Eur. Phys. J. B (2015) 88: 65
DOI: 10.1140/epjb/e2015-50671-y

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

Correlation between centrality metrics and their application to the opinion model

Cong Li[1,a], Qian Li[2], Piet Van Mieghem[1], H. Eugene Stanley[2], and Huijuan Wang[1,2]

[1] Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, The Netherlands
[2] Center for Polymer Studies, Department of Physics, Boston University, Boston, Massachusetts 02215, USA

Received 30 September 2014 / Received in final form 6 February 2015
Published online 19 March 2015

Abstract. In recent decades, a number of centrality metrics describing network properties of nodes have been proposed to rank the importance of nodes. In order to understand the correlations between centrality metrics and to approximate a high-complexity centrality metric by a strongly correlated low-complexity metric, we first study the correlation between centrality metrics in terms of their Pearson correlation coefficient and their similarity in ranking of nodes. In addition to considering the widely used centrality metrics, we introduce a new centrality measure, the degree mass. The mth-order degree mass of a node is the sum of the weighted degree of the node and its neighbors no further than m hops away. We find that the betweenness, the closeness, and the components of the principal eigenvector of the adjacency matrix are strongly correlated with the degree, the 1st-order degree mass and the 2nd-order degree mass, respectively, in both network models and real-world networks. We then theoretically prove that the Pearson correlation coefficient between the principal eigenvector and the 2nd-order degree mass is larger than that between the principal eigenvector and a lower order degree mass. Finally, we investigate the effect of the inflexible contrarians selected based on different centrality metrics in helping one opinion to compete with another in the inflexible contrarian opinion (ICO) model. Interestingly, we find that selecting the inflexible contrarians based on the leverage, the betweenness, or the degree is more effective in opinion-competition than using other centrality metrics in all types of networks. This observation is supported by our previous observations, i.e., that there is a strong linear correlation between the degree and the betweenness, as well as a high centrality similarity between the leverage and the degree.

1 Introduction

Recent research has explored social dynamics [1–3] by using complex networks in which nodes represent people/agents and links the associations between them. Such centrality metrics as degree and betweenness have been studied in dynamic processes [4–7], such as opinion competition, epidemic spreading, and rumor propagation on complex networks. These studies used centrality metrics to identify influential nodes [4–6], such as the source nodes from which a virus spreads and the nodes with high spreading capacity, as well as to select which nodes are to be immunized when a virus is prevalent [7]. Numerous centrality metrics have been proposed. Degree, betweenness, closeness, and principal eigenvector of the adjacency matrix (which is shortly called the principal eigenvector in this work) are the most popular centrality metrics [4,8–13]. Several new centrality metrics have been introduced in a number of different fields recently. Kitsak et al. [5] studied the SIS and SIR spreading models on four real-world networks and proposed that the k-shell index is a better indicator for the most efficient spreaders (nodes) than degree or betweenness. Joyce et al. [14] proposed a new centrality metric – *leverage* – for identifying neighborhood hubs (the most highly-connected nodes) in functional brain networks. Leverage centrality identifies nodes that are connected to more nodes than their nearest neighbors. In addition to considering these widely-used centrality metrics, we here propose a new centrality metric, *degree mass*. The mth-order degree mass of a node is defined as the sum of the weighted degree of its m-hop neighborhood[1]. If the degree of a node and of its neighbors are all high, the node has a high degree mass.

Centrality metrics have been compared in various networks, such as sampled networks, biological networks, food webs, and vocabulary networks in literature [4,15–18]. Comin et al. [4] compared the centrality metrics characterizing the performances of nodes in such dynamic processes as virus spreading. Kim and Jeong [15] compared the reliability of rank orders using centrality

[a] e-mail: licong1986@gmail.com

[1] The m-hop neighborhood of a node i includes the node i and all nodes no further away than m hops from i.

metrics in sampling networks. The correlations between centrality metrics have been studied in biological networks [16,17]. However correlations between centrality metrics are still not well understood. If correlations between centrality metrics were better understood, we might be able to rank the nodes in a network by using the centrality metrics with a low computational complexity instead of the ones with a high computational complexity. To investigate the correlation between any two centrality metrics, we compute their Pearson correlation coefficient and their similarity in ranking nodes in both network models and real-world networks. The two methods have been applied to study the correlation between metrics in references [19–23]. In this work (i) we consider Erdős-Rényi (ER) networks[2] with a binomial degree distribution [24] and scale-free (SF) networks[3] with a power-law degree distribution [25,26]. Studying these two network models allows us to understand how the degree distribution influences correlations between the centrality metrics. (ii) We further explore correlations in 34 real-world networks with differing numbers of nodes and links. (iii) We theoretically compare the Pearson correlation coefficients between the principal eigenvector and the degree masses.

Recently there has been considerable interest in understanding how two competing opinions [27–31] evolve in a population. In this work we apply our centrality metrics to an inflexible contrarian opinion (ICO) model [32] in which only two opinions (denoted A and B) exist, with the goal of helping one opinion (opinion B) as it competes with with the other opinion (opinion A). At the initial time, opinions are randomly assigned to all nodes (with a fraction f of nodes holding opinion A and a fraction $1-f$ of nodes holding opinion B). At each step, each agent simultaneously and in parallel adopts the opinion of the majority of its nearest neighbors and itself, and if there is a tie, the agent does not change its opinion. After the system reaches a steady state, a fraction p_o of agents with opinion A is placed among the inflexible contrarians permanently holding opinion B, which can affect the opinion of their nearest neighbors. It is known that the size of the giant component of agents with opinion A can be decreased or even destroyed by the inflexible contrarians [32]. Li et al. [32] have selected the inflexible contrarians in ER and SF networks either randomly or based on degree. Here we choose inflexible contrarians using all the centrality metrics we have considered in both modelled networks and real-world networks. We compare the efficiencies of these centrality metrics in reducing the size of the largest opinion A cluster and find that strongly correlated centrality metrics have approximately the same efficiency in both modelled networks and real-world networks. Thus a high-complexity centrality metric could be approximated by a strongly correlated low-complexity centrality metric.

[2] An Erdős-Rényi random graph $G_p(N)$ can be generated from a set of N nodes by randomly assigning a link with probability p to each pair of nodes.

[3] A scale-free network is characterized by a power-law degree distribution $\text{Prob}[D=k] \sim k^{-\alpha}$, with $k_{\min} \leq k < k_{\max}$. Here, we choose $k_{\min}=2$, $k_{\max}$ as the natural cutoff and $\alpha = 2.5$.

This paper is organized as follows. In Section 2 we introduce the centrality metrics. In Section 3 we study the Pearson correlation coefficient and the centrality similarity between any two centrality metrics in both network models and real-world networks. In Section 4 the Pearson correlations between the degree masses and the principal eigenvector are theoretically analysed. In Section 5 the centrality metrics are applied in choosing the inflexible contrarians in the ICO model and the efficiencies of the centrality metrics are compared.

2 Definition of network centrality metrics

Centrality metrics quantify node properties in a network. Here we first review some centrality metrics that are widely used or have been recently proposed [4,5,8–12,14,33]. We then propose a new centrality metric, which we call *degree mass*. Let $G(\mathcal{N}, \mathcal{L})$ be a network, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ is the set of links. The number of nodes is denoted by $N = |\mathcal{N}|$ and the number of links by $L = |\mathcal{L}|$. The network G can be represented by an $N \times N$ symmetric adjacency matrix A, consisting of elements a_{ij}, which are either one or zero depending on whether node i is connected to node j or not. The networks mentioned in this paper are simple, unweighted and do not have self-loops or multiple links.

- Principal eigenvector x_1

The largest eigenvalue of the adjacency matrix A is λ_1, also called the spectral radius [34]. The principal eigenvector x_1 corresponding to the spectral radius λ_1 satisfies the eigenvalue equation

$$Ax_1 = \lambda_1 x_1.$$

Component j of the principal eigenvector is denoted by $(x_1)_j$. The X_1 is the element in the principal eigenvector that corresponds to a random node.

- Betweenness B_n

Betweenness was introduced independently by Anthonisse [35] in 1971 and Freeman [9] in 1977. The betweenness of a node i is the number of shortest paths between all possible pairs of nodes in the network that traverse the node

$$b_{ni} = \sum_{s \neq i \neq d \in \mathcal{N}} \frac{\sigma_{sd}(i)}{\sigma_{sd}},$$

where $\sigma_{sd}(i)$ is the number of shortest paths that pass through node i from node s to node d, and σ_{sd} is the total number of shortest paths from node s to node d. The betweenness B_n incorporates global information and is a simplified quantity for assessing the traffic carried by a node. Assuming that a unit packet is transmitted between each node pair, the betweenness b_{ni} is the total number of packets passing through node i [36].

- Closeness C_n

The closeness [37] of a node i is the average hopcount of the shortest paths from node i to all other nodes. It measures how close a node is to all the others. The most commonly used definition is the reciprocal of the total hopcount,

$$c_{ni} = \frac{N-1}{\sum_{j\in\mathcal{N}\setminus\{i\}} H_{ij}},$$

where H_{ij} is the hopcount of the shortest path between nodes i and j, and $\sum_{j\in\mathcal{N}\setminus\{i\}} H_{ij}$ is the sum of the hopcount of the shortest paths from node i to all other nodes. Closeness has been used to identify central metabolites in metabolic networks [38].

- K-shell index K_s

The k-shell decomposition of a network allows us to identify the core and the periphery of the network. The k-shell decomposition proceedure is as follows:

(1) Remove all nodes of degree $d = 1$ and also their links. This may reduce the degree of other nodes to 1.
(2) Remove nodes whose degree has been reduced to 1 and their links until all of the remaining nodes have a degree $d > 1$. All of the removed nodes and the links between them constitute the k-shell with an index $k_s = 1$.
(3) Remove nodes with degree $d = 2$ and their links in the remaining networks until all of the remaining nodes have a degree $d > 2$. The newly removed nodes and the links between them constitute the k-shell with an index $k_s = 2$, and subsequently for higher values of k_s.

The k-shell is a variant of the k-core [39,40], which is the largest subgraph with minimum degree of at least k. A k-core includes all k-shells with an index of $k_s = 0, 1, 2, \ldots, k$. An $O(m)$ algorithm for k-shell network decomposition was proposed in reference [41]. The k-shell index of the original infected node is a better predictor of the infected population in the susceptible-infectious-recovered (SIR) epidemic spreading process than other centrality metrics, such as the degree [5].

- Leverage L_n

Joyce et al. [14] introduced leverage centrality in order to identify neighborhood hubs in functional brain networks. The leverage measures the extent of the connectivity of a node relative to the connectivity of its nearest neighbors. The leverage of a node i is defined

$$l_{ni} = \frac{1}{d_i} \sum_{j\in\mathcal{N}_i} \frac{d_i - d_j}{d_i + d_j},$$

where $\mathcal{N}_i$ is the directly connected neighbors of the node i. With the definition of l_{ni} and the range $[1, N-1]$ of the degree d_i in connected networks, the leverage of a node i is bounded by $-1 + \frac{2d_i}{d_i+(N-1)} \le l_{ni} \le 1 - \frac{2}{d_i+1}$. Hence the range of the leverage l_{ni} is $[-1+2/N, 1-2/N]$ and the equality occurs in star graphs and complete graphs K_N. The leverage of a node is high when it has more connections than its direct neighbors. Thus a high-degree node with high-degree nearest neighbors will probably have a low leverage.

- Degree mass $D^{(m)}$

The degree of a node i in a network G is the number of its direct neighbors,

$$d_i = \sum_{j=1}^{N} a_{ij} = (Au)_i,$$

where $u = (1, 1, \ldots, 1)^T$ is the all-one vector. Here we propose a new set of centrality metrics, the degree mass, which is a variant of degree centrality. The mth-order degree mass of a node i is defined as the sum of the weighted degree of its m-hop neighborhood,

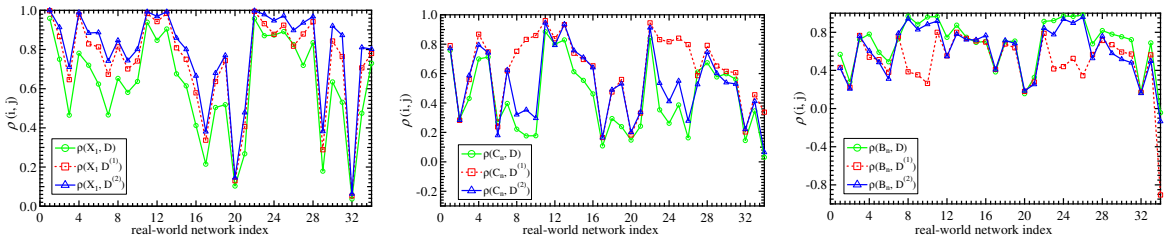
$$d_i^{(m)} = \sum_{k=1}^{m+1} \left(A^k u\right)_i = \sum_{j=1}^{N} \left(\sum_{k=0}^{m} A^k\right)_{ij} d_j,$$

where $m \ge 0$. The weight of the degree d_j is the number of walks[4] of length no longer than m from node i to node j. The weight of d_j is larger than the weight of d_l when node l is farther than node j from node i. The mth-order degree mass vector is defined $d^{(m)} = [d_1^{(m)}, d_2^{(m)}, \ldots, d_N^{(m)}]$. The 0th-order degree mass is the degree centrality. The 1st-order degree mass of node i is the sum of the degree of node i and the degree of its nearest neighbors. When m is large, the mth-order degree mass is proportional to the principal eigenvector.

3 Correlations between centrality metrics

We investigate the correlations between the centrality metrics introduced in Section 2, in both network models and real-world networks. The network models include the Erdős-Rényi (ER) network and the scale-free (SF) network. ER networks are characterized by a binomial degree distribution with $\text{Prob}[D = k] = \binom{N-1}{k} p^k (1-p)^{N-1-k}$, where N is the number of nodes and p is the probability that each node pair is connected. A SF network [25,42] has a power-law degree distribution with $\text{Prob}[D = k] \sim k^{-\alpha}$, $k \in [k_{\min}, k_{\max}]$, where $k_{\min}$ is the smallest degree, $k_{\max}$ is the degree cutoff, and α is the exponent characterizing the broadness of the distribution. In this work we use the natural cutoff at approximately $N^{1/(\alpha-1)}$ and $k_{\min} = 2$. We consider 34 real-world networks, e.g., airline connections, electrical power grids, and coauthorship collaborations. The descriptions and properties of these real-world networks are given in Appendix A. We study the correlations between any two centrality metrics using the Pearson correlation coefficient and the centrality similarity.

[4] A walk from i to j is any sequence of edges that allows back and forth movement and repeated visits to the same node.



(a) (b) (c)

Fig. 1. Pearson correlation coefficients (a) between the principal eigenvector and the degree masses: $\rho(X_1, D)$ (in circle marks), $\rho(X_1, D^{(1)})$ (in rectangle marks), and $\rho(X_1, D^{(2)})$ (in triangle marks); (b) between the closeness and the degree masses: $\rho(C_n, D)$ (in circle marks), $\rho(C_n, D^{(1)})$ (in rectangle marks), and $\rho(C_n, D^{(2)})$ (in triangle marks); (c) between betweenness and degree masses: $\rho(B_n, D)$ (in circle marks), $\rho(B_n, D^{(1)})$ (in rectangle marks), and $\rho(B_n, D^{(2)})$ (in triangle marks), in 34 real-world networks.

3.1 Pearson correlation coefficients between centrality metrics

Here we explore the linear correlation between the centrality metrics using numerical simulations in both ER and SF networks as well as in real-world networks. The results in Appendix B indicate that strong linear correlations do exist between certain centrality metrics in both ER and SF networks, and that network size has little influence on the correlations. Note that the k-shell index is weakly correlated with all the other centrality metrics. This might be the case because the k-shell indices of all nodes are similar to each other in binomial networks. We note the following seemingly universal relations between the degree masses and three centrality metrics, the principal eigenvector x_1, the closeness C_n and the betweenness B_n, as:

$$\begin{cases} \rho\left(X_1, D^{(2)}\right) > \rho\left(X_1, D^{(1)}\right) > \rho\left(X_1, D\right), \\ \rho\left(C_n, D^{(1)}\right) > \rho\left(C_n, D^{(2)}\right) > \rho\left(C_n, D\right), \\ \rho\left(B_n, D\right) > \rho\left(B_n, D^{(2)}\right) > \rho\left(B_n, D^{(1)}\right), \end{cases}$$

in most real-world networks (see Figs. 1a–1c). The same results can be found in both ER and SF networks (see Appendix B). We theoretically prove the inequality $\rho(X_1, D^{(2)}) > \rho(X_1, D^{(1)}) > \rho(X_1, D)$ in ER networks in Section 4.

Almost all of the Pearson correlation coefficients $\rho(X_1, D^{(2)})$, $\rho(C_n, D^{(1)})$, and $\rho(B_n, D)$ are large (>0.95) in both ER and SF networks (see Figs. B.1 and B.2) and are also large (>0.6) in most real-world networks (see Fig. 1). The betweenness of a power-law distributed network also follows a power-law distribution [43]. This supports the strong linear correlation between the betweenness B_n and the degree D in SF networks [17].

3.2 Centrality similarities $M_{A,B}(\Upsilon)$ between centrality metrics

Different centrality metrics rank the nodes in different orders within a network. The centrality similarity was proposed in reference [23] to quantify the similarity of centrality metrics in ranking nodes.

Definition. In a graph $G(N, L)$ assume we obtain two node rankings, $[a_{(1)}, a_{(2)}, \ldots, a_{(N)}]$ and $[b_{(1)}, b_{(2)}, \ldots, b_{(N)}]$, according to centrality metrics A and B, where $a_{(j)}$ or $b_{(j)}$ is the node whose centrality metric A or B is the jth largest in the networks. The centrality similarity $M_{A,B}(\Upsilon)$ is the percentage of the nodes in $[a_{(1)}, a_{(2)}, \ldots\ldots, a_{(\Upsilon N)}]$, which are also in $[b_{(1)}, b_{(2)}, \ldots\ldots, b_{(\Upsilon N)}]$, where $\Upsilon \in [0, 1]$.

The measure $M_{A,B}(\Upsilon)$ gives the percentage of overlapping nodes from the top $100\Upsilon\%$ of nodes, ranked by the centrality metrics A and B, respectively. The range of $M_{A,B}(\Upsilon)$ is between $[0, 1]$. If the $100\Upsilon\%$ of nodes chosen by centrality metric A are not at all in the $100\Upsilon\%$ of nodes chosen by centrality metric B, $M_{A,B}(\Upsilon) = 0$. It means that the most important (top $100\Upsilon\%$) nodes chosen by the two centrality metrics are completely different, i.e., the centrality metrics A and B differ greatly. When all nodes are chosen ($\Upsilon = 1$) there is a full overlap, which indicates that $M_{A,B}(1) = 1$. For a given $\Upsilon < 1$, a larger $M_{A,B}(\Upsilon)$ represents a stronger correlation between the two centrality metrics A and B.

3.2.1 Centrality similarities in network models

We study the centrality similarity $M_{A,B}(\Upsilon)$ between any two centrality metrics[5] in 10^3 network realizations of ER networks and SF networks with $N = 10^4$ and $\Upsilon = [0.001, 0.01, 0.1]$.

We observe that in both ER and SF networks, the $M_{B_n,D}(\Upsilon)$ is notably larger than the centrality similarity between B_n and any other centrality metric; $M_{C_n,D^{(1)}}(\Upsilon) > M_{C_n,D^{(2)}}(\Upsilon) > M_{C_n,D}(\Upsilon)$; and the centrality similarities $M_{x_1,D^{(1)}}(\Upsilon)$ and $M_{x_1,D^{(2)}}(\Upsilon)$ are both large (see Fig. 2). In ER networks, $M_{x_1,D^{(2)}}(\Upsilon) > M_{x_1,D^{(1)}}(\Upsilon) > M_{x_1,D}(\Upsilon)$. The k-shell index has low similarity with other metrics in ER networks for the same reason mentioned in Section 3.1. All these observations agree with what we have found using the Pearson correlation coefficients in Section 3.1.

[5] Our study shows that the centrality similarity $M_{A,B}(\Upsilon)$ increases with the increase of Υ in ER networks, but decreases with the increase of Υ in SF networks. Note that this observation holds only for small Υ and, if Υ is around 1, $M_{A,B}(\Upsilon) = 1$ in all networks.

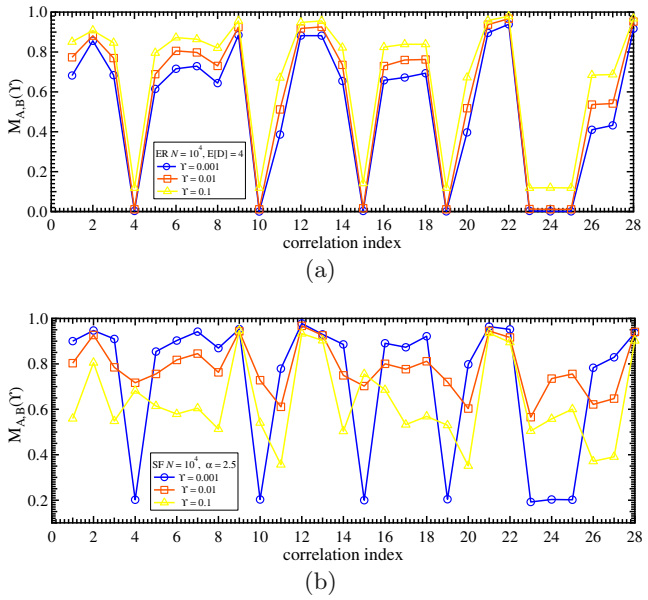


Fig. 2. Centrality similarities between centrality metrics in network models: (a) for ER networks and (b) for SF networks. The x-axis is the correlation index (see Appendix B).

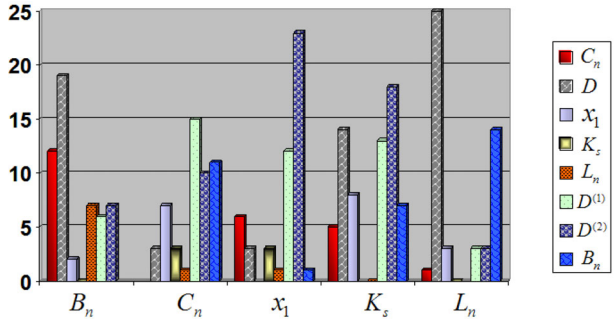


Fig. 3. Number of networks (among the 34 real-world networks) in which $M_{A,B}(\Upsilon)$ is the highest among the centrality similarities between A and all other centrality metrics, when $\Upsilon = 5\%$. The centrality metric A is given by the x-axis label, and B is reflected by the pattern described in the box on right side. Take the betweenness B_n as an example. The centrality similarities between B_n and all the other metrics are compared with each other to find the largest similarity in each real-world network. For instance, the $M_{B_n,C_n}(\Upsilon)$ is the largest centrality similarity in 'Electric_s208' network, so that one is counted into the leftmost bar of B_n (with C_n).

3.2.2 Centrality similarities in real-world networks

For the 34 real-world networks the percentage Υ should be larger than 3%, since the smallest network only has 35 nodes. We compare the similarity between each centrality metric (e.g., B_n) and all other metrics to determine which metric is the closest to the centrality metric (e.g., B_n). In Figure 3 the height of each bar indicates the number of networks in which $M_{A,B}(\Upsilon)$ is the highest among the centrality similarities between A and all the other centrality metrics. The bar chart shows that the D, $D^{(1)}$, and $D^{(2)}$ are, respectively, most similar to B_n, C_n, and x_1 in most real-world networks, which is consistent with what is observed in the network models. We also observe that either $M_{L_n,D}(\Upsilon)$ or $M_{L_n,B_n}(\Upsilon)$ is the largest among the centrality similarities between L_n and all other metrics in most real-world networks.

4 Theoretical analysis

The above simulations indicate that the three lowest-order degree masses, with a low computational complexity, are strongly correlated with the betweenness, the closeness, and the components of the principal eigenvector, all of which are complex to compute. We first prove that the high-order ($m \to \infty$) degree mass is proportional to the principal eigenvector x_1 in any network. Next we prove that when m is small the correlation between degree mass and the principal eigenvector increases with an increase in m, i.e., $\rho(X_1, D^{(2)}) \geq \rho(X_1, D^{(1)}) \geq \rho(X_1, D)$. We then apply the generating function method [44,45] to analyze such statistical properties of the degree masses as expectation and variance (see Appendix C).

Theorem 1. *The mth-order degree mass vector $d^{(m)}$ is proportional to the principal eigenvector x_1 in any network with a sufficiently large spectral gap when $m \to \infty$.*

Proof. The mth-order degree mass vector $d^{(m)}$ is:

$$
\begin{aligned}
d^{(m)} &= \sum_{k=1}^{m+1} \left(A^k u\right) = \sum_{k=1}^{m+1} \sum_{j=1}^{N} \lambda_j^k x_j \left(x_j^T u\right) \\
&= \sum_{j=1}^{N} \left(\lambda_j \frac{\lambda_j^{m+1} - 1}{\lambda_j - 1}\right) \left(x_j^T u\right) x_j \\
&= \left(\lambda_1 \frac{\lambda_1^{m+1} - 1}{\lambda_1 - 1}\right) \left(x_1^T u\right) x_1 + \sum_{j=2}^{N} \left(\lambda_j \frac{\lambda_j^{m+1} - 1}{\lambda_j - 1}\right) \left(x_j^T u\right) x_j \\
&= \left(\lambda_1 \frac{\lambda_1^{m+1} - 1}{\lambda_1 - 1}\right) \left(x_1^T u\right) x_1 \left(1 + O\left(\sum_{j=2}^{N} \left(\frac{|\lambda_j|}{|\lambda_1|}\right)^m\right)\right).
\end{aligned}
$$

Literature [34] has proved that $x_1^T u > x_j^T u$ for all $1 < j \leq N$. Accordingly, the term $\sum_{j=2}^{N} (\lambda_j \frac{\lambda_j^{m+1}-1}{\lambda_j - 1})(x_j^T u) x_j$ is small in the graphs with a large spectral gap ($\lambda_1 - \lambda_2$). When m increases, $d^{(m)} \to (\lambda_1 \frac{\lambda_1^{m+1}-1}{\lambda_1 - 1})(x_1^T u) x_1$. Moreover, when m is large, especially when $m \to \infty$, $O(\sum_{j=2}^{N} (\frac{|\lambda_j|}{|\lambda_1|})^m) \to 0$ in any graph. Thus we find that $d^{(m)}$ tends to be proportional to x_1 when m increases in networks with a large spectral gap, and $d^{(m)} \sim \lambda_1^{(m+1)}(x_1)$ in networks when $m \to \infty$.

Lemma 1. *In large sparse Erdős-Rényi (ER) networks, $\rho(D^{(2)}, X_1) \geq \rho(D^{(1)}, X_1) \geq \rho(D, X_1)$.*

Proof. See Appendix C.

5 Application to the inflexible contrarian opinion (ICO) model

In this section we apply the studied centrality metrics to select the inflexible contrarians in the inflexible contrarian opinion (ICO) model [32] to help one opinion to compete with another. Both network models and three social networks will be considered.

5.1 The ICO model

The ICO model is a variant of the non-consensus opinion (NCO) model [29]. The ICO and NCO models are both opinion competition models in which two opinions exist and compete with each other. In the NCO model opinions are randomly assigned to all agents (nodes). At time $t = 0$ each agent is assigned opinion A with a probability f and opinion B with a probability $1 - f$. At each subsequent time step each agent adopts the opinion of the majority of its nearest neighbors and itself. When there is a tie, the opinion of the agent does not change. All of the updates are made simultaneously in parallel at each step. The system reaches a state in which the opinions A and B coexist and are stable when f is above a critical threshold f_c.

When the NCO model is in the stable state, the ICO model further selects a fraction p_o of agents with opinion A to be the inflexible contrarians who will hold opinion B, will never change their opinion, but will influence the opinion of other agents. The two opinions then compete with each other according to the update rules of the NCO model. The system will reach a new stable state by following these opinion dynamics.

We use S_1 and S_2 to denote the size of the largest and the second largest clusters of agents with opinion A in the new stable state. A phase transition threshold f_c separates two different phases of the stable state. When $f > f_c$, a giant component of agents with opinion A exists and the coexistence of opinions A and B is stable. When $f \leq f_c$, no giant component of agents with opinion A exists ($S_1 = 0$). The f_c depends on p_o. When $p_o = 0$, the ICO model clearly reduces to the classical NCO model and they have the same critical threshold f_c. When $0 < p_o < p^*$, the threshold f_c of the ICO model increases with p_o, but the size S_1 for the finial stable state decreases with p_o. When p is above a certain value p^*, the phase transition no longer occurs, and the giant component of agents with opinion A is completely destroyed ($S_1 = 0$).

5.2 Strategies of selecting inflexible contrarians using centrality metrics

The final stable state of the ICO model is affected not only by the percentage p_o, but also by how inflexible contrarian agents are selected. Here we select the inflexible contrarians based on their centrality metrics. Li et al. [32] studied the ICO model by choosing the inflexible contrarian agents with opinion A either randomly or according to highest degree. The degree strategy is significantly more

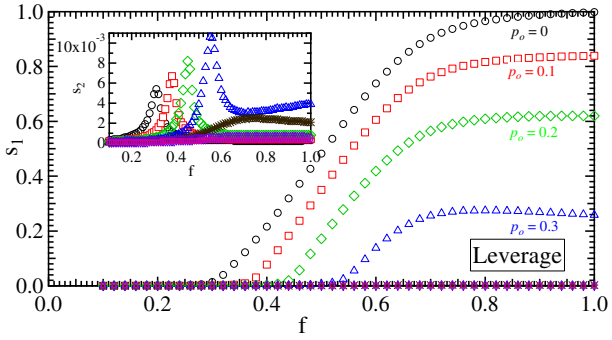


Fig. 4. An example: the results of leverage strategy. Plot of $s_1 \equiv S_1/N$ as a function of f for different values of p_o for ER networks with $E[D] = 4$ and $N = 10^4$. We denote by S_1 the size of the largest A opinion cluster in the steady-state. Different marks show the results of ICO model with different p_o: $p_o = 0(\circ)$, p_o=0.1($\square$), $p_o = 0.2(\diamond)$, $p_o = 0.3(\triangle)$, $p_o = 0.4(*)$, $p_o = 0.5(\Diamond)$, $p_o = 0.6(\boxtimes)$. The insets plot the $s_2 \equiv S_2/N$, where S_2 is the size of the second largest A opinion cluster, as a function of the f for different values of p_o.

effective than the random strategy in reducing the size S_1 of the largest opinion A cluster in the stable state when p_o is the same. Here we want to determine which centrality metric used to pick the inflexible contrarians reduces S_1 most efficiently. We also want to determine whether the S_1 decrease is similar when the inflexible contrarians are chosen based on two strongly correlated (with a large Pearson correlation coefficient or a high centrality similarity) centrality metrics. Here the inflexible contrarians are chosen as nodes with highest (i) betweenness; (ii) degree; (iii) 1st-order degree mass; (iv) 2nd-order degree mass; (v) eigenvector component; (vi) k-shell index; or (vii) leverage or (viii) chosen randomly.

5.3 Comparison of inflexible contrarian selection strategies

We first compare the efficiency in decreasing the size S_1 of the largest opinion A cluster in ER and SF networks when choosing the inflexible contrarians using different centrality metrics. We consider ER networks ($N = 10^4$ or 10^5) with $E[D] = 4$, and SF networks ($N = 10^4$ or 10^5) with $\alpha = 2.5$, and perform all the simulations on 10^3 network realizations. Figure 4 shows a plot of $s_1 = S_1/N$ as a function of f for different values of p_o in ER networks (with $N = 10^4$) using a leverage strategy. The size $s_2 = S_2/N$ shows a sharp peak, a characteristic of a second-order phase transition, in the insets of Figure 4. As p_o increases, f_c shifts to a larger value and the largest cluster becomes significantly smaller. When $p > p^*$, the giant component with opinion A disappears, i.e., $S_1 = 0$. For example, the p^* value for the leverage strategy is between 0.3 and 0.4 (see Fig. 4). A small p^* implies that the inflexible contrarians can efficiently destroy the largest opinion A cluster. We can compare the efficiency of the strategies in decreasing S_1 by the value of p^*. When we compare strategies in the ICO model with the same p_o,

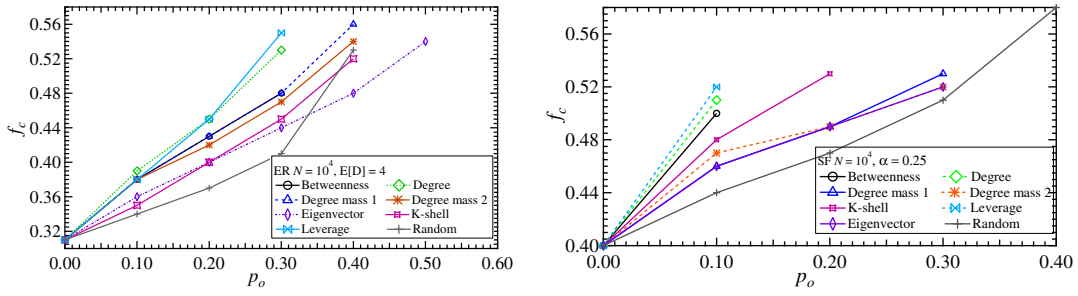


Fig. 5. Plot of f_c as a function of p_o for strategies 1 to 8: (a) in ER graphs with $N = 10^4$, $E[D] = 4$; (b) in SF graphs with $N = 10^4$, $D_{min} = 2$, $\alpha = 2.5$.

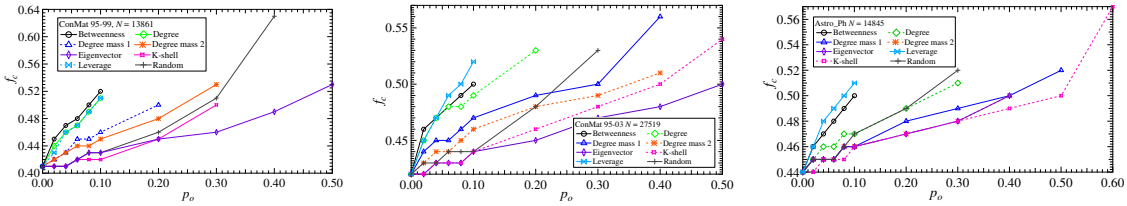


Fig. 6. Plot of f_c as a function of p_o for strategies in social networks: (a) in network of coauthorships between scientists posting preprints on ConMat E-Print Archives between 1995 to 1999; (b) in network of coauthorships between scientists posting preprints on ConMat E-Print Archives between 1995 to 2003; (c) in network of coauthorships between scientists posting preprints on Astrophysics E-Print Archives between 1995 to 1999.

a larger phase transition f_c for a strategy indicates that the inflexible contrarians chosen using this strategy decreases S_1 more efficiently. Figure 5a plots the phase transition f_c as a function of p_o. Note that the efficiency of each strategy is ranked in decreasing order as: Leverage, Degree, Betweenness, 1st-order Degree mass, 2nd-order Degree mass, k-shell index, Principal Eigenvector, and Random. The same result can be also found in ER and SF networks with $N = 10^5$.

We find that all strategies are more efficient in SF networks than in ER networks of the same size. We base this on two observations. First, the relative change of f_c with p_o for all strategies in SF networks is larger than it is in ER networks. Second, the p^* for all strategies in SF is much smaller than it is in ER networks. The reason for this may be that (i) hubs can be readily selected as inflexible contrarians when using centrality metrics in SF networks, and (ii) hubs can strongly influence the opinion of their large number of nearest neighbors.

Figure 6 compares these centrality metrics in real-world networks, i.e., the ConMat 95-99 network, the ConMat 95-03 network, and the Astro_Ph network. Note that the inflexible contrarians selected using the leverage L_n, the betweenness B_n, and the degree D are the most efficient in helping opinion B win the competition. The similar behaviors of the three strategies are supported by the large Pearson correlation coefficient $\rho(B_n, D)$ and the large centrality similarities $M_{B_n,D}(\Upsilon)$, $M_{L_n,D}(\Upsilon)$ and $M_{L_n,B_n}(\Upsilon)$.

In both network models and real-world networks, strongly correlated centrality metrics tend to perform similarly. For example, we have discovered both numerically and theoretically that $\rho(D^{(2)}, X_1) \geq \rho(D^{(1)}, X_1)$. Correspondingly, the principal eigenvector x_1 strategy performs closer to the 2nd-order degree mass $D^{(2)}$ than the 1st-order degree mass $D^{(1)}$ in the ICO model.

6 Conclusion

In this paper we have studied the correlation between widely studied and recently proposed centrality metrics in numerous real-world networks as well as in network models, i.e., as in Erdős-Rényi (ER) random networks and scale-free (SF) networks. A strong correlation between two centrality metrics indicates the possibility of approximating one centrality metric, usually the one with a higher computational complexity, using the other. We study the correlations between the centrality metrics using the Pearson correlation coefficient and the centrality similarity. An important finding is that the degree D, the 1st-order degree mass $D^{(1)}$, and the 2nd-order degree mass $D^{(2)}$ are strongly correlated with the betweenness B_n, the closeness C_n, and the principal eigenvector x_1, respectively. This observation is partially supported by our analytical proof that $\rho(X_1, D^{(2)}) > \rho(X_1, D^{(1)}) > \rho(X_1, D)$.

We have introduced the degree mass $D^{(m)}$ as a new network centrality metric. The 0th-order degree mass is the degree and the high-order ($m \to \infty$) degree mass is proportional to the principal eigenvector x_1. We also find that the influence of network size (the number N of nodes) on the Pearson correlation coefficients is small. In addition, the leverage L_n has high centrality similarities with the degree D and the betweenness B_n. We use these

centrality metrics to select the inflexible contrarians in the ICO model to help one opinion to compete with the other. The leverage L_n turns out to be the most efficient strategy in both network models and real-world networks. We also find that strongly correlated metrics perform similarly in the ICO model. This suggests that the metrics with a low computational complexity, such as the degree D and the leverage L_n, could be used to approximate more complex metrics, e.g., the betweenness B_n, to locate important nodes in complex networks. Examples of important nodes would include inflexible contrarians in opinion propagation networks and nodes that should be immunized in disease transmission networks.

The authors are grateful to Shlomo Havlin for discussion and useful comments. This work has been supported by the European Commission within the framework of the CONGAS project FP7-ICT-2011-8-317672 and the China Scholarship Council (CSC).

References

1. S.H. Strogatz, Nature **410**, 268 (2001)
2. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Phys. Rep. **424**, 175 (2006)
3. A. Barrat, M. Barthelemy, A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008)
4. C.H. Comin, L. da Fontoura Costa, Phys. Rev. E **84**, 056105 (2011)
5. M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse, Nat. Phys. **6**, 888 (2010)
6. J. Borge-Holthoefer, Y. Moreno, Phys. Rev. E **85**, 026116 (2012)
7. R. Pastor-Satorras, A. Vespignani, Phys. Rev. E **65**, 036104 (2002)
8. S.P. Borgatti, Social Netw. **27**, 55 (2005)
9. L.C. Freeman, Social Netw. **1**, 215 (1979)
10. N.E. Friedkin, Am. J. Soc. **96**, 1478 (1991)
11. B. Mullen, C. Johnson, E. Salas, Soc. Netw. **13**, 169 (1991)
12. M.E.J. Newman, in *The New Palgrave Encyclopedia of Economics*, edited by L.E. Blume, S.N. Durlauf (Palgrave Macmillan, Basingstoke, 2008)
13. P. Van Mieghem, `arXiv:1401.4580` (2014)
14. K.E. Joyce, P.J. Laurienti, J.H. Burdette, S. Hayasaka, PLoS One **5**, e12200 (2010)
15. P.-J. Kim, H. Jeong, Eur. Phys. J. B **55**, 109 (2007)
16. D. Koschützki, F. Schreiber, Comparison of centralities for biological networks, in *German Conference on Bioinformatics, 2004*, pp. 199–206
17. E. Estrada, Ecological Complexity **4**, 48 (2007)
18. C. Li, H. Wang, P. Van Mieghem, Degree and principal eigenvectors in complex networks, in *Proceedings of NETWORKING 2012* (Springer, 2012), pp. 149–160
19. M. Faloutsos, P. Faloutsos, C. Faloutsos, ACM SIGCOMM Computer Communication Review **29**, 251 (1999)
20. L. da F. Costa, F.A. Rodrigues, G. Travieso, P.R. Villas Boas, Adv. Phys. **56**, 167 (2007)
21. A. Jamakovic, S. Uhlig, Networks and Heterogeneous Media **3**, 345 (2008)
22. C. Li, H. Wang, W. de Haan, C.J. Stam, P. Van Mieghem, J. Stat. Mech. **2011**, P11018 (2011)
23. S. Trajanovski, J. Martín-Hernández, W. Winterbach, P. Van Mieghem, J. Complex Netw. **1**, 44 (2013)
24. P. Erdős, A. Rényi, Publ. Math. Debrecen **6**, 290 (1959)
25. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
26. R. Cohen, S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, Cambridge, 2010)
27. S. Galam, Europhys. Lett. **70**, 705 (2005)
28. C. Castellano, S. Fortunato, V. Loreto, Rev. Mod. Phys. **81**, 591 (2009)
29. J. Shao, S. Havlin,, H.E. Stanley, Phys. Rev. Lett. **103**, 018701 (2009)
30. Q. Li, L.A. Braunstein, H. Wang, J. Shao, H.E. Stanley, S. Havlin, J. Stat. Phys. **151**, 92 (2013)
31. B. Qu, Q. Li, S. Havlin, H.E. Stanley, H. Wang, `arXiv:1404.7318` (2014)
32. Q. Li, L.A. Braunstein, S. Havlin, H.E. Stanley, Phys. Rev. E **84**, 066101 (2011)
33. P. Van Mieghem, *Performance Analysis of Complex Networks and Systems* (Cambridge University Press, 2014)
34. P. Van Mieghem, *Graph spectra for complex networks* (Cambridge University Press, Cambridge, 2011)
35. J.M. Anthonisse, The rush in a directed graph, *Stichting Mathematisch Centrum. Mathematische Besliskunde*, No. BN 9/71, 1971, pp. 1–10
36. H. Wang, J.M. Hernandez, P. Van Mieghem, Phys. Rev. E **77**, 046105 (2008)
37. D. Koschützki, K.A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, O. Zlotowski, in *Network Analysis: Methodological Foundations* (Springer, 2005), pp. 16–61
38. H.-W. Ma, A.-P. Zeng, Bioinformatics **19**, 1423 (2003)
39. S.B. Seidman, Social Netw. **5**, 269 (1983)
40. B. Pittel, J. Spencer, N. Wormald, J. Combinatorial Theory Ser. B **67**, 111 (1996)
41. V. Batagelj, M. Zaversnik, Adv. Data Anal. Classi. **5**, 129 (2011)
42. R. Cohen, K. Erez, D. Ben-Avraham, S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000)
43. M.P. Joy, A. Brock, D.E. Ingber, S. Huang, BioMed Res. Int. **2005**, 96 (2005)
44. P. Van Mieghem, *Performance Analysis of Communications Networks and Systems* (Cambridge University Press, Cambridge, 2006)
45. M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. E **64**, 026118 (2001)
46. M. Krivelevich, B. Sudakov, Comb. Probab. Comput. **12**, 61 (2003)
47. I.J. Farkas, I. Derényi, A.-L. Barabási, T. Vicsek, Phys. Rev. E **64**, 026704 (2001)

Appendix A: Description of the real-world networks

A.1 Descriptions

Table A.1. Descriptions of real-world networks.

Index	Networks	Descriptions
1	American airline	The direct airport-to-airport American mileage a maintained by the U.S. Bureau of Transportation Statistics.
2	American football	This is the network of American football games between Division IA colleges during regular season Fall 2000, as compiled by M. Girvan and M. Newman.
3	ARPANET80	The Advanced Research Projects Agency Network as seen in 1980.
4	Celegensneural	Network representing the neural network of C. Elegans.
5	Dophins	An undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand.
6	Dutch soccer	Dutch football players represent the nodes. Two nodes are linked if they played together a match.
7	Gnutella 1	Gnutella snapshots. Four different crawls are available.
8	Gnutella 2	
9	Gnutella 3	
10	Gnutella 4	
11	Karate	Social network of friendships between 35 members of a karate club at a US university in the 1970.
12	LesMis	Coappearance network of characters in the novel Les Miserables.
13	Surfnet	SURFNET topology inferred from the switch interface interconnections.
14	Electric s208	ISCAS89 Sequential Benchmark Circuits. Each node represents a logical operation implemented
15	Electric s420	physically. Links between them relate their inputs/outputs.
16	Electric s838	
17	Epowergridl1	Power-grid infrastructure at three different levels of one city-area in Western Europe.
18	Epowergridl2	
19	Epowergridl3	
20	Erailwayl1	Railway infrastructure at two levels of one Western-European country
21	Erailwayl2	
22	WordAdj	Adjacency network of common adjectives and nouns in the novel David Copperfield by Charles Dickens.
23	WordAdjEnglish	Word-adjacency networks of texts in English, French and Japanese separately.
24	WordAdjFranch	
25	WordAdjJapanese	
26	Internet AS (01')	Internet snapshot retrieved from the merge of different data sources (BGP routing tables and updates: Route Views, RIPE, Abilene, CERNET, BGP View).
27	Astro_Ph	Network of coauthorships between scientists posting preprints on the Astrophysics E-Print Archive between Jan 1, 1995 and December 31, 1999.
28	SciMet	Web of Science C. The citation network was created using the Web of Science database SciMet. Networks created with the tool HistCite.
29	HighE-th	High Energy Theory C. Network of coauthorships between scientists posting preprints on the High-Energy Theory E-Print Archive between Jan 1, 1995 and December 31, 1999.
30	CondMat 95-03	Network of coauthorships between scientists posting preprints on the Condensed Matter E-Print
31	CondMat 95-99	Archive. We have two networks corresponding to different periods of time. Periods are Jan 1, 1995-December 31, 1999 and 2003 respectively.
32	Dutch Roadmap	A graph representing the interconnection between cities in the Netherlands.
33	Network Science C	Coauthorship network of scientists working on network theory and experiment, as compiled by M. Newman in May 2006.
34	Next Generation	A typical Next Generation Transport network.

Table A.2. Properties of real-world networks. The real-world network index is shown in Table A.1. N is the number of nodes, L is the number of links. $E[H]$ is the average shortest path, C_G is the clustering coefficient of networks. ρ_D is the degree correlation coefficient (called the assortativity) of networks. λ_1 is the largest eigenvalue (called spectral radius) of the adjacency matrix of the network. μ_{N-1} is the second smallest Laplace eigenvalue (called spectral radius) of the networks. μ_1/μ_{N-1} is the ratio of the largest eigenvalue μ_1 and the second smallest eigenvalue μ_1 of Laplacian matrix. R_G is the effective graph resistance.

Index	N	L	E[H]	C_G	ρ_D	λ_1	μ_{N-1}	μ_1/μ_{N-1}	R_G	$E[D]$	$\sqrt{\mathrm{Var}[D]}$	$H_{\max}$
1	2179	31326	3.0262	0.4849	−0.0409	144.6112	0.2082	$2.0675e^3$	$1.6072e^4$	28.7526	56.6782	8
2	115	613	2.5082	0.4032	0.1624	10.7806	1.4590	10.7350	$1.5086e^3$	10.6609	0.8835	4
3	71	86	6.4849	0.0141	−0.2613	2.7648	0.0374	170.2063	$7.0158e^3$	2.4225	0.7442	17
4	297	2148	2.4553	0.2924	−0.1632	24.3655	0.8485	159.1562	$1.3710e^4$	14.4646	12.9443	5
5	62	159	3.3570	0.2590	−0.0436	7.1936	0.1730	78.7034	$1.8643e^3$	5.1290	2.9319	8
6	685	10 310	4.4583	0.7506	−0.0634	50.8428	0.1613	372.0373	$3.1157e^4$	30.1022	21.1957	11
7	737	803	9.1351	0.0063	−0.1934	4.8913	0.0073	$2.6292e^3$	$1.4181e^6$	2.1791	2.0069	24
8	1568	1906	6.1037	0.0192	−0.0946	13.7828	0.0167	$1.1205e^4$	$4.0212e^4$	2.4311	5.5778	21
9	435	459	6.7085	0.0145	−0.3301	8.2281	0.0110	$5.9278e^3$	$4.2533e^5$	2.1103	5.1534	20
10	653	738	5.4513	0.0232	−0.2459	12.1145	0.0231	$6.2319e^3$	$6.6603e^5$	2.2603	7.0228	15
11	35	134	1.9126	0.3908	−0.5036	9.6253	1.7264	12.6030	221.6283	7.6571	4.7265	3
12	77	254	2.6411	0.5731	−0.1652	12.0058	0.2050	180.9490	$3.0166e^3$	6.5974	6.0006	5
13	65	111	4.1236	0.0359	0.2288	5.0523	0.1137	92.7068	$3.2979e^3$	3.4154	1.9046	10
14	122	189	4.9278	0.0591	−0.0020	4.1036	0.0836	135.2786	$1.3082e^4$	3.0984	1.4395	11
15	252	399	5.8064	0.0651	−0.0059	4.3600	0.0512	297.3970	$5.8313e^4$	3.1667	1.5340	13
16	512	819	6.8585	0.0547	−0.0300	5.0097	0.0285	809.9553	$2.5149e^5$	3.1992	1.6296	15
17	3419	3953	21.1147	0.0120	−0.1283	5.1781	$<e^{-5}$	$>e^{15}$	$4.8953e^7$	2.3124	1.8425	51
18	1205	1384	12.3547	0.0171	0.1082	4.8994	0.0022	$9.1191e^3$	$4.3901e^6$	2.2971	1.3609	31
19	395	441	13.6088	0.0201	−0.0235	4.4854	0.0020	$8.8844e^3$	$7.2535e^5$	2.2329	1.2834	42
20	8710	11 332	79.0448	0.0212	−0.0219	2.9865	$<e^{-5}$	$>e^{15}$	$7.2107e^8$	2.6021	0.7696	213
21	689	778	34.1261	0.0731	0.0980	3.6926	$7.7321e^{-3}$	$1.0526e^4$	$3.9229e^6$	2.2583	0.7658	84
22	112	425	2.5356	0.1728	−0.1293	13.1502	0.6950	72.0767	$3.7941e^3$	7.5893	6.8512	5
23	7377	44205	2.7780	0.4085	−0.2366	109.4416	$<e^{-5}$	$9.1266e^{15}$	$2.2149e^7$	11.9846	60.8260	8
24	8308	23 832	3.2189	0.2138	−0.2330	60.6735	0.1197	$1.5810e^4$	$3.9917e^7$	5.7371	34.8979	9
25	2698	7995	3.0771	0.2196	−0.2590	42.9980	$<e^{-5}$	$5.8851e^{15}$	$4.3489e^6$	5.9266	24.6695	8
26	12 254	25 319	3.6214	0.2992	−0.1903	61.1066	$<e^{-5}$	$4.8974e^{15}$	$1.0349e^8$	4.1324	33.5463	11
27	14 845	11 9652	4.7980	0.6696	0.2277	73.8868	0.0302	$1.1966e^4$	$7.2012e^7$	16.1202	21.7466	14
28	2678	10 368	4.1797	0.1736	−0.0352	20.4290	0.0853	$1.9365e^3$	$2.9549e^6$	7.7431	9.2480	12
29	5835	13 815	7.0264	0.5062	0.1852	18.0442	0.0214	$2.3870e^3$	$2.8800e^7$	4.7352	4.5571	19
30	27 519	11 6181	5.7667	0.6546	0.1657	40.3097	0.0276	$7.3675e^3$	$3.3638e^8$	8.4437	10.8110	16
31	13 861	44 619	6.6278	0.6514	0.1571	24.9822	0.0292	$3.6992e^3$	$1.1613e^8$	6.4381	6.7598	18
32	29 663	34 982	148.7102	0.0443	0.2462	3.4567	$<e^{-5}$	$>e^{15}$	$1.5472e^{10}$	2.3586	0.6823	531
33	379	914	6.0419	0.7412	−0.0817	10.3755	0.0152	$2.3053e^3$	$1.4826e^5$	4.8232	3.9272	17
34	29 902	32 707	7109.8681	0.0306	−0.0355	49.5455	$<e^{-5}$	$>e^{15}$	$2.1188e^{12}$	2.1876	9.7574	14 253

A.2 Properties of the real-world networks

The properties of real-world networks are shown in Table A.2. The definition of these properties has been described in detail in reference [22].

Appendix B: Pearson correlation coefficients between centrality metrics

The correlation indexes mentioned in the following images and tables are the indexes for pairs of centrality metrics: 1. (B_n, C_n); 2. (B_n, D); 3. (B_n, x_1); 4. (B_n, K_s); 5. (B_n, L_n); 6. $(B_n, D^{(1)})$; 7. $(B_n, D^{(2)})$; 8. (C_n, D); 9. (C_n, x_1); 10. (C_n, K_s); 11. (C_n, L_n); 12. $(C_n, D^{(1)})$; 13. $(C_n, D^{(2)})$; 14. (D, x_1); 15. (D, K_s); 16. (D, L_n); 17. $(D, D^{(1)})$; 18. $(D, D^{(2)})$; 19. (x_1, K_s); 20. (x_1, L_n); 21. $(x_1, D^{(1)})$; 22. $(x_1, D^{(2)})$; 23. (K_s, L_n); 24. $(K_s, D^{(1)})$; 25. $(K_s, D^{(2)})$; 26. $(L_n, D^{(1)})$; 27. $(L_n, D^{(2)})$; 28. $(D^{(1)}, D^{(2)})$.

Appendix C: Proof of Lemmas

Lemma 2. *In an Erdős-Rényi (ER) random network $G_p(N)$, when $N \to \infty$, the average 1st-order degree mass is:*

$$E\left[D^{(1)}\right] = N\left(2p + p^2N\right), \tag{C.1}$$

and the variance is:

$$\mathrm{Var}\left[D^{(1)}\right] = N\left(2p + 4p^2N + p^3N^2\right). \tag{C.2}$$

The average and the variance of 2nd-order degree mass are

$$E\left[D^{(2)}\right] = N\left(2p + 3p^2N + p^3N^2\right), \tag{C.3}$$

$$\mathrm{Var}\left[D^{(2)}\right] = N\left(2p + 14p^2N + 17p^3N^2 + 7p^4N^3 + p^5N^4\right). \tag{C.4}$$

Table B.1. Pearson correlation coefficients among the centrality metrics in the real-world networks.

Index	1	2	3	4	5	6	7	8	9
	$\rho(B_n, C_n)$	$\rho(B_n, D)$	$\rho(B_n, x_1)$	$\rho(B_n, K_s)$	$\rho(B_n, L_n)$	$\rho(B_n, D^{(1)})$	$\rho(B_n, D^{(2)})$	$\rho(C_n, D)$	$\rho(C_n, x_1)$
1	0.3667	0.5690	0.4119	0.3377	0.4027	0.4314	0.4224	0.7580	0.7684
2	0.8167	0.2813	0.1450	0.0871	0.3212	0.2230	0.2075	0.2913	0.2462
3	0.7129	0.7235	0.5358	0.3496	0.55585	0.7660	0.7593	0.4308	0.6851
4	0.4271	0.7805	0.5206	0.1822	0.4212	0.5388	0.6044	0.6997	0.7827
5	0.6657	0.5902	0.2835	0.4703	0.5639	0.5131	0.4850	0.7127	0.6979
6	0.3303	0.4909	0.0857	0.1523	0.4170	0.3807	0.3113	0.2701	−0.1604
7	0.4456	0.7292	0.4780	0.5182	0.4556	0.7575	0.7882	0.3973	0.5241
8	0.2196	0.9691	0.7006	0.2677	0.2679	0.3858	0.9416	0.2225	0.5469
9	0.2475	0.8839	0.4926	0.4667	0.4356	0.3533	0.8283	0.1763	0.5112
10	0.2338	0.9603	0.5848	0.3296	0.3880	0.2640	0.8839	0.1774	0.5733
11	0.8699	0.9651	0.8757	0.3782	0.8707	0.7999	0.9166	0.8853	0.9599
12	0.6287	0.7468	0.4231	0.2388	0.5317	0.5534	0.5468	0.7997	0.6812
13	0.7136	0.8743	0.7365	0.6345	0.6985	0.7999	0.7816	0.8290	0.9286
14	0.6408	0.7475	0.5595	0.2147	0.5551	0.7357	0.7227	0.6127	0.7987
15	0.5956	0.6933	0.5514	0.1583	0.4508	0.7084	0.7203	0.5541	0.7178
16	0.5323	0.7044	0.5410	0.1314	0.3913	0.6971	0.7661	0.4623	0.5633
17	0.2349	0.3843	0.1180	0.1189	0.1889	0.4101	0.4082	0.1082	0.0607
18	0.3210	0.7005	0.5517	0.0560	0.2686	0.6772	0.7144	0.2946	0.4627
19	0.3001	0.7081	0.4775	0.1060	0.2945	0.6371	0.6825	0.2395	0.4925
20	0.2664	0.1565	−0.0442	0.1979	0.1112	0.1805	0.1876	0.1477	0.0209
21	0.5022	0.3274	0.0364	0.3836	0.2548	0.2790	0.2540	0.2428	0.1141
22	0.6559	0.9150	0.8226	0.3517	0.6586	0.7891	0.8444	0.8410	0.9245
23	0.1880	0.9225	0.6525	0.2068	0.2642	0.4157	0.7765	0.3535	0.6528
24	0.1874	0.9714	0.8047	0.2729	0.2636	0.4403	0.9385	0.2625	0.6215
25	0.2747	0.9660	0.7859	0.3249	0.3584	0.5266	0.8972	0.3868	0.6880
26	0.1382	0.9826	0.7994	0.3292	0.2290	0.3441	0.9582	0.1631	0.5776
27	0.3764	0.6787	0.4353	0.2869	0.4631	0.5670	0.5270	0.6109	0.4220
28	0.4068	0.8185	0.6959	0.3147	0.4401	0.7143	0.7605	0.6741	0.7030
29	0.4526	0.7798	−0.0109	0.3574	0.5079	0.6700	0.5803	0.5774	0.0119
30	0.3801	0.7534	0.3753	0.3152	0.4488	0.5933	0.5173	0.5989	0.3906
31	0.4002	0.7225	0.2781	0.2607	0.4581	0.5718	0.4816	0.5616	0.3248
32	0.2214	0.1741	−0.0037	0.1619	0.1117	0.1719	0.1608	0.1450	−0.0221
33	0.4302	0.6883	0.1884	0.1917	0.4707	0.5630	0.4997	0.3468	0.2593
34	−0.1342	−0.0436	−0.6295	−0.9718	0.9538	−0.9051	−0.1342	0.0313	0.2446

Index	10	11	12	13	14	15	16	17	18	19
	$\rho(C_n, K_s)$	$\rho(C_n, L_n)$	$\rho(C_n, D^{(1)})$	$\rho(C_n, D^{(2)})$	$\rho(D, x_1)$	$\rho(D, K_s)$	$\rho(D, L_n)$	$\rho(D, D^{(1)})$	$\rho(D, D^{(2)})$	$\rho(x_1, K_s)$
1	0.8174	0.5944	0.7903	0.7712	0.9592	0.8730	0.7259	0.9657	0.9643	0.9254
2	0.1742	0.2704	0.2826	0.2839	0.7501	0.3881	0.9181	0.9619	0.9314	0.2456
3	0.3807	0.2524	0.5598	0.5870	0.4650	0.5127	0.8914	0.9020	0.9079	0.1326
4	0.6861	0.5776	0.8680	0.7951	0.7810	0.5434	0.7886	0.8830	0.9311	0.5572
5	0.7498	0.6094	0.7475	0.7422	0.7196	0.8303	0.9050	0.9574	0.9417	0.5388
6	0.0680	0.2221	0.2381	0.1801	0.6237	0.7300	0.8963	0.9393	0.8801	0.7983
7	0.5073	0.2052	0.6184	0.6248	0.4660	0.5933	0.8117	0.8217	0.8573	0.3912
8	0.4015	0.0017	0.7515	0.3210	0.6523	0.3463	0.3888	0.3594	0.9132	0.1840
9	0.2377	−0.3534	0.8326	0.3544	0.5811	0.3316	0.4651	0.3050	0.9493	0.2032
10	0.2234	−0.2234	0.8594	0.2967	0.6366	0.2492	0.3751	0.2256	0.9481	0.0868
11	0.5492	0.7227	0.9606	0.9463	0.9392	0.5331	0.9390	0.8718	0.9714	0.6221
12	0.5622	0.6340	0.8375	0.7931	0.8467	0.7969	0.8474	0.9455	0.9380	0.8100
13	0.7311	0.3466	0.9330	0.9363	0.9046	0.8289	0.7598	0.9486	0.9391	0.8425
14	0.5670	0.3265	0.7388	0.7574	0.6757	0.4296	0.8184	0.9260	0.9225	0.3108
15	0.5257	0.2675	0.6964	0.7100	0.6147	0.3995	0.7980	0.9078	0.9200	0.2464
16	0.4949	0.1937	0.6534	0.6411	0.4120	0.3738	0.7690	0.8670	0.9055	0.1143
17	−0.0402	−0.0122	0.1651	0.1653	0.2143	0.4102	0.6878	0.7733	0.8456	0.0447
18	0.1582	0.1027	0.4752	0.4902	0.5040	0.1904	0.5901	0.8725	0.8851	0.0638
19	0.2490	0.2137	0.5599	0.5316	0.5183	0.2287	0.5911	0.7611	0.8338	0.0327
20	0.1649	0.0836	0.1829	0.2016	0.1031	0.7905	0.9247	0.9522	0.9241	0.1132
21	0.4880	0.0325	0.3314	0.3382	0.2678	0.4149	0.7508	0.8884	0.8524	0.0835
22	0.8194	0.7371	0.9451	0.9123	0.9575	0.6433	0.8327	0.9390	0.9707	0.7010
23	0.7195	0.3891	0.8312	0.5353	0.8704	0.4649	0.4862	0.6580	0.9504	0.7992
24	0.6355	0.0669	0.8167	0.4111	0.8733	0.4146	0.3627	0.5403	0.9779	0.6980
25	0.6814	0.2080	0.8410	0.5506	0.8911	0.5155	0.5048	0.6631	0.9694	0.7628
26	0.4291	−0.0707	0.7971	0.2788	0.8253	0.3935	0.2696	0.3771	0.9754	0.5413
27	0.5427	0.2819	0.5861	0.5264	0.7188	0.8070	0.5920	0.9352	0.8728	0.5695
28	0.8188	0.5093	0.7923	0.7456	0.8345	0.6962	0.7237	0.9204	0.9236	0.6212
29	0.4884	0.2103	0.6517	0.6022	0.1789	0.7311	0.7080	0.9080	0.8292	0.5171
30	0.6341	0.2404	0.6153	0.5392	0.6346	0.7339	0.6197	0.9035	0.8259	0.5001
31	0.5157	0.2077	0.6067	0.5300	0.5304	0.7166	0.6631	0.8941	0.8021	0.4229
32	0.1465	−0.0170	0.2033	0.2220	0.0364	0.5291	0.7674	0.9271	0.8880	0.0101
33	0.0926	0.0970	0.4562	0.4120	0.4748	0.6803	0.7723	0.8795	0.8415	0.4195
34	0.3609	−0.3531	0.3378	0.0649	0.7297	0.0866	0.0487	0.1570	0.9858	0.6768

Table B.1. Continued.

Index	20	21	22	23	24	25	26	27	28
	$\rho(x_1, L_n)$	$\rho(x_1, D^{(1)})$	$\rho(x_1, D^{(2)})$	$\rho(K_s, L_n)$	$\rho(K_s, D^{(1)})$	$\rho(K_s, D^{(2)})$	$\rho(L_n, D^{(1)})$	$\rho(L_n, D^{(2)})$	$\rho(D^{(1)}, D^{(2)})$
1	0.6327	0.9978	0.9998	0.7122	0.9389	0.9245	0.6604	0.6405	0.9984
2	0.4881	0.8660	0.9134	0.4481	0.3467	0.3274	0.7771	0.7189	0.9929
3	0.1934	0.6460	0.7101	0.5798	0.4773	0.4407	0.6485	0.6530	0.9811
4	0.6130	0.9783	0.9885	0.7710	0.6277	0.5737	0.6605	0.6789	0.9813
5	0.4991	0.8285	0.8842	0.8506	0.8171	0.7668	0.7887	0.7535	0.9913
6	0.3684	0.8132	0.8867	0.6089	0.8563	0.8700	0.7478	0.6517	0.9864
7	0.2262	0.6736	0.7412	0.4906	0.6480	0.5920	0.5024	0.4922	0.9475
8	0	0.8135	0.8463	0.5030	0.3187	0.2061	−0.0050	0.1176	0.4936
9	−0.1161	0.7007	0.7440	0.3782	0.2365	0.2762	−0.3636	0.2134	0.4889
10	−0.1437	0.7414	0.8018	0.5184	0.1290	0.1398	−0.3598	0.1438	0.3751
11	0.8128	0.9837	0.9930	0.5722	0.6484	0.5928	0.7290	0.8623	0.9568
12	0.6520	0.9427	0.9691	0.7984	0.8524	0.8447	0.7713	0.7455	0.9924
13	0.4673	0.9841	0.9927	0.6005	0.8604	0.8512	0.5510	0.5248	0.9969
14	0.3310	0.8087	0.8589	0.2983	0.4885	0.4576	0.6007	0.5809	0.9839
15	0.2497	0.7503	0.8010	0.2684	0.4523	0.4137	0.5520	0.5475	0.9788
16	0.0562	0.5789	0.6656	0.2530	0.4262	0.3673	0.4862	0.4847	0.9533
17	0.0545	0.3371	0.3805	0.7626	0.1513	0.1393	0.2283	0.2760	0.9458
18	0.0501	0.6365	0.6794	0.3420	0.1429	0.1199	0.2204	0.2123	0.9812
19	0.0748	0.7433	0.7697	0.3351	0.1335	0.1077	0.0448	0.1010	0.9619
20	0.0564	0.1303	0.1454	0.6233	0.8541	0.8624	0.7665	0.7184	0.9907
21	0.0347	0.4062	0.4780	0.2918	0.4205	0.3829	0.4013	0.3398	0.9842
22	0.7490	0.9949	0.9983	0.8031	0.7300	0.6910	0.7541	0.7622	0.9888
23	0.6646	0.9320	0.9790	0.7406	0.8912	0.6890	0.6611	0.6156	0.8432
24	0.3912	0.8774	0.9476	0.5641	0.7939	0.5488	0.3408	0.3734	0.6794
25	0.5507	0.9242	0.9721	0.6990	0.8180	0.6646	0.4857	0.5386	0.8112
26	0.1486	0.8169	0.8977	0.4876	0.5646	0.4417	0.0699	0.1943	0.4845
27	0.2248	0.8789	0.9367	0.4761	0.7840	0.7124	0.3996	0.3245	0.9845
28	0.4680	0.9417	0.9682	0.7181	0.7457	0.6886	0.5866	0.5501	0.9877
29	0.0427	0.2885	0.3822	0.5164	0.7657	0.7361	0.4493	0.3477	0.9771
30	0.1765	0.8431	0.9205	0.5016	0.7344	0.6617	0.3726	0.2850	0.9795
31	0.1358	0.7641	0.8725	0.4877	0.7372	0.6597	0.3945	0.2903	0.9731
32	0.0063	0.0524	0.0629	0.3943	0.5167	0.4740	0.4892	0.4156	0.9878
33	0.1267	0.7062	0.8105	0.5701	0.7390	0.6966	0.5089	0.4324	0.9766
34	−0.5920	0.7797	0.8022	−0.9766	0.9347	0.1797	−0.9156	−0.0611	0.2549

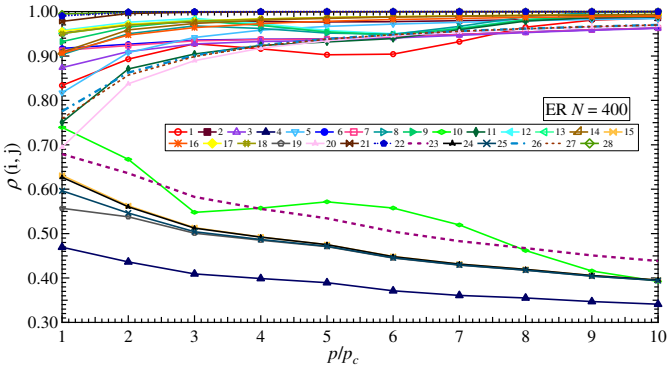


Fig. B.1. Pearson correlation coefficient between any two centrality metrics as a function of the link density p, in ER networks ($N = 400$). The number in the annotation is the correlation index.

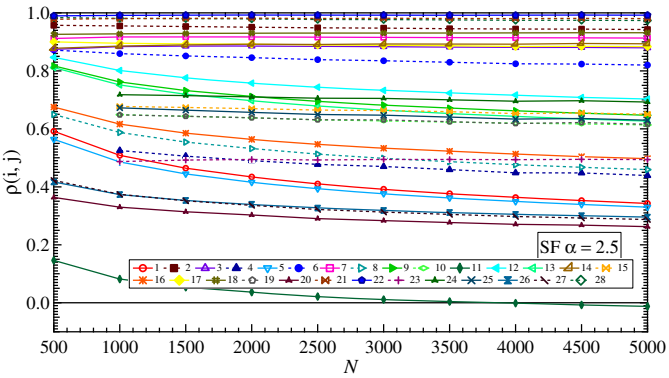


Fig. B.2. Pearson correlation coefficient between any two centrality metrics as a function of the size N of networks, in scale-free networks ($\alpha = 2.5$). The number in the annotation is the correlation index.

Proof. The generating function for the probability distribution of node degree is defined as:

$$\varphi_D(z) = \sum_{k=0}^{N-1} z^k \text{Prob}[D = k],$$

and the generating function of the degree of the node that we arrive at by following a randomly chosen link is:

$$\frac{\sum_k k\text{Prob}[D = k]z^k}{\sum_k k\text{Prob}[D = k]} = z\frac{\varphi'_D(z)}{E[D]}, \qquad \text{(C.5)}$$

where $E[.]$ is the expectation. If we start at a randomly chosen node, the generating function of the degree of a nearest neighbor of this node follows equation (C.5). The 1st-order degree mass $D^{(1)}$ of a node equals the degree sum of the node and its neighbors. The generating function has the "powers" property [45], that the distribution of the 1st-order degree mass of a node obtained from one nearest neighbor is generated by:

$$\varphi_D(z)^* = z^2 \frac{\varphi'_D(z)}{E[D]},$$

then, the distribution of the total of the 1st-order degree mass over k independent realizations (k nearest neighbors) of the node is generated by kth power of $\varphi_D(z)^*$ as:

$$\begin{aligned}\varphi_{D^{(1)}}(z) &= \varphi_D(\varphi_D(z)^*) \\ &= \sum_k \text{Prob}[D=k]\left(z^2\frac{\varphi'_D(z)}{E[D]}\right)^k. \end{aligned} \tag{C.6}$$

For ER networks, $E[D] = (N-1)p$ is the average degree in an ER network $G_p(N)$, and $\varphi_D(z) = (1-p+pz)^{N-1}$, thus,

$$\varphi_{D^{(1)}}(z) = \left((1-p) + z^2 p\,(1-p+pz)^{N-2}\right)^{N-1}, \tag{C.7}$$

In addition, the generating function has the "Moments" property [45], that $E[(D^{(1)})^n] = [(z\frac{d}{dz})^n \varphi_{D^{(1)}}(z)]_{z=1}$. Together with $\text{Var}[D^{(1)}] = E[(D^{(1)})^2] - E[D^{(1)}]^2$, we arrive at the (C.1) and (C.2), when $N \to \infty$.

Similarly, the distribution of the 2nd-order degree mass is generated by $\varphi_D(\varphi_{D^{(1)}}(\varphi_{D^{(1)}}(z)))$. Hence, we obtain the generating function of the 2nd-order degree mass as:

$$\begin{aligned}\varphi_{D^{(2)}}(z) = \Big(&1-p+pz^2\,(1-p+pz)^{N-2} \\ &\times \left(1-p+pz^2\,(1-p+pz)^{N-2}\right)^{N-2}\Big)^{N-1},\end{aligned}$$

Thus, we can obtain (C.3) and (C.4).

C.1 Proof of Lemma 1

Proof. The eigenvalue equation $Ax = \lambda x$ leads to $\lambda_1^k x_1 = A^k x_1$, from which we obtain

$$u^T x_1 \sum_{j=1}^{m} \lambda_1^j = u^T \left(\sum_{j=1}^{m} A^j\right) x_1,$$

where $u^T x_1 = NE[X_1]$ and $u^T \sum_{j=1}^{m+1} A^j = \left(d^{(m)}\right)^T$. Hence, the relation between the principal eigenvector and the mth-order degree mass vector can be expressed as: $E[X_1] N \sum_{j=1}^{m+1} \lambda_1^j = \left(d^{(m)}\right)^T x_1$, leading to:

$$E[D^{(m)} X_1] = E[X_1] \sum_{j=1}^{m+1} \lambda_1^j. \tag{C.8}$$

The Pearson correlation coefficient follows as:

$$\begin{aligned}\rho(D^{(m)}, X_1) &= \frac{E[D^{(m)}X_1] - E[D^{(m)}]E[X_1]}{\sqrt{\text{Var}[D^{(m)}]}\sqrt{\text{Var}[X_1]}} \\ &= \frac{\left(\sum_{j=1}^{m+1}\lambda_1^j - E[D^{(m)}]\right)E[X_1]}{\sqrt{\text{Var}[D^{(m)}]}\sqrt{\text{Var}[X_1]}}.\end{aligned} \tag{C.9}$$

The ratio of the two Pearson correlation coefficients is:

$$\frac{\rho(D^{(1)}, X_1)}{\rho(D, X_1)} = \frac{\sqrt{\text{Var}[D]}}{\sqrt{\text{Var}[D^{(1)}]}}\left(1 + \frac{(\lambda_1^2 - E[D^2])}{(\lambda_1 - E[D])}\right). \tag{C.10}$$

For large ER graphs, $E[D] = (N-1)p \to Np$, $E[D^2] = (N-1)^2p^2 - (N-1)p^2 + (N-1)p \to N^2p^2 - Np^2 + Np$ and $\text{Var}[D] = (N-1)p(1-p) \to Np(1-p)$. From (C.2), we obtain

$$\frac{\sqrt{\text{Var}[D]}}{\sqrt{\text{Var}[D^{(1)}]}} = \sqrt{\frac{(1-p)}{(E[D]+2)^2 - 2}} > \frac{1}{E[D]+2}. \tag{C.11}$$

When $N \to \infty$ and $Np = \varsigma$ (ς is a constant and independent of N), the spectral radius $\lambda_1 \to \varsigma$, in sparse random graphs [46,47]. With (C.10) and (C.11), $\rho(D^{(1)}, X_1) \geq \rho(D, X_1)$ is proved.

The ratio of the two Pearson correlation coefficients is:

$$\frac{\rho(D^{(2)}, X_1)}{\rho(D^{(1)}, X_1)} = \frac{\left(\lambda_1 + \lambda_1^2 + \lambda_1^3 - E[D^{(2)}]\right)\sqrt{\text{Var}[D^{(1)}]}}{(\lambda_1 + \lambda_1^2 - E[D^2] - E[D])\sqrt{\text{Var}[D^{(2)}]}},$$

with (C.3) and $\lambda_1 \to Np$, when $N \to \infty$ we arrive at

$$\frac{\left(\lambda_1 + \lambda_1^2 + \lambda_1^3 - E[D^{(2)}]\right)}{(\lambda_1 + \lambda_1^2 - E[D^2] - E[D])} = 2E[D] + 1.$$

With (C.2) and (C.4), for large sparse random networks, $\rho(D^{(2)}, X_1) \geq \rho(D^{(1)}, X_1)$ is proved.