Spectral centrality measures in complex networks

Nicola Perra[1,2] and Santo Fortunato[3]

[1]*Dipartimento di Fisica, Università di Cagliari, Cagliari, Italy*
[2]*Linkalab, Center for the Study of Complex Networks, Cagliari 09129, Sardegna, Italy*
[3]*Complex Networks Lagrange Laboratory (CNLL), Institute for Scientific Interchange (ISI), Viale S. Severo 65, 10133, Torino, Italy*

Complex networks are characterized by heterogeneous distributions of the degree of nodes, which produce a large diversification of the roles of the nodes within the network. Several centrality measures have been introduced to rank nodes based on their topological importance within a graph. Here we review and compare centrality measures based on spectral properties of graph matrices. We shall focus on PageRank (PR), eigenvector centrality (EV), and the hub and authority scores of the HITS algorithm. We derive simple relations between the measures and the (in)degree of the nodes, in some limits. We also compare the rankings obtained with different centrality measures.

PACS number(s): 89.75.Hc

I. INTRODUCTION

Complex systems can be represented as networks, where the main units of the system become nodes and interacting units are connected by edges. The last years have witnessed an intense research activity on networks by the scientific community, after the discovery that many systems in nature, society, and technology turn into graphs with peculiar properties [1,2]. In particular, many networks are characterized by a heterogeneous distribution of the number of neighbors of a node, or degree, where nodes with low degree coexist with nodes with large degree (hubs). Such heterogeneity is responsible for a number of remarkable features of real networks, such as resilience to random failures and attacks [3], and the absence of a threshold for percolation [4] and epidemic spreading [5]. The presence of nodes with different degrees means that there is a broad diversification of their roles within the graph. Centrality measures are designed to rank graph nodes based on their topological importance [6,7]. Among the most popular centrality measures, we mention degree itself, but also measures depending on shortest paths between nodes pairs, like node betweenness and closeness. There are as well centrality measures that depend on spectral properties of graph matrices. These measures are important because they are usually associated with simple dynamic processes taking place on graphs, like diffusion. In particular, the PageRank algorithm, proposed by the Google founders Brin and Page [8], managed to turn Google into the leading interface between users and the World Wide Web. In recent work spectral properties of graph matrices have also been used to characterize the participation of nodes in network subgraphs (subgraph centrality) [9,10] and to estimate the bipartitivity of graphs [11].

However, spectral centrality measures have not been much investigated in the physics literature. We shall introduce and review four centrality measures: PageRank, eigenvector centrality (EV)[12], and the hub and authority scores introduced by Kleinberg for his HITS algorithm [13]. These measures are usually adopted on directed graphs; we shall as well discuss extensions to the undirected case, where applicable.

In Sec. II we present the measures and describe them in some detail. Analytical and numerical results on particular graphs will be shown in Sec. III, whereas in Sec. IV we shall compare the rankings of nodes of real graphs for different centrality measures. Conclusions will be reported in Sec. V.

II. CENTRALITY MEASURES

The basic matrix of a graph is the *adjacency matrix A*, where the element A_{ij} equals 1 if nodes i and j are connected by a link, 0 if they are not. If the network is directed, the adjacency matrix is not symmetric. In this case, it is necessary to distinguish between two types of links adjacent to a node, i.e., links pointing to the node (incoming) and links pointing outside (outgoing). Therefore, there are two types of degree: *indegree*, i.e., the number of incoming links; *outdegree*, i.e., the number of outgoing links. Likewise, one distinguishes between the *in-neighbors* of a node, i.e., the nodes pointing at the node, and the *out-neighbors*, i.e., the nodes pointed at by the node. The directedness of the links has a number of important implications, involving both some basic structural concepts, like connectivity, and processes taking place on the network. For instance, a random walk is a stationary process on any undirected graph, but it is not in general on a directed graph, due to the possible presence of *dangling ends*, i.e., nodes with zero outdegree, that act as sinks for the process. On the other hand, diffusion leads to a natural definition of centrality, based on the frequency with which a walker stops by a node during the process. In order to make a diffusive process stationary on a directed graph, one needs to give the walker the opportunity to leave from a dangling end. PageRank offers a simple solution, which we describe below.

A. PageRank

PageRank (PR) is the prestige measure used by Google to rank Web pages. It is supposed to simulate the behavior of a user browsing the Web. Most of the time, the user visits pages just by surfing, i.e., by clicking on hyperlinks of the page he is on; otherwise, the user will jump to another page

036107-1

by typing its URL on the browser, or going to a bookmark, etc. On a graph, this process can be modeled by a simple combination of a random walk with occasional jumps toward randomly selected nodes. This can be described by the simple set of implicit relations

$$p(i) = \frac{q}{n} + (1-q) \sum_{j:j \to i} \frac{p(j)}{k_{\text{out}}(j)}, \quad i = 1, 2, \ldots, n. \quad (1)$$

Here, n is the number of nodes of the graph, $p(i)$ is the PR value of node i, $k_{\text{out}}(j)$ the outdegree of node j, and the sum runs over the nodes pointing toward i. The *damping factor q* is a probability that weighs the mixture between random walk and random jump. On practical applications it is usually set to small values (typically 0.15). For any $q > 0$ the process reaches stationarity, as a walker has a finite (no matter how small) probability to escape from a dangling end, whenever it lands there. When $q = 0$, the process may not be stationary and PR is ill defined. When $q = 1$, instead, the jumping process dominates and all nodes have the same PR value $1/n$. PR goes beyond indegree: in order to have a large PR value for a node it is important to have many neighbors pointing at a node, i.e., large indegree, but it is also important that the neighbors have large PR values. So, if two nodes have equal indegree, the node with more "important" neighbors will have larger PR.

Solving the set of equations (1) is equivalent to solving the eigenvalue problem for the transition matrix $\mathcal{M}$, whose element $\mathcal{M}_{ij}$ is given by the following expression:

$$\mathcal{M}_{ij} = \frac{q}{n} + (1-q)\frac{1}{k_{\text{out}}(j)}A_{ji}. \quad (2)$$

PR is just the principal eigenvector of $\mathcal{M}$, and is usually determined with the power method, i.e., by repeatedly multiplying the matrix $\mathcal{M}$ by an arbitrary vector until all the entries of the resulting vector are stable. This is also the procedure we adopted to compute the eigenvectors corresponding to all centrality measures we studied.

The literature on PR is very large, because of its huge impact on Web search. In one of the first theoretical studies [14], the dependence of PR on the damping factor was investigated. In general, the attention has been mostly focused on the graph of the World Wide Web, where Web pages are nodes and the hyperlinks their connections. Comparatively little has been done to study the measure on more general classes of networks. A recent mean field study [15] has shown that the average PR value of nodes with the same indegree is a linear function of indegree in the absence of degree-degree correlations. In another study, some analytical results were found on PR distributions on special classes of graphs [16]. In Sec. III A we shall briefly summarize the results of [16] and build upon them.

B. Eigenvector centrality

The EV is also based on the principle that the importance of a node depends on the importance of its neighbors. In this case the relationship is more straightforward than for PR: the prestige x_i of node i is just proportional to the sum of the prestiges of the neighboring nodes pointing to it,

$$\lambda x_i = \sum_{j:j \to i} x_j = \sum_j A_{ji} x_j = (\mathbf{A}^t \mathbf{x})_i. \quad (3)$$

From Eq. (3) we see that x_i is just the i component of the eigenvector of the transpose of the adjacency matrix with eigenvalue λ. We notice that the trivial eigenvector with all components equal to zero is always a solution of Eq. (3). The true EVC is then associated with the existence of nontrivial solutions of the eigenvalue problem of Eq. (3). From Eq. (3) we see that nodes with indegree zero also have zero centrality: in general, nodes pointed at by nodes with zero centrality also have zero centrality, and this effect will propagate to other nodes, so that in many cases the EVC would not give any information about a big number of nodes. To avoid this, it is useful to make the following modification: to each node we assign a prestige ϵ, which is independent of its relationships with the other nodes. Equation (3) is then modified as follows:

$$x_i = \alpha(\mathbf{A}^t \mathbf{x})_i + \epsilon. \quad (4)$$

The role of the parameter ϵ recalls that of the damping factor q in PR. The parameter α weighs the relative importance of the contribution of the peers versus that of the node itself. The new measure is called α-*centrality* (αEV) [12] and is the one we shall investigate in this paper. We remark that, in contrast to PR, here the solutions do not have a natural interpretation in terms of probability, so the sum of the α-centralities need not be 1. However, we shall normalize the final values by dividing them by their sum, so as to make them add up to 1, for practical purposes.

C. HITS scores

Google's PR was not the first prestige measure for Web pages based on the Web's graph representation. Shortly before the seminal paper by Brin and Page, Kleinberg [13] had proposed another solution to the problem of ranking Web sites based on their importance for the users. This solution was the HITS *algorithm*, which distinguishes two types of Web pages: *hubs* and *authorities*. Let us suppose that a user submits a query through a search engine. If a page is very relevant for this query, one can reasonably expect that it will be pointed at by many other pages. However, the simple indegree would not allow one to discriminate the relevant pages from other pages with similar (large) indegree. An important difference is that pages pointing to a relevant page are likely to point as well to other relevant pages, so to create a sort of bipartite structure where relevant pages (authorities) are cited by special pages or indices (hubs). Such bipartite structures allow the relevant pages for the user query to be identified. Therefore one assigns two scores to a page i of the Web: the *hub score* x_i and the *authority score* y_i. Pages with high authority scores are pointed at by pages with high hub scores. In turn, a good hub points at (very) authoritative pages. This mutually reinforcing mechanism is described by the coupled relations

$$\lambda y_i = \sum_{j:j \to i} x_j = \sum_j A_{ji} x_j = (\mathbf{A}^t \mathbf{x})_i, \quad (5)$$

$$\mu x_i = \sum_{j:i \to j} y_j = \sum_j A_{ij} y_j = (\mathbf{A}\mathbf{y})_i, \qquad (6)$$

which can be rewritten in the form of simple eigenvalue equations for both **x** and **y** by substitution

$$\lambda \mu x_i = (\mathbf{A}\mathbf{A}^t \mathbf{x})_i, \qquad (7)$$

$$\lambda \mu y_i = (\mathbf{A}^t \mathbf{A}\mathbf{y})_i. \qquad (8)$$

From Eqs. (7) and (8) we see that the hub and authority scores are just eigenvectors of the matrices AA^t and A^tA. We stress that both AA^t and A^tA are symmetric, whether A is symmetric or not. The scores **x** and **y** correspond to the principal eigenvectors of these matrices.

III. RESULTS

A. PageRank

In [16] the two main limits of the PR measure, corresponding to $q \to 0$ and $q \to 1$, were investigated. Analytical results can be derived for special graphs, such as graphs grown with popular mechanisms, like preferential attachment [17]. For our proofs we shall focus on the model by Dorogovtsev, Mendes, and Samukhin (DMS) [18], which generates graphs with power-law degree distributions with any exponent larger than 2. In this model, at each time step a new node is added to the graph and m links are set from the new node to preexisting ones. The probability that a new node i gets attached to a node j (with indegree k_j) is

$$\Pi(k_j, a) = \frac{a + k_j}{\sum_{l=1}^{i-1}(a + k_l)}, \qquad (9)$$

where a is a positive constant. When $a=m$ one recovers the recipe of the original preferential attachment formulation of Barabási and Albert [17]. In general, the exponent of the indegree distribution $\gamma = 2 + a/m$. For simplicity, we shall study the special case in which $m=1$, i.e., each node has outdegree 1 and the network is a tree. The results are, however, independent of m.

1. The limit $q \to 0$

We assume that q is very small. To the first order in q, and remembering that each node has outdegree 1 by construction, Eq. (1) takes the following form:

$$p(i) \sim \frac{q}{n} + \sum_{j:j \to i} p(j), \quad i = 1, 2, \ldots, n, \qquad (10)$$

which looks particularly simple, though not generally solvable. From Eq. (10) we see that the PR of a node equals a constant plus the PR of its in-neighbors. This recipe enables us to calculate PR recursively on simple trees, as shown in Fig. 1, where we focus on a subgraph of a tree. Node A is the root of the subgraph as every walk starting on any of the nodes will reach A at some stage. We call any node with this property a *predecessor* of A. The PR value of any node of the graph is determined only by its predecessors. In the case illustrated, the calculation is particularly simple: we start
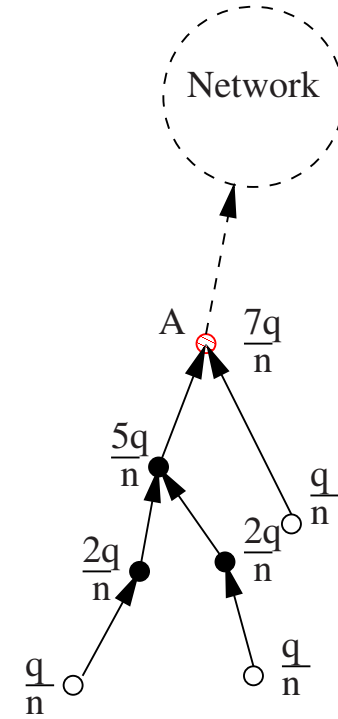

FIG. 1. (Color online) Subgraph of a tree. The PR values of all nodes shown can be simply calculated.

from the leaves of the subgraph (empty circles) whose PR is just q/n because they have no incoming links, and move towards A. For each node, we apply the relation (10). The final values are reported next to the nodes. From this example we can deduce a number of general properties: (1) all PR values are multiples of the elementary unit q/n; (2) PR increases if one moves from a node to another by following a link; (3) the PR of each node i, in units of q/n, equals the number of its predecessors. Since PR takes only discrete values, in the following we shall measure it in units of q/n. We thus indicate the distribution as $P_{PR}(l)$, with $l=1,2,\ldots,n$.

In a dynamic process like network growth, it is crucial to see what happens to the PR values and distribution when a new node comes into the picture. This is shown in Fig. 2, where a new node N is added to the network of Fig. 1. We see that only the nodes encountered along the path from N to A, including A, are affected, while the others retain their PR values. In particular, the presence of the node N determines an increase by q/n in the PR values of the affected nodes.

Now we are ready to build a master equation for the PR distribution $P_{PR}(l)$ on a DMS graph. At time n, the graph has n nodes and $n-1$ links (the root does not generate links); the PR distribution is $P_{PR}^n(l)$. If we add node $n+1$ we get a new distribution $P_{PR}^{n+1}(l)$. As we have seen above, the new node will contribute an additional q/n to the PR of the nodes in the path from $n+1$ to the root of the graph. We need to compute the balance between the nodes passing from PR $l-1$ to l and those passing from l to $l+1$. The probability Π_i^n that the PR of node i, initially equal to l, will be changed by the new node equals the probability that the link set by the new node gets attached to one of the predecessors of i (including i), and is given by
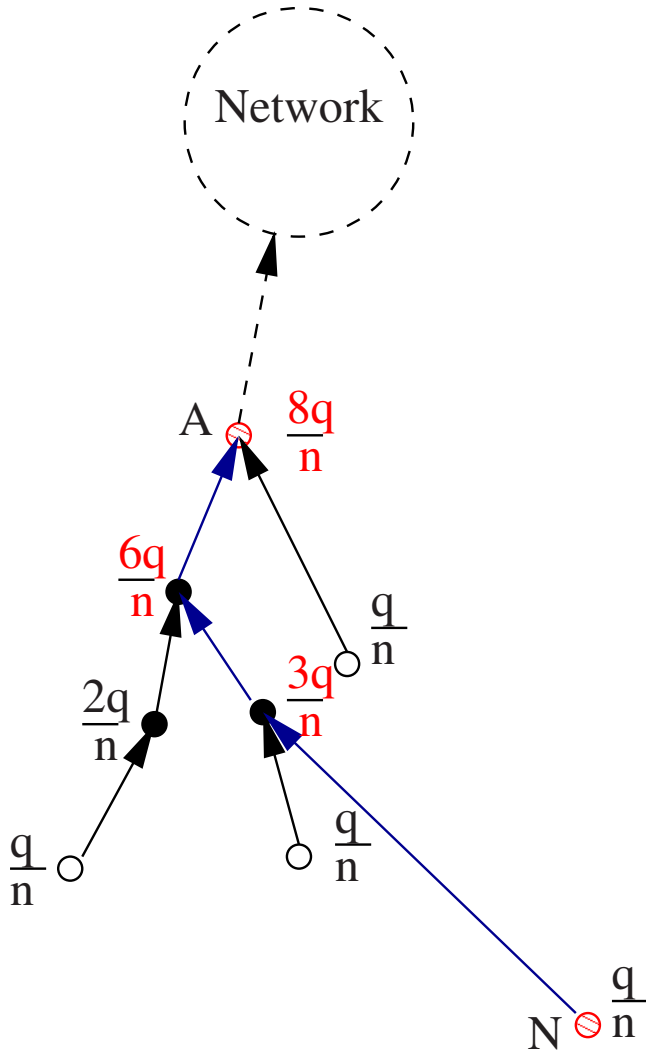
FIG. 2. (Color online) If a new node N gets attached to any node of the subgraph, it adds an equal contribution q/n to the PR of all nodes in the path from N to the root.

$$\Pi_i^n = \sum_{j\Rightarrow i} \frac{a + k_j}{\sum\limits_{t=1}^{n}(a + k_t)} = \sum_{j\Rightarrow i} \frac{a + k_j}{(a+1)n - 1}, \quad (11)$$

where $j \Rightarrow i$ means that j is a predecessor of i. None of the predecessors of i, other than i, can reach PR $l+1$ because of the new node, as their initial values are necessarily smaller than l. The number of predecessors of i (including i) is l, and the total number of adjacent links to the predecessors is $l-1$ (one for each predecessor, except i). So

$$\Pi_i^n = \sum_{j\Rightarrow i} \frac{a + k_j}{(a+1)n - 1} = \frac{(a+1)l - 1}{(a+1)n - 1}. \quad (12)$$

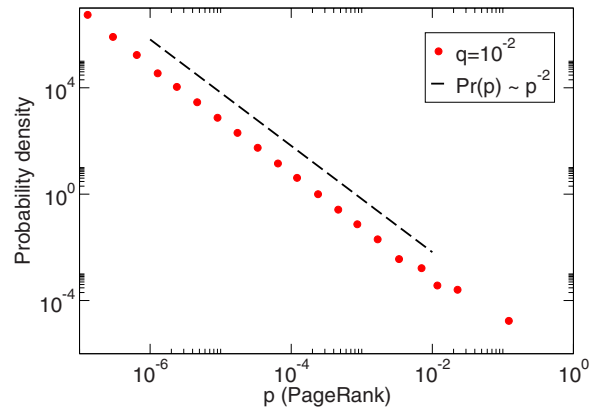The number of nodes with PR l that are affected by the presence of the new node and its link is then

FIG. 3. (Color online) PR distribution for small q on a DMS graph with 10^6 nodes, $m=1$, and $a=1$. In this case the indegree distribution is a power law with exponent $\gamma = 3$.

$$\Pi^n(l) = nP_{PR}^n(l)\Pi_i^n = \frac{(a+1)l - 1}{(a+1) - 1/n}P_{PR}^n(l). \quad (13)$$

and the master equation reads

$$(n+1)P_{PR}^{n+1}(l) - nP_{PR}^n(l) = \Pi^n(l-1) - \Pi^n(l). \quad (14)$$

Equation (14) holds for $l>1$. For $l=1$ a modification is necessary, as there cannot be nodes with zero PR, so the term $\Pi^n(0)$ is not defined. However, since the new node has no incoming links, the number of nodes with PR 1 increases by 1 because of the new node, so we can write

$$(n+1)P_{PR}^{n+1}(1) - nP_{PR}^n(1) = 1 - \Pi^n(1). \quad (15)$$

The stationarity condition of Eqs. (14) and (15) in the limit of large n leads to the relations

$$P_{PR}(l) = \begin{cases} \dfrac{(a+1)l - a - 2}{(a+1)l + a}P_{PR}(l-1) & \text{if } l > 1, \\[2ex] \dfrac{a+1}{2a+1} & \text{if } l = 1, \end{cases} \quad (16)$$

which has the solution

$$P_{PR}(l) = \frac{a(a+1)}{[(a+1)l + a][(a+1)l - 1]} \sim \frac{1}{l^2}, \quad \text{for } l \gg 1. \quad (17)$$

We see that the PR distribution in the limit $q\to 0$ on a DMS tree is a power law with exponent 2, for any value of the parameter a, including the limit case $a\to\infty$, when the indegree distribution becomes exponential. This result is confirmed by numerical simulations (Fig. 3), which also show that the hypothesis of the tree is not necessary, as long as each node has the same outdegree m.

In [16] the same result was found for other models of network growth, like Barabási-Albert preferential attachment [17] and the copying model [19]. It is possible that this property holds for general graphs where the flows converge toward a central root (sink). Indeed, our finding agrees with the more general result on the size distribution of supercritical trees [20]. Moreover, numerical studies have shown that the
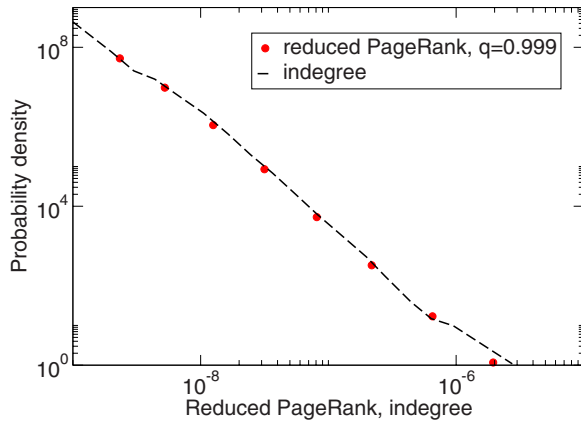
FIG. 4. (Color online) Reduced PR distribution for $q \sim 1$ on a DMS graph with 10^6 nodes, $m=1$, and $a=1$. The curve matches the indegree distribution.

same behavior holds for the graph of the Internet, when one considers the distribution of the size of the basin connected to a given point [21]. Indeed, our calculation follows the same procedure usually adopted for the calculation of the area of basins in river networks.

2. The limit $q \to 1$

The case $q=1$ is well defined, but trivial, as all nodes end up having the same PR value $1/n$. We ask how this limit is reached. If $q \sim 1$, the contribution to PR given by the in-neighbors of a node is very small compared to the constant term, which is close to $1/n$. In order to study the behavior of this term, we define the *reduced* PageRank $p_r(i)$ of a node i as

$$p_r(i) = p(i) - \frac{q}{n}, \quad i = 1, 2, \dots, n. \tag{18}$$

We assume that all nodes have the same outdegree m. In this case, to leading order in the infinitesimal $1-q$ Eq. (1) can be rewritten as

$$p_r(i) = \frac{q(1-q)}{mn} k_{\text{in}}(i), \quad i = 1, 2, \dots, n. \tag{19}$$

where $k_{\text{in}}(i)$ is the indegree of i. We conclude that on any graph the reduced PR of a node in the limit $q \to 1$ is proportional to the indegree of the node, if all nodes have the same outdegree. This result has been derived independently in [22]. As a consequence of Eq. (19), the distribution of the reduced PR for $q \to 1$ has the same trend as that of the indegree, which can be easily verified numerically (Fig. 4).

3. Extension to undirected graphs

PR can be easily extended to undirected graphs as well. The corresponding equation reads

$$p(i) = \frac{q}{n} + (1-q) \sum_{j: j \leftrightarrow i} \frac{p(j)}{k_j}, \quad i = 1, 2, \dots, n, \tag{20}$$

where now k_j is the degree of node j. For the purposes of a random walk, undirected links can be crossed in both direc-
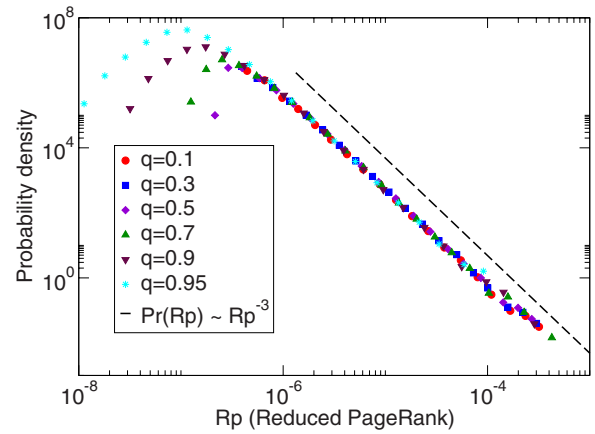

FIG. 5. (Color online) Reduced PR on undirected graphs. Variability of reduced PR distribution with q on a DMS graph with 10^6 nodes, $m=3$, and $a=3$. The degree distribution has a power-law tail with exponent $\gamma=3$.

tions, so a pure random walk now always reaches stationarity due to the absence of dangling ends. In fact, the stationary probability of a random walk on a node of any undirected graph is simply proportional to the degree of the node [23]. However, in Eq. (20) we have still the contribution of random jumping, and it turns out that the mixed process is still hard to solve. We are not aware of a general solution in this case. In the limit $q \to 0$ PR is now well behaved, and its distribution coincides with the degree distribution of the graph. In Fig. 5 we show the distributions of reduced PR for different values of q on a DMS graph with a power-law degree distribution and exponent $\gamma=3$. The reduced PR expresses the contribution to PR given by the random walk. We see that the curves follow the decay of the degree distribution for any value of q. We have computed the reduced PR distribution on many other graphs, and in all cases we found that they follow the same trend as the degree distribution. For example, in Fig. 6 we show the comparison between reduced PR and degree for a sample of the Web link graph.
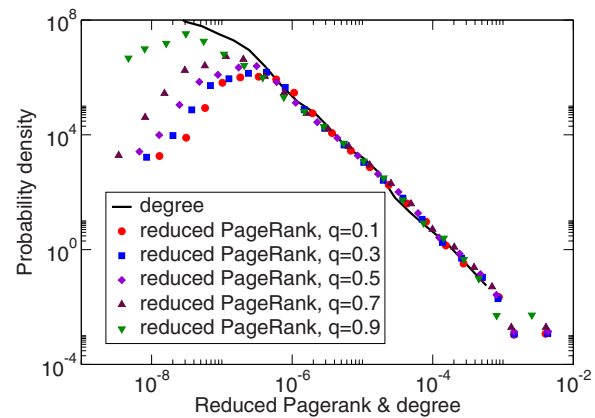

FIG. 6. (Color online) Reduced PR on undirected graphs. Variability of reduced PR distribution with q on the domain *.gov* of the World Wide Web. The degree distribution has a tail which follows fairly well a power law with exponent 2.1. To better show the agreement we have shifted the curves such that the tails overlap.

Here the nodes are Web pages of the domain .*gov* and two pages are connected if there is a hyperlink from one to the other. There are 794 184 nodes and 6 460 903 links. The graph is directed but PR was calculated by neglecting the directedness of the links. As we can see, the decay of the distributions of reduced PR resembles that of the degree distribution. The graph at hand is not simple like the DMS networks, as it presents a large number of loops and community structure. Therefore the result is likely to be general. We can show this with a simple argument. The general equation for reduced PR on undirected graphs is

$$p_r(i) = \frac{(1-q)q}{n} \sum_{j:j\leftrightarrow i} \frac{1}{k_j} + (1-q) \sum_{j:j\leftrightarrow i} \frac{p_r(j)}{k_j}, \qquad (21)$$

which we can solve formally by successive iteration, obtaining the general form

$$p_r(i) = \frac{q}{n} \sum_s (1-q)^s \sum_{i_1} \frac{1}{k_{i_1}} \sum_{i_2} \frac{1}{k_{i_2}} \cdots \sum_{i_s} \frac{1}{k_{i_s}}$$

$$= \frac{q}{n} \sum_s (1-q)^s \prod_{i_1 \leftrightarrow i_2 \cdots \leftrightarrow i_s} \frac{1}{k_{i_s}}, \qquad (22)$$

where i_s indicates the neighbors of the s shell of the node i; so i_1 indicates the nearest neighbors of i, i_2 the next-to-nearest neighbors, and so on. The last sum in the first line of Eq. (22) is, for a given node i_{s-1}, a sum over its neighbors i_s. This sum, which we call T_{i_s}, contains k_{i_s} terms, k_{i_s} being the degree of node i_s. The sum T_{i_s} can be approximated as the product $k_{i_s} \langle 1/k \rangle_{\mathrm{NN}}$, where $\langle 1/k \rangle_{\mathrm{NN}}$ is the expected value of the average of $1/k$ over the neighbors of a node of the network. In general, $T_{i_s} = k_{i_s} \langle 1/k \rangle_{\mathrm{NN}} + \eta_{i_s}$, where η_{i_s} is a random variable with mean zero. In this way, it is easy to see from Eq. (22) that, for any value of s, the product of sums reduces to $k_i \langle 1/k \rangle_{\mathrm{NN}}$ plus the sum of many random variables like η_{i_s}. Due to the central limit theorem, the latter sum, if it includes a large number of terms, yields a very small value with large probability. We can then conclude that, for k_i sufficiently large, each term of the series in Eq. (22) is proportional to k_i with good approximation; therefore $p_r(i)$ is also proportional to k_i, for any value of the damping factor q. We have verified numerically that this assertion is true for many graphs and degree distributions, without finding exceptions.

B. Eigenvector centrality

1. Directed graphs

The defining Eq. (4) is formally analogous to Eq. (10). The only difference is that the eigenvalue α is not 1 as for PR. However, the results of Sec. III A 1 hold as well when the outdegree m is greater than 1 (as long as it is the same for all nodes), and in this case the sum of Eq. (10) would include a multiplicative factor $1/m$, which makes it identical to Eq. (4). We then deduce that all results found for PR in the limit $q \rightarrow 0$ hold for αEV. Here the results are more general, because we did not need to make any approximation to get to Eq. (4) as we instead needed to derive Eq. (10). In particular, it is not necessary that ϵ be very small and the nodes need
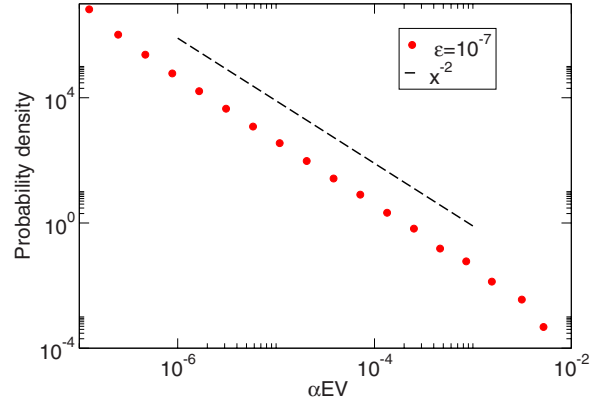

FIG. 7. (Color online) Distribution of αEV on a directed DMS graph with 10^6 nodes, $m=1$, and $a=1$. The dashed line indicates the predicted slope.

not have the same outdegree, although this is the case for the graphs we considered. We conclude that the distribution of αEV on DMS graphs has a power-law tail with exponent 2 (Fig. 7). The same holds for graphs built using preferential attachment and the copying model, just as it happens for PR in the limit $q \rightarrow 0$.

2. Extension to undirected graphs

On undirected graphs, Eq. (4) becomes

$$x_i = \alpha(\mathbf{Ax})_i + \epsilon, \qquad (23)$$

since $A^t = A$. So, the αEV of a node is proportional to the sum of the αEV of its neighbors, modulo an additive constant ϵ. As we have done for PR, we define the *reduced α-centrality* as

$$x_i^r = x_i - \epsilon. \qquad (24)$$

So we can rewrite Eq. (24) as

$$x_i^r = \alpha(\mathbf{Ax}^r)_i + k_i \alpha \epsilon, \qquad (25)$$

where k_i is again the degree of node i. We can apply a similar argument as in Sec. III A 3. The sum over the k_i neighbors of i can be approximated as $k_i \langle x^r \rangle$, where $\langle x^r \rangle$ is the average of the reduced αEV over the whole graph. The approximation is the more valid the larger the number k_i of summands. In this way, from Eq. (25) we see that the reduced αEV of a node is proportional to its degree, if the latter is large enough. This result is independent of the specific graph we consider, and we have verified it numerically for many types of networks. In Fig. 8 we show the distribution of reduced αEV for different choices of the parameter ϵ/α for the sample of the Web graph we analyzed in Fig. 6. The curves closely follow the decay of the degree distribution.

C. HITS scores

The meaning of the eigenvalue equations (7) and (8) is quite simple. The hub score of a node is the sum of the hub scores of the in-neighbors of the out-neighbors of the node. The authority score of a node is the sum of the authority scores of the out-neighbors of the in-neighbors of the node
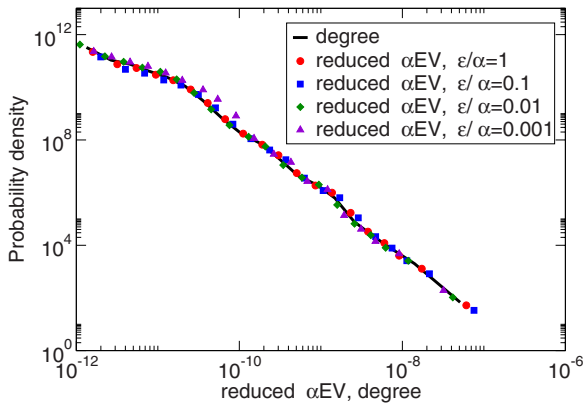
FIG. 8. (Color online) Reduced αEV on undirected graphs. Variability of reduced αEV distribution with ϵ/α on the domain *.gov* of the World Wide Web. The degree distribution has a tail which follows fairly well a power law with exponent 2.1. To better show the agreement we have shifted the curves such that the tails overlap.

FIG. 10. (Color online) Distribution of the authority scores versus indegree distribution. (Left) DMS graph with 10^5 nodes, $m=10$, and $a=1$. (Right) DMS graph with 10^5 nodes, $m=50$, and $a=1$.

(Fig. 9). Let us suppose that the nodes have the same outdegree m. The authority score of a node i is given by the sum of $mk_{in}(i)$ terms, where $k_{in}(i)$ is the indegree of i. In fact, node i has $k_{in}(i)$ in-neighbors, each of them having m out-neighbors. If $k_{in}(i)$ is large, the number of summands is very large, and can be approximated by the average value of the authority score over the whole graph, times $mk_{in}(i)$. This approximation is the more valid the larger m and $k_{in}(i)$. We conclude that on a directed graph with constant outdegree the distribution of the authority scores will have the same tail as the indegree distribution. This is clearly illustrated in Fig. 10. For the hub scores it is not possible to make predictions; the sum that delivers the hub score of a node cannot be approximated through other graph variables in most cases.

The extension of the HITS scores to the case of undirected graphs is not interesting. In this case $A^t=A$, so $A^tA=AA^t=A^2$ and the hub and authority scores are identical. Moreover, they coincide with EV, as the matrices A and A^2 have the same eigenvectors.

IV. RANKINGS

In the previous sections we have investigated the distributions of spectral centrality measures and their similarities. As
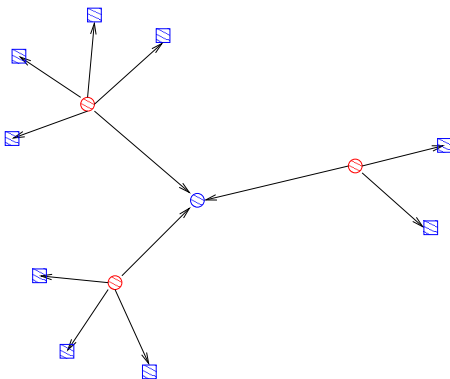

FIG. 9. (Color online) The authority score of the node in the center is proportional to the sum of the authority scores of the out-neighbors (blue squares) of the in-neighbors (red circles) of the node.
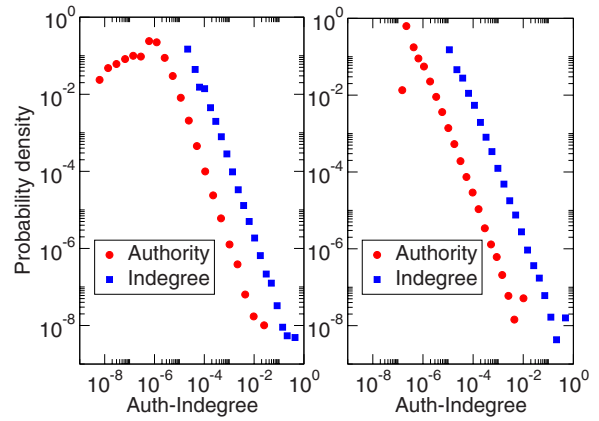
we mentioned in the Introduction, centrality measures are used to rank nodes. In this section we shall compare the rankings obtained with different centrality measures. In order to compare two rankings we adopt Kendall's τ [24], a widely used index in this type of analysis. Kendall's τ ranges from 1 (perfect correlation) to -1 (perfect anticorrelation). In Table I we show the cross comparisons between all centrality measures we discuss in this work, for a DMS directed graph. For completeness we have included the outdegree as well. As we can see, PR, αEV, and the authority scores are well correlated with indegree and with each other, whereas the other coefficients are small or negative; αEV has a strong correlation with outdegree as well.

DMS graphs have a fairly regular structure; we have seen that in this case the behavior of centrality measures is quite regular, and that there are simple relations between their distributions, which may be determined by simple relations be-

TABLE I. Kendall's τ for each pair of centrality measures computed for a DMS directed graph, with $n=10^6$, $m=3$, and $a=3$.

Measure	τ
PR-αEV	0.8192
PR-AUTH	0.5774
PR-HUBS	0.1213
PR-IN	0.6444
PR-OUT	-0.3012
αEV-AUTH	0.5788
αEV-IN	0.6487
αEV-HUBS	0.1220
αEV-OUT	0.5788
AUTH-IN	0.5458
AUTH-HUBS	0.1076
AUTH-OUT	-0.2611
HUBS-IN	0.1142
HUBS-OUT	-0.2126
IN-OUT	-0.2507

TABLE II. Kendall's τ for each pairs of centrality measures for the network of political blogs studied by Adamic and Glance.

Measures	τ
PR-αEV	0.09
PR-AUTH	0.14
PR-HUBS	0.04
PR-IN	0.14
PR-OUT	0.02
αEV-AUTH	0.12
αEV-IN	0.07
αEV-HUBS	0.08
αEV-OUT	0.01
AUTH-IN	0.12
AUTH-HUBS	0.07
AUTH-OUT	0.01
HUBS-IN	0.02
HUBS-OUT	0.07
IN-OUT	0.07

tween a measure and indegree at the level of the single node. Therefore, we cannot deduce general conclusions from Table I and we repeated the analysis for two real-world networks: a network of political blogs and the subset of the Web link graph corresponding to the URLs of the domain .*gov*, that we have studied in the previous sections.

The first network is a citation network consisting of 1490 blogs; 758 are Democratic and 732 Republican. It was first studied by Adamic and Glance [25], who focused on the community structure of the graph, which matches that determined by the two political areas. The correlations now are rather weak (Table II). The small coefficients indicate that the rankings differ considerably with the measure chosen. To have an idea, in Table III we show the top ten blogs in the rankings obtained with all centrality measures. We see that there are clear differences between the listings.

The results are basically the same for the Web graph. Table IV reports the Kendall's τ between the rankings. The values are of the same magnitude as for the network of the blogs. The top ten listings for the Web are shown in Table V and appear again considerably different from each other.

V. CONCLUSIONS

Centrality measures are very important to understand the properties of the nodes of complex networks and their topological roles. We have studied the most important centrality measures based on properties of graph matrices: PageRank, EV, and the hub and authority scores of HITS. All these measures deduce the importance of a node in a self-consistent way from the importance of its nearest neighbors and, in the case of the HITS scores, of its next-to-nearest neighbors. We have summarized some recent results on PageRank distributions on particular types of treelike graphs. On those graphs, the distribution of PageRank in the limit $q \to 0$ decays as a

TABLE III. Top ten of the network of political blogs according to PR, αEV, authorities, hubs, indegree and outdegree. D Democratic; R, Republican.

Rank	PR	αEV	Auth
1°	dailykos.com, D	atrios.blogspot.com, D	dailykos.com, D
2°	atrios.blogspot.com, D	dailykos.com, D	talkingpointsmemo.com, D
3°	instapundit.com, R	talkingpointsmemo.com, D	atrios.blogspot.com, D
4°	blogsforbush.com, R	washingtonmonthly.com, D	washingtonmonthly.com, D
5°	talkingpointsmemo.com, D	talkleft.com, D	talkleft.com, D
6°	michellemalkin.com, R	prospect.org/weblog, D	instapundit.com, R
7°	drudgereport.com, R	juancole.com, D	juancole.com, D
8°	washingtonmonthly.com, D	digbysblog.blogspot.com, D	yglesias.typepad.com/matthew, D
9°	powerlineblog.com, R	pandagon.net, D	pandagon.net, D
10°	andrewsullivan.com, R	yglesias.typepad.com/matthew, D	digbysblog.blogspot.com, D
Rank	Hubs	In	Out
1°	politicalstrategy.org, D	dailykos.com, D	blogsforbush.com, R
2°	madkane.com/notable.html, D	instapundit.com, R	newleftblogs.blogspot.com, D
3°	liberaloasis.com, D	talkingpointsmemo.com, D	politicalstrategy.org, D
4°	stagefour.typepad.com/commonprejudice, D	atrios.blogspot.com, D	madkane.com/notable.html, D
5°	bodyandsoul.typepad.com, D	drudgereport.com, R	cayankee.blogs.com, R
6°	corrente.blogspot.com, D	powerlineblog.com, R	liberaloasis.com, D
7°	aurelientt.blogspot.com, D	blogsforbush.com, R	lashawnbarber.com, D
8°	tbogg.blogspot.com, D	washingtonmonthly.com, D	gevkaffeegal.typepad.com/thealliance, R
9°	newleftblogs.blogspot.com, D	michellemalkin.com, R	presidentboxer.blogspot.com, R
10°	atrios.blogspot.com, D	truthlaidbear.com, R	corrente.blogspot.com, D

TABLE IV. Kendall's τ for each pair of centrality measures for the domain *.gov* of the Web.

Measures	τ
PR-αEV	0.189
PR-AUTH	0.079
PR-HUBS	0.060
PR-IN	0.155
PR-OUT	0.090
αEV-AUTH	0.081
αEV-IN	0.147
αEV-HUBS	0.074
αEV-OUT	0.086
AUTH-IN	0.046
AUTH-HUBS	0.109
AUTH-OUT	0.072
HUBS-IN	0.003
HUBS-OUT	0.056
IN-OUT	0.081

power law with exponent 2. The same is true for α-centrality, because its defining equation is formally equivalent to the equation for PageRank in the limit $q \to 0$. These results on centrality distributions are likely to be true for an extended class of graphs, where there is a flow from the outermost

nodes (leaves) to a sink. We have also seen that, on any graph, in the limit $q \to 1$, the reduced PageRank of a node, i.e., the contribution of the random walk process to the measure, is simply proportional to the indegree of the node, if the nodes have (about) the same outdegree. We have studied the extension of PageRank to the case of undirected networks, finding that the reduced PageRank of a node is proportional to its degree, for large degrees, for any graph and value of q. We proposed a simple explanation of this effect based on the Central Limit Theorem, and verified numerically in several cases that the argument holds. Similarly, the reduced α-centrality of a node is also proportional to its degree, for large degrees, on any graph. With the same type of argument it is possible to show that the authority score of a node is proportional to its indegree, for large indegrees, when the outdegrees of all nodes are (approximately) the same.

We conclude that there are often strong relations between our centrality measure and (in)degree: some relations hold on particular graphs and/or limits, others are more general. These findings imply that the measures are often strongly correlated with each other. We have indeed seen that the rankings of nodes according to the centrality measures we have considered are quite close to each other for indegree, PageRank, EV, and authority score on graphs built with the prescription of Dorogovtsev, Mendes, and Samukhin. We have shown in the paper that these graphs have special properties, and that some measures may be correlated with each other. Instead, on real graphs, like the networks of political

TABLE V. Top ten of the web domain *.gov* according to PR, αEV, authorities, and indegree.

Rank	PR	αEV
1°	www.usgs.gov	polar.wwb.noaa.gov/waves/main_int.js
2°	www.nws.noaa.gov	polar.wwb.noaa.gov/waves/welcome.html
3°	www.naca.larc.nasa.gov/readme.html	polar.wwb.noaa.gov/waves/main_table.html
4°	www.usda.gov	polar.wwb.noaa.gov/waves/products.html
5°	www.nws.noaa.gov/disclaimer.html	polar.wwb.noaa.gov/waves/main_int.html
6°	www.ar.inel.gov/home.htm	www.nws.noaa.gov/disclaimer1.html
7°	www.4woman.gov/search/search.cfm	www.nws.noaa.gov
°	www.nws.noaa.gov/feedback.shtml	polar.wwb.noaa.gov/waves/references.htm
9°	www.access.wa.gov	polar.wwb.noaa.gov/waves/validation.htm
10°	www.usinfo.state.gov/products/pdq/pdq.htm	polar.wwb.noaa.gov/waves/valid_wna.html
Rank	Auth	In
1°	www.srh.noaa.gov/oun/cgi-bin/ wxclick.pl?county=oklahoma	www.usgs.gov
2°	www.srh.noaa.gov/oun/cgi-bin/ wxclick.pl?county=cleveland	www.cdc.gov
3°	www.srh.noaa.gov/oun/cgi-bin/wxclick.pl?county=kiowa	www.usda.gov
4°	www.nws.noaa.gov	www.doi.gov
5°	www.srh.noaa.gov/oun/cgi-bin/wxclick.pl?county=logan	www.nws.noaa.gov
6°	www.srh.noaa.gov/oun/cgi-bin/wxclick.pl?county=payne	www.usgs.gov/disclaimer.html
7°	www.srh.noaa.gov/oun/cgi-bin/wxclick.pl?county=knox	www.usda.gov/news/privacy.htm
8°	weather.noaa.gov	www.abag.ca.gov
9°	weather.noaa.gov/weather/ok_cc_us.html	www.ars.usda.gov/nodisc.html
10°	www.crh.noaa.gov/ddc	www.ars.usda.gov/comm.htm

blogs and the sample of the Web graph we have considered, the structure is less regular and the measures are far less correlated with each other, as confirmed by the small values of the Kendall's τ for each pair of centrality measures. This means that, for practical purposes, and in spite of their similarities, spectral centrality measures look at nodes from different perspectives, and allow to diversify their roles within the network, obtaining in this way more information about the importance of nodes. The scores computed from spectral centrality measures can complement the information about the node's centrality derived from more traditional measures like node betweenness [26]. This is especially important for directed graphs, where node betweenness, as well as other measures based on geodesic paths, like closeness [27], are not well defined.

[1] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[3] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).

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

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

[6] S. Wasserman and K. Faust, *Social Networks Analysis* (Cambridge University Press, Cambridge, U.K., 1994).

[7] J. Scott, *Social Networks Analysis: A Handbook* (Sage Publications, London, 2000).

[8] S. Brin and L. Page, Comput. Netw. **30**, 107 (1998).

[9] E. Estrada and J. A. Rodríguez-Velázquez, Phys. Rev. E **71**, 056103 (2005).

[10] E. Estrada and N. Hatano, Chem. Phys. Lett. **439**, 247 (2007).

[11] E. Estrada and J. A. Rodríguez-Velázquez, Phys. Rev. E **72**, 046105 (2005).

[12] P. Bonacich and P. Lloyd, Soc. Networks **23**, 191 (2001).

[13] J. Kleinberg, in *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* (ACM, New York, 1998), pp. 668–677.

[14] P. Boldi, M. Santini, and S. Vigna, in *Proceedings of the Fourteenth International World Wide Web Conference, Chiba, Japan* (ACM New York, 2005), pp. 557–566.

[15] S. Fortunato, M. Boguñá, A. Flammini, and F. Menczer, e-print arXiv:cs/0511016, Internet Math. (to be published).

[16] S. Fortunato and A. Flammini, Int. J. Bifurcation Chaos Appl. Sci. Eng. **17**, 2343 (2007).

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

[18] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).

[19] J. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, *Computing and Combinatorics*, Lecture Notes in Computer Science Vol. 1627 (Springer, Berlin, 1999), p. 1.

[20] P. De Los Rios, Europhys. Lett. **56**, 898 (2001).

[21] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).

[22] P. Chen, H. Xie, S. Maslov, and S. Redner, J. Informet. **1**, 8 (2007).

[23] B. Bollobás, *Modern Graph Theory* (Springer, New York, 1998).

[24] M. Kendall, Biometrika **30**, 81 (1938).

[25] L. Adamic and N. Glance, in Proceedings of the Third International Workshop on Link Discovery, Los Angeles, 2005 (unpublished), p. 36.

[26] L. C. Freeman, Sociometry **40**, 35 (1977).

[27] G. Sabidussi, Psychometrika **31**, 581 (1966).