

***Matching Exponential-Based and
Resolvent-Based Centrality Measures***

Aprahamian, Mary and Higham,
Desmond J. and Higham, Nicholas J.

2015

MIMS EPrint: **2015.3**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

Matching Exponential-Based and Resolvent-Based Centrality Measures*

MARY APRAHAMIAN

School of Mathematics, The University of Manchester, Oxford Road, Manchester, M13 9PL
mary.aprahamian@manchester.ac.uk

DESMOND J. HIGHAM

Department of Mathematics and Statistics, The University of Strathclyde, Glasgow, G1 1XH
d.j.higham@strath.ac.uk

AND

NICHOLAS J. HIGHAM

School of Mathematics, The University of Manchester, Oxford Road, Manchester, M13 9PL
nick.higham@manchester.ac.uk

The relative importance of nodes in a network can be quantified via functions of the adjacency matrix. Two popular choices of function are the exponential, which is parameter-free, and the resolvent function, which yields the Katz centrality measure. Katz centrality can be the more computationally efficient, especially for large directed networks, and has the benefit of generalizing naturally to time-dependent network sequences, but it depends on a parameter. We give a prescription for selecting the Katz parameter based on the objective of matching the centralities of the exponential counterpart. For our new choice of parameter the resolvent can be very ill conditioned, but we argue that the centralities computed in floating point arithmetic can nevertheless reliably be used for ranking. Experiments on five real networks show that the new choice of Katz parameter leads to rankings of nodes that match those from the exponential centralities well in practice.

Keywords: Katz centrality; Katz parameter; adjacency matrix; matrix exponential; matrix resolvent; network analysis; inverse iteration; condition number.

2000 Math Subject Classification: 65F30, 05C50

1. Introduction

The term *centrality* refers to a real number associated with a node of a network that conveys information about its relative “importance.” Centrality measures came to prominence in social network analysis [46], but have proved to be extremely useful tools across network science [16], [35].

A popular way to define centrality is to quantify the ability of a node to initiate walks around the network, a concept that leads naturally to the use of matrix functions. Using standard notation, we let $A = (a_{ij})$ denote the adjacency matrix for an unweighted network of n nodes, so that $a_{ij} = 1$ if there is an edge from i to j and $a_{ij} = 0$ otherwise. It follows that the number of walks of length k from node i to node j is given by $(A^k)_{ij}$; see, for example, [9]. Resolvent-based centrality measures, first suggested by Katz [29], penalize long walks through multiplication by a fixed factor α for each edge used. This leads to a series of the form $\sum_{k=0}^{\infty} \alpha^k A^k$, where for $i \neq j$ the (i, j) element gives a weighted count of the number of walks of all lengths from i to j . This series converges to the resolvent $(I - \alpha A)^{-1}$ for any

*Version of January 7, 2015.

$\alpha \in [0, 1/\rho(A))$, where $\rho(A)$ denotes the spectral radius of A . The i th row sum of the resolvent therefore summarizes the ability of node i to initiate walks to all nodes in the network. Similarly, the (i, i) element of the resolvent gives a weighted count of *closed* walks, that is, walks that start and finish at node i , with a uniform unit shift. Since we are concerned with the comparative performance across nodes, this shift is not important.

A related centrality concept arises from the suggestion by Estrada and Rodríguez-Velázquez [18] to weight walks of length k by the factor $1/k!$, so that the resolvent is replaced by $\sum_{k=0}^{\infty} A^k/k!$, which is the matrix exponential function, e^A [23, Chap. 10]. They define the subgraph centrality of a node to be the weighted sum of all closed walks originating from it, which can be computed as the diagonal entry (i, i) of e^A . Some justification for this definition is given by Estrada, Hatano, and Benzi [17, Sec. III] using the metaphor of a network as a system of oscillators.

In this work we measure the importance of a node via a weighted sum of both the open and closed walks starting from it; that is, we use the *total subgraph communicability* of a node, introduced by Benzi and Klymko [7], as its measure of centrality. The associated exponential-based centrality measure of node i is thus given by the i th element of the vector

$$\mathbf{c}_e(A) = e^A \mathbf{1}, \quad (1.1)$$

where $\mathbf{1} = [1, 1, \dots, 1]^T$. Similarly the resolvent-based centrality of node i is the i th element of the vector

$$\mathbf{c}_\alpha(A) = (I - \alpha A)^{-1} \mathbf{1}. \quad (1.2)$$

We note that the combinatoric “weighted walk count” interpretation of the matrix resolvent and matrix exponential extends naturally to the case of nonnegative integer weights if we interpret a_{ij} as recording the number of distinct connections between node i to j . For example, in a road network, if there are two distinct roads connecting town A and town B and three distinct roads connecting town B and town C, then there are $2 \times 3 = 6$ distinct ways to get from town A to town C in two hops via town B. The adjacency matrix power A^k therefore continues to count walks in this generalized sense, and the centrality vectors \mathbf{c}_e in (1.1) and \mathbf{c}_α in (1.2) have a clear meaning. We also point out that in the case where A is nonsymmetric computing these centrality measures on the transpose, A^T , quantifies the propensity of nodes to receive, rather than broadcast, information.

The motivation for our work is that there currently seems to be no agreed mechanism for selecting the Katz parameter α , and, as we will show in Section 4, centrality rankings can be strongly dependent on this value. In order to derive and judge an approach for choosing α , we make the assumption that exponential-based total communicability is the “gold standard” and thereby seek to match this measure as closely as possible. Therefore we select α in (1.2) to match closely the centralities in (1.1).

In Section 2 we pursue this approach for selecting a Katz parameter α , both for directed and undirected networks, and propose a new choice of the parameter. In Section 2.2 we give an overview of some particular choices of α that have appeared in the literature. In Section 3 we show that our new choice of Katz parameter can lead to a very ill conditioned resolvent and explain why the ill conditioning is innocuous. Numerical experiments are presented in Section 4 that test the performance of the proposed new value of the Katz parameter for ranking nodes in real networks. In Section 5 we briefly explain why computing resolvent-based centrality measures may be more favorable than the exponential versions for very large and sparse networks and also for time-dependent networks. Concluding remarks are given in Section 6.

2. Katz parameter

The exponential-based centrality measure penalizes longer walks more heavily than the resolvent-based one; for a walk of length k the coefficient in the exponential series is $1/k!$, compared with α^k in the resolvent series. The exponential-based centrality has been found to yield meaningful results for some particular problems, for example those arising from biochemical applications [15]. Furthermore, in social networks and in other human interactions direct acquaintanceship is typically more important, which can be successfully exploited via the exponential-based centrality analysis [16, Chap. 19]. The resolvent measure has the advantage of being flexible, since α can be tuned according to the requirements of the specific problem. This, however, requires good knowledge of the network, which may not always be readily available to the person constructing the model. It is therefore desirable to have a prescription for a Katz parameter that closely matches the node rankings produced by the exponential measure. This will provide a computational alternative to the matrix exponential function for obtaining reliable node rankings.

2.1 A new Katz parameter

We propose a new method for selecting the Katz parameter that aims to minimize the norm of the difference between the centrality vectors \mathbf{c}_e in (1.1) and \mathbf{c}_α in (1.2). This approach naturally ensures that the centralities of the nodes with the highest scores are closely matched. Indeed, in many applications it is only the best ranked nodes that are of practical interest. We would therefore like to find α that solves

$$\min_{\alpha} \text{err}(\alpha) := \min_{\alpha} \|\mathbf{c}_e(A) - \mathbf{c}_\alpha(A)\|_2 \quad \text{subject to} \quad 0 \leq \alpha < 1/\rho(A), \quad (2.1)$$

where the 2-norm is defined by $\|x\|_2 = (x^T x)^{1/2}$. Initially we will make no assumptions about the network except that A is a diagonalizable matrix, so that $A = VDV^{-1}$, where $D = \text{diag}(\lambda_i)$ contains the eigenvalues of A and V is nonsingular. (In fact, our derivation can be modified to use the Jordan canonical form when A is not diagonalizable: our assumption mainly serves to simplify the notation.) Since the matrix A is nonnegative, the Perron–Frobenius theory [27, Thm. 8.4.4] applied to A^T tells us that $\rho(A)$ is an eigenvalue of A with an associated nonnegative left eigenvector y : $y^T A = \rho(A)y^T$. Without loss of generality we can take $\lambda_1 = \rho(A)$ and the first row of V^{-1} to be y^T . We have

$$\begin{aligned} \text{err}(\alpha)^2 &= \|V(e^D - (I - \alpha D)^{-1})V^{-1}\mathbf{1}\|_2^2 \\ &\leq \|V\|_2^2 \|(e^D - (I - \alpha D)^{-1})V^{-1}\mathbf{1}\|_2^2 \end{aligned} \quad (2.2)$$

$$= \|V\|_2^2 \sum_{i=1}^n \left| e^{\lambda_i} - \frac{1}{1 - \alpha \lambda_i} \right|^2 |w_i|^2, \quad (2.3)$$

where $w = V^{-1}\mathbf{1}$. Then

$$\min_{\alpha} \text{err}(\alpha)^2 \leq \text{err}(\alpha_{\min})^2 \leq \|V\|_2^2 \sum_{i=2}^n \left| e^{\lambda_i} - \frac{1}{1 - \alpha \lambda_i} \right|^2 |w_i|^2, \quad (2.4)$$

where α_{\min} is such that $(e^{\lambda_1} - 1/(1 - \alpha_{\min} \lambda_1))^2 w_1^2 = 0$. But $w_1 = y^T \mathbf{1} \neq 0$ as y is a nonzero vector with nonnegative entries, so

$$\alpha_{\min} = \frac{1 - e^{-\lambda_1}}{\lambda_1}. \quad (2.5)$$

The value of the upper bound (2.4) on the minimum is governed both by the distribution of the eigenvalues of A and the sums w_i of the elements of the left eigenvectors of A .

For the special case of normal adjacency matrices, i.e., ones that satisfy $A^T A = A A^T$ and hence are diagonalizable by orthogonal matrices—in particular, symmetric matrices, corresponding to undirected networks—we can take V orthogonal, and (2.2) and the second inequality in (2.4) are then equalities. Some classes of directed networks are known to have normal adjacency matrices. For example, (unweighted) “ring” networks are such that for $i = 1 : n - 1$ there is an edge from node i to node $i + 1$ and an edge from node n to node 1, and for these it is always true that $A^T A = A A^T = I$.

The upper bound (2.4) on $\min_{\alpha} \text{err}(\alpha)$ is attained for certain types of graphs. For example, for unweighted and undirected k -regular graphs, where each node has degree k , it is easy to see that $\mathbf{1}$ is always an eigenvector of the adjacency matrix [9, Chap. 3] and then from the orthogonality of the eigenvectors it follows that $w_i = 0$ for all $i \geq 2$ [13]. In general the upper bound (2.4) provides a good estimate for $\min_{\alpha} \text{err}(\alpha)$ either if A is such that there is a relatively big separation $|\lambda_1 - \text{Re}(\lambda_2)|$ between its two eigenvalues with largest real part, or if $|w_1|$ is significantly larger than $|w_i|$ for all $i > 1$. These cases are common in practice, as we see from the examples in Section 4.

Benzi and Klymko [6, Sec. 9] observe experimentally that for both undirected and directed networks the exponential and resolvent measures differ the most for values of the Katz parameter that satisfy $0 \leq \alpha \leq 0.9/\lambda_1$. Provided $\lambda_1 > \log 10 \approx 2.3026$, α_{\min} avoids this interval. We note that by [27, Thm. 8.1.22] λ_1 lies between the smallest and largest row sums of the adjacency matrix of the network, so in practice such a small value for λ_1 will rarely be observed.

2.2 Other Katz parameters

Many different choices for the Katz parameter in the resolvent-based centrality measure have appeared in the literature, some of them proving more popular than others. In his original paper Katz suggests that a value for α in the interval $[1/(2\lambda_1), 1/\lambda_1]$ should be suitable [29]. Some authors have in particular chosen the value

$$\alpha_{0.5} = \frac{1}{2\lambda_1}$$

to study similarity in texts [4, p. 4411]. This Katz parameter has also successfully been used in the context of supply chain management [10]. We will use this value in our comparison studies in Section 4.

Another favored choice for the Katz parameter is [7]

$$\alpha_{0.85} = \frac{0.85}{\lambda_1}.$$

It arises by analogy with the damping factor of Google’s PageRank sorting algorithm, usually set to 0.85 [31].

In many applications the induced node rankings have been found to be very strongly dependent on the choice of the Katz parameter, so either an α particular to the model has been computed [36] or rankings have been reported for many values of α [11]. In some cases, the Katz parameter has a particularly meaningful interpretation, such as in protein–protein interaction networks [48], where it is indicative of the balance between the influence of the neighbors and the difference in activity levels.

Since λ_1 is bounded by any subordinate norm of the adjacency matrix it is natural to propose a value for α that depends on such a norm. Foster et.al. [19] suggest

$$\alpha_{\text{deg}} = \frac{1}{\|A\|_{\infty} + 1},$$

where the subscript relates to the fact that for networks with undirected and unweighted edges the ∞ -norm is the largest node degree. The upper bound $\lambda_1 \leq \|A\|_\infty$ is known to be attained for several classes of networks. For k -regular (undirected and unweighted) graphs it is always true that $\lambda_1 = \|A\|_\infty = k - 1$. This includes complete graphs and rings.

In Section 4 we present a comparison between the node rankings obtained using the exponential-based centrality measure and its resolvent-based counterpart, computed with the Katz parameter taken as α_{\min} and the other options discussed in this section. But first we look more closely at the properties of α_{\min} .

3. Conditioning

The resolvent-based centralities are the solution of a linear system, so it is of interest to know the conditioning of the coefficient matrix $I - \alpha A$ of that linear system, since this will influence the accuracy of the solution obtained in floating point arithmetic. Upper and lower bounds on the 2-norm condition number $\kappa_2(I - \alpha_{\min}A) = \|I - \alpha_{\min}A\|_2 \|(I - \alpha_{\min}A)^{-1}\|_2$ are given in the next result.

LEMMA 3.1 Let A be nonnegative and let λ_1 be an eigenvalue such that $\lambda_1 = \rho(A)$. Let $\alpha_{\min} = (1 - e^{-\lambda_1})/\lambda_1$. (a) If A is diagonalizable, so that $A = VDV^{-1}$ with $D = \text{diag}(\lambda_i)$ and V nonsingular, then

$$\kappa_2(I - \alpha_{\min}A) \leq \kappa_2(V)^2(2e^{\lambda_1} - 1). \quad (3.1)$$

(b) If A has an eigenvalue with nonpositive real part then $\kappa_2(I - \alpha_{\min}A) \geq e^{\lambda_1}$.

Proof. We have

$$\kappa_2(I - \alpha_{\min}A) = \|V(I - \alpha_{\min}D)V^{-1}\|_2 \|V(I - \alpha_{\min}D)^{-1}V^{-1}\|_2 \leq \kappa_2(V)^2 \kappa_2(I - \alpha_{\min}D).$$

Now

$$\max_i |1 - \alpha_{\min}\lambda_i| \leq \max_i (1 + \alpha_{\min}|\lambda_i|) = 1 + \alpha_{\min}\lambda_1 = 1 + \left(\frac{1 - e^{-\lambda_1}}{\lambda_1}\right)\lambda_1 = 2 - e^{-\lambda_1}.$$

Also, $\min_i |1 - \alpha_{\min}\lambda_i| = |1 - \alpha_{\min}\lambda_1| = e^{-\lambda_1}$. Hence

$$\kappa_2(I - \alpha_{\min}D) \leq \frac{2 - e^{-\lambda_1}}{e^{-\lambda_1}} = 2e^{\lambda_1} - 1.$$

Finally, if λ_k has nonpositive real part then $\|I - \alpha_{\min}A\|_2 \geq |1 - \alpha_{\min}\lambda_k| \geq 1$, and $\|(I - \alpha_{\min}A)^{-1}\|_2 \geq |1 - \alpha_{\min}\lambda_1|^{-1} = e^{\lambda_1}$, which gives the lower bound. \square

The condition in part (b) of the lemma is often satisfied in practice; indeed it is satisfied for the adjacency matrices of all the networks used in the experiments of Section 4, and more generally it is satisfied for any nonnegative A with zero diagonal.

The bounds in Lemma 3.1 are a cause for concern because they suggest that $I - \alpha_{\min}A$ is potentially very ill conditioned when either $\lambda_1 \gg 1$ or V is ill conditioned, the latter case corresponding to A being highly nonnormal. It is certainly possible that $\lambda_1 \gg 1$; indeed λ_1 is as large as 94 in our test problems in Section 4. Therefore $I - \alpha_{\min}A$ can be extremely ill conditioned, and in floating point arithmetic we can expect the computed centrality vector to have a large relative error. However, the ill conditioning is innocuous, as we now explain.

When we solve the linear system $(I - \alpha_{\min}A)x = \mathbf{1}$, we are effectively carrying out inverse iteration according to $(A - \alpha_{\min}^{-1}I)x = -\alpha_{\min}^{-1}\mathbf{1}$, and for large λ_1 , $\alpha_{\min}^{-1} = \lambda_1/(1 - e^{-\lambda_1})$ is a very good approximation to the eigenvalue λ_1 of A . Standard theory of inverse iteration [28, Sec. 6.3], [37, Sec. 4.3], [39, Sec. 2] shows that the error in the computed x will be almost parallel to x , that is, the inaccuracy is concentrated in its length and not its direction. Inverse iteration theory therefore tells us that while the computed centrality vector may be inaccurate the relative sizes of the elements will be accurately determined. Since our interest in centralities is to assess the relative importance of nodes, we conclude that we can safely use α_{\min} in practice.

It is interesting to note that $I - \alpha_{\min}A$ is an M -matrix, since $\alpha_{\min} < 1/\lambda_1$ [8], but the above argument does not depend on any special properties of A .

4. Experiments with Ranking

In our experiments we compare node rankings obtained from centrality vectors computed using exponential- and resolvent-based measures. Our computations are done in IEEE double precision arithmetic, which has unit roundoff $u \approx 1.1 \times 10^{-16}$.

For the resolvent measure we use each of the four choices for the Katz parameter suggested in Section 2: α_{\min} , $\alpha_{0.5}$, $\alpha_{0.85}$, and α_{deg} . The first three depend on the largest eigenvalue λ_1 of A , which we compute using the MATLAB sparse eigensolver via the call `eigs(A, 1)`. Although we state the value of the relative error $\text{err}_{\text{rel}}(\alpha) := \|\mathbf{c}_e(A) - \mathbf{c}_\alpha(A)\|/\|\mathbf{c}_e(A)\|$ for every choice of α , the conclusions of our tests are based on correlation coefficients between the rankings arising from \mathbf{c}_e and \mathbf{c}_α . We compute three types of correlation coefficients: Kendall's τ [30], Spearman's ρ [43], and Pearson's r [38]; see [40, Chap. 14] for a summary of them. The first is a popular statistic used to measure the association between rankings of objects by counting the numbers of concordant and discordant pairs of elements. Similarly Spearman's ρ is a nonparametric statistic indicative of whether the relation between two sets of elements can be expressed as a monotonic function. Spearman's ρ is better suited to lists with repeated values, and hence equal ranks. We also report the values of Pearson's r statistic. It serves as a test for linear dependence which, while not of immediate interest when comparing rankings, can still provide some indication as to how the different centrality measures relate. All three statistics take real values in the interval $[-1, 1]$, where 1 indicates perfect agreement and -1 indicates perfect disagreement between the objects.

In practice, only the top ranked nodes may need to be identified and there are several ways of reflecting this in the reported correlation coefficients. One of them is to apply a weighting to the test statistics, so that a disagreement of the best ranked nodes results in a lower than usual correlation coefficient. For example, Langville and Meyer suggest a weighted version of Spearman's ρ [32]. Another alternative, also described in [32], is to tune the statistics to take into account that the compared lists are only partial. We will use the standard forms of the statistics to compute the correlation coefficients between the rankings obtained from full centrality vectors, and also from the top ranked $k\%$ of the nodes.

As a representative notation, we will use $\tau_{0.05}(\mathbf{c}_e(A), \mathbf{c}_{\alpha_{\min}}(A))$ to mean Kendall's correlation coefficient between the exponential- and resolvent-based rankings of the top 5% of the nodes of network A , obtained using the Katz parameter α_{\min} .

To check the reliability of the centralities computed in floating point arithmetic we compute the quantity

$$\text{err}_{\text{dir}} = \left(1 - \frac{\hat{x}^T x_q}{\|\hat{x}\|_2 \|x_q\|_2}\right)^{1/2} \equiv (1 - \cos \theta)^{1/2},$$

Table 1. Basic characteristics of test networks. “Sparsity” denotes the proportion of nonzeros.

Name	Nodes	Edges	Sparsity	Directed	Weighted	λ_1	λ_2
Karate	34	78	1.3×10^{-2}	No	No	6.7257	4.9771
p53	133	558	3.2×10^{-2}	Yes	No	5.4032	$2.0696 + 0.2998i$
ca-CondMat	108299	93497	1.6×10^{-5}	No	No	37.9541	30.6438
ca-AstroPh	133279	198110	2.2×10^{-5}	No	No	94.4415	75.5007
Strathclyde MUFC	148918	193032	8.7×10^{-6}	Yes	Yes	41.1511	34.2307

where θ is the angle between the computed \hat{x} and a reference vector x_q computed in quadruple precision. We compute x_q using the Advanpix Multiprecision Computing Toolbox for MATLAB [1], which has very efficient IEEE 754-2008-compliant quadruple precision arithmetic. We actually compute err_{dir} from the alternative formula

$$\text{err}_{\text{dir}} = \frac{1}{\sqrt{2}} \left\| \frac{\hat{x}}{\|\hat{x}\|_2} - \frac{x_q}{\|x_q\|_2} \right\|_2,$$

which is more accurately evaluated in floating point arithmetic. A value err_{dir} of order u indicates that the computed and reference solutions are parallel to working precision. For each network we also compute the 1-norm condition number $\kappa_1(I - \alpha A)$ for each α , or, for the three largest networks, an estimate of the condition number computed using the MATLAB function `condest`, which implements the algorithm of [25].

We will use four examples of real networks available in the literature and one which is new and consists of recorded communication on the social networking platform Twitter. Table 1 summarizes the basic features for each network, including the spectral radius λ_1 and the eigenvalue with next largest real part, λ_2 . Figures 1–4 show the sparsity patterns and/or eigenvalue distributions, as appropriate. Sparsity plots for networks ca-CondMat and ca-AstroPh are omitted as they lack a distinctive visual pattern, at least in the node orderings provided.

First we consider a rather well studied example, Zachary’s Karate Club [47]. The network is of dimension 34 and the value of its largest eigenvalue λ_1 is 6.7257. The values of the different Katz parameters and the respective correlation coefficients between $\mathbf{c}_\alpha(A)$ and $\mathbf{c}_e(A)$ are summarized in Table 2. We observed that all the choices for the Katz parameter agree on the best ranked 5% of the nodes. However this is not a very indicative since the network has only 34 elements, so we have shown instead how the parameters perform on the top 20% (7 out of the 34) of the nodes and all the nodes. All four choices for the Katz parameter yield node rankings that are positively correlated with the exponential result. For α_{min} , $\alpha_{0.5}$ and α_{deg} the correlation between the top ranked 20% of the nodes is stronger than between the full rankings. On the contrary $\alpha_{0.85}$ matches the full rankings better than the partial ones for Kendall’s τ and Pearson’s r . This observation emphasizes the sensitivity of node rankings to the exact choice of Katz parameter. For the Karate Club network the resolvent-based measure evaluated with α_{min} yields almost identical node rankings to its exponential-based counterpart. Figure 5 shows the dependence of both $\tau_1(\mathbf{c}_e(A), \mathbf{c}_\alpha(A))$ and $\tau_{0.20}(\mathbf{c}_e(A), \mathbf{c}_\alpha(A))$ on different values of the Katz parameter α .

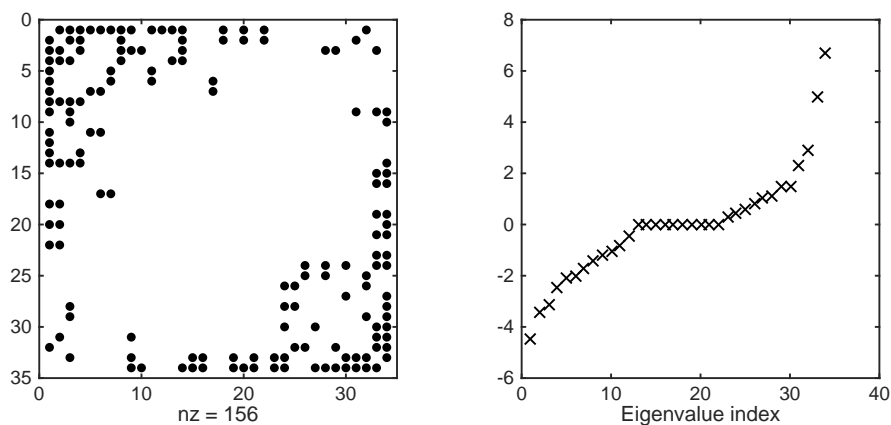


FIG. 1. Sparsity and eigenvalue distribution plots for Zachary's Karate Club network.

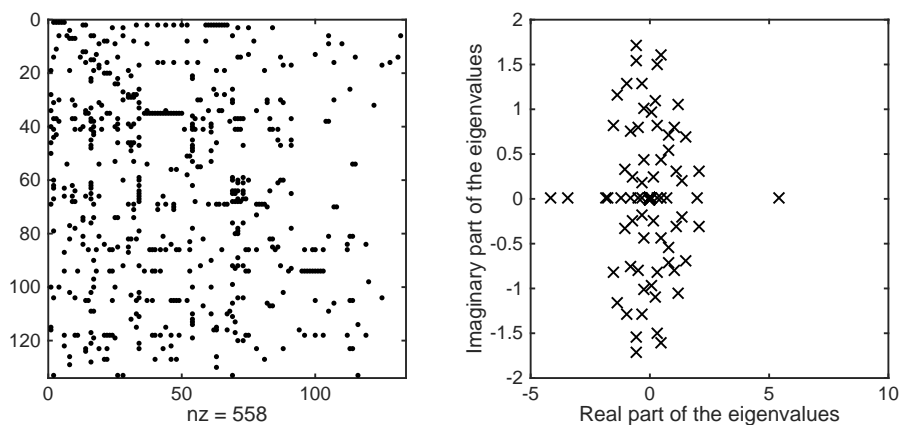


FIG. 2. Sparsity and eigenvalue distribution plots for the p53 network.

Table 2. Correlation coefficients between node rankings (all nodes and top 20%) obtained from exponential-based centrality and resolvent centralities computed using α_{\min} , $\alpha_{0.5}$, $\alpha_{0.85}$, and α_{deg} applied to Zachary's Karate Club network.

Katz parameter	$\tau_{0.20}$	τ_1	$\rho_{0.20}$	ρ_1	$r_{0.20}$	r_1	err_{rel}	$\kappa_1(I - \alpha A)$	err_{dir}
$\alpha_{\min} = 0.1485$	1.0000	0.9964	1.0000	0.9997	1.0000	0.9997	0.0059	5.5e3	1.3e-16
$\alpha_{0.5} = 0.0743$	0.4286	0.1052	0.6429	0.1419	0.1546	0.1474	0.9976	7.1e0	1.9e-16
$\alpha_{0.85} = 0.1264$	0.5238	0.5579	0.6786	0.5866	0.2619	0.5866	0.9920	3.8e1	1.8e-16
$\alpha_{\text{deg}} = 0.0556$	0.4286	0.0624	0.6429	0.0845	0.1489	0.0845	0.9981	4.5e0	1.5e-16

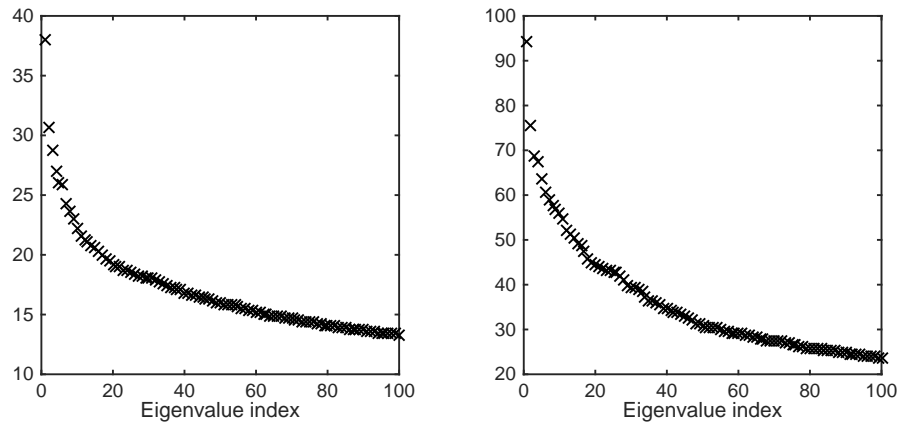


FIG. 3. Eigenvalue distribution (100 largest positive) plots for the ca-CondMat (left) and ca-AstroPh (right) networks.

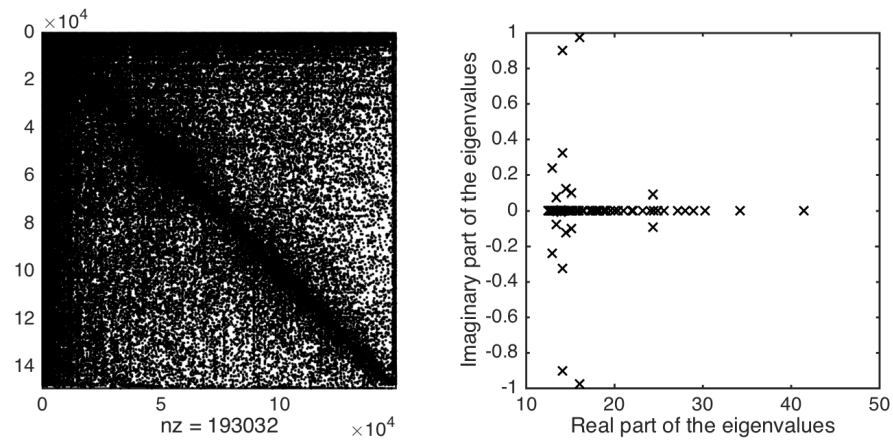


FIG. 4. Sparsity and eigenvalue distribution (100 with largest real part) plots for the Strathclyde MUFC network.

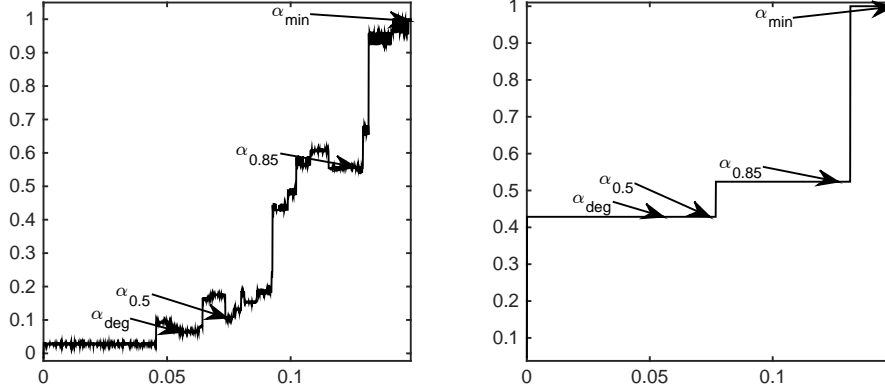


FIG. 5. Kendall correlation coefficients between node rankings obtained from $\mathbf{c}_e(A)$ and $\mathbf{c}_\alpha(A)$ for different α for Zachary's karate network with all nodes (left) and top 20% of nodes (right).

Table 3. Correlation coefficients between node rankings (all nodes and top 10%) obtained from exponential-based centrality and resolvent centralities computed using α_{\min} , $\alpha_{0.5}$, $\alpha_{0.85}$, and α_{deg} applied to the p53 network.

Katz parameter	$\tau_{0.10}$	τ_1	$\rho_{0.10}$	ρ_1	$r_{0.10}$	r_1	err _{rel}	$\kappa_1(I - \alpha A)$	err _{dir}
$\alpha_{\min} = 0.1842$	0.4066	0.3830	0.3978	0.4353	0.4225	0.3665	0.0215	4.2e3	1.8e-16
$\alpha_{0.5} = 0.0925$	0.0769	0.0980	0.1033	0.1608	0.1188	0.1462	0.9924	1.4e1	1.4e-16
$\alpha_{0.85} = 0.1573$	0.2088	0.2796	0.2044	0.3645	0.2131	0.3645	0.9713	9.8e1	1.9e-16
$\alpha_{\text{deg}} = 0.0435$	0.0110	0.1107	-0.0110	0.1333	-0.0606	0.1333	0.9955	4.3e0	9.5e-17

Next we consider a network consisting of 133 nodes arising from recorded levels of the oncogene p53 [45]. The network has 558 directed unweighted edges and an edge from node i to node j exists if i expresses above its usual level while j expresses below its usual level. The network is part of the NESSIE collection of networks [44]. Correlation coefficients between the resolvent- and exponential-based measures are summarized in Table 3 and their variation with α can be seen in Figure 6. For both the best 10% of the nodes (top 14 out of the total 133 nodes) and all nodes, α_{\min} performs better than the other available options for the Katz parameter. The parameter based on maximum node degree, α_{deg} , produces partial ranking negatively correlated to the exponential-based one.

We next consider two undirected and unweighted networks, ca-AstroPh and ca-CondMat, which record research collaborations in the areas of condensed matter and astrophysics, respectively. Both networks are such that $a_{ij} = 1$ if and only if scholars i and j co-authored at least one publication. They are available from the Stanford Network Analysis Project (SNAP) [33] and are described by Leskovec

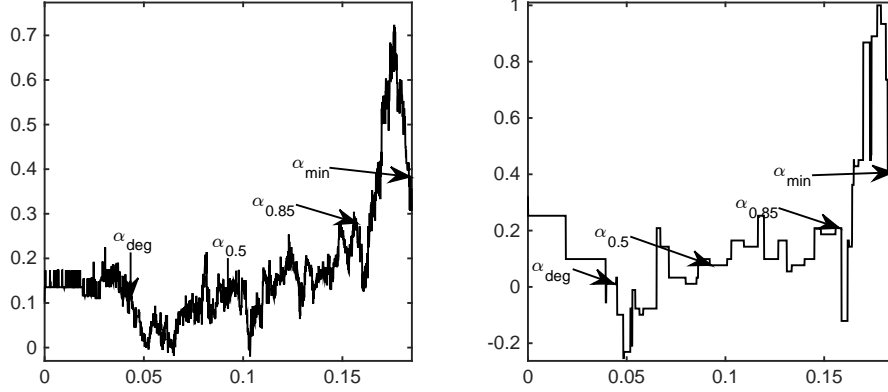


FIG. 6. Kendall correlation coefficients between node rankings obtained from $\mathbf{c}_e(A)$ and $\mathbf{c}_\alpha(A)$ for different α , for network p53 with all nodes (left) and top 10% of nodes (right).

Table 4. Correlation coefficients between node rankings (all nodes and top 1%) obtained from exponential-based centrality and resolvent centralities computed using α_{\min} , $\alpha_{0.5}$, $\alpha_{0.85}$, and α_{deg} applied to the ca-CondMat network.

Katz parameter	$\tau_{0.01}$	τ_1	$\rho_{0.01}$	ρ_1	$r_{0.01}$	r_1	err _{rel}	$\kappa_1(I - \alpha A)$	err _{dir}
$\alpha_{\min} = 0.0263$	0.7166	0.8587	0.7699	0.8925	0.7693	0.8925	0.8015	3.7e16	1.3e-15
$\alpha_{0.5} = 0.0132$	-0.0092	0.7383	-0.0176	0.7938	-0.0199	0.7938	1.0000	3.5e1	7.9e-13
$\alpha_{0.85} = 0.0224$	0.0154	0.7302	0.0316	0.7938	0.0292	0.7938	1.0000	2.6e2	5.3e-14
$\alpha_{\text{deg}} = 0.0036$	-0.0124	0.7376	-0.0194	0.7927	-0.0189	0.7927	1.0000	4.2e0	6.5e-13

et al. [34]. Correlation coefficients between the node rankings are presented in Tables 4 and 5. The dominance of α_{\min} is most appreciable when we compare only the top ranked 1% of the nodes. The alternative choices for the Katz parameter produce rankings which are weakly or even negatively correlated to the exponential ones. This is also true for other choices of the Katz parameter, as can be seen from Figures 7 and 8.

The final real-world network example arises from the online social networking service Twitter. It consists of 148918 nodes and 193032 directed edges. Unlike the previous examples, this network has edges with nonnegative integer weights. The weight of an edge from node i to node j specifies how many times Twitter account i sent a (meaningful) communication to Twitter account j on the newsworthy topic of Sir Alex Ferguson's retirement from his position as manager of Manchester United Football Club in May 2013. The individual time-stamped interactions are available via the Strathclyde MUFC Twitter Data Set at <http://www.mathstat.strath.ac.uk/outreach/twitter/mufc>, and have

Table 5. Correlation coefficients between node rankings (all nodes and top 1%) obtained from exponential-based centrality and resolvent centralities computed using α_{\min} , $\alpha_{0.5}$, $\alpha_{0.85}$, and α_{deg} applied to the ca-AstroPh network.

Katz parameter	$\tau_{0.01}$	τ_1	$\rho_{0.01}$	ρ_1	$r_{0.01}$	r_1	err_{rel}	$\kappa_1(I - \alpha A)$	err_{dir}
$\alpha_{\min} = 0.0106$	0.9434	0.9746	0.9551	0.9826	0.9548	0.9826	1.0000	$5.7e16$	$1.3e-15$
$\alpha_{0.5} = 0.0053$	0.0144	0.8176	0.0248	0.8574	0.0244	0.8574	1.0000	$2.3e1$	$8.6e-13$
$\alpha_{0.85} = 0.0090$	0.0115	0.8224	0.0134	0.8617	0.0135	0.8617	1.0000	$1.8e2$	$2.7e-13$
$\alpha_{\text{deg}} = 0.0020$	-0.0248	0.8212	-0.0372	0.8608	-0.0362	0.8608	1.0000	$4.4e0$	$7.6e-13$

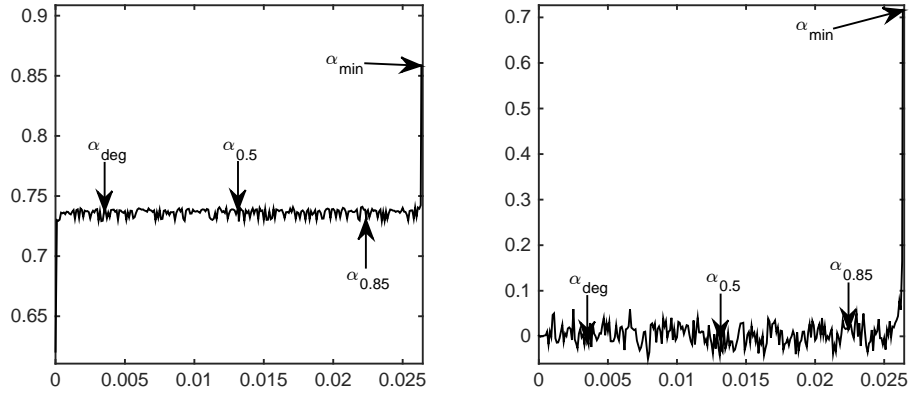


FIG. 7. Kendall correlation coefficients between node rankings obtained from $c_e(A)$ and $c_\alpha(A)$ for different α , for network ca-CondMat with all nodes (left) and top 1% of nodes (right).

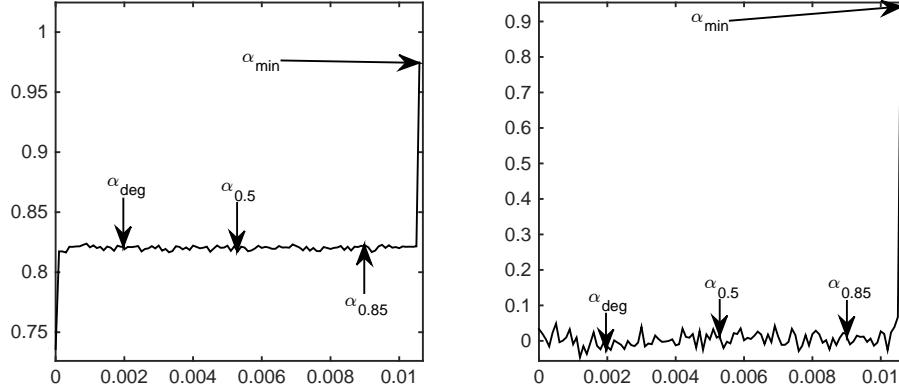


FIG. 8. Kendall correlation coefficients between node rankings obtained from $\mathbf{c}_e(A)$ and $\mathbf{c}_\alpha(A)$ for different α , for network ca-AstroPh with all nodes (left) and top 1% of nodes (right).

Table 6. Correlation coefficients between node rankings (all nodes and top 1%) obtained from exponential-based broadcaster centrality and resolvent broadcaster centralities computed using α_{\min} , $\alpha_{0.5}$, $\alpha_{0.85}$, and α_{deg} applied to the Strathclyde MUFC network.

Katz parameter	$\tau_{0.01}$	τ_1	$\rho_{0.01}$	ρ_1	$r_{0.01}$	r_1	err_{rel}	$\kappa_1(I - \alpha A)$	err_{dir}
$\alpha_{\min} = 0.0242$	0.7850	0.6939	0.8959	0.7558	0.8893	0.7558	0.9997	6.9e17	1.1e-14
$\alpha_{0.5} = 0.0121$	0.0257	0.4524	0.0287	0.5419	0.0150	0.5419	1.0000	1.5e3	1.5e-13
$\alpha_{0.85} = 0.0205$	0.0512	0.4523	0.0620	0.5423	0.0393	0.5423	1.0000	1.2e4	8.9e-13
$\alpha_{\text{deg}} = 0.0003$	0.0317	0.4467	0.0210	0.5401	0.0023	0.5401	1.0000	1.8e0	1.4e-13

also been studied in [22]. Our network was built by aggregating the tweets over the 12 hour period.

We consider ranking the nodes of the network and its transpose, where the top ranked nodes of Strathclyde MUFC and its transpose represent the best broadcasters and receivers, respectively, of information. Correlation coefficients between the rankings of the nodes of Strathclyde MUFC and its transpose obtained using resolvent- and exponential-based measures are presented in Tables 6 and 7, respectively. For the correlations obtained using all the values of the Katz parameter, except α_{\min} , we observe that the full rankings are matched significantly better than the partial ones. So $\alpha_{0.5}$, $\alpha_{0.85}$, and α_{deg} retrieve more successfully the position of the lower ranked nodes. This is usually of less practical interest, especially for the case of very large networks. Only α_{\min} is able to successfully match a greater part of the highly ranked nodes, both in their broadcaster and receiver capacities. Figures 9 and 10 illustrate the variation of the node ranking with respect to parameter α .

Table 7. Correlation coefficients between node rankings (all nodes and top 1%) obtained from exponential-based receiver centrality and resolvent receiver centralities computed using α_{\min} , $\alpha_{0.5}$, $\alpha_{0.85}$, and α_{deg} applied to the transpose of the Strathclyde MUFC network.

Katz parameter	$\tau_{0.01}$	τ_1	$\rho_{0.01}$	ρ_1	$r_{0.01}$	r_1	err _{rel}	$\kappa_1(I - \alpha A)$	err _{dir}
$\alpha_{\min} = 0.0242$	0.7188	0.7015	0.7735	0.7529	0.7714	0.7529	0.9985	6.5e18	2.6e-14
$\alpha_{0.5} = 0.0121$	0.0237	0.5263	0.0340	0.6287	0.0253	0.6287	1.0000	1.5e4	1.9e-13
$\alpha_{0.85} = 0.0205$	0.0409	0.5405	0.0617	0.6336	0.0563	0.6336	1.0000	1.1e5	5.8e-13
$\alpha_{\text{deg}} = 0.0001$	0.0741	0.5365	0.1407	0.6328	0.0991	0.6328	1.0000	1.6e1	2.6e-14

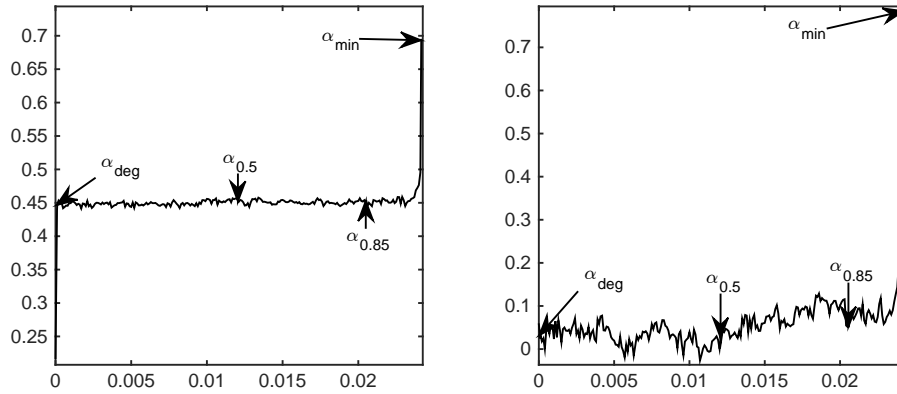


FIG. 9. Kendall correlation coefficients between node rankings obtained from $c_e(A)$ and $c_\alpha(A)$ for different α for network Strathclyde MUFC, with all nodes (left) and top 1% of nodes (right).

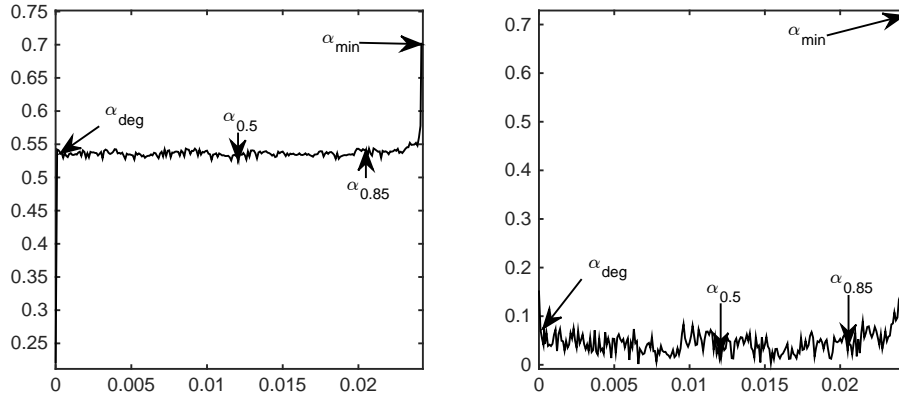


FIG. 10. Kendall correlation coefficients between node rankings obtained from $\mathbf{c}_e(A)$ and $\mathbf{c}_\alpha(A)$ for different α for transpose of network Strathclyde MUFC, with all nodes (left) and top 1% of nodes (right).

Finally, we note that for every network the values of err_{dir} are all less than 10^{-12} . Furthermore, α_{min} , which by its definition tends to lead to more ill conditioned systems than the other choices of α , produced values of err_{dir} sometimes the smallest over all choice of α and never the largest. Our experiments therefore support the conclusions drawn from the analysis of inverse iteration in Section 3 that ill conditioning does not vitiate the rankings obtained.

5. Computational considerations

We now discuss some aspects of the computation of the exponential- and resolvent-based measures, and consider potential advantages and challenges associated with a few different methods.

First, note that to compute the exponential-based centrality scores $\mathbf{c}_e(A) = e^A \mathbf{1}$ we do not require the full matrix exponential: just the action of this matrix on a vector. This opens up the possibility of using algorithms based on matrix–vector products involving A (and optionally A^T). For our numerical examples in Section 4 we used Al-Mohy and N. J. Higham’s `expmv` algorithm for computing the action of the matrix exponential [3], which takes advantage of the nonnegativity of A in its norm estimation phase. It is based on scaling the matrix and applying truncated Taylor series approximations. There is a wide range of alternative algorithms based on Krylov subspace projection techniques [2], [26], [42]. A comparison of available algorithms for computing the action of the matrix exponential on a vector, along with a new implementation of a method based on Leja interpolation, is presented by Caliari et al. [12]. Available software is surveyed by N. J. Higham and Deadman [24].

For the resolvent-based measure the computation of α_{min} requires computation of λ_1 . This can be done, for example, with the power method or the Arnoldi method. These and other methods (see [5] for a broad overview) are able to exploit one or more of the properties of sparsity, nonnegativity, and symmetry associated with adjacency matrices. There has also been interest in approximating λ_1 for directed or undirected, unweighted networks [41].

The resolvent-based centrality vector satisfies the linear system $(I - \alpha A)\mathbf{x} = \mathbf{1}$, which can be solved using direct solvers [14]. Iterative methods are not likely to be successful due to the potential ill condi-

tioning of the coefficient matrix. For undirected networks the linear system is symmetric, which can of course be exploited. For directed networks usually both the receiver and broadcaster scores of the nodes are of interest. A matrix decomposition can be re-used to compute both $\mathbf{c}_\alpha(A)$ and $\mathbf{c}_\alpha(A^T)$.

There are also established resolvent-based methods for computing centrality scores for the nodes of temporally evolving networks [20], [21]. To see how resolvent-based centrality extends naturally to the case of a time-ordered network sequence, let $A^{[0]}, A^{[1]}, A^{[2]}, \dots, A^{[M]}$ represent nonnegative integer-valued adjacency matrices for a fixed set of nodes. So, over a discrete time sequence, $t_0 < t_1 < \dots < t_M$, the matrix $A^{[k]}$ records interactions at time t_k . For example, in the social interaction context, we may have $(A^{[k]})_{ij} = 1$ if individual i contacted individual j at least once in the time period $(t_{k-1}, t_k]$ and $(A^{[k]})_{ij} = 0$ otherwise. In this setting, there is a natural concept of *dynamic walks* through the network—such traversals may use whatever edges are available at one time and then continue at the next time point using the new edge list. The time-ordered product of resolvents

$$(I - \alpha A^{[0]})^{-1} (I - \alpha A^{[1]})^{-1} (I - \alpha A^{[2]})^{-1} \dots (I - \alpha A^{[M]})^{-1}$$

collects together weighted counts of such dynamic walks, where a walk using k edges is downweighted by α^k . This combinatorial interpretation relies on the index law $\alpha^m \times \alpha^n = \alpha^{m+n}$, which allows walks to be correctly pieced together across time points. By contrast, the product of matrix exponentials

$$e^{A^{[0]}} e^{A^{[1]}} e^{A^{[2]}} \dots e^{A^{[M]}}$$

does not allow this simple combinatorial interpretation, since $1/(m!) \times 1/(n!) \neq 1/((m+n)!)$ in general. Hence, from this dynamic walk perspective, the resolvent-based centrality is a more natural choice than the exponential alternative when we wish to extend to time-dependent interactions.

Finally, we present in Table 8 the time required to compute the node rankings of the larger real-world networks introduced in Section 4, using the exponential-based measure $\mathbf{c}_e(A)$ and the resolvent-based measure $\mathbf{c}_\alpha(A)$ with the value of the Katz parameter α_{\min} . The tests were performed in MATLAB 2014b on a MacBook Pro 2.6 GHz Intel Core i7 computer. The total communicability $\mathbf{c}_e(A)$ was computed using function `expmv` from <http://www.mathworks.com/matlabcentral/fileexchange/29576-matrix-exponential-times-a-vector>, which implements the algorithm of [3], and the Katz centrality $\mathbf{c}_\alpha(A)$ was computed using the MATLAB backslash algorithm for sparse matrices. The time required to compute the Katz centrality includes the time required to compute the parameter α_{\min} , for which we used the MATLAB eigensolver `eigs` to compute λ_1 . The Katz centrality is computed faster than the exponential-based centrality. The speed-up which can be obtained by using the resolvent instead of the matrix exponential can be even more significant for networks with heavily weighted edges, as `expmv` can be less efficient for large $\|A\|$. This can be seen in the time required to compute the node centralities of the Strathclyde MUFC network with the adjacency matrix multiplied by a factor of 10.

6. Conclusion

Our aim in this work was to find a value for the damping parameter α such that the centrality scores obtained from the Katz centrality vector $(I - \alpha A)^{-1} \mathbf{1}$ and the action of the exponential of the adjacency matrix $e^A \mathbf{1}$ are similar. By considering an upper bound on the distance between the two vectors we obtained a value for the Katz parameter $\alpha = (1 - e^{-\lambda_1})/\lambda_1$ that performs substantially better in our tests than alternatives that have been previously proposed. The new Katz parameter leads to linear systems that are potentially much more ill conditioned than those corresponding to existing choices,

Table 8. Time (seconds) required to compute the centrality vectors $\mathbf{c}_e(A)$ and $\mathbf{c}_\alpha(A)$ using α_{\min} for the networks ca-CondMat, ca-AstroPh and Strathclyde MUFC from Sec. 4, as well as the Strathclyde MUFC network multiplied by a factor of 10.

	ca-CondMat	ca-AstroPh	Strathclyde MUFC	10× Strathclyde MUFC
\mathbf{c}_e	0.889925	3.246808	1.427309	5.858958
\mathbf{c}_α	0.519708	2.003820	0.911084	0.929030

but ill conditioning has essentially no effect on the suitability of the computed centralities for ranking. A natural extension of this idea is to automate the choice of Katz parameter in the case of temporally evolving networks [21].

Funding

The work of MA and NJH was supported by European Research Council Advanced Grant MATFUN (267526). The work of DJH was supported by a Royal Society/Wolfson Research Merit Award.

REFERENCES

- [1] Advanpix. *Multiprecision Computing Toolbox*. Advanpix, Tokyo. <http://www.advanpix.com>.
- [2] Afanasjew, M., Eiermann, M., Ernst, O. G. & Güttel, S. (2008) Implementation of a restarted Krylov subspace method for the evaluation of matrix functions. *Linear Algebra Appl.*, **429**, 2293–2314.
- [3] Al-Mohy, A. H. & Higham, N. J. (2011) Computing the action of the matrix exponential, with an application to exponential integrators. *SIAM J. Sci. Comput.*, **33**(2), 488–511.
- [4] Amancio, D. R., Oliveira, Jr., O. N. & da F. Costa, L. (2012) Structure-semantics interplay in complex networks and its effects on the predictability of similarity in texts. *Physica A*, **391**, 4406–4419.
- [5] Bai, Z., Demmel, J., Dongarra, J., Ruhe, A. & van der Vorst, H. (2000) *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*. Society for Industrial and Applied Mathematics.
- [6] Benzi, M. & Klymko, C. (2013a) A Matrix Analysis of Functional Centrality Measures. ArXiv preprint. Revised January 2015. <http://arxiv.org/abs/1312.6722>.
- [7] Benzi, M. & Klymko, C. (2013b) Total communicability as a centrality measure. *Journal of Complex Networks*, **1**(2), 124–149.
- [8] Berman, A. & Plemmons, R. J. (1994) *Nonnegative Matrices in the Mathematical Sciences*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA. Corrected republication, with supplement, of work first published in 1979 by Academic Press.
- [9] Biggs, N. (1993) *Algebraic Graph Theory*. Cambridge University Press, second edition.
- [10] Borgatti, S. P. & Li, X. (2009) On social network analysis in a supply chain context. *Journal of Supply Chain Management*, **45**(2), 5–22.
- [11] Bryan, N. J. & Wang, G. (2011) Musical influence network analysis and rank of sampled-based music. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, pages 329–334.
- [12] Caliari, M., Kandolf, P., Ostermann, A. & Rainer, S. (2014) Comparison of software for computing the action of the matrix exponential. *BIT*, **54**, 113–128.
- [13] Cvetković, D. M., Doob, M. & Sachs, H. (1995) *Spectra of Graphs—Theory and Applications*. Johann Ambrosius Barth Verlag, Heidelberg - Leipzig, third edition.

- [14] Davis, T. A. (2006) *Direct Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
- [15] Estrada, E. (2006) Virtual identification of essential proteins within the protein interaction network of yeast. *Proteomics*, **6**(1), 35–40.
- [16] Estrada, E. (2011) *The Structure of Complex Networks*. Oxford University Press.
- [17] Estrada, E., Hatano, N. & Benzi, M. (2012) The physics of communicability in complex networks. *Physics Reports*, **514** (3), 89–119.
- [18] Estrada, E. & Rodríguez-Velázquez, J. A. (2005) Subgraph centrality in complex networks. *Phys. Rev. E*, **71**(5), 056103.
- [19] Foster, K. C., Muth, S. Q., Potterat, J. J. & Rothenberg, R. B. (2001) A faster Katz status score algorithm. *Computational & Mathematical Organization Theory*, **7**(4), 275–285.
- [20] Grindrod, P. & Higham, D. J. (2013) A matrix iteration for dynamic network summaries. *SIAM Rev.*, **55**, 118–128.
- [21] Grindrod, P., Higham, D. J., Parsons, M. C. & Estrada, E. (2011) Communicability across evolving networks. *Physical Review E*, **83**, 046120.
- [22] Higham, D. J., Grindrod, P., Mantzaris, A. V., Otley, A. & Laffin, P. (2014) Anticipating activity in social media spikes. Technical Report 5, Department of Mathematics and Statistics, University of Strathclyde, UK.
- [23] Higham, N. J. (2008) *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
- [24] Higham, N. J. & Deadman, E. (2014) A catalogue of software for matrix functions. Version 1.0. MIMS EPrint 2014.8, Manchester Institute for Mathematical Sciences, The University of Manchester, UK.
- [25] Higham, N. J. & Tisseur, F. (2000) A block algorithm for matrix 1-norm estimation, with an application to 1-norm pseudospectra. *SIAM J. Matrix Anal. Appl.*, **21**(4), 1185–1201.
- [26] Hochbruck, M. & Lubich, C. (2006) On Krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.*, **34**(5), 1911–1925.
- [27] Horn, R. A. & Johnson, C. R. (2013) *Matrix Analysis*. Cambridge University Press, Cambridge, UK, second edition.
- [28] Ipsen, I. C. F. (1997) Computing an eigenvector with inverse iteration. *SIAM Rev.*, **39**(2), 254–291.
- [29] Katz, L. (1953) A new status index derived from sociometric data analysis. *Psychometrika*, **18**, 39–43.
- [30] Kendall, M. G. (1938) A new measure of rank correlation. *Biometrika*, **30** (1–2), 81–93.
- [31] Langville, A. N. & Meyer, C. D. (2006) *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, Princeton, NJ, USA.
- [32] Langville, A. N. & Meyer, C. D. (2012) *Who's #1?: The Science of Rating and Ranking*. Princeton University Press.
- [33] Leskovec, J. Stanford Network Analysis Project. <http://snap.stanford.edu>.
- [34] Leskovec, J., Kleinberg, J. & Faloutsos, C. (2007) Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, **1**(1), 2:1–2:41.
- [35] Newman, M. E. J. (2010) *Networks: An Introduction*. Cambridge University Press.
- [36] Park, J. & Newman, M. E. J. (2005) A network-based ranking system for US college football. *J. Stat. Mech.*, (2005) P10014.
- [37] Parlett, B. N. (1998) *The Symmetric Eigenvalue Problem*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA. Unabridged, amended version of book first published by Prentice-Hall in 1980.
- [38] Pearson, K. (1901) LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **2**(11), 559–572.
- [39] Peters, G. & Wilkinson, J. H. (1979) Inverse iteration, ill-conditioned equations and Newton's method. *SIAM Rev.*, **21**(3), 339–360.
- [40] Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2007) *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK, third edition.
- [41] Restrepo, J. G., Ott, E. & Hunt, B. R. (2007) Approximating the largest eigenvalue of network adjacency matrices. *Phys. Rev. E*, **76**, 056119.

- [42] Sidje, R. B. (1998) Expokit: a software package for computing matrix exponentials. *ACM Trans. Math. Software*, **24**, 130–156.
- [43] Spearman, C. (1904) The proof and measurement of association between two things. *Amer. J. Psychol.*, **15**(1), 72–101.
- [44] Taylor, A. & Higham, D. J. (2010) NESSIE: Network Example Source Supporting Innovative Experimentation. In Estrada, E., Fox, M., Higham, D. J. & Oppo, G.-L., editors, *Network Science: Complexity in Nature and Technology*, pages 85–106. Springer-Verlag.
- [45] Vass, J. K., Higham, D. J., Mudaliar, M. A. V., Mao, X. & Crowther, D. J. (2011) Discretization provides a conceptually simple tool to build expression networks. *PLoS ONE*, **6**(4), e18634.
- [46] Wasserman, S. & Faust, K. (1994) *Social Network Analysis: Methods and Applications*, volume 8. Cambridge University Press.
- [47] Zachary, W. W. (1977) An information flow model for conflict and fission in small groups. *J. Anthropol. Res.*, **33**, 452–473.
- [48] Zhao, J., Yang, T.-H., Huang, Y. & Holme, P. (2011) Ranking candidate disease genes from gene expression and protein interaction: A Katz-centrality based approach. *PLoS ONE*, **6**(9).