

Technical Report

TR-2013-014

A matrix analysis of different centrality measures

by

Michele Benzi, Christine Klymko

MATHEMATICS AND COMPUTER SCIENCE

EMORY UNIVERSITY

A MATRIX ANALYSIS OF DIFFERENT CENTRALITY MEASURES

MICHELE BENZI* AND CHRISTINE KLYMKO†

Abstract. Node centrality measures including degree, eigenvector, Katz and subgraph centralities are analyzed for both undirected and directed networks. We show how parameter-dependent measures, such as Katz and subgraph centrality, can be “tuned” to interpolate between degree and eigenvector centrality, which appear as limiting cases of the other measures. We interpret our finding in terms of the local and global influence of a given node in the graph as measured by graph walks of different length through that node. Our analysis gives some guidance for the choice of parameters in Katz and subgraph centrality, and provides an explanation for the observed correlations between different centrality measures and for the stability exhibited by the ranking vectors for certain parameter ranges. The important role played by the spectral gap of the adjacency matrix is also highlighted.

Key words. centrality, communicability, adjacency matrix, spectral gap, matrix functions, network analysis

AMS subject classifications. 05C50, 15A16

1. Introduction. The mathematical and computational study of complex networks has experienced tremendous growth in recent years. An amazing variety of highly interconnected systems, both in nature and in the man-made world of technology, can be modeled in terms of networks. Models of social interaction have long made use of graphs and linear algebra, and network models are now common place not only in the “hard” sciences but also in economics, finance, anthropology, and even in the humanities. As more and more data has become available, the need for tools to analyze these networks has increased and a new field of Network Science has come of age [1, 2, 7, 16, 17, 19, 25, 27, 47, 48].

Since graphs, which are abstract models of real-world networks, can be described in terms of matrices, it comes as no surprise that linear algebra plays an important role in network analysis. Many problems in this area require the solution of linear systems, the computation of eigenvalue and eigenvectors, and the evaluation of matrix functions. Also, the study of dynamical processes on graphs leads to the need for solving systems of differential and difference equations posed on graphs; the behavior of the solution as a function of time is strongly influenced by the structure (topology) of the underlying graph, which in turn is reflected in the spectral property of matrices associated with the graph.

One of the most basic questions in network analysis is how to determine the “most important” nodes in a network. Examples include essential proteins in Protein-Protein Interaction Networks, keystone species in ecological networks, authoritative web pages on the World Wide Web, influential authors in scientific collaboration networks, leading actors in the International Movies Database, and so forth; see, e.g., [27] for details and many additional examples. When the network being examined is very small (say, on the order of 10 nodes), this determination of importance can often be done visually, but as networks increase in size and complexity, visual analysis becomes impossible. Instead, computational measures of node importance, called *centrality measures*, are used to rank the nodes in a network. There are many different

*Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia 30322, USA (benzi@mathcs.emory.edu). The work of this author was supported by National Science Foundation grant DMS1115692.

†Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia 30322, USA (cklymko@emory.edu).

centrality measures in use (see [5, 6, 11, 12, 13, 14, 15, 20, 33, 37, 38, 40, 42, 43, 46, 49], for example) and the “best” centrality measure to use tends to be application-dependent. However, many authors have noted that different centrality measures often provide rankings that are highly correlated, at least when analysis is restricted to the most highly ranked nodes; see, e.g., [24, 45], and [5, 6] for more recent studies.

In this paper, we analyze the relationship between degree centrality, eigenvector centrality, and four centrality measures based on the diagonal entries (for undirected graphs) and row sums of the exponential and resolvent of the adjacency matrix of the graph: subgraph centrality, total communicability, resolvent subgraph centrality, and Katz centrality. The measures based on the matrix exponential include a parameter β and those based on the matrix resolvent included a parameter α . Note that HITS [40], as a type of eigenvector centrality, is covered by our analysis, as is the extension of subgraph centrality to digraphs given in [5].

As mentioned, there are a number of other ranking methods in use, yet we restrict ourselves to considering the above six. The choice of which of the many centrality measures to study and why is something that must be considered carefully; see the discussion in [18]. In this paper we focus our attention on centrality measures that have been widely tested and that can be expressed in terms of linear algebra (more specifically, in terms of the adjacency matrix of the network). We additionally restricted our scope to centrality measures that we could demonstrate were related to each other. For instance, we did not include in our analysis two popular centrality measures, betweenness centrality [37] and (for directed networks) PageRank [49]. Betweenness centrality does not seem to have a simple expression in terms of the adjacency matrix, while PageRank involves modifications and scalings of the adjacency matrix which make it impossible, at least for us, to find an analytic relation with the other methods. Also, PageRank has a probabilistic interpretation (in terms of Markov chains) that is different from that of other centrality measures.

The paper is organized as follows. The relationship between the various centrality measures for undirected networks is examined in section 3. In section 4, we give an interpretation of this relationship. Section 5 contains experimental comparisons between degree and eigenvector centrality with those based on the matrix exponential and resolvent for various values of α and β on a variety of real-world networks. In section 6, we examine the analytical relationship between the various centrality measures on directed networks. Experiments on directed networks are presented in section 7. We present some concluding remarks in section 8.

2. Background and definitions. In this section we recall some basic concepts from graph theory that will be used in the rest of the paper. A more complete overview can be found, e.g., in [23].

A *graph* $G = (V, E)$ is defined by a set of n nodes (also referred to as vertices) V with $|V| = n$ and a set of edges $E = \{(i, j) | i, j \in V\}$. G is *undirected* if the edges are formed by unordered pairs of vertices. The *degree* of a vertex i , denoted by d_i , is given by the number of edges incident to i in G . If every vertex has degree k , then the graph G is said to be *k-regular* or *degree regular* (or, simply, *regular*).

A *walk* of length k in G is a set of nodes $i_1, i_2, \dots, i_k, i_{k+1}$ such that for all $1 \leq l \leq k$, there is an edge between i_l and i_{l+1} . A *closed walk* is a walk where $i_1 = i_{k+1}$. A *path* is a walk with no repeated nodes. A *cycle* is a path with an edge between the first and last node. A graph G is *walk-regular* if the number of closed walks of length $k \geq 0$ starting at a given node is the same for every node in the graph. A graph is *simple* if it has no *loops* (edges from a node i to itself), no multiple edges, and unweighted

edges. An undirected graph is *connected* if there exists a path between every pair of nodes.

A graph G is *directed* (also called a *digraph*) if the edges are made up of ordered pairs of vertices. That is, $(i, j) \in E \nRightarrow (j, i) \in E$. In a directed graph, each node has two types of degrees. The *in-degree* of node i , d_i^{in} , is given by the number of edges ending at node i (pointing in to i) and the *out-degree*, d_i^{out} , is given by the number of edges starting at node i (pointing out from node i).

A walk in a directed network is a set of nodes $i_1, i_2, \dots, i_k, i_{k+1}$ such that for all $1 \leq l \leq k$, there is an edge from i_l to i_{l+1} . The definitions of closed walks, paths, and cycles are extended to digraphs similarly. The existence of a walk from i_1 to i_{k+1} does not imply the existence of a walk from i_{k+1} to i_1 when G is directed. A digraph G is *strongly connected* if for every pair of vertices i and j , there is both a directed walk from i to j and from j to i .

Every graph G can be represented as a matrix through the use of an adjacency matrix A . It is given by $A = (a_{ij})$ with

$$a_{ij} = \begin{cases} 1, & \text{if } (i, j) \text{ is an edge in } G, \\ 0, & \text{else.} \end{cases}$$

If G is a simple, undirected graph, A is binary and symmetric with zeros along the main diagonal. In this case, the eigenvalues of A will be real. We label the eigenvalues of A in non-increasing order: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. If G is connected, then $\lambda_1 > \lambda_2$ by the Perron-Frobenius theorem [44, page 673]. Since A is a symmetric, real-valued matrix, we can decompose A into $A = Q\Lambda Q^T$ where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ with $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_n$ where $Q = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n]$ is orthogonal and \mathbf{q}_i is the eigenvector associated with λ_i .

If G is a strongly connected digraph, its adjacency matrix A is irreducible. Let $\rho(A) = r$ be the spectral radius of A . Then, again by the Perron-Frobenius theorem, $\lambda_1 = r$ is a simple eigenvalue of A and both the left and right eigenvectors of A associated with λ_1 are positive. If G is also diagonalizable, then there exists an invertible matrix X such that $A = X\Lambda X^{-1}$ where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ with $\lambda_1 \geq |\lambda_i|$ for $2 \leq i \leq n$, $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, and $(X^{-1})^* = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$. The left eigenvector associated with λ_i is \mathbf{y}_i and the right eigenvector associated with λ_i is \mathbf{x}_i . In the case where G is not diagonalizable, A can be decomposed using the Jordan canonical form:

$$A = XJX^{-1} = X \begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & \hat{J} \end{pmatrix} X^{-1},$$

where J is the Jordan matrix of A , except that we place the 1×1 block corresponding to λ_1 first for notational convenience.

In the rest of the paper, I denotes the $n \times n$ identity.

2.1. Common measures of centrality. As we discussed in the Introduction, many measures of node centrality have been developed and used over the years. The higher the measure of node centrality, the more “important” a given node is considered to be in the network. The most common include degree centrality, eigenvector centrality, betweenness centrality [15, 37], Katz centrality [39], and subgraph centrality [32, 33]. More recently, total communicability has been introduced as a centrality measure [6]. In this paper we focus our attention on the following centrality measures:¹

¹Centrality measures for directed networks are discussed in section 6 below.

- degree centrality: $C_d(i) = d_i$;
- eigenvector centrality: $C_{ev}(i) = \mathbf{q}_1(i)$, where \mathbf{q}_1 is the dominant eigenvector of A ;
- exponential subgraph centrality: $SC_i(\beta) = [e^{\beta A}]_{ii}$;
- resolvent subgraph centrality: $RC_i(\alpha) = [(I - \alpha A)^{-1}]_{ii}$,
- total communicability: $TC_i(\beta) = [e^{\beta A} \mathbf{1}]_i = \mathbf{e}_i^T e^{\beta A} \mathbf{1}$;
- Katz centrality: $K_i(\alpha) = [(I - \alpha A)^{-1} \mathbf{1}]_i = \mathbf{e}_i^T (I - \alpha A)^{-1} \mathbf{1}$.

Here $\mathbf{1}$ is the vector of all ones, \mathbf{e}_i is the i th standard basis vector, $0 < \alpha < \frac{1}{\lambda_1}$ (see below), and $\beta > 0$. In [39], the column sums of $(I - \alpha A)^{-1}$ were used to measure centrality but, in the case of an undirected network, this is equivalent to the row sums.

Often, the value $\beta = 1$ is used in the calculation of exponential subgraph centrality and total communicability. The parameter β can be interpreted as an *inverse temperature* and has been used to model the effects of external disturbances on the network. As $\beta \rightarrow 0+$, the “temperature” of the environment surrounding the network increases, corresponding to more intense external disturbances. Conversely, as $\beta \rightarrow \infty$, the temperature goes to 0 and the network “freezes.” We refer the reader to [31] for an extensive discussion and applications of these concepts.

The justification behind using the diagonals of the (scaled) matrix exponential as a centrality measure can be seen by considering the power series expansion of $e^{\beta A}$:

$$e^{\beta A} = I + \beta A + \frac{(\beta A)^2}{2!} + \cdots + \frac{(\beta A)^k}{k!} + \cdots = \sum_{k=0}^{\infty} \frac{(\beta A)^k}{k!}. \quad (2.1)$$

It is well-known that given an adjacency matrix A of an unweighted network, $[A^k]_{ij}$ counts the total number of walks of length k between nodes i and j . Thus, the exponential subgraph centrality of node i counts the total number of closed walks in the network which are centered at node i , weighting walks of length k by a factor of $\frac{\beta^k}{k!}$.

Total communicability is closely related to subgraph centrality. This measure also counts the number of walks starting at node i , scaling walks of length k by $\frac{\beta^k}{k!}$. However, rather than just counting closed walks, total communicability counts all walks between node i and every node in the network. The name stems from the fact that $TC_i(\beta) = \sum_{j=1}^n C_i(\beta)$ where $C_i(\beta) = [e^{\beta A}]_{ij}$, the communicability between nodes i and j , is a measure of how “easy” it is to exchange a message between nodes i and j over the network; see, again, [31] for details. Although subgraph centrality and total communicability are clearly related, they do not always provide the same ranking of the nodes. More information about the relation between the two measures can be found in [6].

The matrix resolvent $(I - \alpha A)^{-1}$ was first used to rank nodes in a network in the early 1950s, when Katz used the column sums to calculate node importance [39]. Since then, the diagonal values have also been used to calculate subgraph centrality, see [32]. The resolvent subgraph centrality score of node i is given by $[(I - \alpha A)^{-1}]_{ii}$ and the Katz centrality score is given by $[(I - \alpha A)^{-1} \mathbf{1}]_i$.

As when using the matrix exponential, these resolvent-based centrality measures count the number of walks in the network, penalizing longer walks. This can be seen by considering the power series expansion of $(I - \alpha A)^{-1}$:

$$(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \cdots + \alpha^k A^k + \cdots = \sum_{k=0}^{\infty} \alpha^k A^k. \quad (2.2)$$

The resolvent subgraph centrality of node i , $[(I - \alpha A)^{-1}]_{ii}$, counts the total number of closed walks in the network which are centered at node i , weighting walks of length k by α^k . Similarly, the Katz centrality (also known as resolvent total communicability, see [6]) of node i counts all walks beginning at node i , penalizing the contribution of walks of length k by α^k . The bounds on α ($0 < \alpha < \frac{1}{\lambda_1}$) ensure that the matrix $I - \alpha A$ is invertible and that the power series in (2.2) converges to its inverse. The bounds on α also force $(I - \alpha A)^{-1}$ to be nonnegative, as $I - \alpha A$ is a nonsingular M -matrix. Hence, both the diagonal entries and the row sums of $(I - \alpha A)^{-1}$ are nonnegative and can thus be used for ranking purposes.

3. Relationship between various centrality measures. One difficulty in measuring the “importance” of a node in a network using centrality is that it is not always clear which of the many centrality measures should be used. Additionally, it is not clear a priori whether two centrality measures will give similar node rankings on a given network. When using exponential or resolvent-based centrality measures, the necessity of choosing a parameter (β and α , respectively) adds another layer of difficulty. Different choices of α and β will generally produce different centrality scores and can lead to different node rankings. However, experimentally, it has been seen that different centrality measures often provide rankings that are highly correlated. Moreover, in most cases, the *rankings* are quite stable, in the sense that they do not appear to change much for different choices of α and β [41], even if the actual *scores* vary by orders of magnitude. This happens in particular when the parameters α and β approach their limits:

$$\alpha \rightarrow 0+, \quad \alpha \rightarrow \frac{1}{\lambda_1} -, \quad \beta \rightarrow 0+, \quad \beta \rightarrow \infty.$$

Noting that the first derivatives of the node centrality measures grow unboundedly as $\alpha \rightarrow \frac{1}{\lambda_1} -$ and as $\beta \rightarrow \infty$, the centrality scores are extremely sensitive to (vary extremely rapidly with) small changes in α (when α is close to $\frac{1}{\lambda_1} -$) and in β (when β is even moderately large). Yet, the rankings produced do not change much, if at all, and the same is observed as $\alpha, \beta \rightarrow 0+$.

The observed correlations and stability of these rankings can be explained by the following two theorems, which relate degree and eigenvector centrality to exponential and resolvent-based subgraph centrality, respectively:

THEOREM 3.1. *Let $G = (V, E)$ be a connected, undirected network with adjacency matrix A . Let $SC_i(\beta) = [e^{\beta A}]_{ii}$ be the subgraph centrality of node i and \mathbf{SC} be the corresponding vector of subgraph centralities. Then,*

- (i) *as $\beta \rightarrow 0+$, the rankings produced by $\mathbf{SC}(\beta)$ converge to those produced by $\mathbf{C_d}$, the vector of degree centralities;*
- (ii) *as $\beta \rightarrow \infty$, the rankings produced by $\mathbf{SC}(\beta)$ converge to those produced by $\mathbf{C_{ev}}$, the vector of eigenvector centralities.*

Proof. To prove (i), consider the Taylor expansion of $SC_i(\beta)$:

$$SC_i(\beta) = [e^{\beta A}]_{ii} = 1 + \beta[A]_{ii} + \frac{\beta^2[A^2]_{ii}}{2!} + \frac{\beta^3[A^3]_{ii}}{3!} + \cdots]_{ii} = 1 + 0 + \frac{\beta^2}{2!}d_i + \frac{\beta^3}{3!}[A^3]_{ii} + \cdots$$

Let $\phi(\beta) = \frac{2!}{\beta^2}[\mathbf{SC}(\beta) - 1]$. The rankings produced by $\phi(\beta)$ will be the same as those produced by $\mathbf{SC}(\beta)$, as the scores for each node have all been shifted and scaled in

the same way. Now, the i th entry of $\phi(\beta)$ is given by

$$\phi_i(\beta) = \frac{2!}{\beta^2}[SC_i(\beta) - 1] = d_i + \frac{2\beta}{3!}[A^3]_{ii} + \frac{2\beta^2}{4!}[A^4]_{ii} + \dots$$

which tends to d_i as $\beta \rightarrow 0+$. Thus, as $\beta \rightarrow 0+$, the rankings produced by the subgraph centrality scores reduce to those produced by the degrees.

To prove (ii), consider the expansion of $SC_i(\beta)$ in terms of the eigenvalues and eigenvectors of A :

$$SC_i(\beta) = \sum_{k=1}^n e^{\beta\lambda_k} \mathbf{q}_k(i)^2 = e^{\beta\lambda_1} \mathbf{q}_1(i)^2 + \sum_{k=2}^n e^{\beta\lambda_k} \mathbf{q}_k(i)^2.$$

Let $\psi(\beta) = \frac{1}{e^{\beta\lambda_1}} \mathbf{SC}(\beta)$. As in the proof of (i), the rankings produced by $\psi(\beta)$ are the same as those produced by $\mathbf{SC}(\beta)$, since the scores for each node have all been rescaled by the same amount. Next,

$$\psi_i(\beta) = \mathbf{q}_1(i)^2 + \sum_{k=2}^n e^{\beta(\lambda_k - \lambda_1)} \mathbf{q}_k(i)^2.$$

Since $\lambda_1 > \lambda_k$ for $2 \leq k \leq n$, as $\beta \rightarrow \infty$ we see that $\psi_i(\beta) \rightarrow \mathbf{q}_1(i)^2$. By the Perron-Frobenius Theorem $\mathbf{q}_1 > 0$, hence the rankings produced by $\mathbf{q}_1(i)^2$ are the same as those produced by $\mathbf{q}_1(i)$. Thus, as $\beta \rightarrow \infty$, the rankings produced by the subgraph centrality scores reduce to those obtained with eigenvector centrality. \square

THEOREM 3.2. *Let $G = (V, E)$ be a connected, undirected network with adjacency matrix A . Let the resolvent subgraph centrality of node i be given by $RC_i(\alpha) = [(I - \alpha A)^{-1}]_{ii}$ and let $\mathbf{RC}(\alpha)$ be the corresponding vector of resolvent subgraph centralities. Then,*

- (i) *as $\alpha \rightarrow 0+$, the rankings produced by $\mathbf{RC}(\alpha)$ converge to \mathbf{C}_d , the vector of degree centralities;*
- (ii) *as $\alpha \rightarrow \frac{1}{\lambda_1}-$, the rankings produced by $\mathbf{RC}(\alpha)$ converge to \mathbf{C}_{ev} , the vector of eigenvector centralities.*

The proof of Theorem 3.2 follows the same arguments as that of Theorem 3.1 and is omitted.

Similar relationships hold between degree and eigenvector centralities and the centrality measures based on row sums (exponential total communicability and Katz centrality). These relationships are made explicit in the following two theorems.

THEOREM 3.3. *Let $G = (V, E)$ be a connected, undirected network with adjacency matrix A . Let $TC_i(\beta) = [e^{\beta A} \mathbf{1}]_i$ be the total communicability of node i and $\mathbf{TC}(\beta)$ be the corresponding vector of total communicabilities. Then,*

- (i) *as $\beta \rightarrow 0+$, the rankings produced by $\mathbf{TC}(\beta)$ converge to those produced by \mathbf{C}_d , the vector of degree centralities;*
- (ii) *as $\beta \rightarrow \infty$, the rankings produced by $\mathbf{TC}(\beta)$ converge to those produced by \mathbf{C}_{ev} , the vector of eigenvector centralities.*

THEOREM 3.4. *Let $G = (V, E)$ be a connected, undirected network with adjacency matrix A . Let the Katz centrality of node i be given by $K_i(\alpha) = [(I - \alpha A)^{-1} \mathbf{1}]_i$ and let $\mathbf{K}(\alpha)$ denote the corresponding vector of Katz centralities. Then,*

- (i) as $\alpha \rightarrow 0+$, the rankings produced by $\mathbf{K}(\alpha)$ converge to \mathbf{C}_d , the vector of degree centralities;
- (ii) as $\alpha \rightarrow \frac{1}{\lambda_1}-$, the rankings produced by $\mathbf{K}(\alpha)$ converge to \mathbf{C}_{ev} , the vector of eigenvector centralities.

The proofs of these theorems are similar to those of Theorem 3.1 above and are therefore omitted. We also note that these results can be obtained as special cases of those given in section 6 below for directed networks.

4. Interpretation. The centrality scores which we are considering in this paper are all based on walks in the network. The degree centrality of a node i counts the number of walks of length one starting at i (the degree of i). In contrast, the eigenvector centrality of node i gives the limit as $k \rightarrow \infty$ of the percentage of walks of length k which start at node i (see [27, p. 127] and [21]). Thus, the degree centrality of node i measures the *local* influence of i and the eigenvector centrality measures the *global* influence of i .

Exponential subgraph centrality and total communicability take both local and global influence into account, weighting walks of length k by $\frac{\beta^k}{k!}$, $\beta > 0$. As β decreases, the weights corresponding to larger k decay faster and shorter walks become more important in the centrality rankings. In the limit as $\beta \rightarrow 0+$, walks of length one dominate the centrality scores and the rankings converge to the degree centrality rankings. As β increases to infinity, given a fixed walk length k , the corresponding weight increases more rapidly than those of shorter walks. In the limit as $\beta \rightarrow \infty$, walks of infinite length dominate and the centrality rankings converge to those of eigenvector centrality.

The situation is similar when resolvent subgraph centrality and Katz centrality are considered. Walks of length k are weighted by a factor of α^k , $0 < \alpha < \frac{1}{\lambda_1}$. Again, as α decreases, shorter walks dominate the rankings and, in the limit as $\alpha \rightarrow 0+$, the rankings converge to those produced by degree centrality. As α approaches $\frac{1}{\lambda_1}$, longer walks dominate and, in the limit as $\alpha \rightarrow \frac{1}{\lambda_1}-$, the rankings converge to those produced by eigenvector centrality.

In these parameterized centrality rankings, the parameters α and β should be viewed as a method for tuning between rankings based on local influence (short walks) and those based on global influence (long walks). In applications where local influence is most important, degree centrality will often be equivalent to any of the parameterized centrality rankings with α or β small. Similarly, when global influence is the only important factor, parameterized centrality rankings with α or β large will often be equivalent to eigenvector centrality.

Exponential subgraph centrality and total communicability, along with resolvent subgraph centrality and Katz centrality, are most useful when both local and global influence need to be considered in the ranking of nodes in a network. In order to achieve this, “moderate” values of β and α must be used; a more precise discussion is given in section 5, where we report on the differences between the rankings produced by these methods with various choices of parameters and those produced by degree and eigenvector centrality in both synthetic and real world networks.

The rate at which these rankings converge to either degree centrality (as $\alpha, \beta \rightarrow 0+$) or eigenvector centrality (as $\alpha \rightarrow \frac{1}{\lambda_1}-$ and $\beta \rightarrow \infty$) depends on the *spectral* gap of the adjacency matrix of the network, $\lambda_1 - \lambda_2$. When the spectral gap is large, all four of the parameterized centrality rankings will converge to eigenvector centrality more quickly as α and β increase than in the case when the spectral gap is small.

Thus, in networks with a large enough spectral gap, eigenvector centrality may as well be used instead of a method based on the exponential or resolvent of the adjacency matrix. However, it's not always easy to tell *a priori* when $\lambda_1 - \lambda_2$ is "large enough"; some guidelines can be found in [26]. We also note that the tuning parameters α and β can be regarded as a way to artificially widen or shrink the (absolute) gap, thus giving more or less weight to the dominant eigenvector (or, equivalently, increasing or decreasing the importance of global influences, represented by long walks in the graph).

4.1. Discriminating power of centrality measures. Consider the eight vertex, 3-regular graph shown in Fig. 4.1. This graph is not walk-regular. This can be verified by noting that nodes 1, 7, and 8 make up a triangle, as do nodes 2, 3, and 4, but nodes 5 and 6 are not involved in any triangles. Consequently, for all nodes $i \neq 5, 6$, there are two closed walks of length 3 beginning at node i while nodes 5 and 6 are not involved in any closed walks of length three.

Due to the regularity of this network, degree centrality and eigenvector centrality are unable to discriminate between the nodes (they all have the same score), thus no actual ranking is produced by either of these two methods. In terms of our interpretation, each node has exactly the same local and global influence than any other node.

Furthermore, since all the powers A^p of the adjacency matrix A of a k -regular graph have constant row sums ($= k^p$), Katz and total communicability centrality are also unable to discriminate between nodes, regardless of α and β .

In contrast, due to the fact that the graph is not walk-regular, subgraph centrality (whether exponential or resolvent-based) is able to discriminate between the nodes and thus to provide rankings. For example, using the diagonals of e^A leads to a 4-way tie at the top (nodes 3,4,7,8), followed by nodes 1 and 2 (tied) followed by nodes 5 and 6 (tied). The same ranking is obtained using the diagonal entries of $(I - \alpha A)^{-1}$ with, say, $\alpha = 0.25$ (the dominant eigenvalue of A is $\lambda_1 = 3$).

Hence, this simple academic example shows that subgraph centrality has greater discriminatory power than the other ranking methods considered here.²

5. Numerical experiments on undirected networks. In this section we present the results of numerical experiments aimed at illustrating the relationship between the centrality measures under consideration on various undirected networks.

The rankings produced by the various centrality measures are compared using the *intersection distance* method (for more information, see [34] and [8, 20]). Given two ranked lists x and y , the top- k intersection distance is computed by:

$$\text{isim}_k(x, y) := \frac{1}{k} \sum_{i=1}^k \frac{|x_i \Delta y_i|}{2i}$$

where Δ is the symmetric difference operator between the two sets and x_k and y_k are the top k items in x and y , respectively. The top- k intersection distance gives the average of the normalized symmetric differences for the lists of the top i items for all $i \leq k$. If the ordering of the top k nodes is the same for the two ranking schemes,

²It is obvious that if a graph is walk-regular, none of the centrality measures considered in this paper can discriminate between its nodes. For subgraph centrality, the converse of this statement is an open conjecture of Estrada; see, e.g., [29, 30, 50, 51] as well as [3] for further discussion of this and related questions.

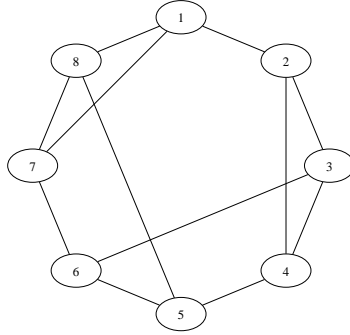


Fig. 4.1: A 3-regular graph on 8 nodes which is not walk-regular.

$\text{isim}_k(x, y) = 0$. If the top k are disjoint, then $\text{isim}_k(x, y) = 1$. Unless otherwise specified, we compare the intersection distance for the full set of ranked nodes.

Unless otherwise specified, all of the numerical experiments were performed in Matlab version 7.9.0 (R2009b) on a MacBook Pro running OS X Version 10.7.5 with a 2.4 GHZ Intel Core i5 processor and 4GB of RAM.

The networks come from a range of sources, although most can be found in the University of Florida Sparse Matrix Collection [22]. The first is the Zachary Karate Club network, which is a classic example in network analysis [52]. The Intravenous Drug User and the Yeast PPI networks were provided by Prof. Ernesto Estrada and are not present in the University of Florida Collection. The three Erdős networks correspond to various subnetworks of the Erdős collaboration network and can be found in the Pajek group of the UF Collection. The ca-GrQc and ca-HepTh networks are collaboration networks corresponding to the General Relativity and High Energy Physics Theory subsections of the arXiv and can be found in the SNAP group of the UF Collection. The as-735 network can also be found in the SNAP group and represents the communication network of a group of Autonomous Systems on the Internet. This communication was measured over the course of 735 days, between November 8, 1997 and January 2, 2000. The final network is the network of Minnesota roads and can be found in the Gleich group of the UF Collection. Basic data on these networks, including the order n , number of nonzeros, and the largest two eigenvalues, can be found in Table 5.1. All of the networks, with the exception of the Yeast PPI network, are simple. The Yeast PPI network has several ones on the diagonal, representing the self-interaction of certain proteins. All are undirected.

5.1. Exponential subgraph centrality and total communicability. We examined the effects of changing β on the exponential subgraph centrality and total communicability rankings of nodes in a variety of undirected real world networks, as well as their relation to degree and eigenvector centrality. Although the only restriction on β is that it must be greater than zero, there is often an implicit upper limit that may be problem-dependent. For the analysis in this section, we impose the following limits: $0.1 \leq \beta \leq 10$. To examine the sensitivity of the exponential subgraph centrality and total communicability rankings, we calculate both sets of scores and rankings for various choices of β . The values of β tested are: 0.1, 0.5, 1, 2, 5, 8 and 10.

The rankings produced by the matrix exponential-based centrality measures for

Table 5.1: Basic data for the networks used in the experiments.

Graph	n	nnz	λ_1	λ_2
Zachary Karate Club	34	156	6.726	4.977
Drug User	616	4024	18.010	14.234
Yeast PPI	2224	13218	19.486	16.134
Pajek/Erdos971	472	2628	16.710	10.199
Pajek/Erdos972	5488	14170	14.448	11.886
Pajek/Erdos982	5822	14750	14.819	12.005
Pajek/Erdos992	6100	15030	15.131	12.092
SNAP/ca-GrQc	5242	28980	45.617	38.122
SNAP/ca-HepTh	9877	51971	31.035	23.004
SNAP/as-735	7716	26467	46.893	27.823
Gleich/Minnesota	2642	6606	3.2324	3.2319

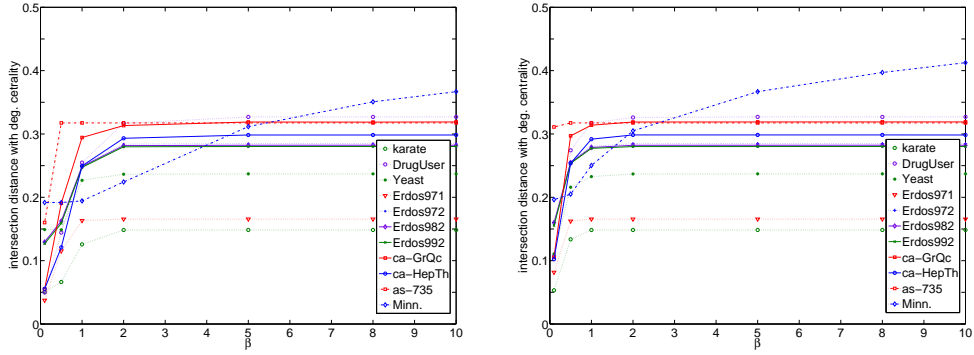


Fig. 5.1: The intersection distances between degree centrality and the exponential subgraph centrality (left) or total communicability (right) rankings of the nodes in the networks in Table 5.1.

all choices of β were compared to those produced by degree centrality and eigenvector centrality, using the intersection distance method described above. Plots of the intersection distances for the rankings produced by various choices of β with those produced by degree or eigenvector centrality can be found in Figs. 5.1 and 5.2. The intersection distances for rankings produced by successive choices of β can be found in Fig. 5.3.

In Figure 5.1, the rankings produced by exponential subgraph centrality and total communicability are compared to those produced by degree centrality. For small values of β , both sets of rankings based on the matrix exponential are very close to those produced by degree centrality (low intersection distances). When $\beta = 0.1$, the largest intersection distance between the degree centrality rankings and the exponential subgraph centrality rankings for the networks examined is slightly less than 0.2 (for the Minnesota road network). The largest intersection distance between the total communicability rankings with $\beta = 0.1$ and the degree centrality rankings is 0.3 (for the as-735 network). In general, the (diagonal based) exponential subgraph centrality rankings tend to be slightly closer to the degree rankings than the (row

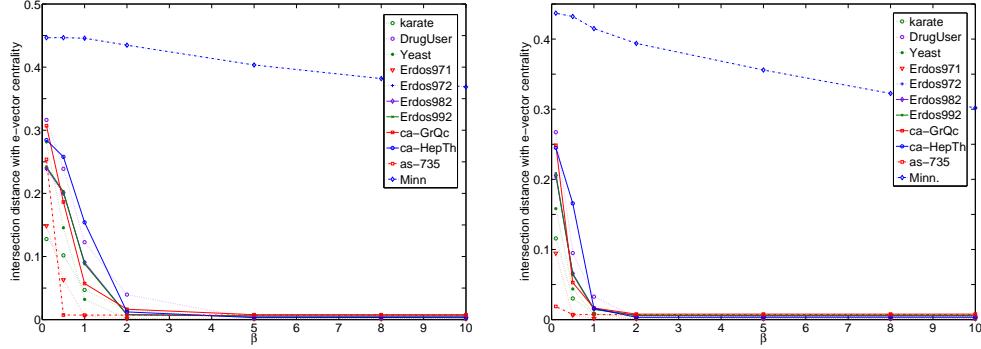


Fig. 5.2: The intersection distances between eigenvector centrality and the exponential subgraph centrality (left) or total communicability (right) rankings of the nodes in the networks in Table 5.1.

sum based) total communicability rankings for low values of β . As β increases, the intersection distances increase, then level off. The rankings of nodes in networks with a very large (relative) spectral gap, such as the karate, Erdos971 and as-735 networks, stabilize extremely quickly, as expected. The one exception to the stabilization is the intersection distances between the degree centrality rankings and exponential subgraph centrality (and total communicability rankings) of nodes in the Minnesota road network. This is expected, as the tiny (< 0.001) spectral gap for the Minnesota road network means that it will take longer for the exponential subgraph centrality (and total communicability) rankings to stabilize as β increases. It is worth noting that the Minnesota road network is quite different from the other ones: it is (nearly) planar, has large diameter and a much more regular degree distribution.

The rankings produced by exponential subgraph centrality and total communicability are compared to those produced by eigenvector centrality for various values of β in Figure 5.2. When β is small, the intersection distances are large but, as β increases, the intersection distances quickly decrease. When $\beta = 2$, they are essentially zero for all but one of the networks examined. Again, the outlier is the Minnesota road network. For this network, the intersection distances between the exponential based centrality rankings and the eigenvector centrality rankings still decrease as β increases, but at a much slower rate than for the other networks. This is also expected, in view of the very small spectral gap. Again, the rankings of the nodes in the karate, Erdos971, and as-735 networks, which have very large relative spectral gaps, stabilize extremely quickly.

In Figure 5.3, the intersection distances between the rankings produced by exponential subgraph centrality and total communicability are compared for successive choices of β . Overall, these intersection distances are quite low (the highest is 0.25 and occurs for the exponential subgraph centrality rankings of the as-735 network when β increases from 0.1 to 0.5). For all the networks examined, the largest intersection distances between successive choices of β occur as β increases to two. For higher values of β , the intersection distance drops, which corresponds to the fact that the rankings are converging to those produced by eigenvector centrality. In general, there is less change in the rankings produced by the total communicability scores for successive values of β than for the rankings produced by the exponential subgraph

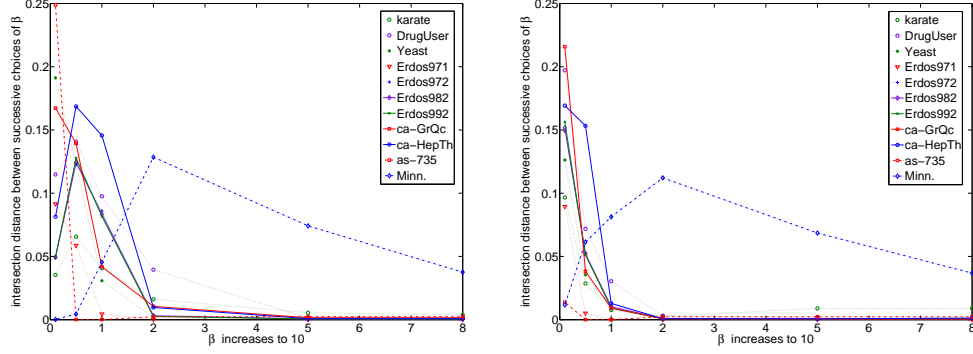


Fig. 5.3: The intersection distances between the exponential subgraph centrality (left) or total communicability (right) rankings produced by successive choices of β . Each line corresponds to a network in Table 5.1.

centrality scores.

If the intersection distances are restricted to the top 10 nodes, they are even lower. For the karate, Erdos992, and ca-GrQc networks, the intersection distance for the top 10 nodes between successive choices of β is always less than 0.1. For the DrugUser, Yeast, Erdos971, Erdos982, and ca-HepTh networks, the intersection distances are somewhat higher for low values of β , but by the time $\beta = 2$, they are all equal to 0 as the rankings have converged to those produced by the eigenvector centrality. For the Erdos972 network, this occurs slightly more slowly. The intersection distances between the rankings of the top 10 nodes produced by $\beta = 2$ and $\beta = 5$ are 0.033 and for all subsequent choices of β are 0. In the case of the Minnesota Road network, the intersection distances between the top 10 ranked nodes never stabilize to 0, as is expected. Detailed results and plots can be found in [41, Appendix B].

For the networks examined, when $\beta < 0.5$, the exponential subgraph centrality and total communicability rankings are very close to those produced by degree centrality. When $\beta \geq 2$, they are essentially identical to the rankings produced by eigenvector centrality. Thus, the most additional information about node rankings (i.e. information that is not contained in the degree or eigenvector centrality rankings) is obtained when $0.5 < \beta < 2$. This supports the intuition developed in Section 4 that moderate values of β should be used to gain the most benefit from the use of matrix exponential based centrality rankings.

5.2. Resolvent subgraph and Katz centrality. In this section, we investigate the effect of changes in α on the resolvent subgraph centrality and Katz centrality in the networks listed in 5.1, as well as the relationship of these centrality measures to degree and eigenvector centrality. We calculate the scores and node rankings produced by \mathbf{C}_d and \mathbf{C}_{ev} , as well as those produced by $\mathbf{RC}(\alpha)$ and $\mathbf{K}(\alpha)$ for various values of α . The values of α tested are given by $\alpha = 0.01 \cdot \frac{1}{\lambda_1}$, $0.05 \cdot \frac{1}{\lambda_1}$, $0.1 \cdot \frac{1}{\lambda_1}$, $0.25 \cdot \frac{1}{\lambda_1}$, $0.5 \cdot \frac{1}{\lambda_1}$, $0.75 \cdot \frac{1}{\lambda_1}$, $0.9 \cdot \frac{1}{\lambda_1}$, $0.95 \cdot \frac{1}{\lambda_1}$, and $0.99 \cdot \frac{1}{\lambda_1}$.

As in section 5.1, the rankings produced by degree centrality and eigenvector centrality were compared to those produced by resolvent-based centrality measures for all choices of α using the intersection distance method. The results are plotted in Figs. 5.4 and 5.5. The rankings produced by successive choices of α are also compared

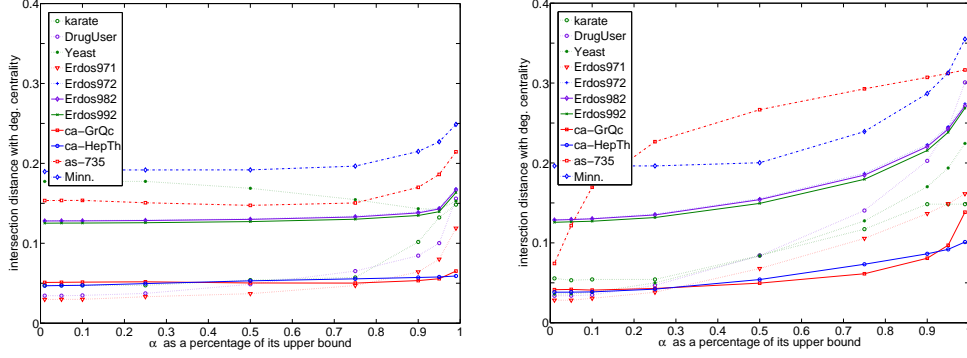


Fig. 5.4: The intersection distances between degree centrality and the resolvent subgraph centrality (left) or Katz centrality (right) rankings of the nodes in the networks in Table 5.1.

and these intersection distances are plotted in Fig. 5.6.

Fig. 5.4 shows the intersection distances between the degree centrality rankings and those produced by resolvent subgraph centrality or Katz centrality for the values of α tested. When α is small, the intersection distances between the resolvent-based centrality rankings and the degree centrality rankings are low. For $\alpha = 0.01 \cdot \frac{1}{\lambda_1}$, the largest intersection distance between the degree centrality rankings and the resolvent subgraph centrality rankings is slightly less than 0.2 (for the Minnesota road network). The largest intersection distance between the degree centrality rankings and the Katz centrality rankings is also slightly less than 0.2 (again, for the Minnesota road network). The relatively large intersection distances for the node rankings on the Minnesota road network when $\alpha = 0.01 \cdot \frac{1}{\lambda_1}$ is due to the fact that both the degree centrality and the resolvent subgraph (or Katz) centrality scores for the nodes are very close. Thus, small changes in the score values can lead to large changes in the rankings. As α increases towards $\frac{1}{\lambda_1}$, the intersection distances increase. This increase is more rapid for the Katz centrality rankings than for the resolvent subgraph centrality rankings.

In Fig. 5.5, the resolvent subgraph centrality and Katz centrality rankings for various values of α are compared to the eigenvector centrality rankings on the networks described in Table 5.1. For small values of α , the intersection distances tend to be large. As α increases, the intersection distances decrease for both resolvent subgraph centrality and Katz centrality on all of the networks examined. This decrease is faster for the (row sum based) Katz centrality rankings than for the (diagonal based) resolvent subgraph centrality rankings. The network with the highest intersection distances between the eigenvector centrality rankings and those based on the matrix resolvent, and slowest decrease of these intersection distances as α increases, is the Minnesota road network. As was the case when matrix exponential based scores were examined, this is expected due to this network's small spectral gap. The node rankings in networks with large relative spectral gaps (karate, Erdos971, as-735) converge to the eigenvector centrality rankings most quickly.

The intersection distance between the resolvent subgraph and Katz centrality rankings produced by successive choices of α are plotted in Fig. 5.6. All of these intersection distances are extremely small (the largest is < 0.08), indicating that the

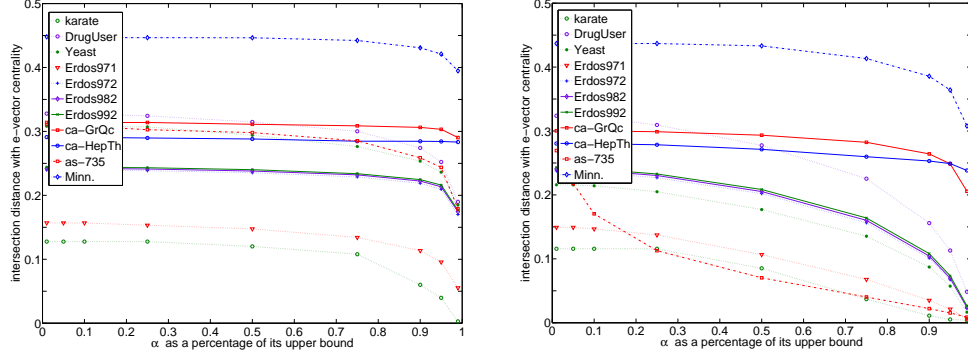


Fig. 5.5: The intersection distances between eigenvector centrality and the resolvent subgraph centrality (left) or Katz centrality (right) rankings of the nodes in the networks in Table 5.1.

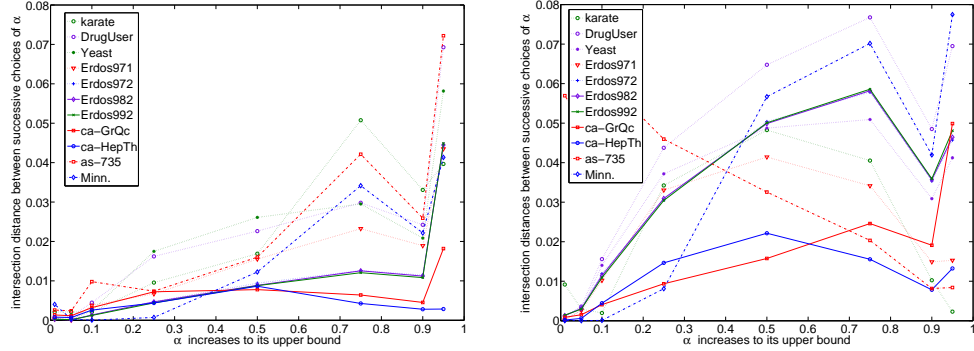


Fig. 5.6: The intersection distances between resolvent subgraph centrality (left) or Katz centrality (right) rankings produced by successive choices of α . Each line corresponds to a network in Table 5.1.

rankings do not change much as α increases. However, as α increases, the rankings corresponding to successive values of α tend to be slightly less similar to each other. The exception to this is the Katz centrality rankings for the as-735 network which become more similar as α increases.

Again, if the analysis is restricted to the top 10 nodes, the intersection distances between the rankings produced by successive choices of α are very small. For the karate, Erdos971, Erdos982, Erdos992, ca-GrQc, and Minnesota road networks, the intersection distances between the top 10 ranked nodes for successive choices of α are always less than or equal to 0.1 and often equal to zero. For the ca-HepTh network, the top 10 ranked nodes are exactly the same for all choices of α . For the DrugUser, Yeast, and Erdos972 networks, they are always less than 0.2. Detailed results can be found in [41].

For the eleven networks examined, the resolvent subgraph and Katz centrality rankings tend to be close to the degree centrality rankings when $\alpha < 0.5 \cdot \frac{1}{\lambda_1}$. It is interesting to note that as α increases, these rankings stay close to the degree

centrality rankings until α is approximately one half of its upper bound. Additionally, the resolvent based rankings are close to the eigenvector centrality rankings when $\alpha > 0.9 \cdot \frac{1}{\lambda_1}$. Thus, the most information is gained by using resolvent based centrality measures when $0.5 \cdot \frac{1}{\lambda_1} \leq \alpha \leq 0.9 \cdot \frac{1}{\lambda_1}$. This supports the intuition from section 4 that moderate values of α provide the most additional information about node ranking beyond that provided by degree and eigenvector centrality.

It is worth noting that similar conclusions have been obtained for the choice of the damping parameter α used in the PageRank algorithm; see [9, 10].

6. Analysis of centrality measures for directed networks. In section 3, we examined the relationship between the rankings produced by centrality measures based on the matrix exponential and resolvent with those produced by degree and eigenvector centrality on undirected networks. However, the measurement of node “importance” becomes more complicated for directed networks. In addition to the difficulties associated with ranking nodes in undirected networks (which centrality measure to use, how various centrality measures are related, etc.), the fact that A is no longer symmetric means that all adjacency matrix based centrality measures can be applied to either A or A^T . In terms of degree centrality, nodes now have both in- and out-degrees and, in terms of eigenvector centrality, A now has both a dominant left and a dominant right eigenvector.

The application of centrality measures to A and A^T correspond to two different types of node importance in directed networks. Since edges can only be traversed in one direction, in terms of information flow there is a difference between the ability of a node to spread information and its ability to gather information. In [40], these different abilities were captured by assigning each node a hub and authority score. In [28], the two aspects of information spread are referred to as broadcast and receive centrality. Broadcast centralities measure the ability of nodes in a network to broadcast information along directed walks. Here, we will examine broadcast centralities based on the row sums of $f(A)$ where f is the matrix exponential or resolvent. Receive centralities measure the ability of nodes in a network to receive information along directed walks. We will examine receive centralities based on the column sums of $f(A)$ (which correspond to the row sums of $f(A^T)$), where f is, again, the matrix exponential or resolvent (or any other suitable matrix function [32]).

We do not consider broadcast and receive centralities based on the diagonal entries of matrix functions, due to the fact that the diagonal entries of $f(A)$ and $f(A^T)$ are the same. Thus, these centralities measures cannot distinguish between the two types of node “importance” in a directed network (but see [5] for an extension of subgraph centrality to digraphs).

The relationship between the broadcast and receive total communicabilities with the in- and out-degrees and left and right eigenvectors of A are described in the following theorem:

THEOREM 6.1. *Let $G = (V, E)$ be a strongly connected, directed network with adjacency matrix A . Let $TC_i^b(\beta) = [e^{\beta A} \mathbf{1}]_i$ be the broadcast total communicability of node i and $\mathbf{TC}^b(\beta)$ be the corresponding vector of broadcast total communicabilities. Furthermore, let $TC_i^r(\beta) = [e^{\beta A^T} \mathbf{1}]_i$ be the receive total communicability of node i and $\mathbf{TC}^r(\beta)$ be the corresponding vector of receive total communicabilities. Then,*

- (i) *as $\beta \rightarrow 0+$, the rankings produced by $\mathbf{TC}^b(\beta)$ converge to those produced by the out-degrees of the nodes in the network;*

- (ii) as $\beta \rightarrow 0+$, the rankings produced by $\mathbf{TC}^r(\beta)$ converge to those produced by the in-degrees of the nodes in the network;
- (iii) as $\beta \rightarrow \infty$, the rankings produced by $\mathbf{TC}^b(\beta)$ converge to those produced by \mathbf{x}_1 , where \mathbf{x}_1 is the dominant right eigenvector;
- (iv) as $\beta \rightarrow \infty$, the rankings produced by $\mathbf{TC}^T(\beta)$ converge to those produced by \mathbf{y}_1 , where \mathbf{y}_1 is the dominant left eigenvector.

Proof. The proofs of (i) and (ii) follow the same procedure as the proof of Theorem 3.1 applied to A and A^T , respectively. To prove (iii), we first assume for clarity of exposition that A is diagonalizable: $A = X\Lambda X^{-1}$. Consider the expansion of $TC_i^b(\beta)$ in terms of X and Λ :

$$\begin{aligned} TC_i^b(\beta) &= [e^{\beta A} \mathbf{1}]_i = [X e^{\beta \Lambda} X^{-1} \mathbf{1}]_i = \sum_{k=1}^n e^{\beta \lambda_k} \mathbf{x}_k(i) (\mathbf{y}_k^* \mathbf{1}) \\ &= e^{\beta \lambda_1} \mathbf{x}_1(i) (\mathbf{y}_1^T \mathbf{1}) + \sum_{k=2}^n e^{\beta \lambda_k} \mathbf{x}_k(i) (\mathbf{y}_k^* \mathbf{1}), \end{aligned}$$

where we have used the fact that \mathbf{y}_1 (like x_1) is necessarily real. Let $\psi^b(\beta) = \frac{1}{e^{\beta \lambda_1} (\mathbf{y}_1^T \mathbf{1})} \mathbf{TC}^b(\beta)$; note that $\mathbf{y}_1^T \mathbf{1} > 0$ by the Perron-Frobenius Theorem. Similarly to the proof of (i), the rankings produced by $\psi^b(\beta)$ are equivalent to those produced by $\mathbf{TC}^b(\beta)$, since the scores for each node have all been scaled by the same constant. Now,

$$\psi_i^b(\beta) = \mathbf{x}_1(i) + \sum_{k=2}^n \frac{e^{\beta(\lambda_k - \lambda_1)}}{\mathbf{y}_1^T \mathbf{1}} \mathbf{x}_k(i) (\mathbf{y}_k^* \mathbf{1}).$$

As $\beta \rightarrow \infty$, $\psi_i^b(\beta) \rightarrow \mathbf{x}_1(i)$. Note that by the Perron-Frobenius Theorem, $\mathbf{x}_1 > 0$, so the rankings produced by $\mathbf{x}_1(i)$ are all real and positive.

If A is not diagonalizable, then from [44, Sec. 7.9] we easily obtain the expression

$$TC_i(\beta) = \sum_{k=1}^s \sum_{j=0}^{l_k-1} \frac{\beta^j e^{\beta \lambda_k}}{j!} [(A - \lambda_k I)^j G_k \mathbf{1}]_i$$

where s is the number of distinct eigenvalues of A , l_k is the geometric multiplicity of the k th distinct eigenvalue, and G_k is the oblique projector onto $\mathcal{N}((A - \lambda_k I)^{l_k})$ along $\mathcal{R}((A - \lambda_k I)^{l_k})$. Due to the fact that λ_1 is simple by the Perron-Frobenius theorem, this becomes:

$$TC_i(\beta) = e^{\beta \lambda_1} \mathbf{x}_1(i) (\mathbf{y}_1^T \mathbf{1}) + \sum_{k=2}^s \sum_{j=0}^{l_k-1} \frac{\beta^j e^{\beta \lambda_k}}{j!} [(A - \lambda_k I)^j G_k \mathbf{1}]_i.$$

Again, let $\psi^b(\beta) = \frac{1}{e^{\beta \lambda_1} (\mathbf{y}_1^T \mathbf{1})} \mathbf{TC}^b(\beta)$. The rankings produced by $\psi^b(\beta)$ will be the same as those produced by $\mathbf{TC}^b(\beta)$. Now,

$$\psi_i^b(\beta) = \mathbf{x}_1(i) + \sum_{k=2}^s \sum_{j=0}^{l_k-1} \frac{\beta^j e^{\beta(\lambda_k - \lambda_1)}}{j! (\mathbf{y}_1^T \mathbf{1})} [(A - \lambda_k I)^j G_k \mathbf{1}]_i, \quad (6.1)$$

Table 7.1: Basic data on the largest strongly connected component of the real-world directed networks examined.

Graph	n	nnz	λ_1	λ_2
Gleich/wb-cs-Stanford	2759	13895	35.618	12.201
SNAP/wiki-Vote	1300	39456	45.145	27.573

which converges to $\mathbf{x}_1(i)$ as $\beta \rightarrow \infty$. \square

An analogous theorem holds for the broadcast and receive Katz centralities. The proof is similar and is therefore omitted.

Looking at the expression (6.1), we note that in the case of directed networks the speed at which the centrality scores $\psi_i^b(\beta)$ stabilize as $\beta \rightarrow \infty$ does not depend solely on the distance between λ_1 and the rest of the spectrum, but also on the Jordan structure of A ; if large Jordan blocks are present, the convergence of the rankings may be slowed down. A similar observation holds for the Katz centrality rankings. Of course, information on the Jordan structure of A is usually not readily available.

7. Numerical experiments on directed networks. In this section, we examine the relationship between the exponential and resolvent-based broadcast centrality measures with the out-degrees and the dominant right eigenvectors of two real world directed networks. A similar analysis can be done on the relationship between the receive centrality measures and the in-degrees and dominant left eigenvectors. For the experiments we use two networks from the University of Florida Sparse Matrix Collection [22]. As was done in section 5, the rankings are compared using the intersection distance method. The first network we examine is wb-cs-Stanford, a network of hyperlinks between the Stanford CS webpages in 2001. It is in the Gleich group of the UF collection. The second network is the wiki-Vote network, which is a network of who votes for whom in elections for Wikipedia editors to become administrators. It is in the SNAP group of the UF collection.

Since the results in section 6 only hold for strongly connected networks with irreducible adjacency matrices, our experiments were performed on the largest strongly connected component of the above networks. Basic data on these strongly connected components can be found in Table 7.1. In both of the networks examined, the two largest eigenvalues of the largest strongly connected component are real. However, this is not always the case. Both networks are simple.

7.1. Total communicability. As in section 5.1, we examine the effect of changing β on the broadcast total communicability rankings of nodes in the networks, as well as their relation to the rankings obtained using the out-degrees and dominant right eigenvectors of the networks. The measures were calculated for the networks described in Table 7.1. To examine the sensitivity of the broadcast total communicability rankings, we calculate the scores and rankings for various choices of β . The values of β tested are: 0.1, 0.5, 1, 2, 5, 8 and 10.

The broadcast rankings produced by total communicability for all choices of β were compared to those produced by the out-degree rankings and the rankings produced by \mathbf{x}_1 using the intersection distance method as described in Section 5. Plots of the intersection distances for the rankings produced by various choices of β with those produced by the out-degrees and right dominant eigenvector can be found in Figs. 7.1

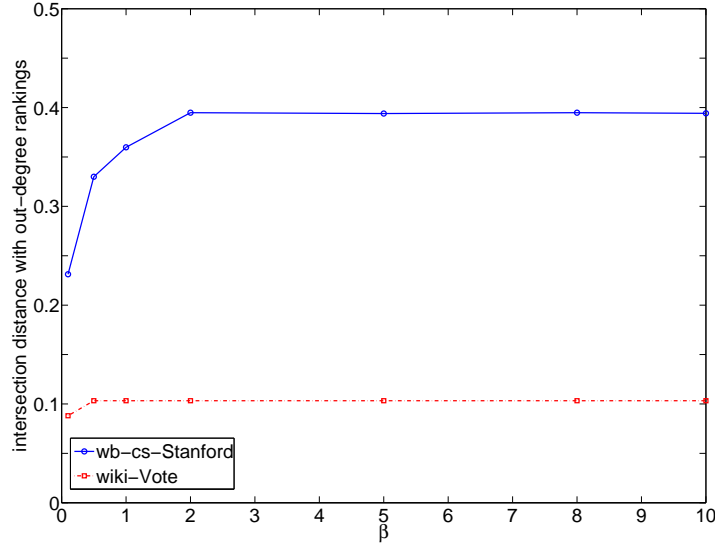


Fig. 7.1: The intersection distances between the out-degree rankings and the broadcast total communicability rankings of the nodes in the networks in Table 7.1.

and 7.2. The intersection distances for rankings produced by successive choices of β can be found in Figure 7.3.

In Fig. 7.1, the intersection distances between the rankings produced by broadcast total communicability are compared to those produced by the out-degrees of nodes in the network. As β approaches 0, the intersection distances decrease for both networks. As β increases to 10, the intersection distances initially increase, then stabilize as the rankings converge to those produced by \mathbf{x}_1 .

The intersection distances between the rankings produced by broadcast total communicability are compared to those produced by \mathbf{x}_1 in Figure 7.2. For both networks, the intersection distances quickly decrease as β increases. In the wiki-Vote network, the intersection distances between the compared rankings are 0 by the time $\beta = 0.5$. For the wb-cs-Stanford network, by the time β has reached five, the intersection distances between the broadcast total communicability rankings and those produced by $\mathbf{x}_1(i)$ have decreased to about 0.04. The rankings then stabilize at this intersection distance. This is due to a group of nodes that have nearly identical total communicability scores.

In Fig. 7.3, the intersection distances between the broadcast total communicability rankings for successive choices of β are plotted. These intersection distances are slightly lower than those observed in the undirected case, with a maximum of approximately 0.14, which occurs in the wb-cs-Stanford network when β increases from 0.01 to 0.05. By the time $\beta = 0.5$, the rankings on the wiki-Vote network have stabilized and all subsequent intersection distances are 0. For both the broadcast total communicability rankings on the wb-cs-Stanford network, the intersection distances decrease (non-monotonically) as β increases until they stabilize at approximately 0.02.

When this analysis is restricted to the top 10 nodes, the intersection distances are extremely small. For the wb-cs-Stanford network, the largest intersection distance

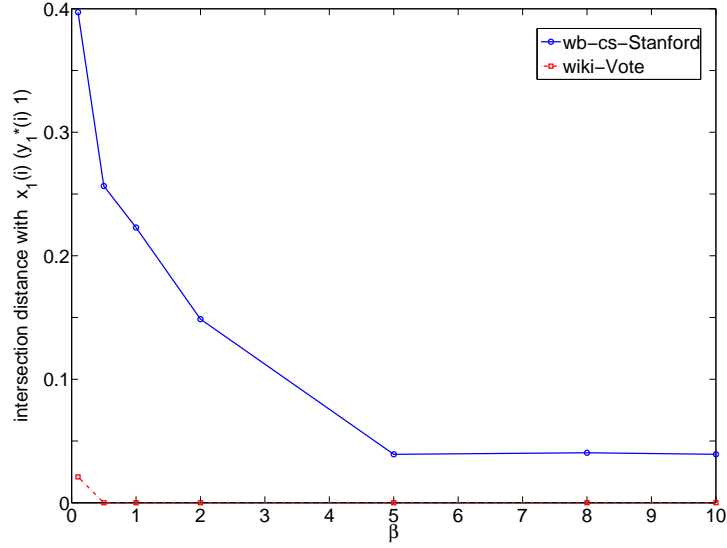


Fig. 7.2: The intersection distances between the rankings produced by \mathbf{x}_1 and the broadcast total communicability rankings of the nodes in the networks in Table 7.1.

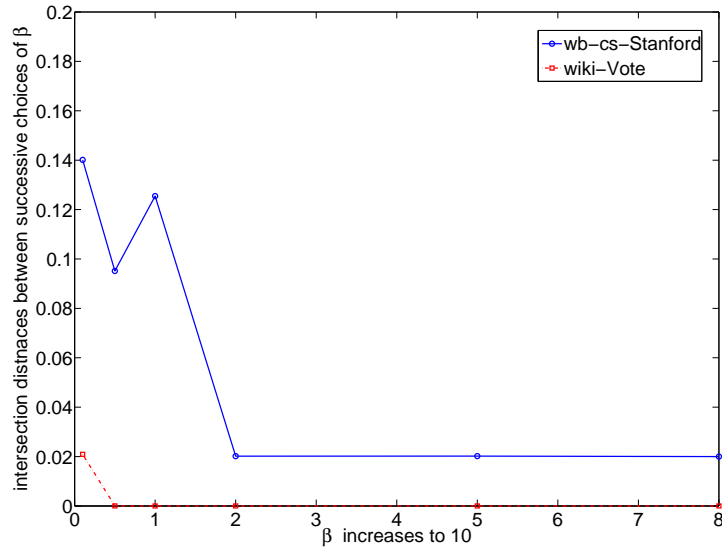


Fig. 7.3: The intersection distances between the broadcast total communicability rankings produced by successive choices of β . Each line corresponds to a network in Table 7.1.

between the top 10 ranked nodes for successive choices of β is 0.11 (when β increases from 0.1 to 0.5). For the wiki-Vote network, the intersection distance between the top 10 total communicability scores is 0.01 when β increases from 0.1 to 0.5, and zero

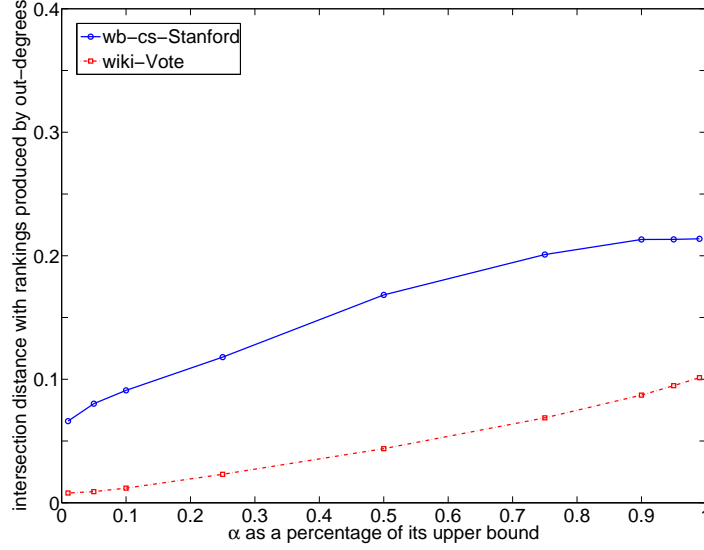


Fig. 7.4: The intersection distances between the rankings produced by the out-degrees and the broadcast Katz centrality rankings of the nodes in the networks in Table 7.1.

otherwise; see [41, Appendix B] for detailed results and plots.

The differences between the out-degree rankings and the broadcast total communicability rankings are greatest when $\beta \geq 0.5$. The differences between the left and right eigenvector based rankings and the broadcast rankings are greatest when $\beta < 2$ (although in the case of the wiki-Vote network, they have converged by the time $\beta = 0.5$). Thus, like in the case of the undirected networks, moderate values of β give the most additional ranking information beyond that provided by the out-degrees and the left and right eigenvalues.

7.2. Katz centrality. In this section, we investigate the effect of changes in α on the broadcast Katz centrality rankings of nodes in the networks listed in Table 7.1 and relationship of these centrality measures to the rankings produced by the out-degrees and the dominant right eigenvectors of the network. We calculate the scores and node rankings produced by $\mathbf{K}^b(\alpha)$ for various values of α . The values of α tested are given by $\alpha = 0.01 \cdot \frac{1}{\lambda_1}$, $0.05 \cdot \frac{1}{\lambda_1}$, $0.1 \cdot \frac{1}{\lambda_1}$, $0.25 \cdot \frac{1}{\lambda_1}$, $0.5 \cdot \frac{1}{\lambda_1}$, $0.75 \cdot \frac{1}{\lambda_1}$, $0.9 \cdot \frac{1}{\lambda_1}$, $0.95 \cdot \frac{1}{\lambda_1}$, and $0.99 \cdot \frac{1}{\lambda_1}$.

The rankings produced by the out-degrees and the dominant right eigenvectors were compared to those produced by Katz centrality for all choices of α using the intersection distance method, as was done in Section 7.1. The results are plotted in Figs. 7.4 and 7.5.

As α increases from $0.01 \cdot \frac{1}{\lambda_1}$ to $0.99 \cdot \frac{1}{\lambda_1}$, the intersection distances between the scores produced by the broadcast Katz centralities and the out-degrees increase. When α is small, the broadcast Katz centrality rankings are very close to those produced by the out-degrees (low intersection distances). On the wb-cs-Stanford network, when $\alpha = 0.01 \cdot \frac{1}{\lambda_1}$, the intersection distance between the two rankings is approximately 0.06. On the wiki-Vote network, it is approximately 0.01. As α increases, the intersection distances also increase. By the time $\alpha = 0.99 \cdot \frac{1}{\lambda_1}$, the intersection dis-

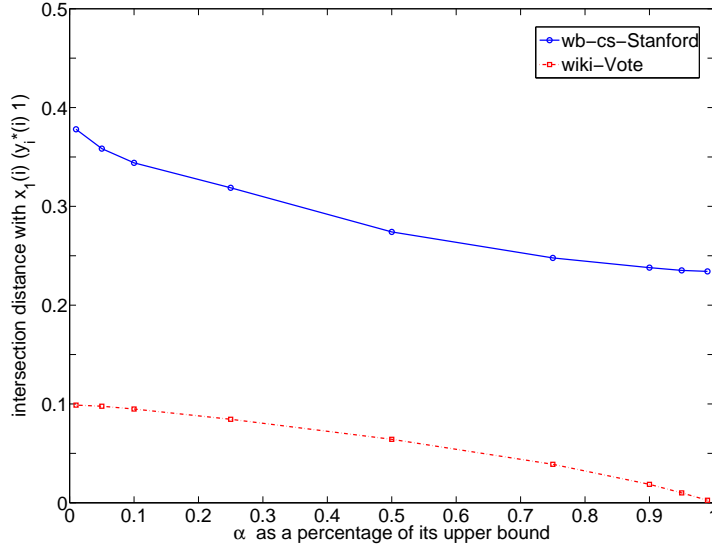


Fig. 7.5: The intersection distances between the rankings produced by \mathbf{x}_1 and the broadcast Katz centrality rankings of the nodes in the networks in Table 7.1.

tance between the two sets of node rankings on the wb-cs-Stanford network is above 0.2 and on the wiki-Vote network it is approximately 0.1.

In Fig. 7.5, the rankings produced by broadcast Katz centrality are compared to those produced by \mathbf{x}_1 . Overall, The intersection distances between the two sets of rankings are lower on the wiki-Vote network than they are on the wb-cs-Stanford network. As α increases from $0.01 \cdot \frac{1}{\lambda_1}$ to $0.99 \cdot \frac{1}{\lambda_1}$, the intersection distances between the two sets of rankings on the wiki-Vote network decrease from 0.1 to essentially 0. On the wb-cs-Stanford network, they decrease from approximately 0.47 to 0.24.

The intersection distances between the rankings produced by the broadcast Katz centralities for successive values of α are plotted in Figure 7.6. As was the case in the undirected networks examined, these rankings are more stable in regards to the choice of α than the total communicability rankings were in regards to the choice of β . Here, the maximum intersection distance is less than 0.1. When only the top 10 ranked nodes are similar, the intersection distances have a maximum of 0.06 (on the wb-cs-Stanford network when α increases from $0.25 \cdot \frac{1}{\lambda_1}$ to $0.5 \cdot \frac{1}{\lambda_1}$). For both networks, the intersection distances between the rankings on the top 10 nodes for successive choices of α are quite small (the maximum is 0.18 and the majority are < 0.1).

The broadcast Katz centrality rankings are only far from those produced by the out-degrees when $\alpha \geq 0.5 \cdot \frac{1}{\lambda_1}$. They are farthest from those produced by the dominant right eigenvector of A when $\alpha < 0.9 \cdot \frac{1}{\lambda_1}$. Thus, as was seen in the case of undirected networks, the most additional information is gained when moderate values of α , $0.5 \cdot \frac{1}{\lambda_1} \leq \alpha < 0.9 \cdot \frac{1}{\lambda_1}$, are used to calculate the matrix resolvent based centrality scores.

8. Conclusions. We have analyzed the relationship between centrality measures based on the diagonal entries and row sums of the matrix exponential and resolvent of the adjacency matrix with degree and eigenvector centrality. We have shown that the parameters α (in the case of resolvent-based centrality rankings) and β (in the case of

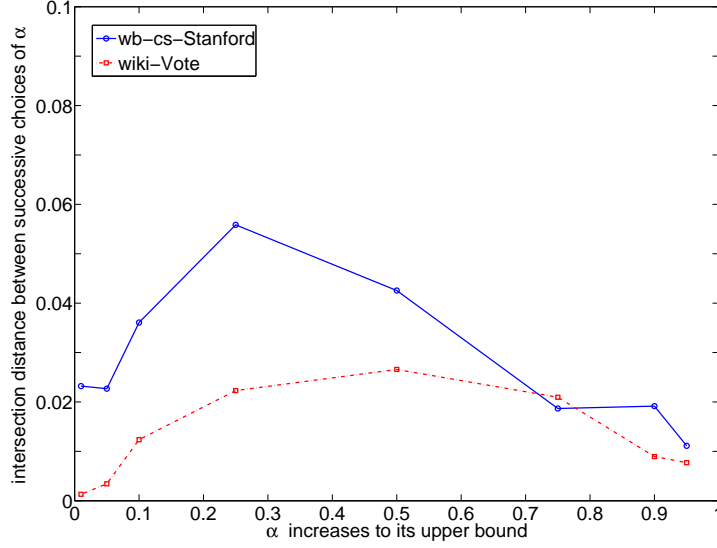


Fig. 7.6: The intersection distances between the broadcast Katz centrality rankings produced by successive choices of α . Each line corresponds to a network in Table 7.1.

exponential-based centrality rankings) act as tuneable parameters between the degree centrality rankings and eigenvector centrality rankings. That is, when α and β tend toward their lower bounds, the rankings produced by exponential subgraph centrality, total communicability, resolvent subgraph centrality, and Katz centrality converge to those produced by degree centrality. Similarly, as α and β tend to their upper bounds, these rankings converge to those produced by eigenvector centrality. We also demonstrated a similar relationship in directed networks between the broadcast and receive centralities (based on the row and column sums of the matrix exponential and resolvent) and the rankings produced by the in- and out-degrees and the left and right eigenvectors. These relationships help explain the observed correlations between the degree and eigenvector centrality rankings on many real-world complex networks, particularly those exhibiting a large spectral gap. They also explain why the rankings tend to be most stable (with respect to variations in α and β) precisely near the extreme values of the parameters. This is at first sight surprising, given that as the parameters approach their upper bounds, the centrality scores and their derivatives diverge, indicating extreme sensitivity.

Additionally, we have presented the results from experiments that used a large variety of choices of α and β to compute rankings on several directed and undirected real-world networks. We computed the intersection distances of these rankings with those produced by degree and eigenvector centrality and with each other. Here, we found that, as expected, the rankings were close to degree centrality when α and β were close to their lower bounds and were close to eigenvector centrality when α and β were close to their upper bounds. The rankings were the least similar to both degree and eigenvector centrality when $0.5 \cdot \frac{1}{\lambda_1} \leq \alpha \leq 0.9 \cdot \frac{1}{\lambda_1}$ and $0.5 < \beta < 2$.

Our results also allow us to provide guidelines for the choice of the parameters α and β in order to produce rankings that are the most different from the degree

and eigenvector centrality rankings and, therefore, most useful in terms of adding more information to the analysis of a complex network. Of course, the larger the spectral gap, the smaller the range of parameter values leading to rankings exhibiting a noticeable difference from those obtained from degree and/or eigenvector centrality. Since degree and eigenvector centrality are considerably less expensive to compute compared to subgraph centrality, for networks with large spectral gap it may be difficult to justify the use of the more expensive centrality measures discussed in this paper.

Similar results were found to hold for directed networks for the two types of centrality (receive/broadcast, or hub/authority) that are typically of interest in this case. We pointed out that in this case the difference in rankings may be affected by the Jordan structure of the adjacency matrix, in addition to the spectral gap.

Finally, in this paper we have mostly avoided discussing computational aspects of the ranking methods under consideration, focusing instead on gaining theoretical understanding of the relationship among the various methods. For recent progress in the computational aspects of subgraph centrality and related methods, see [4, 6, 13, 35, 36].

Acknowledgments. The authors would like to thank Ernesto Estrada for useful discussions.

REFERENCES

- [1] A.-L. BARABASI, *Linked: The New Science of Networks*, Perseus, Cambridge, 2002.
- [2] A. BARRAT, M. BARTHELEMY, AND A. VESPIGNANI, *Dynamical Processes on Complex Networks*, Cambridge University Press, Cambridge, UK, 2008.
- [3] M. BENZI, *A note on walk entropies in graphs*, Linear Algebra Appl., published online, 2013. DOI: 10.1016/j.laa.2013.12.014.
- [4] M. BENZI AND P. BOITO, *Quadrature rule-based bounds for functions of adjacency matrices*, Linear Algebra Appl., 433 (2010), pp. 637–652.
- [5] M. BENZI, E. ESTRADA, AND C. KLYMKO, *Ranking hubs and authorities using matrix functions*, Lin. Algebra Appl., 438 (2013), pp. 2447–2474.
- [6] M. BENZI AND C. KLYMKO, *Total communicability as a centrality measure*, J. Complex Networks, 1(2) (2013), pp. 124–149.
- [7] S. BOCCALETTI, V. LATORA, Y. MORENO, M. CHAVEZ, AND D.-U. HWANG, *Complex networks: structure and dynamics*, Phys. Rep., 424 (2006), pp. 175–308.
- [8] P. BOLDI, *TotalRank: Ranking Without Damping*, in Proceedings of WWW’05, Special interest tracks and posters of the 14th international conference on World Wide Web, Association for Computing Machinery, New York, NY (2005), pp. 898–899.
- [9] P. BOLDI, M. SANTINI, AND S. VIGNA, *PageRank as a function of the damping factor*, in Proceedings of the 14th international conference on World Wide Web, Association for Computing Machinery, New York, NY (2005), pp. 557–566.
- [10] P. BOLDI, M. SANTINI, AND S. VIGNA, *PageRank: Functional dependencies*, ACM trans. Inf. Sys., 27(4) (2009), pp. 1–23.
- [11] P. BONACICH, *Power and centrality: a family of measures*, Amer. J. Sociology, 92 (1987), pp. 1170–1182.
- [12] P. BONACICH AND P. LLOYD, *Eigenvector-like measures of centrality for asymmetric relations*, Social Networks, 23 (2001), pp. 191–201.
- [13] F. BONCHI, P. ESFANDIAR, D. F. GLEICH, C. GREIF, AND L. V. S. LAKSHMANAN, *Fast matrix computations for pair-wise and column-wise commute times and Katz scores*, Internet Math., 8 (2012), pp. 73–112.
- [14] S. P. BORGATTI AND M. G. EVERETT, *A graph-theoretic perspective on centrality*, Social Networks, 28 (2006), pp. 466–484.
- [15] U. BRANDES, *On variants of shortest-path betweenness centrality and their generic computation*, Social Networks, 30 (2008), pp. 136–145.
- [16] U. BRANDES AND T. ERLEBACH, eds., *Network Analysis: Methodological Foundations*, Lecture Notes in Computer Science Vol. 3418, Springer, New York, 2005.

- [17] G. CALDARELLI, *Scale-free Networks*, Oxford University Press, Oxford, UK, 2007.
- [18] T. P. CHARTIER, E. KREUTZER, A. N. LANGVILLE, AND K. E. PEDINGS, *Sensitivity and stability of ranking vectors*, SIAM J. Sci. Comput., 33 (2011), pp. 1077–1102.
- [19] R. COHEN AND S. HAVLIN, *Complex Networks: Structure, Robustness and Function*, Cambridge University Press, Cambridge, UK, 2010.
- [20] P. CONSTANTINE AND D. GLEICH *Random Alpha PageRank*, Internet Math., 6 (2009), pp. 199–236.
- [21] D. CVETKOVIĆ, P. ROWLINSON, AND S. SIMIĆ, *Eigenspaces of Graphs*, Cambridge University Press, Cambridge, 1997.
- [22] T. DAVIS AND Y. HU, *The University of Florida Sparse Matrix Collection*, <http://www.cise.ufl.edu/research/sparse/matrices/>
- [23] R. DIESTEL, *Graph Theory*, Springer-Verlag, Berlin, 2000.
- [24] C. H. Q. DING, H. ZHA, X. HE, P. HUSBANDS, AND H. D. SIMON, *Link analysis: hubs and authorities on the World Wide Web*, SIAM Rev., 46 (2004), pp. 256–268.
- [25] S. N. DOROGVTSEV AND J. F. F. MENDES, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press, Oxford, 2003.
- [26] E. ESTRADA, *Spectral scaling and good expansion properties in complex networks*, Europhys. Lett., 73 (2006), pp. 649–655.
- [27] E. ESTRADA, *The Structure of Complex Networks*, Oxford University Press, Oxford, UK, 2011.
- [28] E. ESTRADA, *Communicability in temporal networks*, Phys. Rev. E 88 (2013), 042811.
- [29] E. ESTRADA, *About the discriminant power of the subgraph centrality and other centrality measures (Working paper)*, arXiv:1305.6836.
- [30] E. ESTRADA, J. A. DE LA PEÑA, AND N. HATANO, *Walk entropies in graphs*, Linear Algebra Appl., 443 (2014), pp. 235–244.
- [31] E. ESTRADA, N. HATANO, AND M. BENZI, *The physics of communicability in complex networks*, Phys. Rep., 514 (2012), pp. 89–119.
- [32] E. ESTRADA AND D. J. HIGHAM, *Network properties revealed through matrix functions*, SIAM Rev., 52 (2010), pp. 671–696.
- [33] E. ESTRADA AND J. A. RODRÍGUEZ-VELÁZQUEZ, *Subgraph centrality in complex networks*, Phys. Rev. E, 71 (2005), 056103.
- [34] R. FAGIN, R. KUMAR, AND D. SIVAKUMAR, *Comparing top k lists*, SIAM J. Discr. Math., 17 (2003), pp. 134–160.
- [35] C. FENU, D. MARTIN, L. REICHEL, AND G. RODRIGUEZ, *Network analysis via partial spectral factorization and Gauss quadrature*, SIAM J. Sci. Comput., 35(4) (2013), pp. A2046–A2068.
- [36] C. FENU, D. MARTIN, L. REICHEL, AND G. RODRIGUEZ, *Block Gauss and anti-Gauss quadrature with application to networks*, SIAM J. Matrix Anal. Appl., 34(4) (2013), pp. 1655–1684.
- [37] L. FREEMAN, *A set of measures of centrality based on betweenness*, Sociometry, 40 (1977), pp. 35–41.
- [38] P. GRINDROD AND D. HIGHAM, *A matrix iteration for dynamic network summaries*, SIAM Rev., 55 (2013), pp. 118–128.
- [39] L. KATZ, *A new status index derived from sociometric data analysis*, Psychometrika, 18 (1953), pp. 39–43.
- [40] J. KLEINBERG, *Authoritative sources in a hyperlinked environment*, J. ACM, 46 (1999), pp. 604–632.
- [41] C. F. KLYMKO, *Centrality and Communicability Measures in Complex Networks: Analysis and Algorithms*, PhD thesis, Emory University, Atlanta, GA (2013).
- [42] A. N. LANGVILLE AND C. D. MEYER, *Google’s PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ, 2006.
- [43] A. N. LANGVILLE AND C. D. MEYER, *Who’s #1? The Science of Rating and Ranking*, Princeton University Press, Princeton, NJ, 2012.
- [44] C. D. MEYER, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2000.
- [45] M. MIHAIL AND C. PAPADIMITRIOU, *On the eigenvalue power law*, in J. D. P. Rolim and S. Vadhan (Eds.), *Proceedings of RANDOM 2002*, Lectures Notes in Computer Science, 2483 (2002), pp. 254–262.
- [46] J. L. MORRISON, R. BREITLING, D. J. HIGHAM, AND D. R. GILBERT, *GeneRank: Using search engine technology for the analysis of microarray experiments*, BMC Bioinformatics, 6(233), 2005.
- [47] M. E. J. NEWMAN, *The structure and function of complex networks*, SIAM Rev., 45 (2003), pp. 167–256.
- [48] M. E. J. NEWMAN, *Networks: An Introduction*, Cambridge University Press, Cambridge, UK, 2010.

- [49] L. PAGE, S. BRIN, R. MOTWANI, AND T. WINOGRAD, *The PageRank Citation Ranking: Bringing Order to the Web*, Tech. Report, Stanford Digital Libraries Technology Project, 1998.
- [50] M. PUCK ROMBACH AND M. PORTER, *Discriminating power of centrality measures*, arXiv:1305.3146.
- [51] D. STEVANOVIĆ, *Comment on “Subgraph centrality in complex networks”*, Phys. Rev. E, 88 (2013), 026801.
- [52] W. W. ZACHARY, *An information flow model for conflict and fission in small groups*, J. Anthropol. Res., 33 (1977), pp. 452–473.