$ nodes and $E = 93,439$ edges, which give $\langle k \rangle = 8.1$. This network is composed of the largest connected component with 21,363 nodes and 566 small components with 3.1 nodes on average. The average path length and the network diameter are $\langle l \rangle = 5.4$ and $l_{\rm max} = 15$, respectively. The Spearman's correlation coefficient of the coauthorship network is $\rho = 0.26$, which means a positive NNDC.

For these two real-world networks, we first calculate the average lth neighbor degree $k_l(k)$ given by Eq. (15). The results are presented by symbols in Fig. 3. The continuous curves in this figure indicate $k_l(k)$ for random networks with the same degree sequences as the real networks, which is calculated from Eq. (15) by replacing $P(k'|k,l)$ with $P_0(k'|k,l)$. The symbols for various l in Fig. 3(a) are approximately fitted by the corresponding curves. This implies that the Gnutella network has almost no LRDCs other than those by the finite-size effect. On the contrary, $k_l(k)$ for the coauthorship network [Fig. 3(b)] considerably deviates from the curves, and the discrepancy becomes more pronounced at the higher degrees. This result clearly demonstrates the LRDC in the coauthorship network in which the average lth neighbor degree is always larger than that expected for random networks.

We also evaluate, for these two networks, the average shortest path length $\langle l(k,k')\rangle$ between nodes of degrees k and k', which is defined by

$$\langle l(k,k')\rangle = \sum_l lP(l|k,k'). \qquad (30)$$

Figure 4 represents the results for the Gnutella and the coauthorship networks. The vertical axis indicates the average shortest path length rescaled by that for random networks with the same degree sequence, namely, $\langle l(k,k')\rangle_{\rm res} = \sum_l lP(l|k,k')/\sum_l lP_0(l|k,k')$. Although the maximum degrees $k_{\rm max}$ of these networks are larger than the range of k in Fig. 4 ($k_{\rm max} = 103$ for the Gnutella network and 279 for the coauthorship network), we depict the results only for $k, k' \le 30$, in which 99.1% and 96.6% of nodes in the Gnutella network and the coauthorship network are included, respectively. This is because $\langle l(k,k')\rangle_{\rm res}$ for large degrees becomes quite bumpy due to poor statistics by the less number of high degree nodes. We see from Fig. 4 that $\langle l(k,k')\rangle_{\rm res}$ for the Gnutella network is close to 1 independently of k and k'. This means that the network has almost no intrinsic LRDCs, which is consistent with the result shown in Fig. 3(a). In contrast, $\langle l(k,k')\rangle_{\rm res}$ for the coauthorship network is larger than unity. This clearly indicates repulsive correlations among nodes. In fact, the average path length $\langle l \rangle = 5.4$ for the coauthorship network is greater than $\langle l \rangle = 4.3$ for random networks with the same degree sequence, whereas $\langle l \rangle = 4.6$ for the Gnutella network does not change so much from $\langle l \rangle = 4.5$ for the corresponding random networks. The fact that, for the coauthorship network, $\langle l(k,k')\rangle_{\rm res}$ for small degrees is larger than that for large degrees demonstrates the LRDC in which small degree nodes strongly repel each other in this network.

V. CONCLUSIONS

In this paper, we have provided a general framework to analyze pairwise correlations between degrees of nodes at an arbitrary distance from each other in a complex network. In order to fully describe such long-range degree correlations (LRDCs) between degrees k and k' of two nodes separated by l in the sense of the shortest path length, we introduced the joint probability distribution $P(k,k',l)$ and four conditional probability distributions $P(l|k,k')$, $P(k'|k,l)$, $P(k,k'|l)$, and $P(k',l|k)$. These distribution functions are not independent, and several relations between them have been presented with the aid of the Bayes' theorem. It has also been shown that the above distributions include the probability distributions $P_{\rm nn}(k,k')$ and $P_{\rm nn}(k'|k)$ describing nearest neighbor degree correlations as a special case. Furthermore, we have analytically calculated these five distribution functions for random networks with a given degree distribution under the local tree assumption and the mean-field approximation. The results for Erdős-Rényi random graphs and scale-free random networks agree well with numerical ones. The probability distributions for random networks enable us to judge the existence of intrinsic LRDCs other than correlations caused by the finite-size effect in a given network and capture the feature of correlations. Finally, we analyzed LRDCs in real-world networks within the present framework and found that the coauthorship network possesses LRDCs in which small degree nodes strongly repel each other.

Although we have just prepared tools for analyzing LRDCs, it is quite interesting to study relations between LRDCs and many network properties such as the robustness of a network, fractality, synchronization, to name a few. Our joint and conditional probability distribution functions have three variables and are not easy to handle. Thus, it is also important to develop intuitive indices characterizing LRDCs, like a measure of the strength of the repulsive correlation between hub nodes, $\sum_l \sum_{k,k'>\langle k\rangle} lP(k,k',l)/\sum_l \sum_{k,k'>\langle k\rangle} lP_0(k,k',l)$, or the average degree of a terminal node of an l-chain, $\sum_k kQ(k|l)$, on the basis of these probability distributions.

ACKNOWLEDGMENTS

The authors thank S. Mizutaka and T. Hasegawa for fruitful discussions. This work was supported by a Grant-in-Aid for Scientific Research (No. 16K05466) from the Japan Society for the Promotion of Science. T.T. acknowledges the financial support through JST ERATO Grant Number JPMJER1201, Japan.

Appendix: Equation (16) for random networks

We numerically investigated whether Eq. (16) holds for finite random networks. To this end, the conditional distributions $P(k,k'|l)$ and $Q(k|l)$ have been computed individually for 10^5 realizations of Erdős-Rényi random graphs with $N = 1,000$ and $\langle k \rangle = 5.0$. The average shortest path length $\langle l \rangle$ of the networks is 4.47. Figure 5 compares the k and k' dependence of $P(k,k'|l)$ with that of the product $Q(k|l)Q(k'|l)$ for several values of l. The distribution $P(k,k'|l)$ coincides with $Q(k|l)Q(k'|l)$ for $l \ll \langle l \rangle$, which demonstrates the validity of Eq. (16) for small l. On the contrary, $P(k,k'|l)$ deviates from $Q(k|l)Q(k'|l)$ for $l \gg \langle l \rangle$. These results clearly show that even random networks are not rigorously long-range uncorrelated in the sense of Eq. (16) if the network size is finite. However, LRDCs in random networks are caused only by the finite-size effect, and there is no intrinsic degree correlations other than them.

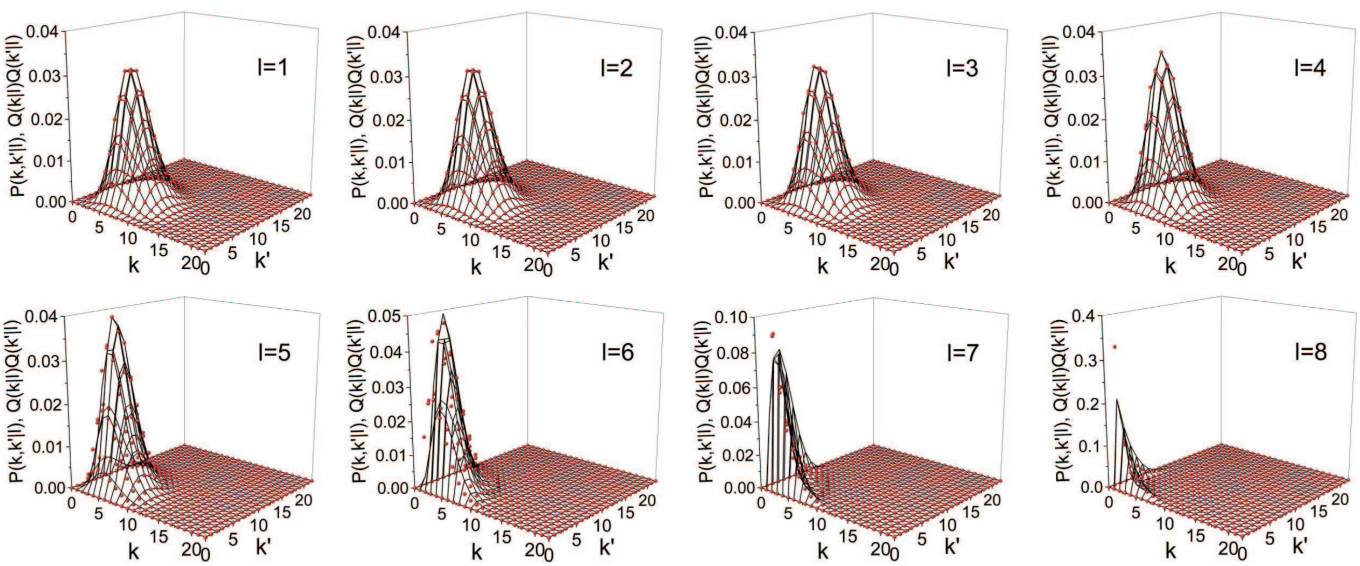

FIG. 5. (Color online) Probability distributions $P(k,k'|l)$ (wireframes) and $Q(k|l)Q(k'|l)$ (symbols) as functions of k and k' for Erdős-Rényi random graphs with $N = 1,000$ and $\langle k \rangle = 5.0$. Each panel represents the result for $l = 1$ to 8 from the top left to the bottom right.

[1] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[2] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[3] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[4] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[5] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
[6] J. Park and M. E. J. Newman, Phys. Rev. E **68**, 026112 (2003).
[7] M. Catanzaro, G. Caldarelli, and L. Pietronero, Phys. Rev. E **70**, 037101 (2004).
[8] M. A. Serrano, M. Boguñá, and R. Pastor-Satorras, Phys. Rev. E **74**, 055101 (2006).
[9] B. Fotouhi and M. G. Rabbat, Eur. Phys. J. B **86**, 510 (2013).
[10] N. Litvak and R. van der Hofstad, Phys. Rev. E **87**, 022801 (2013).
[11] S. N. Soffer and A. Vázquez, Phys. Rev. E **71**, 057101 (2005).
[12] M. A. Serrano and M. Boguñá, Phys. Rev. E **72**, 036133 (2005).
[13] J. C. Miller, Phys. Rev. E **80**, 020901 (2009).
[14] R. van der Hofstad, A. J. E. M. Janssen, J. S. H. van Leeuwaarden, and C. Stegehuis, Phys. Rev. E **95**, 022307 (2017).
[15] J. Menche, A. Valleriani, and R. Lipowsky, Phys. Rev. E **81**, 046103 (2010).
[16] R. Xulvi-Brunet and I. M. Sokolov, Phys. Rev. E **70**, 066102 (2004)
[17] S.-H. Yook, F. Radicchi, and H. Meyer-Ortmanns, Phys. Rev. E **72**, 045105 (2005).
[18] C. Song, S. Havlin, and H. A. Makse, Nat. Phys. **2**, 275 (2006).
[19] Z.-W. Wei and B.-H. Wang, Phys. Rev. E **94**, 032309 (2016).
[20] V. M. Eguíluz and K. Klemm, Phys. Rev. Lett. **89**, 108701 (2002).
[21] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **66**, 047104 (2002).
[22] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003).
[23] T. Gross, Carlos J. Dommar D'Lima, and B. Blasius, Phys. Rev. Lett. **96**, 208701 (2006).
[24] J. Hindes and I. B. Schwartz, Phys. Rev. E **95**, 052317 (2017).
[25] F. Sorrentino, M. di Bernardo, G. H. Cuéllar, and S. Boccaletti, Physica D **224**, 123 (2006).
[26] M. di Bernardo, F. Garofalo, and F. Sorrentino, Int. J. Bifurcation Chaos **17**, 3499 (2007).
[27] M. D. LaMar and G. D. Smith, Phys. Rev. E **81**, 046206 (2010).
[28] C. E. La Rocca, L. A. Braunstein, and P. A. Macri, Physica A **390**, 2840 (2011).
[29] V. Avalos-Gaytan, J. A. Almendral, D. Papo, S. E. Schaeffer, and S. Boccaletti, Phys. Rev. E **86**, 015101 (2012).
[30] S. Jalan, A. Kumar, A. Zaikin, and J. Kurths, Phys. Rev. E **94**, 062202 (2016).
[31] Z. Rong, X. Li, and X. Wang, Phys. Rev. E **76**, 027101 (2007).
[32] Z. Rong and Z.-X. Wu, Europhys. Lett. **87**, 30001 (2009).
[33] S. Devlin and T. Treloar, Phys. Rev. E **80**, 026105 (2009).
[34] R. J. La, IEEE/ACM Trans. Netw. **25**, 2484 (2017).
[35] J. D. Noh, Phys. Rev. E **76**, 026116 (2007).
[36] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, Phys. Rev. E **78**, 051105 (2008).
[37] C. M. Schneider, A. A. Moreira, J. S. Andrade Jr., S. Havlin, and H. J. Herrmann, Proc. Natl. Acad. Sci. **108**, 3838 (2011).
[38] T. Tanizawa, S. Havlin, and H. E. Stanley, Phys. Rev. E **85**, 046109 (2012).
[39] S. Watanabe and Y. Kabashima, Phys. Rev. E **94**, 032308 (2016).
[40] H. D. Rozenfeld, S. Havlin, and D. ben-Avraham, New J. Phys. **9**, 175 (2007)
[41] Y. Fujiki, S. Mizutaka, and K. Yakubo, Eur. Phys. J. B **90**, 126 (2017).
[42] C. Orsini, M. M. Dankulov, P. Colomer-de-Simón, A. Jamakovic, P. Mahadevan, A. Vahdat, K. E. Bassler, Z. Toroczkai, M. Boguñá, G. Caldarelli, S. Fortunato, and D. Krioukov, Nat. Commun. **6**, 8627 (2015).
[43] B. Tadić, S. Thurner, and G. J. Rodgers, Phys. Rev. E **69**, 036102 (2004).
[44] M. Boguñá, C. Castellano, and R. Pastor-Satorras, Phys. Rev. Lett. **111**, 068701 (2013).
[45] R. E. Boulos, A. Arneodo, P. Jensen, and B. Audit, Phys. Rev. Lett. **111**, 118102 (2013).
[46] L. W. Swanson, O. Sporns, and J. D. Hahn, Proc. Natl. Acad. Sci. **113**, E5972 (2016).
[47] D. Rybski, H. D. Rozenfeld, and J. P. Kropp, Europhys. Lett. **90**, 28002 (2010).
[48] M. Mayo, A. Abdelzaher, and P. Ghosh, Comput. Soc. Netw. **2**, 4 (2015).
[49] A. Arcagni, R. Grassi, S. Stefani, and A. Torriero, Eur. J. Oper. Res. **262**, 708 (2017).
[50] A. Allen-Perkins, J. M. Pastor, and E. Estrada, Appl. Math. Comput. **311**, 262 (2017).
[51] More precisely, $k_l(k)$ is the average degree of nodes separated by l from all nodes of degree k. In other words, $k_l(k)$ is the average lth neighbor degree of a *typycal* node of degree k, where the typical node of degree k is a node of degree k with the average number of lth neighbors.
[52] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[53] S. Melnik and J. P. Gleeson, Preprint arXiv:1604.05521.
[54] While Ref. [53] calculates the distance distribution for a network consisting of a connected component, our treatment covers networks composed of multiple disconnected components.
[55] http://snap.stanford.edu/data/p2p-Gnutella04.html
[56] http://snap.stanford.edu/data/ca-CondMat.html
[57] W.-Y. Zhang, Z.-W. Wei, B.-H. Wang, and X.-P. Han, Physica A **451**, 440 (2016).