Scaling of Degree Correlations and Its Influence on Diffusion in Scale-Free Networks

Lazaros K. Gallos, Chaoming Song, and Hernán A. Makse

Levich Institute and Physics Department, City College of New York, New York, New York 10031, USA

(Received 3 May 2007; revised manuscript received 3 January 2008; published 19 June 2008)

Connectivity correlations play an important role in the structure of scale-free networks. While several empirical studies exist, there is no general theoretical analysis that can explain the largely varying behavior of real networks. Here, we use scaling theory to quantify the degree of correlations in the particular case of networks with a power-law degree distribution. These networks are classified in terms of their correlation properties, revealing additional information on their structure. For instance, the studied social networks and the Internet at the router level are clustered around the line of random networks, implying a strongly connected core of hubs. On the contrary, some biological networks and the WWW exhibit strong anticorrelations. The present approach can be used to study robustness or diffusion, where we find that anticorrelations tend to accelerate the diffusion process.

DOI: 10.1103/PhysRevLett.100.248701

PACS numbers: 89.75.Hc, 87.23.Ge, 89.75.Da

Despite their importance, a theoretical framework to describe and characterize degree correlations in scale-free networks is still missing. Here, we find that the degree correlations in the studied scale-free networks can be characterized in terms of a correlation exponent \( \epsilon \), which we calculate using a renormalization approach. This allows us to propose a classification of a set of dissimilar networks according to the degree of correlations into a small number of different classes in a “phase diagram.” For example, biological networks and the WWW are in the strong anticorrelations part of the diagram, while social networks and the Internet are clustered near the region of random networks. We show how we can use these ideas to explore more network properties, such as diffusion and robustness, which depend on the degree of correlations in the network.

We start by recalling the renormalization of a network under a scale transformation. The renormalization procedure tiles a network according to the box-covering algorithm [10], with the minimum number of boxes where the maximum distance in any box is less than \( \ell_B \). Each box is subsequently replaced by a node, and links are established between these new “supernodes” if at least one node included in a box was connected to any node of the other box. These boxes are treated as the nodes of the renormalized network. Renormalization is a reliable method for determining how the network behaves at different length scales. Self-similarity is then obtained if the network structure remains invariant under the renormalization.

Alternatively, we can retain multiple links between the boxes when we renormalize a network [11]. As we show below, though, this is not a strong effect, mainly due to preservation of the self-similar structure under renormalization. In contrast, during a random rewiring process, degree correlations are destroyed, but bias is introduced if multiple links are forbidden.

We use renormalization and scaling theory to determine the form of \( P(k_1, k_2) \). Since the self-similarity of a scale-free network requires the invariance of the degree distribution \( P(k) \), a power-law distribution of \( P(k) \sim k^{-\gamma} \),
where $\gamma$ is the degree exponent, is the only form that can satisfy this condition [5]. Taking this idea one step further, it is interesting to clarify whether correlations between degrees, as expressed by the joint degree distribution $P(k_1, k_2)$, also remain invariant. In Fig. 1, we present an example of this distribution before and after renormalization for the WWW and the Internet at the router level (similar results are derived for other networks, as well). Allowing multiple links between boxes does not significantly modify the result. The statistical similarity of the corresponding plots suggests the invariance of $P(k_1, k_2)$. Accordingly, this suggests that the $k_1$ and $k_2$ dependence can be separated, and the behavior of the tail of the joint degree distribution is

$$P(k_1, k_2) \sim k_1^{-(\gamma - 1)} k_2^{-\left(\frac{1-\epsilon}{\gamma}\right)} \quad (k_1 > k_2).$$

The value of the first exponent $\gamma - 1$ in Eq. (1) is obtained from the density conservation law: $\int P(k_1, k_2) dk_2 = k_1 P(k_1) \sim k_1^{-(\gamma - 1)}$. Equation (1) is also consistent with the known result for completely random networks

$$P(k_1, k_2) \sim k_1 P(k_1) k_2 P(k_2) \sim k_1^{1-\gamma} k_2^{1-\gamma};$$

i.e., the exponent $\epsilon$ for these networks is $\epsilon_{\text{rand}} = \gamma - 1$, as expected from the symmetry in this case. Equation (1) is also consistent with previous findings in Ref. [1]. The probability distribution for the neighbor connectivity in the yeast protein interaction network was there shown to behave differently for low-degree nodes, where the $k$-dependence was $k^{1-\gamma}$, and for large-degree nodes, with a $k^{-\epsilon}$ dependence. Using Eq. (1), we can see that integrating over $k_2$ for low-degree nodes ($k_1 > k_2$), we retrieve the $k^{1-\gamma}$ dependence. For the case of hubs, where integration is over $k_1$, the dependence on the degree is $k^{-\epsilon}$, and for the yeast protein interaction network, we have calculated that $\epsilon = \gamma$ (see Fig. 3). These results are in full agreement with the observed behavior in [1].

Next, we introduce a scale-invariant quantity $E_b(k)$ to simplify the estimation of $\epsilon$, even for small networks. We are motivated to introduce this quantity by asking whether a node is significantly linked to more connected nodes; i.e., a node considers another node as a “hub” if its degree is much larger than its own. We define the ratio

$$E_b(k) = \frac{\int_0^a \frac{k}{k'} P(k|k')dk'}{\int_0^a P(k')dk'},$$

as the measure of a node’s preference to connect to neighbors with degree larger than $bk$ ($b$ is an arbitrary positive number, and large $b$ corresponds to the identification of the hubs) [12]. From the scaling of $E_b(k)$ with $k$, we are able to obtain the exponent $\epsilon$ in a simpler way than using $P(k_1, k_2)$, which presents more fluctuations than the average quantity $E_b(k)$. The conditional probability is $P(k|k') = P(k, k')/\int P(k, k') dk = P(k, k') k^{1-\gamma} = k^{-(\gamma-1)} k'^{-(1+\epsilon-\gamma)}$. We find for a scale-free distribution,

$$E_b(k) \sim k^{-(\epsilon-\gamma)}.$$  

We have verified that the scaling of $E_b(k)$ remains invariant under renormalization. The same scaling exponents are recovered for the renormalized networks, even when multiple links are allowed between two boxes (Fig. 2 inset). In the latter case, the renormalized nodes have in general larger degrees, which means that deviations appear in nodes of smaller degree. Additionally, $\epsilon$ was found to be independent of the value of $b$. We notice that other quantities derived from $P(k_1, k_2)$ may not be invariant under renormalization, such as $r$ or $k_m$, and therefore are not suitable to distinguish fractal from nonfractal networks.

In Fig. 2, we present the behavior of $E_b(k)$ for the WWW, protein homology, Internet (router level), and cond-mat authorship. The existence of a scaling relation over a $k$ range, combined with the invariance of this curve, support Eq. (4) and the form used for $P(k_1, k_2)$ in Eq. (1). The WWW and the protein homology network have been shown to have a fractal topology. The slope of $E_b(k)$ with $k$ is small or negative in these cases with values of $\epsilon = 2.5$ and $\epsilon = 2.4$, respectively. This behavior is in contrast with the two nonfractal networks in the figure, i.e., the Internet at the router level and the cond-mat coauthorship network, where $E_b(k)$ increases almost linearly with increasing $k$. For these networks, we find that $\epsilon = 1.2$ and $\epsilon = 1.6$, respectively.

In order to interpret the values of $\epsilon$, we now turn to the renormalization scheme [5]. After renormalization, the number of nodes $N$ in the network and the degree of a node $k$ scale with $\ell_B$ as power laws with fractal exponent $d_B$ and degree exponent $d_k$, respectively (we use a prime to describe quantities measured in the renormalized network):

$$N \rightarrow N' \sim \ell_B^{-d_B} N, \quad k \rightarrow k' \sim \ell_B^{-d_k} k.$$  

If $d_B$ and $d_k$ are finite, the network is fractal. If $d_B \rightarrow \infty$ and $d_k \rightarrow \infty$ (or equivalently the decay is exponential or faster), the network is not fractal.
After tiling the network with boxes of diameter \( \ell_{B} \), each of these boxes have one unique local hub (i.e., the largest degree node in the box). Considering all possible pairs of boxes, we introduce the probability \( \mathcal{E}(\ell_{B}) \) that there exists a direct connection between the two hubs of any two boxes. We have shown (see, e.g., Figs. 2e, 3d of Ref. [7]) that the probability \( \mathcal{E} \) scales with the length \( \ell_{B} \) as

\[
\mathcal{E}(\ell_{B}) \sim \ell_{B}^{-d_{c}}.
\]

Below, we relate the exponent \( \epsilon \) to the hub-hub repulsion through the hub correlation exponent \( d_{c} \), which is crucial for fractality.

The conservation of links in the renormalized network leads to the expression

\[
NP(k_{1}, k_{2})dk_{1}dk_{2} = \mathcal{E}(\ell_{B})N!P(k')dk'dk'_2.
\]

Using Eqs. (1) and (5)−(7), we get the relation

\[
\frac{\ell_{B}^{d_{c}}}{\ell_{B}^{d_{k}}} = \frac{1}{N!}P(k')dk'dk'_2 \implies \epsilon = 2 + d_{c}/d_{k} = 2 + (\gamma - 1)\frac{d_{c}}{d_{k}},
\]

(8)

where we have substituted the value \( \gamma = 1 + d_{B}/d_{k} \). This relation of \( \epsilon \) with \( d_{c} \) shows that correlations between the hubs of the boxes determine the correlations for all degrees, in accordance with the invariance under renormalization.

The direct determination of \( \epsilon \) through the slope of \( E_{b}(k) \) vs \( k \) enables us to construct a “phase diagram” in the plane \( (\epsilon, \gamma) \), shown in Fig. 3. This plot is classifying the studied networks in classes according to their degree of correlations, even though they correspond to dissimilar systems in biology, sociology, or technology.

As shown in Eq. (2), the exponent for a random network corresponds to the random line \( \epsilon_{\text{rand}} = \gamma - 1 \), which is verified in the plot for different \( \gamma \) values of the configuration model. In random network models, correlations arise because links are selected for connecting with each other equiprobably, so that the probability of two hubs being connected is large [13]. Thus, networks that are close to the line \( \epsilon_{\text{rand}} = \gamma - 1 \) exhibit hub-hub correlations. The random line separates the diagram in two main parts: (a) above the line where the hub correlations tend to become weaker, and (b) below the line where networks have even larger correlations (hubs are connected to each other with even higher probability than the one corresponding to a randomly created structure).

In the diagram, the social networks and the Internet at the router level are clustered around the line \( \epsilon_{\text{rand}} = \gamma - 1 \). This is an indication that there is a strongly connected core of hubs in these systems, consistent with previous studies. The biological networks and the WWW, on the other side, are far away from the random line. This implies that there is a richer structure in these networks with hubs separated from each other. The distance in the plot from \( \epsilon_{\text{rand}} \) quantifies how different from randomness the network structure is, in terms of degree correlations.

As \( \epsilon \) increases from \( \epsilon_{\text{rand}} \), we expect that at some point the networks will become fractal, due to increased hub-hub correlations. The point of emergent fractality is found through Eq. (8), where the borderline case of \( d_{B} \rightarrow \infty \) yields \( \epsilon = 2 \). Indeed, we have verified via direct measure-
ments of $d_y$ that all the networks above the line $\epsilon = 2$ in Fig. 3 are fractals. Thus, starting from $\epsilon_{\text{rand}}$, we can separate the phase space into areas where the hub correlations are stronger than in random models (area I) or weaker than that (areas II and III). The weak correlation areas II and III are further divided by the line $\epsilon = 2$ which determines whether the anticorrelations are strong enough to result in a fractal network (III) or not (II).

An immediate result from this diagram is the different position of the Internet at the router level compared to the AS level [14]. Although the degree distribution of these two networks is the same ($\gamma = 2.2$), the correlation exponent $\epsilon$ reveals that there are more hub-hub connections at the router level, similar to the case of a random network. Contrary to that, the AS level exhibits a structure with less correlations deviating from that of a simple random model. This difference may hint on different design principles or requirements at varying levels of the Internet.

The above approach can be directly applied to explore many interesting properties, such as network robustness, synchronization, or diffusion processes. Until now, theoretical studies have been limited to using the uncorrelated version of $P(k' | k) \sim k'P(k')$. The introduction of Eq. (1) enables us to substitute this form and generalize the problem for networks with known correlation exponents. For example, we can study the effect of correlations on diffusion by starting with the master equation for the density of particles $\rho(k, t)$ on nodes with degree $k$ at time $t$

$$\frac{d\rho(k, t)}{dt} = -\rho(k, t) + k \sum_{k'=k_{\text{min}}}^{\infty} P(k' | k) \frac{\rho(k', t)}{k'}, \quad (9)$$

and substitute $P(k' | k)$ with a form derived through Eq. (1). The Laplace transform of the above equation leads to a Fredholm integral equation of the second kind with separable kernel that can be solved analytically. We define the quantity $K(t) = \langle k^x(t) \rangle^{1/x}$, where $x = \gamma - 1 - \epsilon$, and $K(t)$ serves as a measure of the diffusing particles preference to larger or smaller degrees $k$. The analytical result for $\kappa_c(t)$, defined as $\kappa_c(t) = \langle k^2 \rangle(t)$, is $\kappa_c(t) = \kappa_\infty + (\kappa_0 - \kappa_\infty)e^{-ct}$, where $\kappa_\infty = (\gamma - 2)/(|\epsilon - 1|)\epsilon$, and $\kappa_0 = (\gamma - 1)/|\epsilon$ are constants depending on $\gamma$ and $\epsilon$, while $c = (\epsilon - 1)^2/[\epsilon(\gamma - 2)(\epsilon - \gamma)]$. The result for $K(t)$ is displayed in Fig. 4 for networks with $\gamma = 2.75$. The exponential convergence to the asymptotic steady-state configuration depends on the value of $c$, which increases with the exponent $\epsilon$. Networks that are close to the random case $\epsilon_{\text{rand}} = \gamma - 1$ require longer times for reaching the equilibrium state and the diffusing particles prefer to occupy larger degree nodes. A larger $\epsilon$ value enhances anticorrelations in the network, and the particles move faster occupying smaller degree nodes on the average. We can infer, thus, that stronger correlations tend to speed up the diffusion process. The mechanism behind this behavior is as follows: when the hubs are directly connected to each other, the particles tend to remain localized in the neighborhood around these hubs, so that it takes longer for them to explore wider areas. On the contrary, when hub anticorrelations are important, the particles spend most of their time in the intermediate areas which are formed by smaller degree nodes and which connect indirectly the hubs to each other. We acknowledge valuable discussions with Shlomo Havlin and support from NSF grants.

[12] Notice that we could have defined $E_x(k)$ without using the denominator in Eq. (3), i.e., $E_x(k) = \int_{0}^{\infty} P(k')dk'$, in which case we get $E_x(k) \sim k^{(1 - x)}$. However, due to the large fluctuations occurring at the tails of the degree distribution in real-life networks, the calculation of $\epsilon$ through Eq. (3) was proved to be more robust.
[15] A description of the databases used in Fig. 3 and citations to their source can be found at http://jamlab.org.