Hierarchy in directed random networks

Enys Mones*
Department of Biological Physics, Eötvös Loránd University, Pázmány Péter Sétány 1/A, H-1117 Budapest, Hungary
(Received 4 September 2012; published 25 February 2013)

In recent years, the theory and application of complex networks have been quickly developing in a markable way due to the increasing amount of data from real systems and the fruitful application of powerful methods used in statistical physics. Many important characteristics of social or biological systems can be described by the study of their underlying structure of interactions. Hierarchy is one of these features that can be formulated in the language of networks. In this paper we present some (qualitative) analytic results on the hierarchical properties of random network models with zero correlations and also investigate, mainly numerically, the effects of different types of correlations. The behavior of the hierarchy is different in the absence and the presence of giant components. We show that the hierarchical structure can be drastically different if there are one-point correlations in the network. We also show numerical results suggesting that the hierarchy does not change monotonically with the correlations and there is an optimal level of nonzero correlations maximizing the level of hierarchy.

DOI: 10.1103/PhysRevE.87.022817 PACS number(s): 89.75.Hc, 89.65.−s, 87.23.Ge

I. INTRODUCTION

The application of complex networks in a broad range of social and biological systems has been a subject of much interest recently [1–5]. These applications involve, for example, the description of small-world properties of networks [6] or the consequences of scale-free degree distributions [7]. Many aspects of the real systems can be studied in the framework of undirected networks [8–11]. The most important property of the network is the degree distribution p_k, which is the probability of a randomly chosen node having k edges [7]. Other features of the network can be understood through the degree distribution (average length of the shortest path between nodes, average number of edges between a node's neighbors [12], characteristics of epidemics on the network [10], or robustness against failures and attacks [13]).

However, most of the real networks are directed, i.e., the connection between two units of the system is not symmetric. Many structural properties of a directed network can be derived from undirected networks in a straightforward way [8], but the appearance of directionality also opens the door to features that are essentially different from those in undirected graphs (Fig. 1).

In the presence of directed edges, the organization level of the nodes on a large scale can be very complex and flow hierarchy can emerge. It is the global structure of the network that results from the different roles of the nodes (see Fig. 2 for a comparison of different hierarchy types).

As the growing number of findings show, hierarchy is a frequently appearing property of real networks, especially of networks describing social interactions [17–21]. The concept of hierarchy has led to different definitions (Fig. 2) and also to algorithms for measuring it in both directed and undirected networks [22–27]. In this paper we investigate flow hierarchy by a corresponding measure, the global reaching centrality G_R, that has the intent to quantify that [28]. In the following sections we show that the behavior of G_R is different below and above the critical average degree k_c, i.e., in the absence and presence of giant components. We also give an approximation to its dependence on the average degree for different random network models when there are no degree correlations. Finally, we study the effects of degree correlations.

II. REACHING CENTRALITIES IN UNCORRELATED RANDOM NETWORKS

A. Local and global reaching centralities

Given a directed graph of the set of vertices V and edges E [$G(V,E)$ with $N = |V|$ vertices and $M = |E|$ edges], the local reaching centrality of node i is defined as the number of reachable nodes via out-edges divided by the total number of nodes

$$c_R(i) = \frac{|S_i|}{N-1} = \frac{C_R(i)}{N-1}, \tag{1}$$

where $S_i = \{j \in V | 0 < d^{\text{out}}(i,j) < \infty\}$ is the set of nodes that has a finite, nonzero out-distance from node i. We will denote the size of the reachable set (i.e., the local reaching centrality without normalization) by C_R. The global reaching centrality of the graph is the normalized sum of the distances from the maximum local reaching centrality

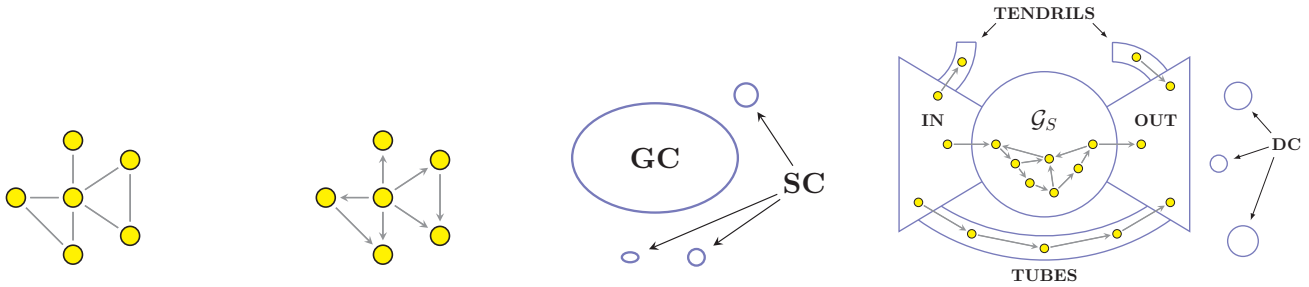
$$G_R = \frac{1}{N-1}\sum_i \left[c_R^{\max} - c_R(i)\right], \tag{2}$$

where $c_R^{\max}$ denotes the largest local reaching centrality in the network. The normalization factor is the maximum possible value of the sum in a graph with N nodes (this can be achieved in a star graph). The definition of the G_R can be written in a more expressive form

$$G_R = \frac{N}{(N-1)^2}\left[C_R^{\max} - \langle C_R\rangle\right]. \tag{3}$$

Here $\langle\cdot\rangle$ is the average over the nodes. Thus G_R is proportional to the difference between the average and the maximum size of the reachable sets.

*enys@hal.elte.hu



(a) **Undirected graph** (b) **Directed graph** (c) **Giant component** (d) **Giant components**

FIG. 1. (Color online) Difference between (a) undirected and (b) directed graphs. (c) On a large scale, the structure of the undirected graph can be described by the largest connected component [the giant component (GC)] and the small components (SC) [14]. (d) In the case of a directed graph, the giant components are best summarized in the bow-tie diagram [15,16]: The primary core is the giant strongly connected component $\mathcal{G}_S$. Inside $\mathcal{G}_S$, each node can reach every other node. There are nodes that can reach the whole $\mathcal{G}_S$, but not vice versa (IN), and together with $\mathcal{G}_S$ they form the giant in-component $\mathcal{G}_{\text{in}}$. The situation is similar for the giant out-component $\mathcal{G}_{\text{out}}$, but in the reverse direction. There are nodes that connect the $\mathcal{G}_{\text{in}}$ and $\mathcal{G}_{\text{out}}$ components, but are not in $\mathcal{G}_S$; they form the tubes. There are also tendrils that attach to the $\mathcal{G}_{\text{in}}$ and $\mathcal{G}_{\text{out}}$ components. The rest of the nodes are in the small disconnected components (DC).

B. Generating the function method in directed networks

The calculation of the reachable set of a node is equivalent to the problem of finding the out-component, which is the union of the reachable set and the node itself. The out-component can be determined by the generalization of the generating function formalism developed by Newman *et al.* [8,12] to directed networks. Assuming that our graph has a joint degree distribution p_{ij}, which is the probability of a randomly chosen node having i in-degrees and j out-degrees, the corresponding double generating function can be defined as

$$g_{00}(x,y) = \sum_{i,j=0}^{\infty} p_{ij} x^i y^j \tag{4}$$

and the generating functions for the excess in- and out-degree distributions [8]

$$g_{10}(x,y) = \frac{1}{\langle k \rangle} \partial_x g_{00}(x,y), \tag{5}$$

$$g_{01}(x,y) = \frac{1}{\langle k \rangle} \partial_y g_{00}(x,y), \tag{6}$$

where $\langle k \rangle$ is the average degree of the network. Let π_s^{out} denote the probability that a randomly chosen node has an out-component of size s and ρ_s^{out} the probability that a randomly chosen edge points to an out-component of size s. Their generating functions are

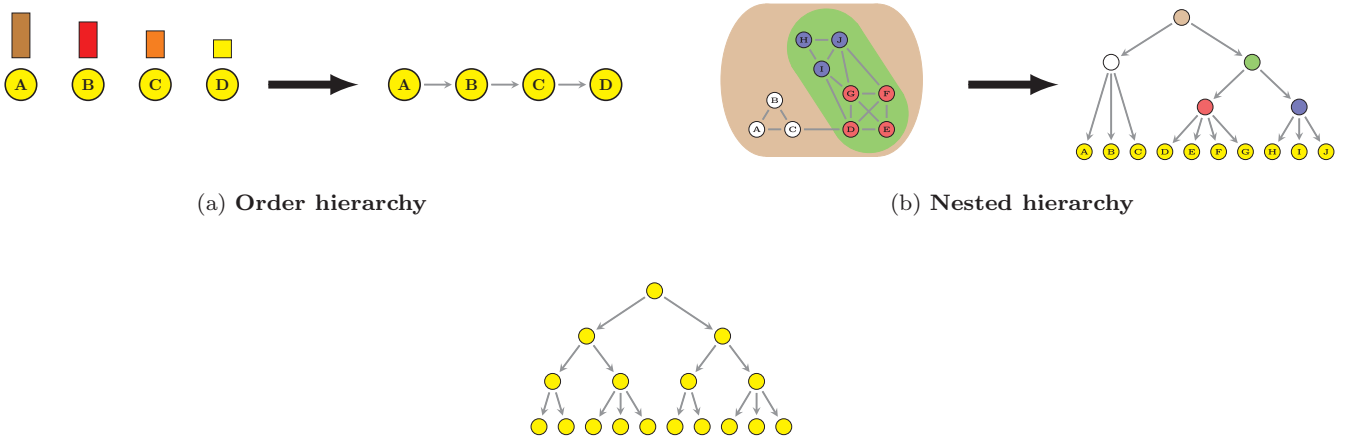
$$h_0(y) = \sum_{s=1}^{\infty} \pi_s^{\text{out}} y^s, \tag{7}$$

$$h_1(y) = \sum_{s=1}^{\infty} \rho_s^{\text{out}} y^s. \tag{8}$$

If we assume that the graph is locally treelike (i.e., loops are infrequent), these functions satisfy the following equations [8]:

$$h_0(y) = y g_{00}[1, h_1(y)], \tag{9}$$

$$h_1(y) = y g_{10}[1, h_1(y)]. \tag{10}$$



(a) **Order hierarchy** (b) **Nested hierarchy**

(c) **Flow hierarchy**

FIG. 2. (Color online) Three different types of hierarchy. (a) The order hierarchy is simply a rank assigned to each unit making them an ordered set. (b) The nested hierarchy is the hierarchy of the nested clusters of nodes. (c) In a (complete) flow hierarchy, nodes can be layered in different levels so that the nodes that are influenced by other nodes (via an out-edge) are at lower levels. As the figures illustrate, each type can be transformed into a flow hierarchy. In the order hierarchy one can introduce a directed edge between every pair of adjacent nodes in the hierarchy. In the nested hierarchy one can assign a virtual node to each cluster and link a cluster to its contained clusters.

Using these equations for the generating functions, it is possible to derive a closed formula for the out-components in directed graphs (for details see Appendix A):

$$\pi_s^{\text{out}} = \frac{\langle k \rangle}{(s-1)!}\left[\frac{d^{s-2}}{dy^{s-2}}(g_{01}(1,y)g_{10}(1,y)^{s-1})\right]_{y=0}. \quad (11)$$

Thus, if we know the joint degree distribution, we can calculate the $P(C_R)$ distribution of the reachable sets [and thus the distribution of the local reaching centralities as well since it differs from $P(C_R)$ only by a scale factor]. In order to calculate $P(C_R)$ we have to determine the out-components and apply the $s = C_R + 1$ substitution. Knowing $P(C_R)$, we are able to calculate G_R at different average degrees.

It is important to note that Eq. (11) describes only the small reaching centralities, i.e., the distribution of local reaching centralities of the nodes outside the giant components. Above the critical average degree, where the giant components are already present, a significant fraction of the nodes can reach a finite fraction of the whole network. In this regime, there are three type of nodes besides the ones with small reaching centrality: the nodes inside $\mathcal{G}_S$, $\mathcal{G}_{\text{in}}$, and $\mathcal{G}_{\text{out}}$; each type can reach a different fraction of the graph. Thus, in the distribution of the local reaching centrality above the percolation threshold, a peak with finite width appears at large values. Before applying these equations to random networks with different degree distributions, we determine G_R for the case of a hierarchical tree.

C. Hierarchical tree

1. Local reaching centralities

In a tree graph with N vertices and a branching number d, nodes in the same level have the same local reaching centrality. Let us denote the size of reachable set of a node in the ℓth level by $C_R^{(\ell)}$. The number of nodes and the size of the reachable set in the ℓth level are (see Fig. 3)

$$k_\ell = d^\ell, \quad (12)$$

$$C_R^{(\ell)} = \frac{N - \sum_{j=0}^{\ell} d^j}{d^\ell} = \frac{N - \frac{d^{\ell+1}-1}{d-1}}{d^\ell}. \quad (13)$$

The probability that a randomly chosen node has a reachable set of size $C_R^{(\ell)}$ is k_ℓ/N. Substituting Eq. (12) in Eq. (13) gives

$$P\left(C_R^{(\ell)}\right) = \frac{N(d-1)+1}{N(d-1)C_R^{(\ell)} + Nd}. \quad (14)$$

In the asymptotic limit of $N \to \infty$, we get the following approximation for the probability of the normalized local reaching centrality (we also omit the ℓ index since in the mentioned limit the allowed discrete values of $C_R^{(\ell)}/(N-1)$ are becoming dense in the interval $[0,1]$ and $c_R^{(\ell)}$ becomes continuous):

$$P(c_R) \approx \frac{1}{Nc_R + \frac{d}{d-1}}. \quad (15)$$

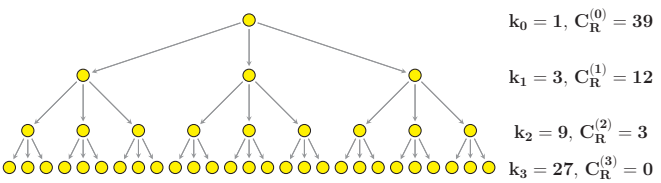


FIG. 3. (Color online) Number of nodes and size of reachable sets in the different levels of a hierarchical tree.

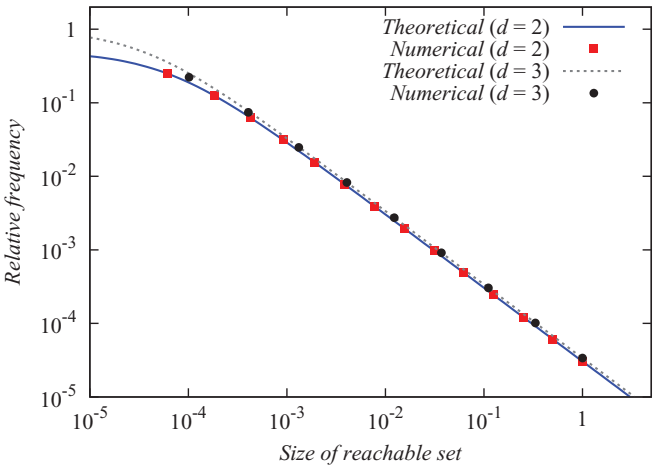


FIG. 4. (Color online) Theoretical and measured distributions of the local reaching centrality in the hierarchical tree. The dots are the results from graphs with $d = 2$ (red squares) and $d = 3$ (black circles) the corresponding numbers of nodes are $N_2 = 32\,767$ and $N_3 = 29\,524$.

A comparison with numerical results is plotted in Fig. 4. In the simulations, trees with branching numbers of 2 and 3 are used; the corresponding numbers of levels are 15 and 10. The numerical results show satisfactory agreement with Eq. (15).

2. Global reaching centrality

Using the equations for the number of nodes and size of reachable sets in the ℓth level, we can easily calculate the size of the reachable sets

$$\langle C_R \rangle = \sum_{\ell=0}^{L-1} P\left(C_R^{(\ell)}\right) C_R^{(\ell)} = \sum_{\ell=0}^{L-1}\left(1 - \frac{d^{\ell+1}-1}{N(d-1)}\right), \quad (16)$$

where L denotes the number of levels (the level of the root is zero). This number can be determined by the constraint that the sum of the nodes in all levels gives the number of nodes:

$$\sum_{\ell=0}^{L-1} d^\ell = \frac{d^L - 1}{d-1} = N. \quad (17)$$

Arranging for L and simplifying, we obtain

$$L = \frac{\ln[N(d-1)+1]}{\ln d}. \quad (18)$$

Calculating the sum on the right-hand side of Eq. (16) gives

$$\langle C_R \rangle = L\left(1 + \frac{1}{N(d-1)}\right) - \frac{d}{(d-1)}. \tag{19}$$

Substituting L into Eq. (19) and using Eq. (3) we get

$$G_R = \frac{N}{(N-1)^2}\Bigg[\underbrace{N-1}_{C_R^{\max}} - \underbrace{\frac{\ln[N(d-1)+1]}{\ln d}\left(1 + \frac{1}{N(d-1)}\right) + \frac{d}{(d-1)}}_{\langle C_R \rangle}\Bigg]. \tag{20}$$

In the asymptotic case of $N \to \infty$,

$$G_R = 1 + \frac{2d-1}{N(d-1)} - \frac{\ln[N(d-1)+1]}{N \ln d}. \tag{21}$$

In Fig. 5 we show a comparison of Eq. (20) with the simulations. In the numerical results, the number of nodes is 10^5 for every branching number. Thus Eq. (17) for the number of levels does not hold exactly: There are fewer nodes in the last level than expected. This fact is taken into account by summing up only to $L-2$ in Eq. (17):

$$\sum_{\ell=0}^{L-2} d^\ell = N. \tag{22}$$

This means that in the expression of L, the number of nodes is replaced by a reduced number (which is the number of nodes at the bottom level). This modification gives better agreement with the numerical calculations for large branching numbers. From Eq. (17) it is also clear that the approximation is more accurate if a branching number has an integer power close to 10^5. This can be observed in Fig. 5 as well.

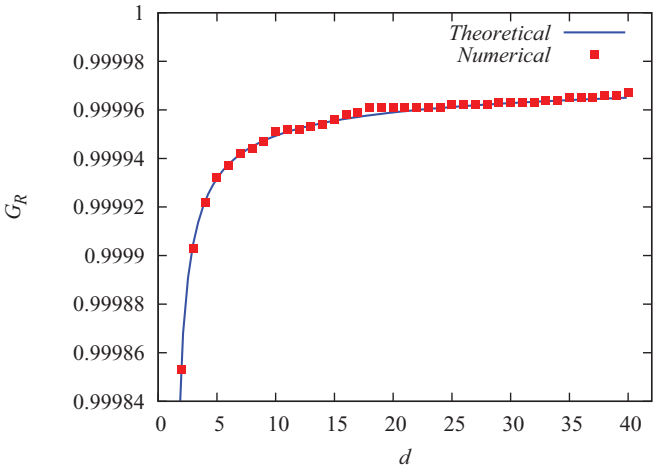


FIG. 5. (Color online) Comparison of the measured G_R with the theoretical prediction. The dots are the G_R of trees with $N = 10^5$ nodes. Because of the fixed number of nodes, the number of levels is approximated by Eq. (22), which neglects the bottom nodes in the total number of nodes. The deviations from the theoretical curve depend on the branching number. For branching numbers with an integer power close to 10^5, the theoretical curve is more accurate.

D. Erdős-Rényi graph

1. Local reaching centralities

In the case of uncorrelated in- and out-degrees, the joint degree distribution of an Erdős-Rényi (ER) graph [14,29] is the product of two independent Poisson distributions

$$p_{ij} = e^{-2\langle k \rangle} \frac{\langle k \rangle^{i+j}}{i!\,j!}, \tag{23}$$

where $\langle k \rangle$ denotes the average degree. The double generating function has the form

$$g_{00}(x,y) = e^{\langle k \rangle (x+y-2)}, \tag{24}$$

which is also the generating function of the excess degree distribution in this case. Now using Eq. (11) we get

$$\pi_s^{\text{out}} = \frac{\langle k \rangle}{(s-1)!}\left[\frac{d^{s-2}}{dy^{s-2}}(e^{\langle k \rangle s(y-1)})\right]_{y=0} = \frac{\langle k \rangle^{s-1} s^{s-2}}{(s-1)!} e^{-\langle k \rangle s}. \tag{25}$$

For the size of small reachable sets (without the giant components)

$$P(C_R) = \frac{\langle k \rangle^{C_R} (C_R+1)^{C_R-1}}{C_R!} e^{-\langle k \rangle (C_R+1)}. \tag{26}$$

Figure 6 shows this result compared with the numerical distributions. It is important to understand the limits of Eq. (26). It is only the distribution for the size of the small reachable sets, i.e., it does not contain $\mathcal{G}_S$ [30]. This can be seen on the plots. Before the transition, all of the nodes are in separated small components and the reaching centrality distribution vanishes very quickly. Above the transition point, they start to aggregate in $\mathcal{G}_S$. Note that in this regime, where the giant component appears, a large number of nodes have a large reachable set. More precisely, the nodes in the $\mathcal{G}_{\text{in}}$ component

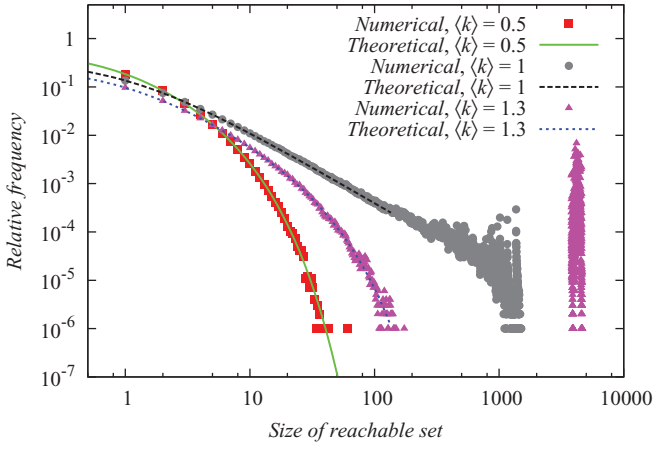


FIG. 6. (Color online) Distribution of the small reachable set sizes for the ER graph, based on the assumption that the graph is locally treelike. Below the transition point (red squares), there is no node that can reach any finite part of the graph. This changes when $\mathcal{G}_S$ appears at $\langle k \rangle = 1$ (gray circles). The numerical results are the averages of 1000 independent calculations on networks with $N = 10^4$. Note that Eq. (26) describes the distribution of the small components, thus the emerging $\mathcal{G}_S$ (the peak at large reachable sets) is beyond its scope.

can reach every node in $\mathcal{G}_{\text{out}}$. This is clearly seen in Fig. 6: The peak has a finite width, corresponding to the three different giant components.

2. *Global reaching centrality*

We have to distinguish between the graphs without $\mathcal{G}_S$ and with $\mathcal{G}_S$ [8]. Below the transition point k_c, there are only small components. All of the nodes can reach only an infinitesimal part in the graph, thus having an average reachable set size of order unity [$\langle C_R\rangle = O(1)$]. In this regime, G_R is dominated by the maximum value of the local reaching centrality. Since most of the nodes can reach very few other nodes and the distribution vanishes quickly, we can assume that the largest reachable set belongs to only one node (and that the corresponding out-component has the smallest relative frequency). Given a graph with N nodes, this condition translates as $P(C_R) \approx 1/N$ for $C_R = C_R^{\max}$. Thus, finding the reachable set size that has only one realization can lead us to find the largest component. In the $C_R^{\max} \gg 1$ limit we can use the Stirling formula to approximate $P(C_R)$:

$$P(C_R) \approx e^{1-\langle k\rangle}\frac{(\langle k\rangle e^{1-\langle k\rangle})^{C_R}}{\sqrt{2\pi}(C_R+1)^{3/2}}. \tag{27}$$

Writing the $P(C_R) \approx 1/N$ condition and rearranging it for C_R we get the following exponential equation:

$$\left(\frac{\sqrt{2\pi}}{Ne^{1-\langle k\rangle}}\right)^{2/3}(C_R+1) = \exp\left[\frac{2}{3}(\ln\langle k\rangle + 1 - \langle k\rangle)C_R\right]. \tag{28}$$

This equation can be solved in terms of the Lambert W function [31]. The equation of the form

$$A(x-R) = e^{-Bx} \tag{29}$$

has the solution of $x = R + \frac{1}{B}W[Be^{-Br}/A]$. Now we have

$$A = \left(\frac{\sqrt{2\pi}}{Ne^{1-\langle k\rangle}}\right)^{2/3}, \tag{30}$$

$$B = \frac{2}{3}(\ln\langle k\rangle + 1 - \langle k\rangle), \tag{31}$$

$$R = -1, \tag{32}$$

so the final expression for the largest local reaching centrality is

$$C_R^{\max} = -\frac{W\left[\frac{2}{3}(\ln\langle k\rangle + 1 - \langle k\rangle)e^{-(2/3)(\ln\langle k\rangle+1-\langle k\rangle)}\left(\frac{\sqrt{2\pi}}{Ne^{1-\langle k\rangle}}\right)^{-2/3}\right]}{\frac{2}{3}(\ln\langle k\rangle + 1 - \langle k\rangle)} - 1. \tag{33}$$

Now we turn our attention to the case when $\mathcal{G}_S$ is already present. In this case, to a good approximation, we can ignore those nodes that are not in the bow-tie (they are relevant only in the average local reaching centrality, but they have only an infinitesimal contribution). Thinking of the bow-tie picture, we can assume that there are some nodes that can reach the whole bow-tie. Thus $C_R^{\max} \approx |\mathcal{G}_{\text{in}}| + |\mathcal{G}_{\text{out}}| - |\mathcal{G}_S|$. The average is slightly different and nontrivial, but let assume that it is equivalent to the size of the giant out-component $\mathcal{G}_{\text{out}}$. If we assume that most of the nodes gather in $\mathcal{G}_S$, it is also reasonable that the average size of reachable sets is dominated by the nodes in $\mathcal{G}_S$ and they can reach the whole $\mathcal{G}_{\text{out}}$ component [see Fig. 1(d)]. Using these assumptions, the size of the reachable sets is approximately $|\mathcal{G}_{\text{out}}|$. The relative sizes of these components are the same as the local reaching centrality. In the generating function formalism they are given by

$$\frac{|\mathcal{G}_S|}{N} = 1 - g_{00}(u,1) - g_{00}(1,v) + g_{00}(u,v), \tag{34}$$

$$\frac{|\mathcal{G}_{\text{in}}|}{N} = 1 - g_{00}(u,1), \tag{35}$$

$$\frac{|\mathcal{G}_{\text{out}}|}{N} = 1 - g_{00}(1,v), \tag{36}$$

where u and v are the smallest nonzero solutions of the following equations [8]:

$$u = g_{01}(u,1), \tag{37}$$

$$v = g_{10}(1,v). \tag{38}$$

Since the generating function is symmetrical in its variables, we have $u = v$. Using Eq. (24), the solution can be written in terms of the Lambert W function, giving

$$u(\langle k\rangle) = -\frac{1}{\langle k\rangle}W(-\langle k\rangle e^{-\langle k\rangle}). \tag{39}$$

For G_R above the phase transition,

$$G_R = \frac{|\mathcal{G}_{\text{in}}|}{N} - \frac{|\mathcal{G}_S|}{N} = g_{00}[u(\langle k\rangle),1] - g_{00}[u(\langle k\rangle),u(\langle k\rangle)]. \tag{40}$$

The theoretical curve and the measurements are shown in Fig. 7. The theoretical curve in the $k_c < \langle k\rangle < 2$ range is a little below the numerical results. This is in good agreement with the assumptions used in deriving Eq. (40): The average reachable set is approximated by the $\mathcal{G}_{\text{out}}$ component; however, at small average degrees, there are many small out-components that decrease the average. The relative sizes of the giant components are depicted in Fig. 8. It is clearly seen that near the critical average degree, an appreciable portion of the nodes is outside the giant components and they all have a very small reachable set. Thus the difference between the largest and average reaching centralities is larger in the real network, resulting in a larger G_R. The theoretical curve below k_c predicts lower G_R, which indicates that the distribution of $P(C_R)$ described by Eq. (26) is no longer valid for the whole network when approaching the critical average degree.

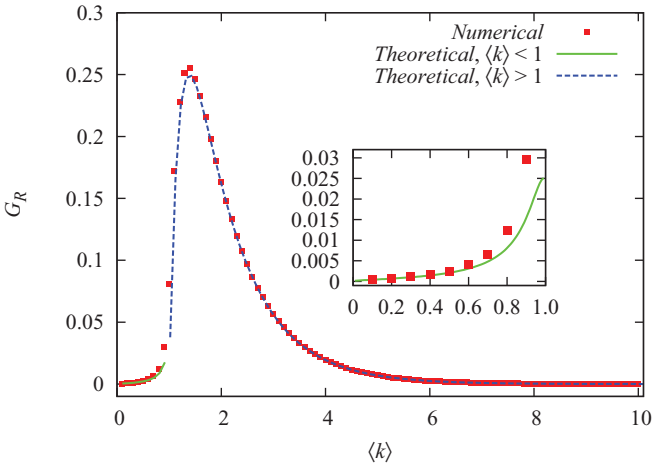


FIG. 7. (Color online) Global reaching centrality in the ER graph. The dots are the average of 1000 independent calculations on networks with $N = 10^4$. The errors are comparable to the size of the dots. The lines show the different approximations in the two regimes: below the percolation threshold (solid green line) and above it (dashed blue line). Both approximations tend to deviate from the numerical values near the transition point.

E. Exponential network

1. Local reaching centralities

In this section we calculate $P(C_R)$ and G_R for the exponential network, motivated by the finding that the distribution of many real-world networks can be well fitted by an exponential [32]. An uncorrelated exponential network has a joint degree distribution of the form

$$p_{ij} = (1 - e^{-1/\kappa})^2 e^{-(i+j)/\kappa}, \tag{41}$$

where the average degree can be obtained from the κ parameter: $\langle k\rangle = (e^{1/\kappa} - 1)^{-1}$. The double generating function and

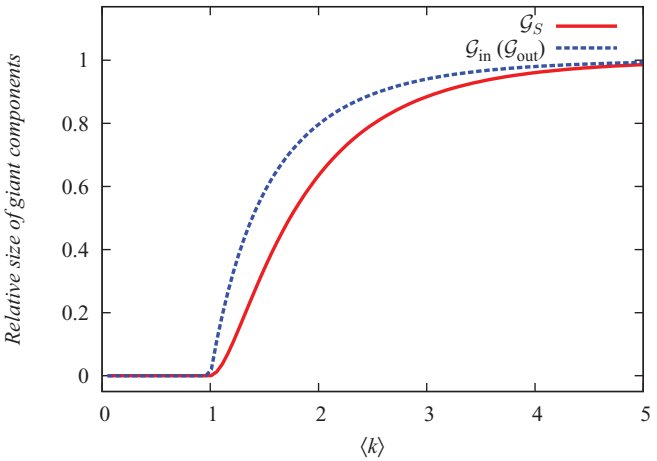


FIG. 8. (Color online) Relative sizes of the giant components of the bow-tie diagram in the ER network [Fig. 1(d)]. Although $\mathcal{G}_{\rm out}$ (dashed blue line) grows quickly with the average degree, at low edge densities ($\langle k\rangle < 2$) both components contain less than 80% of the nodes.

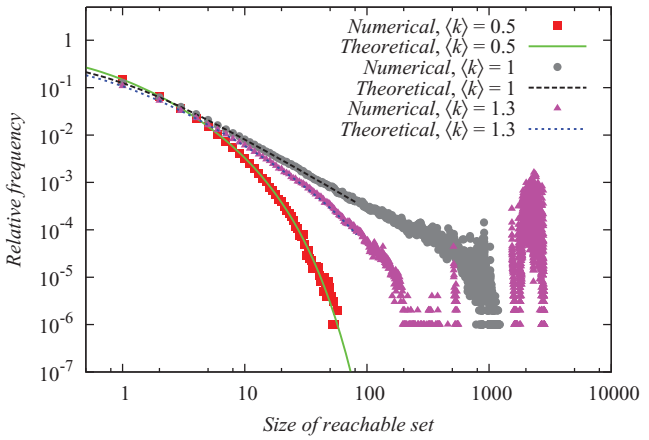


FIG. 9. (Color online) Distribution of the small reachable set sizes in the exponential graph. Numerical results are the average of 1000 measurement on networks with $N = 10^4$. The emergence of the giant components (which are not included by the analytical distribution) can be clearly seen at $\langle k\rangle = 1$ (gray circles) and $\langle k\rangle = 1.3$ (purple triangles).

its derivatives are

$$g_{00}(x,y) = \frac{(e^{1/\kappa} - 1)^2}{(e^{1/\kappa} - x)(e^{1/\kappa} - y)}, \tag{42}$$

$$g_{10}(x,y) = \frac{1}{\langle k\rangle}\frac{(e^{1/\kappa} - 1)^2}{(e^{1/\kappa} - x)^2(e^{1/\kappa} - y)}, \tag{43}$$

$$g_{01}(x,y) = \frac{1}{\langle k\rangle}\frac{(e^{1/\kappa} - 1)^2}{(e^{1/\kappa} - x)(e^{1/\kappa} - y)^2}. \tag{44}$$

By the formula for the out-components we have

$$\pi_s^{\rm out} = \frac{1}{\langle k\rangle^s (s-1)!}\frac{(2s-2)!}{s!}e^{-(2s-1)/\kappa}. \tag{45}$$

Substituting $e^{1/\kappa} = \frac{\langle k\rangle+1}{\langle k\rangle}$ we get

$$\pi_s^{\rm out} = \frac{(2s-2)!}{\langle k\rangle^s (s-1)!s!}\left(\frac{\langle k\rangle}{\langle k\rangle+1}\right)^{2s-1}. \tag{46}$$

Translating this to the small reachable sets (see Fig. 9 for a comparison with numerical calculations) results in

$$P(C_R) = \frac{(2C_R)!}{\langle k\rangle^{C_R+1} C_R!(C_R+1)!}\left(\frac{\langle k\rangle}{\langle k\rangle+1}\right)^{2C_R+1}. \tag{47}$$

2. Global reaching centrality

The first task is to find the smallest solution of the equations $u = g_{01}(u,1)$ and $v = g_{10}(1,v)$. Since we assume uncorrelated exponential distributions, we have $u = v$. The equation

$$u = g_{01}(u,1) = \frac{1}{\langle k\rangle(e^{1/\kappa} - u)} \tag{48}$$

is quadratic in u and has two solutions $u_1 = 1$ and $u_2 = 1/\langle k\rangle$, thus for $\langle k\rangle < 1$ there are no giant components and we can use the rarest component assumption as before with the ER graph. To do this we have to solve the following equation for C_R:

$$\frac{(2C_R)!}{\langle k\rangle^{C_R+1} C_R!(C_R+1)!}\left(\frac{\langle k\rangle}{\langle k\rangle+1}\right)^{2C_R+1} = \frac{1}{N}. \tag{49}$$

Using the Stirling formula (without going into the details), this equation can be approximated by

$$\left(\frac{(\langle k\rangle+1)\sqrt{\pi}}{Ne}\right)\left(\frac{C_R+1}{C_R}\right)^{C_R}(C_R+1)^{3/2} = \exp\left[\frac{2}{3}\ln\left(\frac{4\langle k\rangle}{(\langle k\rangle+1)^2}\right)C_R\right]. \quad (50)$$

We assume that the rarest component is much larger than one, thus $(\frac{C_R+1}{C_R})^{C_R}\approx e$. This reduces our equation and we can rewrite it as

$$AC_R = e^{-BC_R}, \quad (51)$$

where we used the shorthand notation

$$A = \left(\frac{(\langle k\rangle+1)\sqrt{\pi}}{N}\right)^{2/3}, \quad (52)$$

$$B = -\frac{2}{3}\ln\left(\frac{4\langle k\rangle}{(\langle k\rangle+1)^2}\right) \quad (53)$$

and neglected the additional 1 on the left-hand side. The solution of Eq. (50) is then

$$C_R^{\max} = \frac{1}{B}W\left(\frac{B}{A}\right). \quad (54)$$

If $\langle k\rangle > 1$, this approximation fails because of the appearance of the giant components. Using the solution of Eq. (48), the relative sizes of the parts in the bow-tie diagram are

$$\frac{|\mathcal{G}_S|}{N} = \left(1-\frac{1}{\langle k\rangle}\right)^2, \quad (55)$$

$$\frac{|\mathcal{G}_{\text{in}}|}{N} = 1-\frac{1}{\langle k\rangle}, \quad (56)$$

$$\frac{|\mathcal{G}_{\text{out}}|}{N} = 1-\frac{1}{\langle k\rangle}. \quad (57)$$

By the same argument as with the ER graph, we get for G_R

$$G_R = 1-\frac{1}{\langle k\rangle}-\left(1-\frac{1}{\langle k\rangle}\right)^2. \quad (58)$$

This result is shown in Fig. 10 along with the numerical results. The same argument can be applied for the exponential network as in the preceding section. The predicted curves fit well in the very small and in the large average degree regimes.

F. Scale-free network

Without an exponential cutoff, the probability distribution and double generating function of a scale-free network [33,34] are

$$p_{ij} = \frac{(ij)^{-\gamma}}{\zeta(\gamma)^2}, \quad (59)$$

$$g_{00}(x,y) = \frac{\text{Li}_\gamma(x)\text{Li}_\gamma(y)}{\zeta(\gamma)^2}, \quad (60)$$

where $\text{Li}_n(x)$ is the polylogarithm function of the nth order and $\zeta(x)$ is the Riemann zeta function. The generating functions

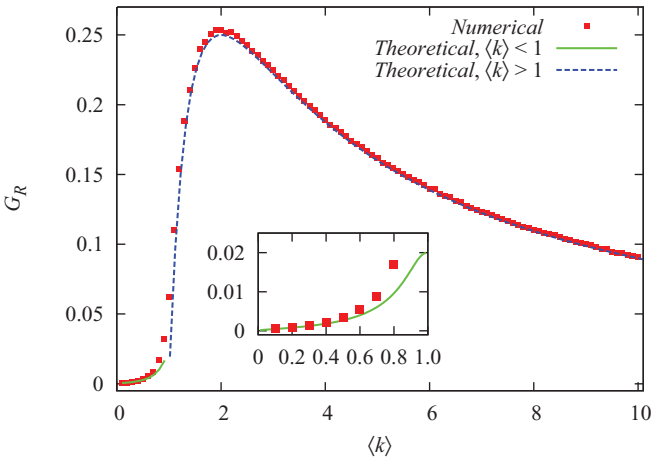


FIG. 10. (Color online) Global reaching centrality of the exponential graph and the predicted curves in the different regimes of $\langle k\rangle$. Each numerical point is the average of 1000 independent runs on networks with $N = 10^4$. The magnitude of the errors is on the order of the dot sizes.

of the excess degree distributions are

$$g_{10}(x,y) = \frac{\text{Li}_{\gamma-1}(x)\text{Li}_\gamma(y)}{\langle k\rangle x\zeta(\gamma)^2}, \quad (61)$$

$$g_{01}(x,y) = \frac{\text{Li}_\gamma(x)\text{Li}_{\gamma-1}(y)}{\langle k\rangle y\zeta(\gamma)^2}. \quad (62)$$

Newman showed in Ref. [8] that the condition for the existence of the giant components is

$$\partial_x\partial_y g_{00}(x,y)|_{x,y=1} > \langle k\rangle. \quad (63)$$

For the scale-free network, this equation reads

$$\left.\frac{\text{Li}_{\gamma-1}(x)\text{Li}_{\gamma-1}(y)}{xy\zeta(\gamma)^2}\right|_{x,y=1} > \langle k\rangle. \quad (64)$$

Now, if make use of the relation between the exponent and the average degree $\langle k\rangle = \frac{\zeta(\gamma-1)}{\zeta(\gamma)}$ and substitute $x=1$ and $y=1$ we get

$$\langle k\rangle^2 > \langle k\rangle \quad (65)$$

or equivalently

$$\langle k\rangle > 1. \quad (66)$$

Thus there is a giant component if the average degree is larger than one. However, if we look at the function

$$\frac{\zeta(\gamma-1)}{\zeta(\gamma)} \quad (67)$$

we can conclude that this condition gives giant components for any $\gamma \geqslant 2$. Numerical simulations show that this is not the case (see Fig. 11).

If we substitute $g_{00}(x,y)$ and its derivatives in Eqs. (3), (37), and (38) we observe that there are giant components of unit size for every exponent larger than 2 since the formula

$$\pi_s^{\text{out}} = \frac{\langle k\rangle}{(s-1)!}\left[\frac{d^{s-2}}{dy^{s-2}}(g_{01}(1,y)g_{10}(1,y)^{s-1})\right]_{y=0} = \frac{1}{\zeta(\gamma)^s(s-1)!}\left[\frac{d^{s-1}}{dy^{s-1}}(\text{Li}_\gamma(y)^s)\right]_{y=0} \quad (68)$$

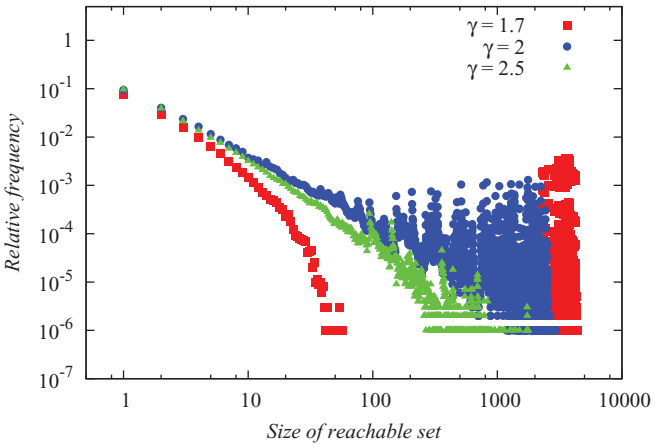


FIG. 11. (Color online) Distribution of the reachable set sizes in the scale-free graph. Each data point is the average of 1000 distributions on networks with $N = 10^4$. The giant components emerge at $\gamma = 2$ (blue circles) and vanish above this threshold (green triangles).

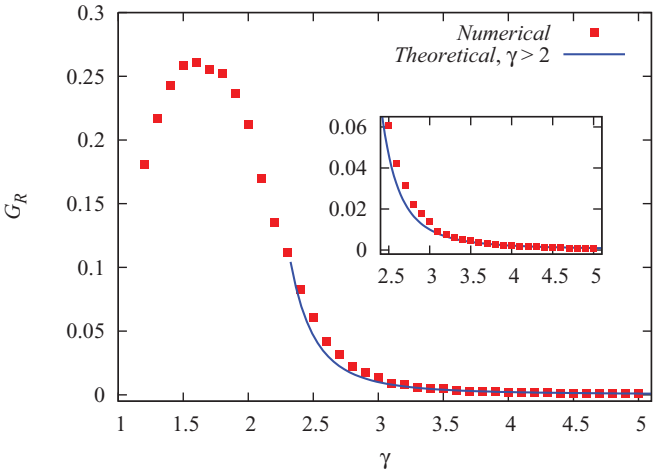


FIG. 12. (Color online) Global reaching centrality of the scale-free graph. Numerical results are the average of 1000 independent calculations on networks with $N = 10^4$. Errors have a magnitude of the dot sizes. The largest degree approximation fits well above $\gamma > 3$ and gives a lower value below it.

gives exactly zero for any s and the equation

$$u = g_{01}(u,1) = \frac{\mathrm{Li}_\gamma(u)}{\zeta(\gamma)} \tag{69}$$

has always two solutions $u_1 = 0$ and $u_2 = 1$.

These show the limitations of the generating function formalism in directed networks. The limits of the method are already well known in some cases of undirected scale-free networks as well [9]. In the undirected case, networks with small exponents tend to have many hubs and the clustering coefficient also increases remarkably [9]. A large clustering coefficient means that the graph is not locally treelike, which is the main assumption of the applied method. Equations (68) and (69) point out that, for directed networks, the method can barely be applied to scale-free networks.

However, it is possible to give a qualitative approximation for G_R in the case of $\gamma > 2$ (this is the most important regime in terms of real networks [7]). We use our observation from the simulations that there is no $\mathcal{G}_S$ and in scale-free networks very large degrees can appear. Since a large number of the nodes have few out-degrees, G_R is obviously dominated by $C_R^{\max}$. Let us assume that the network breaks down into small components that are the neighborhoods of the nodes with large out-degrees (i.e., every component gathers around a hub). In this case, the largest reachable set is the largest out-degree. In a scale-free network with a degree distribution of $p_k \propto k^{-\gamma}$, the largest degree is well approximated by $k_{\max} \approx N^{1/(\gamma-1)}$ [35]. Using this approximation for the out-degrees and not taking into account the in-degrees we get

$$G_R \simeq N^{(2-\gamma)/(\gamma-1)}. \tag{70}$$

A comparison of the real G_R and this approximation is shown in Fig. 12. For large exponents, Eq. (70) agrees well with the numerical results; below $\gamma = 3$, it becomes less accurate. Note that the lower the exponent, the larger the number of nodes that have a large degree. The predicted G_R becomes larger than the real value at some point ($\gamma \approx 2.3$), that is, the number of large hubs increases, which results in a larger average reaching centrality as well.

It is worth pausing here to understand the nonmonotonic dependence of G_R on the exponent. For large exponents, the scale-free network tends to behave similarly to the Erdős-Rényi graph. Thus it becomes more homogeneous. Its average degree also decreases and instead of the hubs, more and more small disconnected components appear. As the network becomes more fragmented, both the average and maximum reaching centrality decrease, which results in a very small G_R. In the case of small exponents, it is known that the clustering coefficient of undirected scale-free networks significantly increases [9]. This phenomenon is likely to be present in the directed case as well. With large clustering, the number of directed circles also increases. However, the nodes in a directed circle have exactly the same reachable set and thus similar local reaching centrality. This tendency of equalization in the local reaching centralities affects the heterogeneity of the distribution and enables multiple nodes to reach (or approach closely) the maximum reaching centrality. The result is again a decreasing G_R.

III. EFFECTS OF CORRELATIONS

A. Degree correlations

In the preceding sections we assumed that the joint probability distribution of the in- and out-degrees is simply the product of two independent distributions, i.e., there are no degree correlations. However, it is known that there are such correlations in real networks [36]. It is also known that these correlations have remarkable effects on the different properties of networks (percolation thresholds, epidemic thresholds, etc.) [37,38]. In this paper, we numerically study the effect of two types of correlations on G_R: one-point correlations and directed assortative mixing. The one-point correlation is the Pearson correlation between the in- and out-degree of the

TABLE I. One- and two-point correlations of real networks along with their references. With the exception of the metabolic networks, most of them have small correlations.

Network	$\langle k\rangle$	ρ	r
	Food networks		
Ythan [2]	4.452	0.168	−0.249
LittleRock [39]	13.628	−0.138	−0.394
Grassland [2]	1.557	−0.179	−0.233
	Electric		
s1488 [40]	2.085	−0.274	0.218
s5378 [40]	1.467	−0.137	0.151
s35932 [40]	1.683	−0.074	0.088
	Trust		
WikiVote [41]	14.573	0.318	−0.083
College [42,43]	3	0.053	−0.159
Prison [42,43]	2.716	0.201	0.129
	Regulatory		
TRN-Yeast-1 [44]	2.899	0.025	−0.173
TRN-Yeast-2 [45]	1.568	−0.236	−0.220
TRN-EC [45]	1.239	−0.082	0.085
	Metabolic		
C. elegans [3]	2.442	0.924	−0.174
E. coli [3]	2.533	0.923	−0.167
S. Cerevisiae [3]	2.537	0.923	0.182

nodes

$$\rho = \frac{\langle k_{\rm in}k_{\rm out}\rangle_V - \langle k_{\rm in}\rangle_V\langle k_{\rm out}\rangle_V}{\sigma_V(k_{\rm in})\sigma_V(k_{\rm out})}. \tag{71}$$

The angular brackets denote averages over the nodes and $\sigma_V(k) = \sqrt{\langle k^2\rangle_V - \langle k\rangle_V^2}$ is the standard deviation of the corresponding degrees. Somewhat similarly, one can define the Pearson correlation between the out-degree of the start of an edge and the in-degree of its end term [36]

$$r = \frac{\langle j_{\rm in}k_{\rm out}\rangle_E - \langle j_{\rm in}\rangle_E\langle k_{\rm out}\rangle_E}{\sigma_E(j_{\rm in})\sigma_E(k_{\rm out})}. \tag{72}$$

Here $j_{\rm in}$ is the in-degree of the node an edge points to, $k_{\rm out}$ is the out-degree of the node that an edge comes from, and $\sigma_E(k) = \sqrt{\langle k^2\rangle_E - \langle k\rangle_E^2}$, where the averages run over the edges. We will refer to this quantity as the two-point correlation or directed assortativity since it is a plausible generalization of the assortativity in undirected networks. Table I shows the two correlations in several real networks.

B. One-point correlations

1. Numerical results

In order to study the effect of the one-point correlation, we generated the in- and out-degree lists for the network with a given distribution and average degree. After randomizing both lists, we fixed the out-degree list while we successively swapped randomly chosen elements in the in-degrees. After every swap, we calculated $\rho(k_{\rm in},k_{\rm out})$ and accepted the new list whenever the correlation increased (decreased). We measured G_R when the difference between the correlation of the current state and the last measured state was larger than 0.01 [this is the resolution of the measured $G_R(\rho)$ function]. The dependence of G_R on the one-point correlation is shown in Fig. 13. It can be clearly seen that the behavior of G_R varies and depends strongly on the average degree. In the ER and exponential graphs, even the monotonicity of the curves changes. In the scale-free network, G_R saturates at some level of correlation and the exponent effects only the threshold above which further correlations do not change G_R.

2. Qualitative approach

In this section we present a simple argument to understand the direction of change if small correlations are present in the graph. It is important to emphasize that the analytic expression for the effects of correlations is beyond the scope of this paper and we intend to give only a rather qualitative insight of the behavior. For this we consider the following joint degree distribution:

$$p_{jk} = p_j p_k + \rho\sigma_p^2 m_{jk}, \tag{73}$$

where the only criteria for the m_{jk} matrix are

$$\sum_{j=0}^{\infty} m_{jk} = \sum_{k=0}^{\infty} m_{jk} = 0, \tag{74}$$

$$\sum_{j,k=0}^{\infty} jkm_{jk} = 1. \tag{75}$$

With these conditions, the one-point degree correlation of the above joint degree distribution is exactly ρ [36]. Note that we assumed the same distribution for the in- and out-degrees [$\sigma_V(k_{\rm in}) = \sigma_V(k_{\rm out}) = \sigma_p$]. The generating function of the modified distribution

$$g_{00}(x,y) = g_{00}^p(x,y) + \rho\chi_{00}(x,y), \tag{76}$$

where the g_{00}^p is the generating function of the original joint degree distribution and we introduced the following function:

$$\chi_{00}(x,y) = \sigma_p^2 \sum_{j,k=0}^{\infty} x^j y^k m_{jk}. \tag{77}$$

Above the critical average degree, we have to solve

$$u = g_{01}^p(u,1) + \frac{\rho}{\langle k\rangle}\partial_y\chi_{00}(u,1) \tag{78}$$

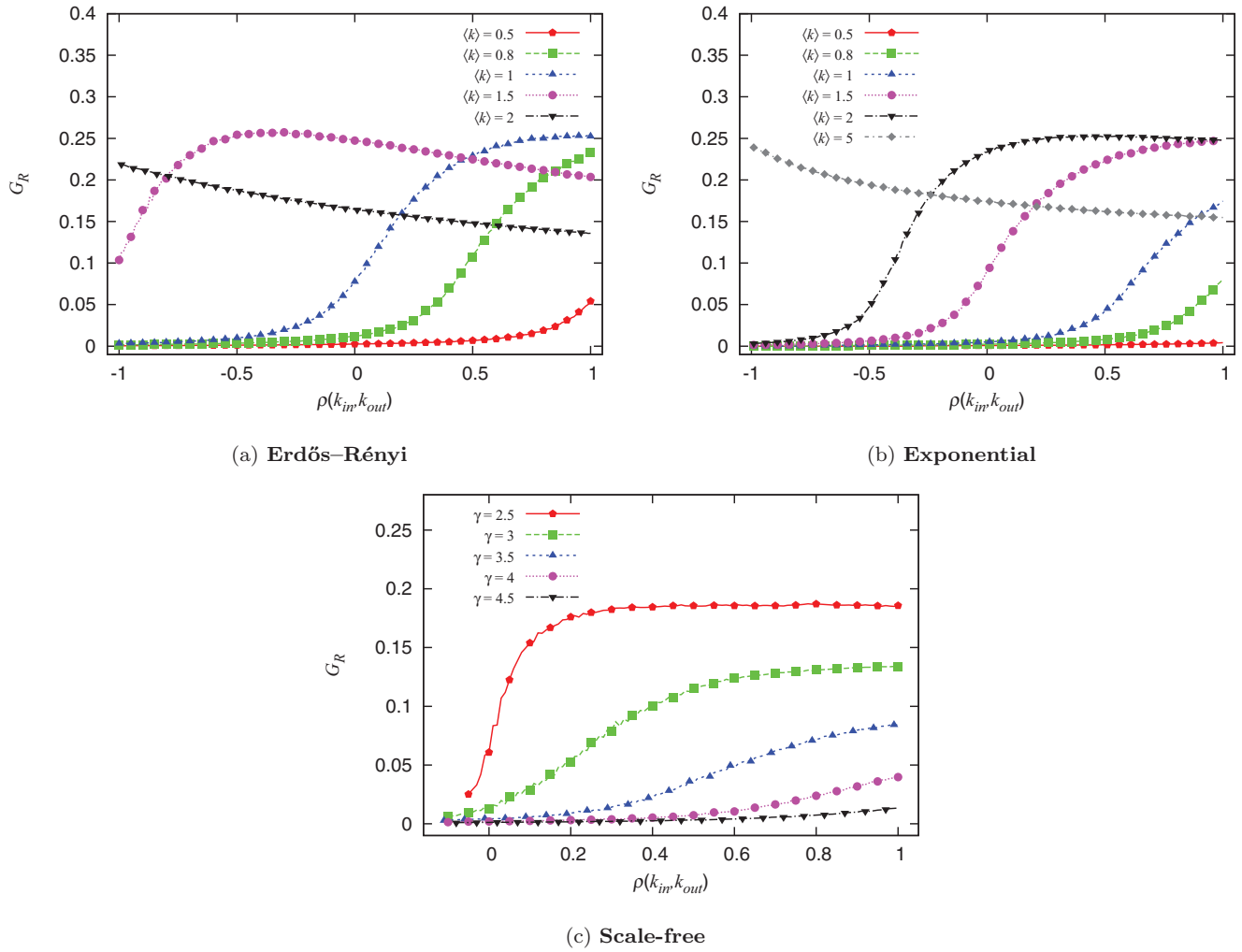
and G_R is given by the modified generating function

$$G_R = g_{00}(u,1) - g_{00}(u,u). \tag{79}$$

The exact solution of Eq. (78) depends on the actual choice of m_{jk} and in most cases it is not possible to find in closed form. However, for small correlations ($\rho \ll 1$), we can describe the change in G_R at a given average degree. In this case, the solution of Eq. (78) is very close to the solution of the original equation without the correlations: $u = u_0 + u_\rho$, where u_0 is the solution of Eq. (37) and u_ρ is the small change. Expanding both sides and keeping only the linear terms in ρ and u_ρ, we get

$$u_\rho = \frac{1}{\langle k\rangle}\frac{\partial_y\chi_{00}(u_0,1)}{1-\partial_x g_{01}^p(u_0,1)}\rho = \beta(u_0)\rho. \tag{80}$$

Here we encapsulated the coefficient of ρ in $\beta(u_0)$. This coefficient is negative above the transition point for the model



(a) **Erdős–Rényi** (b) **Exponential**

(c) **Scale-free**

FIG. 13. (Color online) The G_R versus the one-point correlations for different average degrees in the Erdős-Rényi and exponential graphs and for different exponents in the scale-free network. Each point on the plots is the average of at least 200 independent simulations on networks with $N = 10^4$ nodes. In the case of the scale-free network, only a small negative degree correlation is accessible through random optimization.

networks we are interested in (for the proof, see Appendix B). In the linear approximation, G_R looks like

$$G_R = G_R^{(0)} + G_R^{(\rho)}, \tag{81}$$

where

$$G_R^{(0)} = g_{00}^p(u_0,1) - g_{00}^p(u_0,u_0) \tag{82}$$

and

$$G_R^{(\rho)} = \left[\beta(u_0)(1-2u_0)\frac{dg_0^p(x)}{dx}\bigg|_{x=u_0} - \chi_{00}(u_0,u_0)\right]\rho. \tag{83}$$

In the derivation of $G_R^{(\rho)}$ we used the observation that when the joint degree distribution factorizes, the generating function is also a product of the single distributions, i.e.,

$$g_{00}^p(x,y) = g_0^p(x)g_0^p(y) \tag{84}$$

and also

$$g_0^p(u_0) = \sum_{j=0}^{\infty} p_j u_0^j = \frac{1}{\langle k\rangle}\sum_{j,k=0}^{\infty} k u_0^j p_{jk} = g_{01}^p(u_0,1) = u_0. \tag{85}$$

Now we have to interpret the result we obtained for the change in G_R. First, for simplicity, let us choose m_{jk} as proposed in Ref. [36]:

$$m_{jk} = \frac{(p_j - \phi_j)(p_k - \phi_k)}{(\langle k\rangle - \langle k\rangle_\phi)^2}, \tag{86}$$

where ϕ_j is an arbitrary normalized distribution and $\langle k\rangle_\phi$ is its average. Moreover, we choose the ϕ_j distribution such that $\langle k\rangle_\phi = 1$, which is the critical point of the ER and exponential networks. In this case $\chi_{00}(u_0,u_0) \geqslant 0$ and also note that

$$\beta(u_0) < 0, \tag{87}$$

$$\frac{dg_0^p(x)}{dx}\bigg|_{x=u_0} > 0 \tag{88}$$

above the transition point. The behavior of $G_R^{(\rho)}$ is governed by the relation of the two terms on the right-hand side of Eq. (83). It is easy to check that near the critical point ($u_0 \approx 1$), the second term is very close to zero and $G_R^{(\rho)}$ is dominated by the first term, which is positive. This means that when the

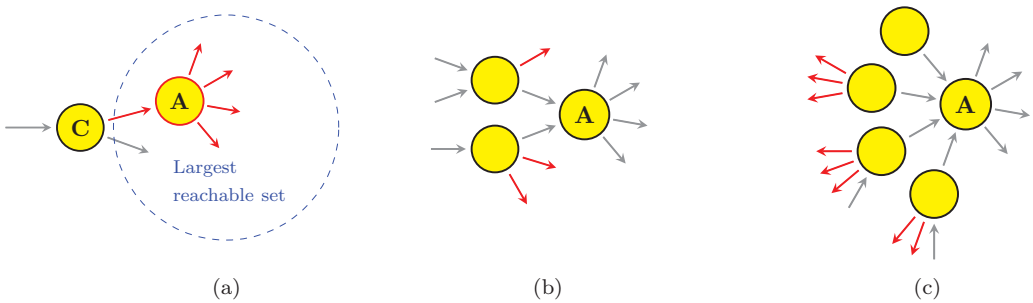


FIG. 14. (Color online) (a) Largest reachable set below the transition point and one of the nodes with large out-degree in the largest reachable set (A). (b) Without degree correlations, the expected number of in-degrees of A is the average degree and the maximum out-degree of its in-neighbors is also moderate. (c) With degree correlations, the nodes with a large out-degree have a large in-degree as well; with increasing number of in-neighbors the expected maximum out-degree of its neighbors increases as well.

average degree is small but the graph already has giant components of finite size, small one-point correlations increase its hierarchical structure, as in Fig. 13. For large average degrees, $u_0 \to 0$ and both terms in Eq. (83) becomes negative, resulting in a decreasing effect of small correlations, in good agreement with the numerical results [see Figs. 13(a) and 13(b)]. In the regime where G_R has a maximum ($u_0 = \frac{1}{2}$), Eq. (83) predicts a negative coefficient for ρ, but as Fig. 13(b) shows, this is not the case for the exponential network. This discrepancy suggests that the change in the slope near this point is not trivial and strongly depends on the details of the addition of correlations.

Below the transition point, there are no giant components and G_R is dominated by the largest reachable set. We can assume that the node with the largest reachable set [node C in Fig. 14(a)] or some of its out-neighbors have a large out-degree since this increases the probability of reaching more nodes. Consider this neighbor with the large out-degree [node A in Fig. 14(a)]. In the case of negative or zero correlations, the expected maximum of the out-degree of its in-neighbors is small [Fig. 14(b)]. However, in the presence of positive one-point correlations, the number of in- and out-edges tend to be similar; in other words, the node in question has a large in-degree as well. In this case, the expected maximum of the out-degree of its in-neighbors (and thus of the candidate node for the largest local reaching centrality) increases and there are more reachable nodes. The scale-free network is a special case in the sense that the largest reachable set is to a good approximation the neighborhood of a hub. Because of the correlations, this hub tends to have many in-neighbors as well and is more likely to have a neighbor with many links [Fig. 14(d)]. The numerical results in Fig. 13(c) suggest that the maximum out-degree of the nearest neighbors reaches its saturation level at relatively small correlations.

C. Two-point correlations

The two-point correlation of the type we investigate in this paper is the generalization of the assortativity to directed networks [36]. In order to introduce this correlation in our calculations, let $\mathcal{P}(j_{\text{in}}, j_{\text{out}}; k_{\text{in}}, k_{\text{out}})$ denote the probability that a randomly chosen edge starts at a node with j_{in} in- and j_{out} out-degrees and points to a node with k_{in} in- and k_{out} out-degrees. This probability satisfies the following equations:

$$\sum_{j_{\text{in}}, j_{\text{out}}=0}^{\infty} \mathcal{P}(j_{\text{in}}, j_{\text{out}}; l_{\text{in}}, l_{\text{out}}) = \sum_{k_{\text{in}}, k_{\text{out}}=0}^{\infty} \mathcal{P}(l_{\text{in}}, l_{\text{out}}; k_{\text{in}}, k_{\text{out}}) = p_{l_{\text{in}} l_{\text{out}}}, \tag{89}$$

which also creates its connection to the joint degree distribution. Without two-point correlations (and other correlations), the two-node joint distribution factorizes:

$$\mathcal{P}^0(j_{\text{in}}, j_{\text{out}}; l_{\text{in}}, l_{\text{out}}) = p_{j_{\text{in}}} p_{j_{\text{out}}} p_{k_{\text{in}}} p_{k_{\text{out}}}. \tag{90}$$

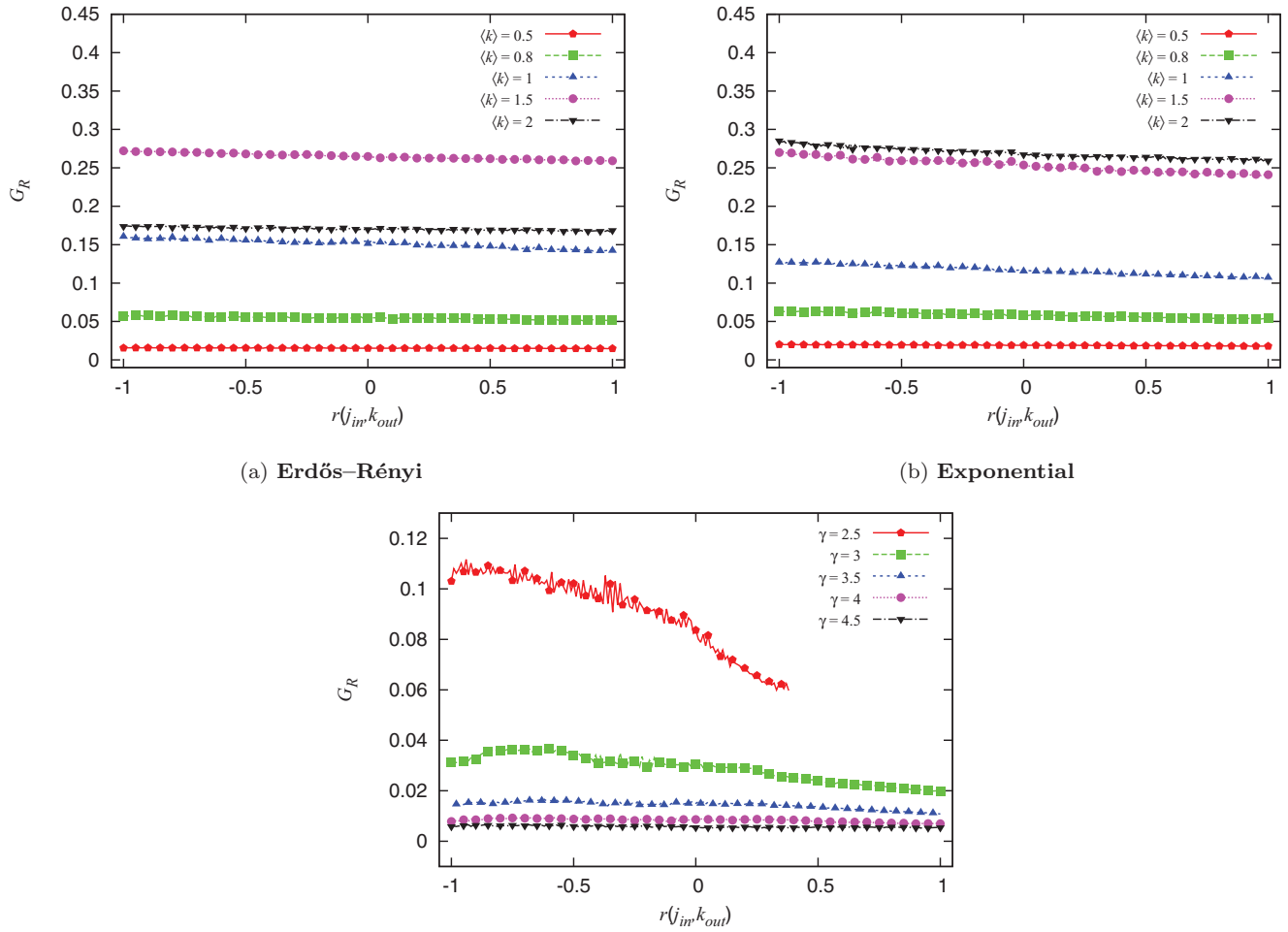
We can modify it by adding a correlation between $p_{j_{\text{out}}}$ and $p_{k_{\text{in}}}$:

$$\mathcal{P}^r(j_{\text{in}}, j_{\text{out}}; k_{\text{in}}, k_{\text{out}}) = p_{j_{\text{in}}} p_{k_{\text{out}}} \left[p_{j_{\text{out}}} p_{k_{\text{in}}} + r\sigma(p_{j_{\text{out}}})\sigma(p_{k_{\text{in}}}) m_{j_{\text{out}} k_{\text{in}}} \right]. \tag{91}$$

However, if we recover the joint degree distribution, the added term vanishes:

$$p^r_{l_{\text{in}} l_{\text{out}}} = \sum_{j_{\text{in}}, j_{\text{out}}=0}^{\infty} \mathcal{P}_r(j_{\text{in}}, j_{\text{out}}; l_{\text{in}}, l_{\text{out}}) = p^0_{l_{\text{in}} l_{\text{out}}} \tag{92}$$

because the sum of any row or column of the matrix $m_{j_{\text{out}} k_{\text{in}}}$ must be zero [Eq. (83)]. This means that the assortativity defined by Eq. (72) does not affect G_R directly (via the joint degree distribution). Figure 15 depicts the numerical results that were produced by the same protocol as in the preceding section. In the case of the ER and exponential networks, a small decrease in G_R can be observed, regardless of the average degree. The scale-free network has a nonmonotonic behavior, but for larger assortativity, a decrease in G_R can be seen. These results point out that, although the two-point correlation does not affect the joint degree distribution directly, it changes the hierarchical structure of the network a little. This small effect can be understood if we look at the different typical structures around an edge (Fig. 16). In the case of large two-point correlations, the nodes with large out-degree (those that are expected to have a larger reachable set) have out-neighbors with large in-degrees [Fig. 16(a)]. This means that the nodes they can directly reach are also reachable from many other nodes (that also have large out-degrees). This reduces the



(a) **Erdős–Rényi** (b) **Exponential**

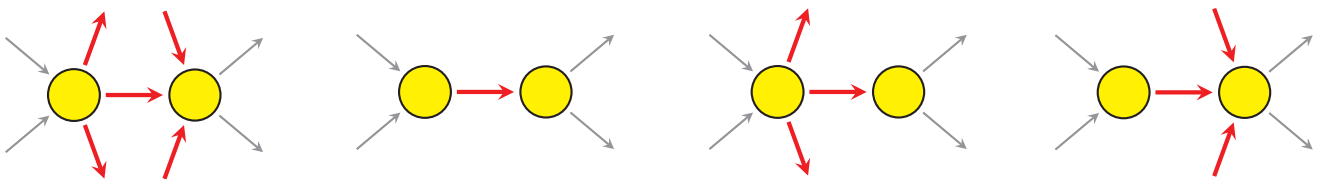
(c) **Scale-free**

FIG. 15. (Color online) Effect of two-point correlations to G_R for the random graph models. Each point is an average of at least 100 independent measure on networks with $N = 1000$.

difference between the local reaching centralities among nodes with large C_R or, in other words, introduces higher $c_R(i)$ in the definition of G_R [see Eq. (2)]. The positive two-point correlation also introduces bottlenecks in the network [Fig. 16(b)]. These structures accumulate many nodes on the in-edges that have a similar reachable set, thus also reducing the differences. In contrast, if the graph has negative two-point correlations, nodes with a large out-degree can reach their out-neighbors uniquely [Fig. 16(c)]. Nodes that are reachable from independent directions also emerge [Fig. 16(d)]. These nodes have in-neighbors with a small out-degree. Both effects decrease the ratio of overlapping reachable sets, which results in a reduced similarity in the C_R values.

IV. CONCLUSION

The global reaching centrality G_R is a measure constructed to provide a characteristic number for the flow hierarchy of a directed network. In this paper we investigated the



(a) $r > 0$ (b) $r > 0$ (c) $r < 0$ (d) $r < 0$

FIG. 16. (Color online) Illustration of the two limiting cases of two-point correlations r defined by Eq. (72). The edges that are correlated are shown in thick red. In the case of $r > 0$, the reachable sets of hubs tend to be similar (a) and bottlenecks emerge (b), causing an increase in the overlap of reachable sets. In negatively correlated networks ($r < 0$), hubs can have their own reachable sets (c) and nodes with a large in-degree tend to be reachable from independent parts of the network (d), decreasing the overlap between reachable sets.

behavior of this measure and its dependence on the joint degree distribution. We have shown that the hierarchical structure of a random network strongly depends on the existence of the giant strongly connected component $\mathcal{G}_S$ and other giant components in the network. In the regime of low average degree, G_R is dominated by the node with the largest reachable set that can be approximated from the distribution of the local reaching centralities. This distribution is connected to the distribution of small out-components in a network and can be analytically determined using the generating function method generalized to directed networks. In the presence of $\mathcal{G}_S$, G_R can be well approximated by the difference in the sizes of the giant in-component $\mathcal{G}_{\text{in}}$ and $\mathcal{G}_S$ and both can be calculated exactly for some network models with a given degree distribution. Using the approximations for G_R, we calculated its dependence on the average degree for different network models: a directed random tree, the Erdős-Rényi graph, and the exponential and scale-free networks. The results show that in the sense of hierarchy, there is an optimal average degree at which G_R has a maximum value. It is a result of the competition between the size of the largest reachable set (which can be found by nodes in $\mathcal{G}_{\text{in}}$) and the average size of reachable sets (which is dominated by the nodes in $\mathcal{G}_S$). Near the transition point, the size of $\mathcal{G}_{\text{in}}$ is close to the size of $\mathcal{G}_S$ and the difference is increasing since $\mathcal{G}_{\text{in}}$ grows faster in the beginning. In the limit of large average degree, the sizes of both giant components tend to become equal since even the two set of nodes tend to be the same.

We also investigated the dependence of G_R on two types of degree correlations: the one-point correlation (the in-degree and out-degree of a node) and the two-point correlation (the out-degree of the source and the in-degree of the target of a directed edge). Numerical results show that the hierarchical structure changes systematically with the one-point correlation. This is in accord with the fact that G_R can be expressed with the joint degree distribution and later is directly affected by the one-point correlations. We have pointed out that in the two limiting cases of the average degree, the qualitative effect of small correlations can be understand by the addition of a correlation term to the uncorrelated joint degree distribution. When the average degree is very low, small correlations have a positive effect on G_R. This is not true for the denser graph, in which correlations decrease G_R. There is also a regime of average degree in which a given level of correlation can maximize the hierarchy.

Both numerical and analytical results suggest that the two-point correlation (which is the generalization of the assortativity for directed networks) does not effect G_R directly. However, a small negative effect can be observed. It can be understood by looking at the effect of the two-point correlation on the neighborhood of an edge.

The results on the random network models have shown a deeper insight into the behavior of G_R, but we have to keep in mind that they are only the first steps in the understanding of hierarchy. The main message of the results is that hierarchy is sensitive to the edge density of a network and tends to emerge more likely in sparse networks. This is in good agreement with the observation that most of the real-world networks are in the low-density range and many of them have an inherent hierarchical structure. The results also point out that correlations can have a large effect and can change the hierarchical structures fundamentally (they can produce a finite magnitude of G_R even if the network would have an infinitesimal G_R otherwise). This is another indicator of the well-known fact that one has to take into account the presence of correlations when dealing with real networks.

ACKNOWLEDGMENTS

The author would like to thank Tamás Vicsek, Péter Pollner, and Gergely Palla for fruitful conversations that facilitated the interpretation of the results and for their advice during the preparation of the manuscript. This research was fully supported by the EU ERC FP7 COLLMOT Grant No. 227878.

APPENDIX A: OUT-COMPONENTS IN DIRECTED GRAPHS

Using the definition of the generating function $h_0(y)$, we can obtain the probabilities for the out-components:

$$\pi_s^{\text{out}} = \frac{1}{(s-1)!}\left[\frac{d^{s-1}}{dy^{s-1}}\left(\frac{h_0(y)}{y}\right)\right]_{y=0}. \tag{A1}$$

Substituting the expression behind the derivation from Eq. (9), we get

$$\begin{aligned}\pi_s^{\text{out}} &= \frac{1}{(s-1)!}\left[\frac{d^{s-1}}{dy^{s-1}} g_{00}[1,h_1(y)]\right]_{y=0} \\ &= \frac{1}{(s-1)!}\left[\frac{d^{s-2}}{dy^{s-2}}\{\partial_y g_{00}[1,h_1(y)]h_1'(y)\}\right]_{y=0}.\end{aligned} \tag{A2}$$

This equation can be transformed into an integral by the Cauchy formula:

$$\begin{aligned}\pi_s^{\text{out}} &= \frac{1}{2\pi i(s-1)}\oint \frac{\partial_y g_{00}[1,h_1(y)]}{y^{s-1}}\frac{dh_1}{dy}dy \\ &= \frac{1}{2\pi i(s-1)}\oint \frac{\partial_y g_{00}[1,h_1]}{y^{s-1}}dh_1.\end{aligned} \tag{A3}$$

The contour goes around the origin and has an infinitesimal radius, ensuring that it does not enclose poles. In our calculations $y_0 = 0$. In Eq. (A3) we just changed variables. We have to note that when $y \to 0$, h_1 also converges to zero. This can be easily seen from Eq. (10). We can also eliminate y from the argument using Eq. (10):

$$\begin{aligned}\pi_s^{\text{out}} &= \frac{1}{2\pi i(s-1)}\oint \frac{\partial_y g_{00}[1,h_1]g_{10}[1,h_1]^{s-1}}{h_1^{s-1}}dh_1 \\ &= \frac{\langle k\rangle}{2\pi i(s-1)}\oint \frac{g_{01}[1,h_1]g_{10}[1,h_1]^{s-1}}{h_1^{s-1}}dh_1.\end{aligned} \tag{A4}$$

A second application of the Cauchy formula gives us the final form of the out-components:

$$\pi_s^{\text{out}} = \frac{\langle k\rangle}{(s-1)!}\left[\frac{d^{s-2}}{dy^{s-2}}(g_{01}(1,y)g_{10}(1,y)^{s-1})\right]_{y=0}. \tag{A5}$$

APPENDIX B: PROOF OF $\beta(u) < 0$

We show that the coefficient

$$\beta(u) = \frac{1}{\langle k\rangle}\frac{\partial_y \chi_{00}(u,1)}{1-\partial_x g_{01}^p(u,1)} \tag{B1}$$

that appears in Eq. (80) is negative for any $\langle k\rangle > k_c$. The numerator has the form of

$$\partial_y \chi_{00}(u,1) = \sigma_p^2 \sum_{j,k=0}^{\infty} x^j k m_{jk}, \tag{B2}$$

where the only criteria for m_{jk} are Eqs. (74) and (75). They can be satisfied by the following choice [36]:

$$m_{jk} = \frac{(p_j - \phi_j)(p_k - \phi_k)}{(\langle k\rangle - \langle k\rangle_\phi)^2}. \tag{B3}$$

Here ϕ_j is an arbitrary normalized distribution and $\langle k\rangle_\phi$ is the corresponding average. With this form of m_{jk}, the numerator looks like

$$\partial_y \chi_{00}(u,1) = \sigma_p^2 \frac{g_0^p(u) - g_0^\phi(u)}{\langle k\rangle - \langle k\rangle_\phi}, \tag{B4}$$

where $g_0^p(u)$ and $g_0^\phi(u)$ are the corresponding generating functions for the distributions p_j and ϕ_j. Assume that $\langle k\rangle > \langle k\rangle_\phi$ and fix the ϕ_j distribution such that $\langle k\rangle_\phi = k_c$. It can be easily seen that for the ER and exponential networks, $g_0^p(u) < g_0^\phi(u)$ for any u.

We only have to show that $\partial_x g_{01}^p(u,1) < 1$ above the transition point. For this to see, let us consider the function in question as a function of the average degree

$$F(\langle k\rangle) = \partial_x g_{01}^p[u(\langle k\rangle),1] \tag{B5}$$

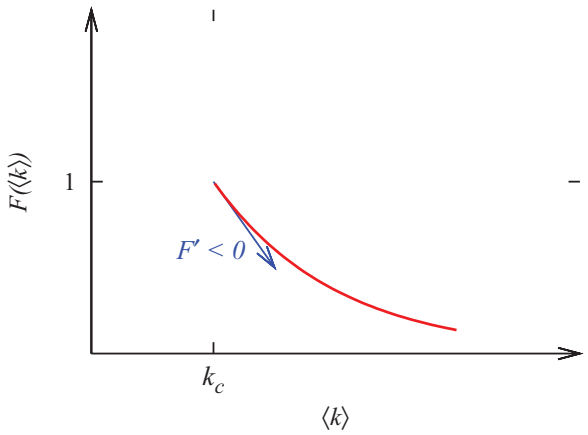


FIG. 17. (Color online) Illustration of $F(\langle k\rangle)$ for the proof of $\beta(u_0) < 0$. Its value at k_c follows directly from the definition and its derivative is examined in the text.

and note that $F(k_c) = 1$. Furthermore, $F(x)$ is a decreasing function of its argument

$$\frac{dF(\langle k\rangle)}{d\langle k\rangle} = \underbrace{\partial_x^2 g_{01}^p(x,1)\big|_{x=u(\langle k\rangle)}}_{>0} \underbrace{\frac{du(\langle k\rangle)}{d\langle k\rangle}}_{<0}. \tag{B6}$$

The first term on the right-hand side is positive for all $0 < u < 1$ because it is a sum of positive numbers. In contrast, the second term is negative in the ER and exponential networks because the size of the giant components increases with the average degree. Thus $F(\langle k\rangle) < 1$ (see Fig. 17).

[1] C. Castellano, S. Fortunato, and V. Loreto, Rev. Mod. Phys. **81**, 591 (2009).
[2] J. A. Dunne, R. J. Williams, and N. D. Martinez, Proc. Natl. Acad. Sci. USA **99**, 12917 (2002).
[3] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[4] L. Négyessy, T. Nepusz, L. Kocsis, and F. Bazsó, Eur. J. Neurosci. **23**, 1919 (2006).
[5] R. F. Cancho and R. V. Solé, Proc. R. Soc. London Ser. B **268**, 2261 (2001).
[6] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[7] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[8] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010), pp. 428–473.
[9] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[10] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[11] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet* (Cambridge University Press, Cambridge, 2004), pp. 112–208.
[12] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 26118 (2001).
[13] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[14] B. Bollobás, *Random Graphs* (Cambridge University Press, Cambridge, 2001), pp. 96–130.
[15] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Comput. Networks **33**, 309 (2000).
[16] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **64**, 025101 (2001).
[17] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, Nature (London) **464**, 890 (2010).
[18] H. Fushing, M. P. McAssey, B. Beisner, and B. McCowan, PLoS ONE **6**, e17817 (2011).
[19] H.-W. Ma, J. Buer, and A.-P. Zeng, BMC Bioinf. **5**, 199 (2004).
[20] L. Huseyn and D. A. Whetten, Social Networks **6**, 31 (1984).
[21] *Hierarchy in Natural and Social Sciences*, edited by D. Pumain (Springer, Dordrecht, 2006), pp. 1–12.
[22] D. Lane, in *Hierarchy in Natural and Social Sciences* (Ref. [21]), pp. 81–120.
[23] E. T. Wimberley, *Nested Ecology. The Place of Humans in the Ecological Hierarchy* (Johns Hopkins University Press, Baltimore, 2009), pp. 4–13.
[24] L. Carmel, D. Haren, and Y. Koren, *Drawing Directed Graphs Using One-Dimensional Optimization* (Springer, Berlin, 2002), pp. 193–206.
[25] A. Trusina, S. Maslov, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **92**, 178702 (2004).
[26] R. Rowe, G. Creamer, S. Hershkop, and S. J. Stolfo, *WebKDD/SNA-KDD 07*, Proceedings of the Ninth WebKDD

and First SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (ACM, New York, 2007), pp. 109–117.

[27] N. Memon, H. L. Larsen, D. L. Hicks, and N. Harkiolakis, *Proceedings of the IEEE ISI 2008 PAISI, PACCF, and SOCO International Workshops on Intelligence and Security Informatics* (Springer, Heidelberg, 2008), pp. 477–489.

[28] E. Mones, L. Vicsek, and T. Vicsek, PLoS ONE **7**, e33799 (2012).

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

[30] However, the integrated size of the giant components is encoded in the distribution since it is normalized to $N - |\mathcal{G}_W|$, where $\mathcal{G}_W$ is the giant weakly connected component.

[31] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, Adv. Comput. Math. **5**, 329 (1996).

[32] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. USA **97**, 11149 (2000).

[33] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).

[34] F. Chung and L. Lu, Ann. Combinatorics **6**, 125 (2002).

[35] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).

[36] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).

[37] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).

[38] M. Boguná, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 28701 (2003).

[39] N. Martinez, Ecolog. Monogr. **61**, 367 (1991).

[40] http://courses.engr.illinois.edu/ece543/iscas89.html

[41] J. Leskovec, D. Huttenlocher, and J. Kleinberg, *Proceedings of the 28th International Conference on Human Factors in Computing Systems* (ACM, New York, 2010), pp. 1361–1370.

[42] M. A. J. Van Duijn, M. Huisman, F. N. Stokman, F. W. Wasseur, and E. P. H. Zeggelink, J. Math. Sociol. **27**, 153 (2003).

[43] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science **303**, 1538 (2004).

[44] S. Balaji, M. M. Babu, L. M. Iyer, N. M. Luscombe, and L. Aravind, J. Mol. Biol. **360**, 213 (2006).

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