Received <day> <Month>, <year>; | Revised <day> <Month>, <year>; | Accepted <day> <Month>, <year>

DOI: xxx/xxxx

ORIGINAL PAPER

Matrix Functions in Network Analysis

Michele Benzi*[1] | Paola Boito[2]

[1]Classe di Scienze, Scuola Normale Superiore, Pisa, Italy
[2]Dipartimento di Matematica, Università di Pisa, Pisa, Italy

Correspondence
*Michele Benzi, Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy.
Email: michele.benzi@sns.it

Abstract

We review the recent use of functions of matrices in the analysis of graphs and networks, with special focus on centrality and communicability measures and diffusion processes. Both undirected and directed networks are considered, as well as dynamic (temporal) networks. Computational issues are also addressed.

KEYWORDS:
matrix functions, networks, centrality, communicability, connectivity, diffusion, numerical methods

1 | INTRODUCTION

Network science is a mature interdisciplinary field that has proved essential to the analysis of a multitude of phenomena in virtually every area of knowledge, from the natural and life sciences to the social sciences, from engineering to the humanities [15, 30, 32, 52, 111, 112]. It provides conceptual, qualitative tools for the theoretical understanding of general, complex interconnected systems, as well as quantitative methods of analysis for real-world networks arising in sociology, biology, engineering, information sciences, economics, finance, and many other fields.

Since the early days of social network analysis, the field has been making use of mathematical tools from graph theory and linear algebra, as well as probability and statistics [120, 95, 128]. In particular, powerful linear algebraic methods have been developed to define computable metrics for the structural analysis of both static and dynamic networks, such as centrality and communicability measures, and to tackle fundamental problems in community detection, clustering, and network robustness. Many of these network problems are formulated in matrix terms, most often involving either the adjacency matrix of the network or the graph Laplacian, and they lead to the solution of classical problems of numerical linear algebra: sparse linear systems, the computation of eigenvalues and eigenvectors, and the evaluation of functions of matrices. These problems can be of very large size and require sophisticated techniques for their efficient solution.

In this review paper we focus on applications to network problems that can be cast in terms of matrix functions. We are especially interested in the use of matrix functions to define node centrality and communicability measures. Functions of the network adjacency matrix are closely related to graph walks, which represent a basic tool for exploring network structure. Diffusive processes on networks (closely related to continuous and discrete random walks) also provide important structural information, with the graph Laplacian now being the central object. Due to space limitations we focus mainly use of matrix functions as a modeling tool: while we include a discussion of the numerical aspects, we refer mostly to the literature for information about algorithms, their implementation, and available software. For an earlier survey of matrix functions in network analysis we refer the reader to the 2010 paper by Estrada and Higham [66].

The paper is organized as follows. Section 2 is devoted to preliminaries and background. Node centrality measures are covered in section 3, while section 4 discusses communicability between nodes for static networks. The temporal case is discussed in section 5. In section 6 we describe other matrix functions that have found application in network analysis, including hyperbolic functions (used to study graph bipartivity) as well as functions arising in the modeling of dynamic processes on networks via differential equations. Numerical aspects are briefly discussed in section 7. Section 8 provides a few examples of centrality computations, and section 9 contains some concluding remarks and future directions.

2 | PRELIMINARIES AND BACKGROUND

In this section we review a few basic notions about graph theory and matrix analysis that will be used later. Good general references are Diestel [47], Horn and Johnson [93], and Higham [90].

2.1 | Graphs and matrices

A *directed graph* or *digraph*, G is defined by a set of n nodes (also called vertices) V and a set of edges $E = \{(i, j)|i, j \in V\}$. If $(i, j) \in E$ implies $(j, i) \in E$, then G is *undirected* and the edges are formed by unordered pairs of vertices.

In a directed graph, the *out-degree* of a vertex i, denoted by d_i^{out}, is the number of edges with i as the starting node, i.e., the number of edges in E of the form (i, k). The *in-degree* of node i is the number d_i^{in} of edges of the form (k, i). If G is undirected, then $d_i^{\text{out}} = d_i^{\text{in}} = d_i$, the *degree* of node i. An undirected graph G is *k-regular* if all the nodes have the same degree k.

A *walk* of length k in G is a list of nodes $i_1, i_2, \dots, i_k, i_{k+1}$ such that for all $1 \le j \le k$ there is a (directed) edge from i_j to i_{j+1}. A walk is *closed* when $i_1 = i_{k+1}$. A walk with no repeated nodes is called a *path*, and a *cycle* is a closed walk with no repeated nodes, except for the first and last one.

In a *weighted graph*, a weight $w_{ij} \in \mathbb{R}^+$ is assigned to each edge (i, j). If all the weights are taken equal to 1, the graph is *unweighted*. In this paper we will generally assume to work with unweighted graphs, although many results extend to the weighted case.

A *loop* in a graph is an edge from a node i to itself. A graph is *simple* if it has no loops, no multiple edges, and its edges are unweighted. An undirected graph is *connected* if there exists a path between every pair of nodes. A directed graph is *strongly connected* if there exists a directed path between every pair of nodes.

A graph is called *bipartite* if its set of nodes V is the disjoint union of two non-empty sets V_1 and V_2 such that each edge connects a node from V_1 and a node from V_2. Given a directed graph, we can define an associated undirected bipartite graph as follows. Take two copies V and V' of the set of nodes of the original graph, and define edges according to the following rule: a node $i \in V$ and a node $j' \in V'$ are connected if and only if there is an edge from node i to node j in the directed graph.

The *adjacency matrix* $A = (a_{ij})$ associated with an unweighted graph G is an $n \times n$ matrix such that

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

Recall that a square matrix A is *reducible* if there exists a permutation matrix P such that

$$PAP^T = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

with B and D square matrices. If no such P exists, A is said to be *irreducible*. It is well known that the adjacency matrix of a graph G is irreducible if and only if G is strongly connected.

Note that the in- and out-degrees of a node i can be computed as the i-th entries of the vectors $A^T\mathbf{1}$ and $A\mathbf{1}$, respectively, where $\mathbf{1}$ denotes the vector of all ones.

For an unweighted graph, the powers of the associated adjacency matrix A have an important property: each entry $[A^k]_{ij}$ counts the number of walks of length k from node i to node j. In particular, the diagonal entries $[A^k]_{ii}$ count the number of *closed* walks of length k passing through node i.

If G is simple and undirected, then its adjacency matrix A is symmetric and binary, with zeros on the main diagonal. In this case, the eigenvalues of A are real; we will label them in nondecreasing order as $\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_n$. The matrix A can be diagonalized as $A = Q\Lambda Q^T$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, the matrix $Q = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n]$ is orthogonal and $\mathbf{q}_i$ is an eigenvector associated with λ_i. If G is connected, the Perron-Frobenius theorem [93] implies that $\lambda_1 > \lambda_2$. Moreover, the *dominant eigenvector* $\mathbf{q}_1$ can be chosen to have positive entries: we write $\mathbf{q}_1 > \mathbf{0}$.

If G is a strongly connected digraph, its adjacency matrix A is irreducible. Let $\rho(A)$ be the spectral radius of A. Then, again by the Perron-Frobenius theorem, $\lambda_1 = \rho(A)$ is a simple eigenvalue of A and both the right and left eigenvectors associated with λ_1 can be chosen to be positive. If A is 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 $\lambda_1 \ge |\lambda_i|$ for $2 \le i \le n$. Let us write $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ and $(X^{-1})^T = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$; the left eigenvector associated with λ_i is $\mathbf{y}_i$ and the right eigenvector associated with λ_i is $\mathbf{x}_i$. If, on the other hand, A is not

diagonalizable, it can still be decomposed according to the Jordan canonical form:

$$A = XJX^{-1} = X \begin{bmatrix} \lambda_1 & \mathbf{0}^T \\ \mathbf{0} & \hat{J} \end{bmatrix} X^{-1}.$$

Here J is the Jordan matrix of A, where for ease of notation we place the 1×1 block corresponding to the eigenvalue λ_1 first. The first column $\mathbf{x}_1$ of X is the dominant right eigenvector of A and the first column $\mathbf{y}_1$ of $(X^{-1})^T$ is the dominant left eigenvector of A (or, equivalently, the dominant right eigenvector of A^T).

Besides the diagonal form and the Jordan form of A, we will also use the *singular value decomposition* [84]. Any matrix $A \in \mathbb{R}^{m\times n}$ can be decomposed as $A = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{m\times n}$ is zero except for its diagonal entries $\sigma_1, \dots, \sigma_{\min(m,n)}$. Here $\sigma_1, \dots, \sigma_{\min(m,n)}$ are nonnegative real numbers and are called the *singular values* of A. The columns $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ of U are the *left singular vectors* of A, whereas the columns $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ of V are the *right singular vectors* of A. The singular values are uniquely determined by A, but the singular vectors, in general, are not. When A has rank $r < \min(mn, n)$ it is often useful to consider the *compact singular value decomposition* of A. This is given by $A = U_r\Sigma_r V_r^T$, where $U_r = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r] \in \mathbb{R}^{n\times r}$, $V_r = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r] \in \mathbb{R}^{n\times r}$, and $\Sigma_r = \mathrm{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \in \mathbb{R}^{r\times r}$ is the leading $r \times r$ block of Σ.

If the undirected graph G is bipartite, with $V = V_1 \cup V_2$, there exists an ordering of the nodes such that the adjacency matrix of G takes the form

$$A = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix},$$

where B is $|V_1| \times |V_2|$. It is well known (see, e.g., [84]) that the nonzero eigenvalues of A are of the form $\lambda_i = \pm\sigma_i$, where the σ_i are the singular values of B.

If G is a digraph with n nodes, the associated undirected bipartite graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ has $2n$ nodes and its adjacency matrix is of the form

$$\mathcal{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix},$$

where A is the $n \times n$ adjacency matrix of G.

The *Laplacian matrix* associated with an undirected graph G is defined as $L = D - A$, where A is the adjacency matrix of G and $D = \mathrm{diag}(d_1, \dots, d_n)$ is the *degree matrix* of G. The graph Laplacian can be written as $L = BB^T$, where B is the *node-edge incidence matrix* of G; if G has m edges, B is the $n\times m$ matrix in which each column contains exactly two nonzero entries, equal to 1 and -1, in position i and j where the corresponding edge is (i, j).

For directed graphs, one may replace D by either the matrix of in-degrees $D^{\text{in}} = \mathrm{diag}(d_1^{\text{in}}, \dots, d_n^{\text{in}})$ or the matrix of out-degrees $D^{\text{out}} = \mathrm{diag}(d_1^{\text{out}}, \dots, d_n^{\text{out}})$, resulting in two distinct notions of Laplacian matrix, L_{in} and L_{out}, the second one being usually preferred.

For an undirected graph, the Laplacian L is a singular, symmetric, positive semidefinite M-matrix [28]. Note that $L\mathbf{1} = \mathbf{0}$. The algebraic multiplicity of the zero eigenvalue is equal to the number of connected components of the graph. In particular, the graph is connected if and only if the zero eigenvalue of the Laplacian is simple (i.e., if and only if L has rank $n - 1$).

For a directed graph, both L_{in} and L_{out} are singular M-matrices, with $\mathbf{1}^T L_{\text{in}} = \mathbf{0}^T$ and $L_{\text{out}}\mathbf{1} = \mathbf{0}$. If G is strongly connected, then L_{in} and L_{out} have rank $n-1$; the converse, in general, is not true. A complete characterization of strongly connected digraph in terms of the associated Laplacian L_{out} requires the notion of *uniform rank* [110]. We say that a matrix A has uniform rank r if it has rank r and every $r \times r$ minor of A is nonzero. The latter condition is equivalent to requiring that every subset of r rows and every subset of r columns of A is linearly independent. Then, as shown in [110], a digraph is strongly connected if and only if the corresponding Laplacian L_{out} has uniform rank $n - 1$.

2.2 | Functions of matrices

Next, we review the notion of *function of a matrix*. Several, fundamentally equivalent definitions are possible [90]. We will give a definition based on the diagonal or Jordan form. Let A be an $n \times n$ matrix (real or complex) and let us suppose for now that A is diagonalizable. Therefore, we can write $A = X\Lambda X^{-1}$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ is a diagonal matrix containing the eigenvalues of A, and X is invertible. Let $f(z)$ be a (real or complex) function that is well-defined on the eigenvalues of A. Then $f(A)$ is defined as

$$f(A) = Xf(\Lambda)X^{-1},$$

where $f(\Lambda) = \mathrm{diag}(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n))$.

If A is not diagonalizable, let $A = XJX^{-1}$ be its Jordan decomposition. The Jordan matrix J has a block diagonal structure $J = \text{diag}(J_1, J_2, \dots, J_m)$, with one or more blocks J_k associated with each distinct eigenvalue of A. Let J_k be one such block of size $\nu \times \nu$, with eigenvalue λ; then $f(J_k)$ is defined as

$$f(J_k) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \dots & \frac{f(\lambda)^{(\nu-1)}}{(\nu-1)!} \\ & f(\lambda) & f'(\lambda) & \dots & \frac{f(\lambda)^{(\nu-2)}}{(\nu-2)!} \\ & & \ddots & \ddots & \vdots \\ & & & f(\lambda) & f'(\lambda) \\ & & & & f(\lambda) \end{bmatrix}.$$

Now set $f(J) = \text{diag}(f(J_1), f(J_2), \dots, f(J_m))$ and

$$f(A) = Xf(J)X^{-1}.$$

Note that here $f(z)$ must be assumed to be differentiable a sufficient number of times at each eigenvalue of A. More precisely, denote with m_i the *index* of λ_i, i.e., size of the largest Jordan block in which the eigenvalue λ_i appears. If, for each distinct eigenvalue λ_i of A, the function f and its derivatives of order up to $m_i - 1$ are defined for $\lambda = \lambda_i$, we say that f *is defined on the spectrum of* A, and $f(A)$ is well-defined.

If $f(z)$ has a Taylor series expansion

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$$

with radius of convergence R, this property can be used to define $f(A)$. Suppose that each eigenvalue λ_i of A is either contained in the interior of the convergence disk $|z - z_0| < R$, or is on the boundary of the disk and the series for $f^{(m_i-1)}(\lambda)$ is convergent at $\lambda = \lambda_i$, with m_i the index of λ_i. Then $f(A)$ is well-defined and it is given by

$$f(A) = \sum_{k=0}^{\infty} a_k (A - z_0 I)^k.$$

Note that the converse is also true, i.e., if $f(A)$ can be expressed through the Taylor series of $f(z)$, then the eigenvalues of A satisfy the conditions stated above.

The notion of matrix function can be extended to rectangular matrices via the singular value decomposition [89]. Take $A \in \mathbb{R}^{m\times n}$ and let $A = U_r \Sigma_r V_r^T$ be its compact singular value decomposition. If f is a scalar function such that $f(\sigma_i)$ is defined for $i = 1, \dots, r$, then the *generalized matrix function* $f^\diamond$ induced by f is defined as

$$f^\diamond(A) = U_r f(\Sigma_r) V_r^T,$$

where $f(\Sigma_r) = \text{diag}(f(\sigma_1), \dots, f(\sigma_r))$. If A is square, $f^\diamond(A)$ does not necessarily coincide with the standard matrix function $f(A)$; however, if A is symmetric positive definite (or semidefinite and $f(0) = 0$), then $f^\diamond(A)$ reduces to $f(A)$.

3 | MATRIX FUNCTIONS FOR CENTRALITY MEASURES

One of the most basic questions in the structural analysis of networks is that of the *centrality* of nodes. A node centrality measure quantifies the relative importance, or "well-connectedness", of each node in G, and can be used to establish a *ranking* of the nodes. In the directed case, one needs to distinguish between the role of each node as "broadcaster" or as "receiver" of information in the network. Nodes with high broadcast centrality are known as *hubs*, whereas nodes with high receive centrality are called *authorities*.

Many definitions of node centrality have been proposed in the literature. The simplest and easiest to compute is *degree centrality*, which is just the degree d_i of node i. The underlying idea is that a node is well-connected if it has many neighbors, but this does not account for the importance of such neighbors. If the graph is directed, simple measures of broadcast and receive centrality of node i are provided by the out-degree d_i^{out} and by the in-degree d_i^{in}, respectively.

Two centrality measures popular in social network analysis are betweenness centrality and closeness centrality. Both focus on communication in a graph V through shortest paths. *Betweenness centrality* [77] is defined for any node i as

$$C_B(i) = \sum_{j\neq i} \sum_{k\neq i} \delta_{jk}(i),$$

where $\delta_{jk}(i)$ is the fraction of all shortest paths in the graph between nodes j and k which contain node i:

$$\delta_{jk}(i) = \frac{\#\text{ of shortest paths between } j,k \text{ containing } i}{\#\text{ of shortest paths between } j,k}.$$

Hence, a node that has high betweenness centrality is a node that is likely to be involved in the exchange of information whenever two nodes communicate with each other.

Closeness centrality [78] for node i is defined as

$$C_C(i) = \frac{1}{\sum_{j\in V\setminus\{i\}} d(i,j)},$$

where $d(i,j)$ is the *geodesic distance* between nodes i and j, that is, the length of a shortest path that connects the two nodes. Thus, a node has high closeness centrality if it is at a relatively short distance from most other nodes in the network. Recall that the *diameter* of an undirected connected graph $G = (V,E)$ is the maximum distance between any two nodes $i,j \in V$. Many real-world complex networks have small diameter; hence, in such a network many nodes are likely to have a high closeness centrality score, making it difficult to rank them.

We now turn to node centrality measures that can be expressed in linear algebra terms. Recall from Section 2 that the adjacency matrix A of a simple, strongly connected graph G satisfies the hypotheses of the Perron-Frobenius theorem and therefore has a positive dominant eigenvector $\mathbf{x} = [x_1,\dots,x_n]^T$. *Eigenvector centrality* [31] takes entry x_i as a measure of centrality of node i. Observe that if A is *primitive* (see [28, page 28]) the dominant eigenvector $\mathbf{x}$ is, up to normalization, the limit for $k\to\infty$ of the fixed-point iteration

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}, \qquad k = 0,1,\dots \tag{1}$$

If A is imprimitive, it suffices to replace it with $\hat{A} = (1-\alpha)I + \alpha A$ with $\alpha\in(0,1)$ and apply the fixed-point iteration to $\hat{A}$. Assume G is an unweighted graph. It is easy to see [42] that x_i (assumed to be positive and normalized with respect to the ℓ^1-norm) satisfies

$$x_i = \lim_{k\to\infty} \frac{\#\text{ of walks of length } k \text{ in } G \text{ through node } i}{\#\text{ of walks of length } k \text{ in } G}.$$

As we will see, walks play an important role also in several other centrality measures, particularly those that can be expressed in terms of matrix functions.

Iteration (1) corresponds to the intuitive notion that "a node is important if it is linked to many important nodes". Its rate of convergence depends on the spectral gap $\gamma = \lambda_1 - \lambda_2$ of the adjacency matrix A: the larger γ, the faster the convergence. Note that Google's PageRank algorithm [34] is a variant of eigenvector centrality applied to the (directed) graph of the World Wide Web, suitably modified so as to guarantee the existence and uniqueness of a positive dominant eigenvector [101].

For directed networks, the dominant left and right eigenvectors of A provide authority and hub scores, respectively. On this topic we mention Kleinberg's HITS (Hyperlink-Induced Topic Search) algorithm [97], originally developed for ranking Web pages. In this iterative ranking scheme, each node i is given an authority score x_i and a hub score y_i, both initially set to 1. In matrix terms, each iteration computes

$$\mathbf{x}^{(k)} = A^T\mathbf{y}^{(k-1)} \qquad \text{and} \qquad \mathbf{y}^{(k)} = A\mathbf{x}^{(k)},$$

where $\mathbf{x}^{(k)}$ and $\mathbf{y}^{(k)}$ are the vectors of authority and hub scores, respectively, at the k-th step. If we rewrite this process as

$$\mathbf{x}^{(k)} = A^TA\mathbf{x}^{(k-1)} \qquad \text{and} \qquad \mathbf{y}^{(k)} = AA^T\mathbf{y}^{(k-1)},$$

we see that HITS is essentially a power method that computes the dominant eigenvector of A^TA and of AA^T, yielding authority and hub rankings. Note that these matrices are nonnegative; however, even if the underlying digraph G is strongly connected, they may be reducible. Hence, the Perron–Frobenius theorem may not be applicable, and the limits of the sequences $\{\mathbf{x}\}^{(k)}$ and $\{\mathbf{y}^{(k)}\}$ may depend on the initial vectors $\mathbf{x}^{(0)}$ and $\mathbf{y}^{(0)}$. Conditions ensuring the well-posedness of the HITS computation can be found in [70]. As shown in [22], HITS can be reformulated in terms of eigenvector centrality for the augmented matrix

$$\mathcal{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}, \tag{2}$$

the adjacency matrix of the bipartite graph associated with the original digraph.

Subgraph centrality, introduced by Estrada and Rodríguez-Velásquez [67], measures the centrality of a node by taking into account the number of subgraphs the node "participates" in. More precisely, let G be an undirected graph with adjacency matrix A and define $SC(i)$, the subgraph centrality of node i, as the total number of walks in G that begin and end at node i, where the

number of walks of length k is scaled by a factor $\frac{1}{k!}$. Now recall that, for any fixed index i, the entry $[A^k]_{ii}$ counts the number of walks of length k that begin and end at node i. Therefore $SC(i)$ can be written in terms of a matrix function as

$$SC(i) = \left[I + A + \frac{A^2}{2} + \frac{A^3}{3!} + \ldots + \frac{A^k}{k!} + \ldots \right]_{ii} = [\mathrm{e}^A]_{ii}. \tag{3}$$

Note that off-diagonal entries $[\mathrm{e}^A]_{i,j}$ provide instead a measure of *communicability* between nodes i and j: the presence of many walks joining node i and node j suggests that information flows easily between the two nodes. We will return to this notion in section 4.

When modeling networks that are subject to some form of external influence (such as a condition of tension, or stress, on a social network, or heightened risk on a financial network) it can be useful to introduce a tuning parameter $\beta > 0$, sometimes called the *inverse temperature* by analogy with physical systems [61]. In this case the penalizing weight for walks of length k is set to $\frac{\beta^k}{k!}$ and (3) becomes

$$SC(i,\beta) = \left[I + \beta A + \frac{\beta^2}{2} A^2 + \frac{\beta^3}{3!} A^3 + \ldots + \frac{\beta^k}{k!} A^k + \ldots \right]_{ii} = [\mathrm{e}^{\beta A}]_{ii}. \tag{4}$$

The sum of subgraph centralities over all the nodes of the graph is called the *Estrada index* of G:

$$EE(G,\beta) = \sum_{i=1}^{n} SC(i,\beta) = \sum_{i=1}^{n} [\mathrm{e}^{\beta A}]_{ii} = \mathrm{Tr}(\mathrm{e}^{\beta A}) = \sum_{i=1}^{n} \mathrm{e}^{\beta \lambda_i}.$$

The normalization of subgraph centralities by the Estrada index defines a discrete probability distribution on the nodes of the graph:

$$p(i) = SC(i,\beta)/EE(G,\beta), \qquad i = 1, \ldots, n.$$

Here $p(i)$ can be interpreted as the probability of selecting any weighted closed walk that starts and ends at node i among all the weighted closed walks that start and end at the same node. The *walk entropy* of a graph is defined as

$$S^V(G,\beta) = -\sum_{i=1}^{n} p(i) \log_2 p(i).$$

Besides the entropy, other statistical mechanical quantities that are useful in network analysis are the *total energy*

$$H(G,\beta) = -\sum_{i=1}^{n} \lambda_i p_i$$

and the *Helmholtz free energy*

$$F(G,\beta) = -\beta^{-1} \ln EE(G,\beta),$$

see [61]. In particular, the Helmholtz free energy is closely related to the "natural connectivity" of a graph, see [65].

Unlike degree, which is a local centrality index, subgraph centrality $SC(i,\beta)$ takes into account the influence of all nodes on node i. The choice of factorial weights tends to emphasize the role of short walks. If a different balance of short- and long-range influence is needed, we may select a different set of real, nonnegative weights $\{w_k\}_{k=0,1,\ldots}$. When these weights are the coefficients in the power series expansion of a suitable function $f(z)$, the associated definition of subgraph centrality can be written analogously to (3):

$$SC_f(i) = \left[w_0 I + w_1 A + w_2 A^2 + \ldots + w_k A^k + \ldots \right]_{ii} = [f(A)]_{ii}, \tag{5}$$

provided that the series is convergent. For a discussion of *admissible functions*, i.e., types of analytic functions (possibly dependent on a parameter) that can be used to define node centrality measures, see [25]. A notable property of the class of admissible functions considered in [25] is that they form a *positive cone*, i.e., positive linear combinations of admissible functions are also admissible.

Besides the matrix exponential, one popular choice for $f(A)$ is the *resolvent* function:

$$(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \cdots = \sum_{k=0}^{\infty} \alpha^k A^k, \quad 0 < \alpha < 1/\lambda_1,$$

from which we obtain the *resolvent-based subgraph centrality*

$$SC_{res}(i,\alpha) = [(I - \alpha A)^{-1}]_{ii}.$$

In fact, row sums of $(I - \alpha A)^{-1}$ were originally proposed by Katz [95] as a notion of centrality. *Katz centrality* is defined as

$$K(i, \alpha) = [(I - \alpha A)^{-1}\mathbf{1}]_i = \mathbf{e_i}^T (I - \alpha A)^{-1}\mathbf{1},$$

where $\mathbf{e}_i$ denotes the i-th vector of the canonical basis of $\mathbb{R}^n$. The bounds on α ensure that $I - \alpha A$ is invertible and that $(I - \alpha A)^{-1}$ is nonnegative. As a consequence, both the diagonal entries and the row/column sums of $(I - \alpha A)^{-1}$ are positive and can be used for ranking purposes. In the case of directed graphs, the row sums of $(I - \alpha A)^{-1}$ yield the hub scores while the column sums yield the authority scores of each node in G. Hence, hubs and authorities can be ranked by solving the two linear systems $(I - \alpha A)\mathbf{x} = \mathbf{1}$ and $(I - \alpha A^T)\mathbf{x} = \mathbf{1}$, respectively. Note that in place of the vector of all ones it is possible to use any (nonnegative) reference vector as the right-hand side.

A similar idea can be applied to $e^{\beta A}$ by defining a measure of centrality known as *total communicability* [24]:

$$TC(i, \beta) = [e^{\beta A}\mathbf{1}]_i = \mathbf{e_i}^T e^{\beta A}\mathbf{1}.$$

A detailed discussion of this centrality measure is postponed to section 4.

It has long been observed that the rankings obtained with different centrality measures, while not identical, are often positively correlated. In the case of centrality measures that depend on a tuning parameter, something more precise can be said. In particular, degree and eigenvector centrality can be seen as limit cases of Katz and subgraph centrality as the parameters tend to specific values.

Theorem 1. [25] Let A be the adjacency matrix for a simple, connected, undirected graph G. Then:

(i) as $\beta \to 0^+$, the rankings produced by $SC(i, \beta)$ and $TC(i, \beta)$ converge to the degree ranking of node i,

(ii) as $\alpha \to 0^+$, the rankings produced by $SC_{res}(i, \alpha)$ and $K(i, \alpha)$ converge to the degree ranking of node i,

(iii) as $\beta \to \infty$, the rankings produced by $SC(i, \beta)$ and $TC(i, \beta)$ converge to the eigenvector centrality ranking of node i,

(iv) as $\alpha \to \frac{1}{\lambda_1}^-$, the rankings produced by $SC_{res}(i, \alpha)$ and $K(i, \alpha)$ converge to the eigenvector centrality ranking of node i,

(v) the results for $SC(i, \beta)$ and $K(i, \alpha)$ remain valid if the vector $\mathbf{1}$ is replaced by an arbitrary preference vector $\mathbf{v} > \mathbf{0}$.

For generalizations of this result to more general matrix functions and to directed and weighted graphs, see [25] and the references therein. We also mention that in some cases it is possible to prescribe the value of the parameter α in Katz centrality so as to obtain rankings that generally match those obtained by subgraph centrality (with $\beta = 1$), see [6]. This can be advantageous for example in the case of large directed graphs, for which computation of the matrix exponential can be challenging.

For directed graphs, notions of authority and hub centrality and communicability based on the exponential matrix function have been proposed in [22]. The authors consider the adjacency matrix $\mathcal{A}$ of the associated bipartite graph, defined in (2), and give an interpretation of the entries of $e^{\mathcal{A}}$ in terms of directed centrality and communicability scores: see [22] and section 4.5.

In recent years, some authors have suggested that a certain type of walks on graphs, called backtracking walks, should not be taken into account in the definition and computation of walk-based centrality measures. A walk is said to be *backtracking* if it contains at least a sequence of nodes of the form (i, j, i), and *non-backtracking* otherwise. The exclusion of backtracking walks, besides reflecting the intuitive notion that back-and-forth patterns appear less relevant in the transmission of information [3], also helps reducing localization effects. See [11, 109] and references therein for the definition and analysis of new centrality measures based on non-backtracking walks on directed and dynamic graphs. In [12], the authors derive an explicit formula for the exponential generating function of non-backtracking walks for both undirected and directed graphs. Using these results, they show how the new measures may be interpreted in terms of the standard exponential centrality computation on a certain *multilayer network*.

In this section we have focused on centrality measures for the nodes, but measures of edge centrality are also of great interest; for example, an edge betweenness centrality measure has been proposed by Girvan and Newman [80]. Edge centrality measures can also be induced by node centrality measures. As an example, in [8] the centrality $C(i, j)$ of the edge connecting two nodes i and j in an undirected network with centrality $C(i)$ and $C(j)$ was defined as the product of the two node centralities: $C(i, j) = C(i)C(j)$. As usual, the choice of an appropriate edge centrality measure is likely to be problem-dependent.

4 | MATRIX FUNCTIONS FOR COMMUNICABILITY MEASURES

Another fundamental problem in the structural analysis of networks is that of quantifying how well different parts of a network, such as individual nodes or groups of nodes, are connected to one another. This problem is of great importance, for instance, in the design and analysis of robust infrastructural networks, in the study of urban traffic, in anatomy and medicine (especially brain networks), in the study of vulnerability of financial networks, in bibliometry, in community detection, and in many other fields.

4.1 | Network communicability

The notion of *communicability* between two nodes, first introduced by Estrada and Hatano in [62], provides an effective measure of the connectivity between pairs of nodes. The underlying intuition is that two nodes i and j are well-connected if there are many walks in G connecting i and j, again with shorter walks being given more weight than longer ones. Thus, Estrada and Hatano define the communicability between the two nodes as

$$[\mathrm{e}^{\beta A}]_{ij} = \sum_{k=0}^{\infty} \frac{\beta^k}{k!} [A^k]_{ij}\,, \tag{6}$$

where $\beta > 0$, the inverse temperature, can be used to tune the relative importance of short and long walks. As in the case of walk-based centrality, other functions besides the matrix exponential can be used. For example, the off-diagonal entries of the resolvent $f(A) = (I - \alpha A)^{-1}$, with $0 < \alpha < \frac{1}{\lambda_1}$, can be used to define the *Katz communicability* between nodes. An *eigenvector-based communicability* can be obtained by taking the product $x_1(i)x_1(j)$ of the ith and jth entries of the (positive) dominant eigenvector $\mathbf{x}_1$ of a (strongly) connected network. Note that if $\mathbf{x}_1$ has been normalized so that $\|\mathbf{x}_1\|_2 = 1$, then this is just the (i, j) entry of the projector $\mathbf{x}_1\mathbf{x}_1{}^T$ onto the eigenspace spanned by $\mathbf{x}_1$; if G is undirected, this is also a matrix function $\phi(A)$, corresponding to any scalar function $\phi(\lambda)$ which takes the value 1 for $\lambda = \lambda_1$ and the value zero on the remaining eigenvalues of A. As in the case of eigenvector centrality, this can be interpreted as a measure of communicability that takes into account longer walks: for a connected undirected graph, it can be obtained as the limit of the expression (6), normalized by $\mathrm{e}^{\beta\lambda_1}$, as $\beta \to \infty$. The same applies to the Katz communicability (multiplied by $(1 - \alpha\lambda_1)$) in the limit $\alpha \to \frac{1}{\lambda_1}^-$. At the other extreme we have the simplest and most local notion of communicability between nodes, taking the values 1 or 0 depending on whether or not there is an edge between nodes i and j. This can be seen as the limit, as $\beta \to 0$ or $\alpha \to 0$, of the two parameter-dependent communicability measures based on $\mathrm{e}^{\beta A}$ and $(I - \alpha A)^{-1}$, respectively, similar to Theorem 1 in section 3.

The communicability function (6) has proved very useful in a number of applications. Some recent examples include:

1. The analysis of brain networks, particularly the identification of anomalies in the brains of patients affected by different conditions, ranging from stroke to Type 2 diabetes [29, 40, 41, 85, 46, 135];
2. The modeling of cancer networks [81];
3. The study of the spread of contagion in financial networks [16, 129];
4. The study of fracture networks in rocks [127];
5. The study of influence of papers in collaboration and citation networks [140];
6. The modeling of concept learning as weighted networks [100];
7. The characterization of robustness in infrastructural and transportation networks [8, 9, 108];
8. The analysis of urban traffic flow [1];
9. The community detection problem [48, 63, 138];
10. The identification of sparse and dense subgraphs of large graphs [76].

Computing the communicability between pairs of nodes requires the evaluation of bilinear forms of the type $\mathbf{e}_i{}^T f(A)\mathbf{e}_j$ for a given matrix function $f(A)$ of the adjacency matrix. While this is computationally expensive for a large network if *all* the communicability values are needed, computing the communicability between a few selected pairs of nodes can be efficiently carried out, even for very large networks; we refer to section 7 for details. For an extensive survey of network communicability up to the year 2012, see [65].

Network	$EE_n(G)$	$TC(G)$	e^{λ_1}
Zachary Karate Club	30.62	608.79	833.81
Drug User	1.12e05	1.15e07	6.63e07
Yeast PPI	1.37e05	3.97e07	2.90e08
Pajek/Erdos971	3.84e04	4.20e06	1.81e07
Pajek/Erdos972	408.23	1.53e05	1.88e06
Pajek/Erdos982	538.58	2.07e05	2.73e06
Pajek/Erdos992	678.87	2.50e05	3.73e06
SNAP/ca-GrQc	1.24e16	8.80e17	6.47e19
SNAP/ca-HepTh	3.05e09	1.06e11	3.01e13
SNAP/as-735	3.00e16	3.64e19	2.32e20
Gleich/Minnesota	2.86	14.13	35.34

TABLE 1 Comparison of the normalized Estrada index $EE_n(G) = \mathrm{Tr}(e^A)/n$, the normalized total network communicability $TC(G)$, and e^{λ_1} for various real-world networks.

4.2 | Total node communicability as a centrality measure

A communicability-based node centrality measure can be defined considering the *total communicability* of a node, that is, the sum of the communicabilities of a node with all the nodes in the network, including itself [24]:

$$TC(i) = \sum_{j=1}^{n} [e^{\beta A}]_{ij} = [e^{\beta A}\mathbf{1}]_i \,. \tag{7}$$

Here we have used the fact that the row sums of a matrix function $f(A)$ are the components of the vector $f(A)\mathbf{1}$. A significant advantage of this centrality measure over those based on the diagonal entries of $f(A)$ ("self-communicabilities") is the fact that for large networks it is much easier to compute the action of a matrix function on a vector than the diagonal entries $[f(A)]_{ii}$. Indeed, Krylov subspace methods can efficiently compute the vector $f(A)\mathbf{1}$ for a large and sparse matrix A without the need to compute any of the entries of $f(A)$; see the discussion in section 7.

On the other hand, centrality measures based on the diagonal entries $[f(A)]_{ii}$ have a higher discrimination power than those based on the row sums $[f(A)\mathbf{1}]_i$. Indeed, if a graph G is k-regular, then $A\mathbf{1} = k\mathbf{1}$ and therefore also $f(A)\mathbf{1} = f(k)\mathbf{1}$. Hence, all nodes in G get the same centrality score and no measure of this kind is able to discriminate between the nodes (the same holds, of course, for degree and eigenvector centrality). On the other hand, there are examples of regular graphs for which subgraph centrality ($C(i) = [e^A]_{ii}$) is able to produce a ranking of the nodes [67]; we refer to [18, 99] and [94] for detailed discussions of various theoretical issues concerning the discriminating power of centrality measures based on closed walks. Generally speaking, however, real world networks are far from being regular and, in most cases, subgraph centrality and total node communicability tend to produce similar rankings. Thus, in the case of large networks the use of the more easily computable total communicability is recommended.

Summing up all the total communicabilities of the nodes and in G and normalizing by the number of nodes results in the *total network communicability* (see [24]):

$$TC(G) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} [e^{\beta A}]_{ij} = \frac{1}{n}\mathbf{1}^T e^{\beta A}\mathbf{1} \,. \tag{8}$$

It is easy to see that this quantity, which can be interpreted as the *average communicability* of a node, is maximized by K_n, the complete graph with n nodes. The total network communicability is a measure of network connectivity, and in this regard it plays a role similar to that of the Estrada index. It is, however, much easier to compute, since it consists of a single quadratic form whereas computing the Estrada index usually requires evaluating the n quadratic forms $\mathbf{e}_i^T e^{\beta A}\mathbf{e}_i$, for $i = 1, \ldots, n$. Again, $TC(G)$ can be computed without having to compute any of the individual entries of the matrix exponential, see section 7.

In Table 1 we report (for $\beta = 1$) the normalized Estrada index $EE_n(G) = \frac{1}{n}\mathrm{Tr}(e^A)$, the total network communicability (8), and the exponential of the dominant eigenvalue for the adjacency matrices of several real-world undirected networks. The first two are social networks, the third one is the protein-protein interaction network of beer yeast (*Saccharomyces cerevisiae*), the following six are different types of scientific collaboration (coauthorship) networks, and the remaining two are infrastructure networks (a computer server network and a road network). The first thing to observe is that all three measures of connectivity

are consistent, in that they all agree on which networks are highly connected and which are not. This is not surprising given the close relationship between the three metrics. The second thing to observe is the great variability of these connectivity measures for different types of networks. This can only partially be explained by the difference in the number of nodes: for example, the Zachary Karate Club network consists of just 34 nodes, while the Minnesota Road Network has 2642 nodes. We mention that all these networks are extremely sparse, with average degree less than 5 for most networks. According to these three measures the most highly connected network is SNAP/as-735, an autonomous server network which has certainly been designed with high connectivity and robustness in mind. The least connected network, not surprisingly, is the (grid-like) road network of the state of Minnesota. The network diameters are also strongly correlated with these measures of connectivity.

4.3 | Communicability distance

Communicability between nodes can also be used to define a notion of distance on undirected graphs. Estrada [49, 51] has shown that letting

$$d_c(i,j) = \sqrt{C(i,i) + C(j,j) - 2C(i,j)}\,, \quad \text{where} \quad C(i,j) = [\mathrm{e}^A]_{ij}\,, \tag{9}$$

defines a distance on the set of nodes of an undirected graph G. To see this, let $A = QDQ^T$ be the spectral decomposition of the adjacency matrix A, and let $\mathbf{u}_j$ denote the jth column of Q^T. Letting $\mathbf{x}_j = \mathrm{e}^{D/2}\mathbf{u}_j$, we have

$$d_c(i,j)^2 = (\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2\,,$$

showing that $d_c(i,j)$ is just the Euclidean distance of the vectors $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^n$. For a similar distance based on the resolvent $(I - \alpha A)^{-1}$ instead of the matrix exponential, see [35].

Communicability distance has been shown to have a number of interesting applications. In [56], it has been applied to the problem of *triadic closure prediction*, producing the best available results so far. Applications to urban traffic modeling can be found in [1, 130], while the related notion of *communicability angles* is able to identify the critical (i.e., most diffusive) links in consensus dynamics problems [69]. A notion of *spatial efficiency* based on the communicability geometry is introduced and used to analyze real-world networks in [64]. Applications to visualization and machine learning can be found in [116]. Finally, an extension of communicability geometry to multiplex networks can be found in [54], where it is applied to the study of the neuronal network of *C. elegans*.

4.4 | Network modification problems

Network communicability is closely related to the notions of *network efficiency* and *network robustness*, which are of great importance in the design and realization of actual networks in different fields of engineering. A network with high communicability is likely to be efficient in the routing of messages or in the transport of goods, and also to be robust against localized loss of connectivity (e.g., edge removal) due to random accidents and possibly even to deliberate attacks. There are also situations where high connectivity has negative effects: for example, in the spreading of contagion, whether on a social network, or a financial network, or a computer network. Another example is in law enforcement or anti-terrorism operations, where it is desirable to disrupt or weaken an existing highly connected network. Highly sparse networks that display excellent connectivity properties are known in the computer science literature as *expander graphs*; we refer to [92] for an extensive survey.

Total network communicability is a global measure which can be used to compare different network designs before networks are actually constructed, or to modify an existing network (via edge addition, deletion or rewiring) towards a desired goal. Consider, for instance, the following problems:

1. *Network updating*: given a sparse network $G = (V, E)$ with $|E| = m$, and a budget of $k \ll m$ edges, choose which k pairs of unconnected nodes in G to connect so as to increase the total network communicability as much as possible.

2. *Network downdating*: given G as above, choose which k existing edges ($k \ll m$) to delete so as to decrease the total communicability as little or as much as possible; the first problem corresponds to the goal of having similar connectivity at a lower cost (fewer edges), the second to the goal of disrupting or weakening an existing network in order to drastically reduce its effectiveness with minimal effort.

3. *Network rewiring*: given G as above, choose which k existing edges ($k \ll m$) to rewire so as to increase, or decrease, the total network communicability as much as possible.

In all cases, the total network communicability can be used as a figure of merit to evaluate the relative effectiveness of different updating/downdating/rewiring strategies. Of course, this is not the only measure of connectivity available but, together with the one given by e^{λ_1}, it has the advantage of being much easier to compute, in the case of large sparse networks, than most other measures, often at a cost of $O(n)$ (i.e., linear in the number of nodes). In [8], a number of different heuristics are compared to solve the updating/downdating problems listed above (the rewiring problem is similar and can be dealt with using the same techniques). These heuristics are much faster, sometimes by orders of magnitude, than previously proposed methods, and require only the identification of a small percentage of nodes in G having the highest centrality (for example, eigenvector or total communicability centrality). Connecting highly central nodes that are not already connected tends to cause the largest increase in total communicability (as well as in other measures); reciprocally, removing edges that connect highly central nodes tends to induce the largest drop in total communicability. If the goal is to delete as many edges as possible without disconnecting the network and with minimal decrease in the total communicability, then the best strategy is to remove edges between nodes outside of the set of highly central nodes. Checking the connectivity of the network after each downdating can also be done in time that is typically linear in n, for a sparse network. We refer the reader to [8] and the references therein for a detailed discussion of this class of network modification problems.

We note that edge addition, deletion or rewiring all lead to low-rank modifications of the adjacency matrix. The important problem of updating/downdating a matrix function after a low-rank change has been investigated in [17] and [118], where applications to node centrality measures are also considered.

4.5 | Communicability in directed networks

In the case of directed networks, several types of node communicabilities that take into account the directionality of communication can be introduced. The first option is to simply define the communicability of two nodes as was done in the undirected case, via (6). The lack of symmetry in A, and therefore in e^A (or, more generally, in $f(A)$) is just a reflection of the fact that in a directed graph, traveling from node i to node j may be faster (or slower) than traveling from node j to node i. Likewise, sending a message on a telecommunication network from node i to node j may be easier (or harder) than sending the same message from node j to node i. Taking row sums of $f(A)$ will then define a kind of total communicability measure that corresponds to the centrality of nodes as broadcasters of information (*hubs*), while taking column sums of $f(A)$ will result in a centrality measure of nodes as receivers of information (*authorities*).

A more refined measure of communicability would distinguish further between four types of communicability: broadcaster-to-receiver, broadcaster-to-broadcaster, receiver-to-receiver, and receiver-to-broadcaster. Such measures have been first studied in [22], based on the bipartite model of a directed graph described in section 2. Let

$$\mathcal{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \tag{10}$$

be the adjacency matrix of the bipartite graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ obtained from the original digraph represented by A. In the bipartite graph the first n nodes contained in $\mathcal{V}$ can be seen as the original nodes of the digraph when they play the role of broadcasters of information, while the last n copies contained in $\mathcal{V}$ represent the original nodes in the role of receivers. It is worth mentioning that the eigenvector of $\mathcal{A}$ corresponding to the leading eigenvalue $\lambda_1(\mathcal{A}) = \sigma_1$ is the vector $\mathbf{q}_1 = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{v}_1 \end{bmatrix}$, thus revealing the connection between the HITS ranking scheme described in section 3 and eigenvector centrality applied to the bipartite graph $\mathcal{G}$.

Proposition 1. [22, Proposition 1] Let $\mathcal{A}$ be as in (10) and let $A = U\Sigma V^T$ be an SVD of A. Then

$$e^{\mathcal{A}} = \begin{bmatrix} \cosh(\sqrt{AA^T}) & U\sinh(\Sigma)V^T \\ V\sinh(\Sigma)U^T & \cosh(\sqrt{A^TA}) \end{bmatrix}. \tag{11}$$

An important feature of this matrix is that its entries are nonnegative. Thus, these quantities can be used to describe the importance of nodes and how well they communicate when they are acting as broadcasters or receivers of information in the graph [22]. Indeed, the entries of the two diagonal blocks $\cosh(\sqrt{AA^T})$ and $\cosh(\sqrt{A^TA})$ provide centrality and communicability indices for nodes and pairs of nodes when they are all seen as playing the same role in the network. More in detail, the diagonal entries of the first diagonal block give the centralities for the nodes in the original network when they are seen as broadcasters of information (hubs). Likewise, the diagonal of the second block contains the centralities for the nodes in their role of receivers (authorities). Similarly to the off-diagonal entries of the matrix exponential of an undirected graphs, the off-diagonal entries of these diagonal blocks measure how well two nodes, both acting as broadcasters (resp., receivers), exchange information.

As for the off-diagonal blocks in (11), they contain information concerning how nodes exchange information when one node is playing the role of broadcaster (resp., receiver) and the other is acting as a receiver (resp., broadcaster).

Thus, the *total hub communicability* and *total authority communicability* of G, defined as $T_h C(A) = \mathbf{1}^T \cosh(\sqrt{AA^T})\mathbf{1}$ and $T_a C(A) = \mathbf{1}^T \cosh(\sqrt{A^T A})\mathbf{1}$, respectively, account for the overall ability of the network of exchanging information when all its nodes are playing the same role of broadcasters ($T_h C(A)$) or receivers ($T_a C(A)$). These measures can be interpreted and justified in terms of *alternating walks* on the original digraph G; for this and other details we refer to [22] and to [9], as well as to the discussion of generalized matrix functions in subsection 6.7. The efficient computation of these quantities for large graphs is discussed in [10] and [13].

5 | TEMPORAL COMMUNICABILITY

The notion of complex network we have used so far is of a static nature: a network represents a snapshot of the interactions taking place in a system at a certain time. But many real-life complex systems evolve in time: this additional feature can be described through *temporal networks*. The presentation given in this section is mostly based on [36, 86].

Given a set of n nodes, consider an ordered sequence $\{G^{[k]}\}_{k=1,\dots,M}$ of unweighted, undirected simple graphs defined over the given nodes. We assume here to have chosen an ordered sequence $t_1 < t_2 < \dots < t_M$ of time points. Each graph $G^{[k]}$ records the state of the network at time t_k and may be represented by the associated symmetric adjacency matrix $A^{[k]}$.

Centrality and communicability measures based on counting walks can be extended to temporal networks through the use of *dynamic walks*. A dynamic walk is a sequence of edges connecting nodes in the network, with the added constraint that the sequence of edges must respect the time ordering. As in the static case, the reasoning here is that nodes participating in many dynamic walks are capable of communicating well with other nodes in the network, and can therefore spread or receive information effectively.

Let $\rho(A^{[k]})$ be the spectral radius of the kth adjacency matrix, and denote by $\sigma^* = \max_k\{\rho(A^{[k]})\}$ the maximum spectral radius over all the adjacency matrices. The dynamic communicability matrix is defined as

$$Q = (I - \alpha A^{[1]})^{-1}(I - \alpha A^{[2]})^{-1} \dots (I - \alpha A^{[M]})^{-1}, \tag{12}$$

where the parameter α satisfies $0 < \alpha < 1/\sigma^*$, thus ensuring that each resolvent matrix in (12) can be expressed as a power series in the corresponding adjacency matrix. If we write each resolvent as a power series and expand the product in (12), we see that Q contains all products of the form

$$\alpha^w (A^{[t_{k_1}]} A^{[t_{k_2}]} \dots A^{[t_{k_w}]}), \tag{13}$$

where $t_{k_1} \le t_{k_2} \le \dots \le t_{k_w}$. Observe that entry (i, j) in the product $A^{[t_{k_1}]} A^{[t_{k_2}]} \dots A^{[t_{k_w}]}$ counts the number of dynamic walks of length w from node i to node j, where the k-th edge of the walk comes from time-step t_k. Consequently, Q_{ij} is a weighted sum of dynamic walks of all possible lengths from node i to node j, where walks of length w are weighted by a factor α^w. Hence, Q_{ij} is a measure of communicability between nodes i and j, with node i in the role of broadcaster and node j as the receiver. The i-th row sum

$$BC(i) = \sum_{j=1}^{n} Q_{ij} = Q \cdot \mathbf{1}$$

is a measure of how well node i broadcasts information to the rest of the network; therefore it provides a notion of *broadcast centrality*. On the other hand, the j-th column sum

$$RC(j) = \sum_{i=1}^{n} Q_{ij} = Q^T \cdot \mathbf{1}$$

measures how well node j receives information from all other nodes in the network and provides a notion of *receive centrality*. Note that, in general, the broadcast and receive centrality of each node are different, as the matrix Q is not symmetric. Keeping in mind that each adjacency matrix is symmetric, we have

$$\begin{aligned} Q^T &= \left((I - \alpha A^{[1]})^{-1}(I - \alpha A^{[2]})^{-1} \dots (I - \alpha A^{[M]})^{-1}\right)^T = \\ &= (I - \alpha A^{[M]})^{-1}(I - \alpha A^{[M-1]})^{-1} \dots (I - \alpha A^{[1]})^{-1}, \end{aligned}$$

that is, broadcast and receive centralities are related by a reversal of the time ordering.

We mentioned that the parameter α should be chosen in the interval $(0, 1/\sigma^*)$. It is interesting to study the limit behavior of the BC and RC rankings when $\alpha \to 0$ or $\alpha \to 1/\sigma^*$. Let us write each resolvent in (12) as a power series, expand the product and collect terms w.r.t. powers of α:

$$Q = I + \alpha \left(\sum_{k=1}^{M} A^{[k]} \right) + \mathcal{O}(\alpha^2).$$

The vector of BC measures can now be expressed as

$$BC = Q \cdot \mathbf{1} = I \cdot \mathbf{1} + \alpha \left(\sum_{k=1}^{M} A^{[k]} \right) \mathbf{1} + \mathcal{O}(\alpha^2). \tag{14}$$

The quantity $AD = \left(\sum_{k=1}^{M} A^{[k]} \right) \mathbf{1}$ above is a measure of *aggregated degree*: it ranks nodes according to the number of distinct contacts weighted by the duration of contact time (the longer the contact time between two nodes, the higher their corresponding AD values). From (14) we see that, for $\alpha \to 0$, BC rankings converge to AD rankings; the same is true for RC rankings. Therefore the parameter α should be chosen sufficiently far away from 0 so as not to merely replicate aggregate degree. On the other hand, α should not be taken too close to $1/\sigma^*$. Indeed, if $\alpha \approx 1/\sigma^*$, the matrix $(I - \alpha A^{[k]})$ such that $\rho(A^{[k]}) = \sigma^*$ will be close to singular, and the corresponding resolvent factor will dominate the computation of Q, thus making the computation of Q very sensitive to small changes in α.

In some applications, one may want to use a modified definition of Q where only walks that use at most one link per time-window are taken into account. This can be an appropriate choice if one needs to eliminate the possibility of long walks, or in ultra-high frequency regimes, where it is physically unfeasible for information to pass through more than one link per time step. As discussed in [86], in such cases one may define

$$\hat{Q} = (I - \alpha A^{[1]})(I - \alpha A^{[2]}) \cdots (I - \alpha A^{[M]}),$$

that is, replace each resolvent factor in (12) by its first-order approximation.

One can also use the matrix exponential instead of the resolvent in the definition of dynamic communicability, as suggested by Estrada [53]:

$$\tilde{Q} = \mathrm{e}^{A^{[1]}} \mathrm{e}^{A^{[2]}} \cdots \mathrm{e}^{A^{[M]}}.$$

In the exponential-based definition, the number of walks of length w is weighted by a factor $1/w!$ instead of α^w, so that long walks are much more penalized than in the resolvent-based definition. This can sometimes lead to an inability to distinguish between nodes in the ranking process, especially when working with many time steps and many adjacency matrices. However, as seen in the static case, it is possible to include an "inverse temperature" parameter β in the exponential, leading to weight factors $\beta^k/k!$. The parameter β can then be tuned to better capture the behavior of the network.

For an application of dynamic communicability to a real-world temporal network arising in the study of a contagion model, see [36].

6 | OTHER MATRIX FUNCTIONS USED IN NETWORK ANALYSIS

In this section we provide a brief overview of some other matrix functions that have found applications in network analysis.

6.1 | Spectral measure of graph bipartivity

Bipartivity is an important graph property that reflects important structural features of real-world networks. It plays an important role, for instance, in the study of granular materials [136] and in the study of the transportation efficiency of European airlines [58]. Many real world networks, while not exactly bipartite, are "nearly bipartite"; to make this notion precise, it is desirable to have a quantitative measure of how "far" a graph is from being bipartite.

It is obvious that in a bipartite graph any closed walk must have even length. The hyperbolic sine and cosine, defined by

$$\sinh(x) = \frac{1}{2}(\mathrm{e}^x - \mathrm{e}^{-x}) \quad \text{and} \quad \cosh(x) = \frac{1}{2}(\mathrm{e}^x + \mathrm{e}^{-x})$$

obviously satisfy

$$\sinh(x) + \cosh(x) = \mathrm{e}^x,$$

have the power series expansions

$$\sinh(x) = \sum_{k=1}^{\infty} \frac{x^{2k+1}}{(2k+1)!}, \qquad \cosh(x) = \sum_{k=1}^{\infty} \frac{x^{2k}}{(2k)!},$$

and are, respectively, odd and even functions. Both are entire, and therefore the matrix functions $\sinh(A)$ and $\cosh(A)$ are defined for any square matrix A.

Define now the following quantity:

$$b_s(G) = \frac{\mathrm{Tr}(\cosh(A))}{\mathrm{Tr}(\mathrm{e}^A)}.$$

Note that $b_s(G)$ is always between $\frac{1}{2}$ and 1. If A is the adjacency matrix of a bipartite graph, it must be $\left[\mathrm{e}^A\right]_{ii} = [\cosh(A)]_{ii}$ for every $i = 1, \ldots, n$, and therefore $b_s(G) = 1$. Based on these simple observations, in [68] it was proposed to use $b_s(G)$ as a measure of graph bipartivity: the closer $b_s(G)$ to 1, the closer G to being bipartite. If, on the other hand, $b_s(G)$ is close to $\frac{1}{2}$, the graph is far from being bipartite. How close $b_s(G)$ must be to 1 for G to be considered "nearly bipartite" is, of course, context-dependent and involves a certain degree of subjectivity. It is also possible to devise edge manipulation techniques aimed at increasing (or decreasing) the degree of bipartivity of a network: to this end, $b_s(G)$ provides a way to monitor the effect of different edge addition/deletion/rewiring strategies.

6.2 | Functions of the graph Laplacian

So far we have limited our discussion to functions of the adjacency matrix of a network. In many important problems, however, it is appropriate to study functions of $L = D - A$, the graph Laplacian of G. For an undirected graph, the Laplacian matrix L is symmetric positive semidefinite, with zero row and column sums. The smallest nonzero eigenvalue of the Laplacian of a connected graph is known as its *algebraic connectivity* [73], and the corresponding eigenvector as its *Fiedler vector*. Loosely speaking, the algebraic connectivity of G measures how hard it is to disconnect G into two subgraphs of comparable size by removing selected edges: the larger the algebraic connectivity, the harder it is to disconnect G into two roughly equal subgraphs by removing a small number of edges. The sign of the entries of the Fiedler vector can be used to assign nodes of G to either subgraph. Hence, the graph Laplacian plays a role in the important problem of graph partitioning.

The Laplacian is also central to the study of diffusion processes and random walks on graphs. These are used sometimes to model dynamic phenomena taking place on networks, and sometimes as tools for exploring, or probing, network structure by studying the behavior of the solution to certain evolution equations posed on the network. An important example is that of *consensus in multi-agent systems*; see, e.g., [115]. The simplest model of diffusion on a graph is probably the initial value problem for the heat equation on G:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = -L\,\mathbf{x}, \quad t > 0; \quad \mathbf{x}(0) = \mathbf{x}_0. \tag{15}$$

This is a system of linear ordinary differential equations (ODEs). Here $\mathbf{x}(t)$ can be interpreted as the concentration of some substance on the nodes of G, or perhaps as the "temperature" of each node, at time t. The initial distribution $\mathbf{x}_0$ is usually a nonnegative vector, which can be assumed to be normalized so that $\|\mathbf{x}_0\|_1 = 1$. In this case, it can be interpreted as a probability distribution on the nodes of G. It is well known that the solution of (15) is unique and is given for all $t \geq 0$ by

$$\mathbf{x}(t) = \mathrm{e}^{-tL}\mathbf{x}_0 = S(t)\,\mathbf{x}_0\,,$$

where we have set $S(t) = \mathrm{e}^{-tL}$. The family $\{S(t)\}_{t\geq 0}$ forms a continuous one-parameter commutative semigroup, i.e.,

$$S(0) = I, \quad S(t_1 + t_2) = S(t_1)\,S(t_2) = S(t_2)\,S(t_1)\,, \ \forall t_1, t_2 \geq 0\,.$$

The graph Laplacian L satisfies

$$S'(0) = \lim_{t\to 0+} \frac{1}{t}\,(S(t) - I) = -L\,;$$

we say that $-L$ is the *infinitesimal generator* of the semigroup. The matrix-valued function $S(t)$ is also called the *heat kernel* on G. Another notion of heat kernel based on the normalized Laplacian will be briefly discussed in the next subsection.

It is easy to prove that for all $t \geq 0$, $S(t)$ is a doubly stochastic matrix; i.e., it has entries between 0 and 1, and row and columns sums equal to 1 (recall that $S(t)$ is symmetric, since G is undirected). It follows that for all t the solution vector $\mathbf{x}(t) = S(t)\mathbf{x}_0$ is nonnegative with entries adding up to 1: the ℓ^1-norm of $\mathbf{x}(t)$ is constant and equal to 1. Hence, for all $t > 0$, the solution $\mathbf{x}(t)$ is a probability distribution on the nodes of G, and its evolution describes a continuous time Markov process. Since G is assumed to be connected, L is irreducible and there is a unique steady-state distribution $\mathbf{x}_* = \lim_{t\to\infty} \mathbf{x}(t)$. This equilibrium is easily

determined to be the unique positive solution to the homogeneous linear system $L\mathbf{x} = \mathbf{0}$, normalized so that $\|\mathbf{x}_*\|_1 = 1$, i.e., the constant vector $\mathbf{x}_* = \frac{1}{n}\mathbf{1}$. In other words, the steady state is just "thermal equilibrium". The rate of convergence of the solution vector $\mathbf{x}(t)$ reflects structural properties of the graph and is governed by the algebraic connectivity of G. In more detail, let

$$L = Q\Lambda Q^T = \sum_{k=1}^{n} \lambda_k \, \mathbf{q}_k \mathbf{q}_k^T \tag{16}$$

be the eigendecomposition of the Laplacian of the (connected) graph G. Recalling that the eigenvalues of L satisfy $0 = \lambda_1 < \lambda_2 \le \ldots \le \lambda_n$ and that the eigenvector associated with the simple zero eigenvalue, normalized in the ℓ^2-norm, is $\mathbf{q}_1 = \frac{1}{\sqrt{n}}\mathbf{1}$, we obtain

$$\mathrm{e}^{-tL} = \frac{1}{n}\mathbf{1}\mathbf{1}^T + \sum_{k=2}^{n} \mathrm{e}^{-t\lambda_k}\mathbf{q}_k\mathbf{q}_k^T \, .$$

Therefore, we have

$$\mathrm{e}^{-tL}\mathbf{x}_0 = \frac{1}{n}(\mathbf{1}^T\mathbf{x}_0)\mathbf{1} + \sum_{k=2}^{n} \mathrm{e}^{-t\lambda_k}(\mathbf{q}_k^T\mathbf{x}_0)\mathbf{q}_k \longrightarrow c\mathbf{1} \quad \text{for} \quad t \to \infty \, , \tag{17}$$

where $c = \frac{1}{n}(\mathbf{1}^T\mathbf{x}_0)$ is positive if the initial (nonzero) vector $\mathbf{x}_0$ is nonnegative; in particular, $c = \frac{1}{n}$ if $\mathbf{x}_0$ is normalized in the 1-norm. We note that if the graph G is k-regular we have $L = kI - A$, and for $t = 1$ the heat kernel is given by $\mathrm{e}^{-L} = \mathrm{e}^{-k}\mathrm{e}^{A}$. Hence, apart from the constant factor e^{-k}, for a regular graph the exponential of the adjacency matrix of G is closely related to the exponential of the (negative) graph Laplacian.

It is clear from (17) that the rate of convergence to equilibrium is governed primarily by the smallest nonzero eigenvalue λ_2 of L, i.e., on the algebraic connectivity of G. The larger is λ_2, the faster the convergence to equilibrium. As an extreme example, consider K_n, the complete graph on n nodes, in which every node is connected to any other node. The Laplacian of this graph has only two distinct eigenvalues, 0 and n, and equilibrium is reached extremely fast. This is not surprising since the complete graph has the smallest diameter possible, and is "maximally connected". More interesting examples of highly connected graphs are given by the already mentioned expander graphs, i.e., graphs that are at the same time highly sparse and have a relatively large value of λ_2. Informally, such graphs are characterized by the property that it is impossible to disconnect them into two roughly equal-size subgraphs without cutting a similarly large number of edges. Expander graphs have important applications, for example in the design of robust communication and computer networks; we refer again to the extensive survey [92] for precise definitions and results.

Recently, the use of the *fractional graph Laplacian* has attracted some attention due to its potential to generate random walks that can efficiently explore large networks [123, 124]. For $\alpha \in (0, 1)$, the fractional graph Laplacian is the matrix

$$L^\alpha = Q\Lambda^\alpha Q^T \, , \tag{18}$$

where Λ^α is the diagonal matrix with entries given by $\lambda_1^\alpha = 0, \lambda_2^\alpha, \ldots, \lambda_n^\alpha$ (cf. (16)). In this definition, the principal (positive) branch of the αth root λ_i is chosen. If G is connected, L^α is generally a full matrix: in other words, L^α is a *non-local operator*. It can be shown that, like L, the matrix L^α is a singular M-matrix. If Δ represents the diagonal matrix containing the diagonal entries of L^α, the matrix $S = I - L^\alpha\Delta^{-1}$ is stochastic and defines a random walk on G. Since S is dense, the random walk on G is not constrained to move from node to neighboring node, but has a nonzero probability of "jumping" from one node to any other node. In particular, there is a non-negligible probability of making long-range jumps ("superdiffusion"); hence, the random walk associated with $S = I - L^\alpha\Delta^{-1}$ can be expected to more efficiently explore a large, sparse network than the "local" random walk associated with $I - LD^{-1}$, whenever the diameter of G is not small. The continuous-time process

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = -L^\alpha\mathbf{x}, \quad t > 0, \quad \mathbf{x}(0) = \mathbf{x}_0 \, , \tag{19}$$

or the analogous one based on the normalized fractional Laplacian, has similar advantages over the standard one (15). For an extension to directed graphs, based on the nonsymmetric Laplacian $L_{\text{out}} = D - A$ (where D is the diagonal matrix of outdegrees), see [19]. The development of efficient numerical methods for approximating the solution of the initial value problem (19) remains an open problem and is an active area of current research.

Finally, we mention that a different proposal for speeding up the navigation of large graphs has recently been made in [59, 60] based on the notion of k-path Laplacian, a natural generalization of the graph Laplacian; see [50].

6.3 | Heat Kernel PageRank

If G is a directed network in which every node i has nonzero outdegree d_i^{out}, then the matrix $W = A^T D^{-1}$, where A is the adjacency matrix and D is the diagonal matrix of outdegrees, is column stochastic, and for $t > 0$ the matrix exponential $K(t) = e^{-t(I-W)} = e^{-t}e^{tW}$ is called the *heat kernel* of G in [37]; see also [139]. The matrix $K(t)$ is itself stochastic (for all t) and therefore defines a Markov chain on G. It is easy to see that for a strongly connected graph, the solution $\mathbf{x}(t) = K(t)\mathbf{x}_0$ of the diffusion-like equation

$$\frac{d\mathbf{x}}{dt} = -(I - W)\,\mathbf{x}, \quad t > 0, \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{20}$$

where (for example) $\mathbf{x}_0 = \frac{1}{n}\mathbf{1}$, converges to the dominant eigenvector of W (the PageRank vector) as $t \to \infty$. The convergence is faster than that of the discrete-time random walk defined by W. Applications and efficient algorithms for approximating the heat kernel PageRank of large networks are discussed in [38]; see also [98].

6.4 | Other differential equations on graphs

While the most commonly studied evolution equations on graphs are related to diffusion, other kinds of equations have also attracted some interest [26, 107]. In this subsection we briefly discuss matrix functions arising in the solution of the wave and Schrödinger equation on undirected graphs.

The simplest form of the wave equation on an undirected graph G takes the form of the following initial value problem for a second-order system of ODEs:

$$\frac{d^2\mathbf{x}}{dt^2} = -L\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{x}_0, \quad \mathbf{x}'(0) = \mathbf{v}_0\,, \tag{21}$$

where $\mathbf{x}_0$ and $\mathbf{v}_0$ are given vectors. Equation (21) can be recast as a first-order system with twice the number of unknown functions, as follows. Recall that the graph Laplacian can be written as $L = BB^T$, where B is the node-edge incidence matrix of the network; note that B is, in general, rectangular. Letting $\mathbf{v}'(t) = B^T\mathbf{x}(t)$, problem (21) can be written as the Hamiltonian-like initial value problem

$$\frac{d\mathbf{w}}{dt} = \mathcal{B}\mathbf{w}, \quad \mathbf{w}(0) = \mathbf{w}_0\,, \tag{22}$$

where

$$\mathbf{w}(t) = \begin{bmatrix} \mathbf{u}(t) \\ \mathbf{v}(t) \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} 0 & -B \\ B^T & 0 \end{bmatrix}, \quad \mathbf{w}_0 = \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{v}_0 \end{bmatrix}.$$

The solution of (22) can be written, for all $t \in \mathbb{R}$, as

$$\mathbf{w}(t) = e^{t\mathcal{B}}\,\mathbf{w}_0 = \mathcal{Q}(t)\,\mathbf{w}_0\,,$$

where the matrix $\mathcal{Q}(t) = e^{t\mathcal{B}}$ is orthogonal for all $t \in \mathbb{R}$ (since $\mathcal{B}$ is skew-symmetric) . The explicit form of $\mathcal{Q}(t)$ can be given in terms of *generalized matrix functions* of B; see subsection 6.7. The efficient computation of the solution $\mathbf{w}(t) = e^{t\mathcal{B}}\,\mathbf{w}_0$ for large networks has been discussed in [13], see also [44] for related work.

The Schrödinger equation has also been proposed for the analysis of networks. One of the simpler forms of this equations is

$$i\,\frac{d\mathbf{u}}{dt} = L\mathbf{u}, \quad \mathbf{u}(0) = \mathbf{u}_0\,, \tag{23}$$

where $i = \sqrt{-1}$ is the imaginary unit and the initial vector $\mathbf{u}_0$ satisfies $\|\mathbf{u}_0\|_2 = 1$; see [132].

The solution of (23) is given, for all $t \in \mathbb{R}$, by

$$\mathbf{u}(t) = e^{-itL}\,\mathbf{u}_0 = U(t)\,\mathbf{u}_0\,,$$

where now $U(t) = e^{-itL}$ is unitary for all t, since $-itL$ is skew-Hermitian. The family $\{U(t)\}_{t\in\mathbb{R}}$ is a continuous one-parameter group of unitary operators: note that $\|\mathbf{u}(t)\|_2 = 1$ for all t, since the Euclidean norm is unitarily invariant. The state vector $\mathbf{u}(t)$ yields the probability distribution vector

$$\mathbf{p}(t) = [\,|u_1(t)|^2\,, |u_2(t)|^2\,, \ldots\,, |u_n(t)|^2\,]\,,$$

where $|u_i(t)|^2$ is the probability of finding the system in state $\mathbf{e}_i$ at time t. This can be interpreted as a random walk on G for a particle obeying the laws of quantum mechanics; to say that the system is in state $\mathbf{e}_i$ at time t means that the particle (or, rather, its wavefunction) described by the Hamiltonian L is localized at node i at time t, starting from the probability distribution given by $\mathbf{u}_0$. The matrix formed with the *amplitudes* $|[U(t)]_{ij}]|^2$ is doubly stochastic and gives the transition probabilities from one

state to another. In other words, its entries represent the probability of finding the "quantum walker" at node j at time t given that it was at node i at time 0. The interest in the solution of the Schrödinger equation on graphs stems from the fact that it can be used to characterize graph structure, see [132]; it has also important applications in quantum information science [96].

6.5 | Graph energy

The concept of *graph energy* originates from quantum chemistry (Hückel's molecular orbital theory). In this theory the energy levels of conjugated hydrocarbons are given by the eigenvalues of the Hamiltonian matrix, which is of the form $H = \alpha I + \beta A$, where A is the adjacency matrix of the *skeleton graph* describing the arrangement of the carbon atoms in the molecule and $\alpha, \beta < 0$ (see, e.g., [131] for an accessible treatment). The smallest eigenvalue (ground state) of the system is given by $\alpha + \beta\lambda_{\max}(A)$; it is non-degenerate (that is, simple) as long as the graph is connected, by the Perron-Frobenius Theorem. The sum of the eigenvalues of H corresponding to the positive eigenvalues of A (the *occupied states*) is the energy of the system. The parameter α is largely irrelevant and can be set to zero. Hence, the energy of the system is essentially determined by the sum of the negative eigenvalues of $H = \beta A$. Since $\beta < 0$, the energy is proportional to the sum of the positive eigenvalues of A.

The *energy* of an undirected graph G with adjacency spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$ is defined as

$$\mathcal{E}(G) = \sum_{i=1}^{n} |\lambda_i| = \mathrm{Tr}(|A|)\,,$$

where $|A|$ is the *absolute value* of A; see [102] for a comprehensive treatment. This matrix function is obtained from the eigendecomposition of A by taking the absolute values of the eigenvalues, and can also be defined as $|A| = (A^2)^{\frac{1}{2}}$, where the positive square root is taken. Note that since A has zero trace, we have that

$$\mathcal{E}(G) = 2\sum_{\lambda_i > 0} \lambda_i\,, \tag{24}$$

i.e., the energy of G is twice the sum of the positive eigenvalues of the adjacency matrix. Hence, up to the constant factor 2β this notion of energy is essentially that encountered in Hückel's molecular orbital theory (with α set to zero), thus justifying its name.

By expanding the matrix function $(A^2)^{\frac{1}{2}}$ in a binomial series, it is easy to see that only closed walks of even length contribute to the energy of G; a chemical interpretation of this fact can be found in [57], together with bounds on the graph energy for certain types of graphs. Recently, it has been suggested that the vertex energies, i.e., the diagonal entries of $|A|$, could be used to define node centrality measures in undirected networks [7]. Here we observe that in the case of directed networks there are two candidate matrices to represent the matrix function $f(A) = |A|$, namely,

$$|A| = (AA^T)^{\frac{1}{2}} \qquad \text{and} \qquad |A| = (A^TA)^{\frac{1}{2}}.$$

These two matrices have the same trace:

$$\mathrm{Tr}\left[(AA^T)^{\frac{1}{2}}\right] = \mathrm{Tr}\left[(A^TA)^{\frac{1}{2}}\right] = \sum_{i=1}^{r} \sigma_i(A)\,,$$

where $r = \mathrm{rank}\,(A)$ and $\sigma_i(A)$ denotes the ith singular value of A; therefore, they define the same notion of graph energy for a digraph (which is also known as the *nuclear norm* of A, see [113]). The diagonal entries of these two matrices, however, differ; a natural interpretation of the diagonal entries of the first one is to regard them as hub scores, while those of the second one can be interpreted as authority scores. As we have seen, this idea has been proposed in [22] for the exponential and the resolvent.

The lack of smoothness of the function $f(x) = |x|$ for $x = 0$ implies that computing $|A|$ (or even just selected entries of it, such as the diagonal ones) will lead to difficulties. In particular, since the adjacency matrix of a (non-empty graph) is always indefinite, $0 \in (\lambda_{\min}(A), \lambda_{\max}(A))$ and methods based on polynomial approximations of $|x|$ will converge slowly: the best approximation error of the function $f(x) = |x|$ with polynomials of degree n is $O(n^{-1})$ for $n \to \infty$. Similar difficulties can be expected in the computation of $\sqrt{B}$ with $B = A^2$, since the square root is not differentiable at $x = 0$; note that adjacency matrices are very often singular (and therefore $0 \in \sigma(B)$). Rational approximation methods, in contrast, can converge much faster (see [117, 134],) but they have higher cost. This is an area of research still under development.

We also observe that (24) can be written as

$$\mathcal{E}(G) = 2\,\mathrm{Tr}\,[h(A)]\,, \tag{25}$$

where $h(\cdot)$ is the step function that takes the value $h(x) = 0$ for $x \le 0$ and $h(x) = 1$ for $x > 0$. Note that $h(A)$ is just the orthogonal projector onto the invariant subspace of A spanned by the eigenvectors associated with the positive eigenvalues. If

the first positive eigenvalue $\lambda_k(A)$ of A is not too close to the eigenvalue to its immediate left, this discontinuous function can be well approximated by a smooth (sigmoid) function, which itself can be approximated by a variety of methods. The wider the "spectral gap" is, the faster the convergence of polynomial approximations to $h(A)$ will be; we refer to [21] for a discussion of this problem in the context of electronic structure computations in quantum chemistry.

6.6 | Gaussian matrix function

Most of the matrix functions considered so far tend to depend heavily on the extreme eigenvalues of A (or of L, in the case of functions of the graph Laplacian). Thus, most of the information contained in the remaining eigenvalues (and corresponding eigenvectors) is lost. There are, however, situations where the interior eigenvalues play an important role: these include the study of magnetic properties of certain molecules arising as a result of spin interactions, and next-near neighbor (NNN) interactions modeling the competition between agents that share the same resources in a complex system. In [2, 55], the following function of the adjacency matrix A of an undirected graph G has been studied:

$$f(A;\beta;\lambda_{\text{ref}}) = \sum_{k=0}^{\infty}(-\beta)^k \frac{\left((\lambda_{\text{ref}}I - A)^2\right)^k}{k!} = \exp\left[-\beta\,(\lambda_{\text{ref}}I - A)^2\right], \tag{26}$$

where $\beta > 0$ is a parameter and λ_{ref} denotes a reference value, such as an interior eigenvalue of A or an approximation of such an eigenvalue. This matrix function is designed to emphasize the contribution coming from the eigenvalues and eigenvectors closest to the reference value λ_{ref}; the larger is β, the more weight this contribution is given relative to the rest of the spectrum. Values like $\lambda_{\text{ref}} = 0$ and $\lambda_{\text{ref}} = -1$ are of particular interest in applications, see [55] and [2]. The choice of imaginary values $\beta = it$, with $t \in \mathbb{R}$ representing time, is also useful in certain physical applications; see [2].

6.7 | Generalized matrix functions

Generalized matrix functions have been introduced at the end of subsection 2.2. Let

$$A = U_r \Sigma_r V_r^T$$

be the compact SVD of A. Let $f : \mathbb{R}^+ \to \mathbb{R}$ be a scalar function such that $f(\sigma_i)$ is defined for all $i = 1, 2, \ldots, r$. Recall that the *generalized matrix function induced by* f is defined as

$$f^\diamond(A) = U_r f(\Sigma_r) V_r^T, \tag{27}$$

where

$$f(\Sigma_r) = \begin{bmatrix} f(\sigma_1) & & & \\ & f(\sigma_2) & & \\ & & \ddots & \\ & & & f(\sigma_r) \end{bmatrix}.$$

We observe that $f^\diamond(A)$ reduces to the standard matrix function $f(A)$ whenever A is symmetric positive definite, or when A is symmetric positive semidefinite and f satisfies $f(0) = 0$. We also note that the rank of $f^\diamond(A)$ can never exceed r, the rank of A, a property which does not hold, in general, for standard matrix functions.

Generalized matrix functions were first introduced in [89] but they have long been overlooked. Only recently they have undergone intensive study, largely motivated by applications in data science and network analysis; see [5, 9, 23, 114] for recent work on the theory and applications of this kind of matrix functions. Efficient numerical methods for the evaluation of generalized matrix functions and related quantities have been developed in [10, 13]; see also the related paper [14]. Here we discuss some situations where generalized matrix functions occur naturally, with particular reference to network-related applications. It should be mentioned that in several of these papers the authors were not aware of the notion of generalized matrix function, a connection which was only noted in [9].

In [44], the authors address the problem of computing functions of real skew-symmetric matrices, in particular the evaluation of the product $\mathrm{e}^A b$ for a given skew-symmetric matrix A and vector b using the Lanczos algorithm. The authors observe that any $A \in \mathbb{R}^{2n\times 2n}$ with $A^T = -A$ is orthogonally similar to a matrix of the form

$$\mathcal{B} = \begin{bmatrix} 0 & -B \\ B^T & 0 \end{bmatrix},$$

where B is lower bidiagonal of order n. As a consequence, if $B = U\Sigma V^T$ is an SVD of B, the matrix exponential e^A is orthogonally similar to the matrix

$$\begin{bmatrix} U\cos(\Sigma)U^T & -U\sin(\Sigma)V^T \\ V\sin(\Sigma)U^T & V\cos(\Sigma)V^T \end{bmatrix}, \tag{28}$$

where the matrix in the upper right block is precisely $-\sin^\diamond(B)$. Note that this formula retains meaning even when B is rectangular. The authors of [44] develop computational techniques for the matrix exponential based on (28). We also mention that in the same paper the authors derive a similar expression, also found in [22], for the exponential of the symmetric matrix

$$\mathcal{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}. \tag{29}$$

These expressions are extended to more general matrix functions in [45], where it is shown how these ideas can be used to develop efficient geometrical integrators for the numerical solution of certain Hamiltonian differential systems. Here we observe that these techniques can be applied to the integration of the wave equation (21) on networks already discussed in subsection 6.4. The same equation is discussed in [13], where generalized Chebyshev polynomials are used to approximate the solution; the results in [13] show that for large networks, the Chebyshev-based approach is much faster and requires much less storage than the Lanczos-based one for this type of problem.

In [39], the authors consider the problem of detecting (approximate) directed bipartite communities in directed networks. Consideration of alternating walks in the underlying graph leads them to introducing a "non-standard matrix function" of the form

$$f(A) = I - A + \frac{AA^T}{2!} - \frac{AA^TA}{3!} + \frac{AA^TAA^T}{4!} - \cdots,$$

where A is the adjacency matrix of the graph. Using $A = U\Sigma V^T$ this expression is readily recognized to be equivalent to

$$f(A) = U\cosh(\Sigma)U^T - U\sinh(\Sigma)V^T,$$

which is a "mixture" of the standard matrix function $\cosh(\sqrt{AA^T})$ and the generalized matrix function $\sinh^\diamond(A)$. This appears to be the first paper in which a generalized matrix function (albeit not recognized as such) occurs in connection to a network problem.

Generalized hyperbolic matrix functions were also considered in [22] in the context of directed networks, again based on the notion of alternating walks in directed graphs. The main goal of that paper was the extension of centrality measures for undirected networks (such as subgraph centrality) to directed networks by means of the bipartite model of a digraph. We recall that in this model the bipartite graph $\mathcal{G}$ associated with a digraph G is the undirected graph whose adjacency matrix is given by (29), where A is the (nonsymmetric) adjacency matrix of G. The centrality measures introduced in [22] extend the HITS hub/ranking method by Kleinberg [97] in the same way that subgraph and Katz centrality can be regarded as extensions of eigenvector centrality.

Based on this work, in [9], the action of generalized matrix functions on a vector of all ones was used to define certain centrality measures for nodes in directed graphs, as follows.

Definition 1. Let $A = U_r\Sigma_r V_r^T$ be the adjacency matrix of a directed network. We call *total hub communicability* of node i the quantity

$$C_h(i) = \mathbf{e}_i^T \sinh^\diamond(A)\,\mathbf{1} = \sum_{k=1}^{r} \sinh(\sigma_k)(\mathbf{v}_k^T\mathbf{1})u_k(i)$$

and *total authority communicability* of node j the quantity

$$C_a(j) = \mathbf{1}^T \sinh^\diamond(A)\,\mathbf{e}_j = \sum_{k=1}^{r} \sinh(\sigma_k)(\mathbf{u}_k^T\mathbf{1})v_k(j)$$

These quantities correspond to row or column sums of the off-diagonal blocks of $e^{\mathcal{A}}$; therefore, $C_h(i)$ quantifies the ability of node i — playing the role of hub — to communicate with all the nodes in the network, when they are all acting as receivers of information. Similarly, $C_a(j)$ accounts for the ability of node j as an authority to receive information from all the nodes in the graph, when they are acting as broadcasters of information. In [9] it was shown how to use these centrality measures in order to modify a given directed network so as to tune its communicability properties, a problem already studied in [8] for the case of undirected networks. Fast algorithms for estimating these network metrics have been presented in [10] and [13].

Finally, we mention that generalized matrix functions also arise when *filter factors* are used to regularize discrete ill-posed problems. Several additional applications of generalized matrix functions (under the name of "singular functional calculus") can be found in [5]. For an extension to tensors, see [105].

7 | NUMERICAL METHODS

For small to moderate size networks (say, up to a few thousand nodes), it is possible to compute each individual entry of the matrix function $f(A)$ or $f(L)$ using a variety of methods. For undirected graphs, using the eigendecomposition of the matrix is usually the easiest way to carry out the desired calculation. For directed graphs the computation of matrix functions is somewhat more delicate, but a number of techniques are available for handling small to moderate size problems, see [90]. Similarly, for generalized matrix functions one can simply compute the (compact) SVD of A to obtain $f^\circ(A)$.

In the case of large-scale networks, on the other hand, it is generally impossible to compute every entry of a matrix function. Even if A is sparse, the matrix $f(A)$ is generally a dense matrix and therefore impossible to store for large n. Fortunately, in the case of large-scale graphs it is seldom required to compute all the entries of a matrix function; as we have seen, typical calculations require computing the *action* of a matrix function on a vector, or a subset of selected entries (for example the ones on the main diagonal). For these tasks there are efficient algorithms which use the matrix A (or L) only in the form of matrix-vector products. Moreover, such computations may not need to be very accurate: for instance, centrality measures are generally used to identify the top-ranked nodes, so there is no need to compute the centrality indices of all the nodes with high accuracy. Bounds on selected entries of $f(A)$ are often sufficient. In this section we review some numerical methods that are especially well-suited to network analysis.

7.1 | Krylov subspace methods

Krylov subspace methods are widely used for solving many large-scale linear algebra problems, including linear systems, eigenvalue problems, matrix equations and, last but not least, approximation of $f(A)\mathbf{v}$. The main idea behind this class of methods goes as follows:

- define a sequence $\{\mathcal{K}_k\}$ of suitable, low-dimensional nested subspaces of $\mathbb{R}^n$; these are called *Krylov subspaces*;
- consider the projection of the original problem onto these subspaces, yielding a sequence of approximations to the original problem that are more and more accurate as the subspace dimension grows larger;
- solve the projected problems "exactly" on subspaces of increasing dimension, until sufficient accuracy is attained;
- expand the computed solution back to $\mathbb{R}^n$.

The classical definition of Krylov subspaces associated with a square matrix A and a nonzero vector $\mathbf{v}$ is

$$\mathcal{K}_k(A, \mathbf{v}) = \text{span}\{\mathbf{v},\ A\mathbf{v},\ \dots,\ A^{k-1}\mathbf{v}\},$$

for $k = 1, 2, \dots$. With this choice, Krylov approximations of $f(A)\mathbf{v}$ take the form $p(A)\mathbf{v}$, where $p(x)$ is a polynomial, usually of degree much less than n. In fact it holds $\mathcal{K}_k(A, \mathbf{v}) = \{q(A)\mathbf{v} \mid q(x) \text{ is a polynomial of degree} \le k-1\}$.

Let $\psi_{A,\mathbf{v}}$ be the *minimal polynomial of A with respect to* $\mathbf{v}$, that is, the monic polynomial of lowest degree such that $\psi_{A,\mathbf{v}}(A)\mathbf{v} = \mathbf{0}$, and set $d = \deg \psi_{A,\mathbf{v}}$. Then we have

$$\mathcal{K}_1(A, \mathbf{v}) \subset \mathcal{K}_2(A, \mathbf{v}) \subset \cdots \subset \mathcal{K}_d(A, \mathbf{v}) = \mathcal{K}_{d+1}(A, \mathbf{v}) = \cdots = \mathcal{K}_n(A, \mathbf{v}).$$

Equivalently, d is the smallest integer such that $\mathcal{K}_d(A, \mathbf{v}) = \mathcal{K}_{d+1}(A, \mathbf{v})$.

Well-known examples of Krylov subspace methods designed to solve linear systems include the conjugate gradient method and its various generalizations and extensions, such as MINRES for the symmetric case, GMRES and BiCG for the nonsymmetric case [103, 125].

Arnoldi's method is an efficient way to compute an orthonormal basis of the Krylov subspace $\mathcal{K}_k(A, \mathbf{v})$, with $\mathbf{v} \neq \mathbf{0}$, as well as the projection of A with respect to this subspace. Arnoldi's method is based on the classical Gram-Schmidt orthogonalization process and can be summarized as follows:

- Set $\mathbf{q}_1 = \mathbf{v}/\|\mathbf{v}\|_2$;
- For $j = 1, \dots, k$ do:
  - $h_{i,j} = \mathbf{q}_j{}^T A\mathbf{q}_i, \ \text{for } i = 1, 2, \dots, j$
  - $\mathbf{u}_j = A\mathbf{q}_j - \sum_{i=1}^{j} h_{i,j}\mathbf{q}_i$

- $h_{j+1,j} = \|\mathbf{u}_j\|_2$
- If $h_{j+1,j} = 0$ then STOP;
- $\mathbf{q}_{j+1} = \mathbf{u}_j / \|\mathbf{u}_j\|_2$

If the algorithm does not stop before the k-th step, then the Arnoldi vectors $\{\mathbf{q}_1, \dots, \mathbf{q}_k\}$ form an orthonormal basis of $\mathcal{K}_k(A, \mathbf{v})$. The Arnoldi process produces the factorization

$$AQ_k = Q_k H_k + h_{k+1,k}\mathbf{q}_{k+1}\mathbf{e}_k{}^T$$

where $Q_k = [\mathbf{q}_1, \dots, \mathbf{q}_k] \in \mathbb{R}^{n\times k}$ and H_k is the $k \times k$ upper Hessenberg matrix defined as $H_k = (h_{i,j})_{i,j=1}^k$. Moreover, the following relation holds:

$$Q_k^T A Q_k = H_k,$$

implying that H_k is the orthogonal projection of A onto $\mathcal{K}_k(A, \mathbf{v})$. If A is real symmetric, the associated H_k is also symmetric and therefore tridiagonal. In this case, the Arnoldi process reduces to the *Lanczos process*, which is based on a three-term recurrence and is therefore cheaper than Arnoldi in complexity and storage.

A crucial feature of the Arnoldi/Lanczos process is that the matrix A is not needed explicitly; what is needed is the action of A on a given vector. This property makes Krylov methods especially well-suited to solving problems where the matrix A is large and sparse, so that matrix-vector products can be computed cheaply. It should be mentioned that the practical implementation of the Arnoldi process requires great care to avoid loss of orthogonality of the Arnoldi vectors and to select appropriate restart techniques and stopping criteria; discussion of these issues, however, is outside the scope of this work.

One widespread application of the Lanczos/Arnoldi processes is in large-scale eigenvalue computation. Since k is small, the eigenvalues of H_k, called *Ritz values*, can be computed explicitly. A few of the Ritz values typically give a good approximation of the extreme eigenvalues of A, with the quality of the approximation improving as k increases; see e.g., [126] for a presentation of Lanczos/Arnoldi eigenvalue convergence theory.

Here we focus on application of the Arnoldi/Lanczos process to the approximation of $f(A)\mathbf{v}$. Once k steps of Arnoldi's process have been carried out, an approximation of $f(A)\mathbf{v}$ can be computed as

$$f(A) \approx \|\mathbf{v}\|_2 Q_k f(H_k)\mathbf{e}_1 = Q_k f(H_k) Q_k^T \mathbf{v}. \tag{31}$$

Since $f(A)\mathbf{v} \in \mathcal{K}_d(A, \mathbf{v})$, the approximation (31) is exact if $k = \deg \psi_{A,\mathbf{v}}$. But in practice a good approximation of $f(A)\mathbf{v}$ may be obtained for a suitable choice of $k \ll n$, even though $k < d$. Since k is small, the explicit computation of $f(H_k)$ is inexpensive. Note that, in general, $f(H_k)$ might not be defined; however, sufficient conditions are available to ensure that it is [90]. In particular, if $f(z)$ and its first n derivatives are defined at all points z within the field of values of A, then $f(H_k)$ is well-defined. We recall that the field of values of a matrix $A \in \mathbb{C}^{n\times n}$ is the set of all complex numbers of the form $z = \mathbf{x}^* A\mathbf{x}$ as $\mathbf{x}$ varies among all vectors in $\mathbb{C}^n$ with $\|\mathbf{x}\|_2 = 1$.

Deciding when (31) is a sufficiently accurate approximation of $f(A)\mathbf{v}$ is a nontrivial task. Roughly speaking, the quality of the approximation depends heavily on the spectrum of A and on how accurately it is approximated by the first few steps of the Arnoldi process. A result of Saad generalized by Higham [90] gives some insight into the convergence process: it states that

$$\|\mathbf{v}\|_2 Q_k f(H_k)\,\mathbf{e}_1 = \tilde{p}_{k-1}(A)\mathbf{v}.$$

where $\tilde{p}_{k-1}$ is the unique polynomial of degree at most $k-1$ that interpolates f on the spectrum of H_k.

For the case $f(A) = \mathrm{e}^A$, Saad [126] suggests the error estimate

$$\|\mathrm{e}^A\mathbf{v} - Q_k \mathrm{e}^{H_k} Q_k^T \mathbf{v}\|_2 \approx \|\mathbf{v}\|_2 \mathbf{e}_k^T \mathrm{e}^{H_k}\mathbf{e}_1\,.$$

Hence, the approximation error can be estimated by monitoring the $(k, 1)$ entry of e^{H_k}. See also [91] and [137] for more results on approximation of $\mathrm{e}^A\mathbf{v}$. We refer to [90, 79] for a general discussion of polynomial Krylov method for computing matrix functions.

As previously mentioned, polynomial approximations may converge very slowly for certain types of functions, in which case one may want to consider rational approximations [117, 134]. The approach outlined so far can be generalized to a rational approximation framework. Fix a sequence of complex or infinite poles $\xi_1, \dots, \xi_m$; then the associated *rational Krylov subspace* of order $m+1$ is defined as $\mathcal{K}_{m+1}(A, \mathbf{v}, q_m) = q_m(A)^{-1}\mathrm{span}\{\mathbf{v},\ A\mathbf{v},\ \dots,\ A^m\mathbf{v}\}$, where $q_m(x) = \prod_{j=1,\xi_j\neq\infty}^m (x - \xi_j)$. Similarly to the standard (polynomial) case, a rational Krylov method computes an orthonormal basis matrix V_{m+1} for the space $\mathcal{K}_{m+1}(A, \mathbf{v}, q_m)$. This basis matrix satisfies a rational Arnoldi decomposition of the form

$$AV_{m+1}K_m = V_{m+1}H_m,$$

with respect to an unreduced upper Hessenberg pencil (H_m, K_m) of size $(m+1) \times m$. See, e.g., [88] for a review of rational Krylov methods for the computation of $f(A)\mathbf{v}$, and the MATLAB toolbox RKToolbox [27] for an efficient implementation.

Finally, we mention that yet another approach for estimating $f(A)\mathbf{v}$, based on *vector extrapolation*, has been recently investigated in [106].

7.2 | Methods based on quadrature formulas

As was mentioned above, the evaluation or estimation of selected entries or averages of entries of $f(A)$ is often required in network analysis. These quantities can be written in the form

$$\mathbf{u}^T f(A)\mathbf{v},$$

for specific choices of the vectors $\mathbf{u}$ and $\mathbf{v}$. For instance, since $f(A)_{ij} = \mathbf{e}_i^T f(A)\mathbf{e}_j$, we have for exponential-based centrality and communicability measures:

- subgraph centrality: $SC(i,\beta) = [\mathrm{e}^{\beta A}]_{i,i} = \mathbf{e}_i^T \mathrm{e}^{\beta A}\mathbf{e}_i$,
- subgraph communicability: $[\mathrm{e}^{\beta A}]_{i,j} = \mathbf{e}_i^T \mathrm{e}^{\beta A}\mathbf{e}_j$,
- hub centrality: $\mathbf{e}_i^T \cosh\left(\sqrt{AA^T}\right)\mathbf{e}_i$,
- authority centrality: $\mathbf{e}_i^T \cosh\left(\sqrt{A^TA}\right)\mathbf{e}_i$,
- total node communicability: $TC(i,\beta) = \mathbf{e_i}^T \mathrm{e}^{\beta A}\mathbf{1}$,

and similar expressions for resolvent- and Katz-based measures. In this section we focus on the case of *undirected*, possibly weighted networks.

Golub and Meurant developed in [82, 83] an elegant technique for estimating $\mathbf{u}^T f(A)\mathbf{v}$, that exploits the connection between Gauss-type quadrature rules, the symmetric Lanczos process and the properties of orthogonal polynomials. The first application of this methodology to computations of centrality and communicability measures for networks goes back to [20]. We recall here the main ideas. We will first need to estabiish a condition on the derivatives of f.

Definition 2. A real function $f(x)$ is *strictly completely monotonic* (s.c.m.) on an interval $I \subset \mathbb{R}$ if $f^{(2j)}(x) > 0$ and $f^{(2j+1)}(x) < 0$ on I for all $j \geq 0$.

For instance, $f(x) = 1/x$ is s.c.m. on $(0,\infty)$ and $f(x) = \mathrm{e}^{-x}$ is s.c.m. on $\mathbb{R}$, although $f(x) = \mathrm{e}^x$ is not. Therefore the functions that are most popular in matrix-based network analysis, that is, the exponential and the resolvent function, are s.c.m. on suitable subsets of the real line, or can be rewritten as such, for example using $\mathrm{e}^x = \mathrm{e}^{-(-x)}$.

Now, suppose that the real $n \times n$ matrix A is symmetric and that the function $f(x)$ is s.c.m. on an interval containing the spectrum of A. Consider the spectral decompositions $A = Q\Lambda Q^T$ and $f(A) = Qf(\Lambda)Q^T$. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ we have

$$\mathbf{u}^T f(A)\mathbf{v} = \mathbf{u}^T Q f(\Lambda) Q^T \mathbf{v} = \mathbf{p}^T f(\Lambda)\mathbf{q} = \sum_{i=1}^{n} f(\lambda_i) p_i q_i,$$

where $\mathbf{p} = Q^T\mathbf{u}$ and $\mathbf{q} = Q^T\mathbf{v}$. This expression can be reformulated as a Riemann-Stieltjes integral with respect to the discrete (signed) measure $\mu(\lambda)$:

$$\mathbf{u}^T f(A)\mathbf{v} = \int_a^b f(\lambda)\, d\mu(\lambda), \quad \mu(\lambda) = \begin{cases} 0 & \lambda < a = \lambda_1 \\ \sum_{j=1}^{i} p_j q_j & \lambda_i \leq \lambda < \lambda_{i+1} \\ \sum_{j=1}^{n} p_j q_j & b = \lambda_n \leq \lambda. \end{cases}$$

Here we are assuming that the eigenvalues of A are numbered in non-decreasing order: $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Note that for $\mathbf{u} = \mathbf{v}$, the measure $\mu(\lambda)$ is positive. The general Gauss-type quadrature rule with N nodes is written as

$$\int_a^b f(\lambda)\, d\mu(\lambda) = \sum_{j=1}^{N} w_j f(t_j) + \sum_{k=1}^{M} v_k f(z_k) + R[f],$$

where $R[f]$ denotes the remainder, for which explicit formulas depending on higher derivatives of $f(x)$ are available. The nodes $\{t_j\}_{j=1}^N$ and the weights $\{w_j\}_{j=1}^N$ are unknown, whereas the M nodes $\{z_k\}_{k=1}^M$ are prescribed. In particular, if $M = 0$ we obtain

the Gauss rule; the choice $M = 1$ and $z_1 = a$ or $z_2 = b$ yields the Gauss-Radau rule, whereas if $M = 2$, $z_1 = a$ and $z_2 = b$ we have the Gauss-Lobatto rule.

The s.c.m. hypothesis on $f(x)$ implies that the sign of $R[f]$ is known. Therefore these rules yield upper/lower bounds on $\mathbf{u}^T f(A)\mathbf{v}$. Indeed, the Gauss rule gives a lower bound, the Gauss-Lobatto rule gives an upper bound, whereas the Gauss-Radau rule can be used to obtain both a lower and an upper bound. If, on the other hand, $f(x)$ is not s.c.m, Gauss-type quadrature rules can only provide estimates of bilinear forms.

It is not necessary to compute Gauss nodes and weights explicitly. The evaluation of these quadrature rules is reduced to computation of orthogonal polynomials via three-term recurrence, or, equivalently, computation of entries and spectral information of the associated tridiagonal matrix via Lanczos' method. Indeed, Golub and Meurant [83] prove the following relation:

$$\sum_{j=1}^{N} w_j f(t_j) = \mathbf{e}_1^T f(J_N)\mathbf{e}_1 = [f(J_N)]_{11},$$

where the tridiagonal matrix J_N corresponds to the three-term recurrence relationship satisfied by the set of polynomials orthonormal with respect to $d\mu$:

$$J_N = \begin{bmatrix} \omega_1 & \gamma_1 & & & \\ \gamma_1 & \omega_2 & \gamma_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \gamma_{N-2} & \omega_{N-1} & \gamma_{N-1} \\ & & & \gamma_{N-1} & \omega_N \end{bmatrix}$$

The eigenvalues of J_N are the Gauss nodes, whereas the Gauss weights are given by the squares of the first entries of the normalized eigenvectors of J_N.

Consider for instance the case $\mathbf{u} = \mathbf{v} = \mathbf{e}_i$, which corresponds to computing the (i, i) entry of $f(A)$. The entries of J_N can be computed using the symmetric Lanczos algorithm:

$$\begin{aligned} &\gamma_j \mathbf{x}_j = \mathbf{r}_j = (A - \omega_j I)\mathbf{x}_{j-1} - \gamma_{j-1}\mathbf{x}_{j-2}, \qquad j = 1, 2, \dots \\ &\omega_j = \mathbf{x}_{j-1}^T A \mathbf{x}_{j-1}, \\ &\gamma_j = \|\mathbf{r}_j\|_2 \end{aligned}$$

with initial vectors $\mathbf{x}_{-1} = \mathbf{0}$ and $\mathbf{x}_0 = \mathbf{e}_i$.

The case $\mathbf{u} \neq \mathbf{v}$ can be handled using the *polarization identity*:

$$\mathbf{u}^T f(A)\mathbf{v} = \frac{1}{4}\left[\mathbf{w}^T f(A)\mathbf{w} - \mathbf{z}^T f(A)\mathbf{z}\right], \quad \mathbf{w} = \mathbf{u} + \mathbf{v}, \quad \mathbf{z} = \mathbf{u} - \mathbf{v}.$$

Setting $\mathbf{u} = \mathbf{e}_i$, $\mathbf{v} = \mathbf{e}_j$ we obtain bounds for the (i, j) entry of $f(A)$ at the cost of evaluating two quadratic forms. Other possible approaches are discussed in [83].

One way of obtaining *a priori* upper and lower bounds on $[f(A)]_{ij}$ consists in carrying out a single step of Lanczos' algorithm by hand; for quantities like the Estrada index, these bounds are orders of magnitude better than those previously found in literature [20]. Alternatively, one may explicitly perform a certain number of Lanczos iterations: each additional Lanczos step corresponds to adding another node to the chosen Gauss-type rule, resulting in tighter and tighter bounds. A MATLAB implementation is provided in the MMQ toolbox by Meurant [104]. If A is sparse, arithmetic complexity per iteration is typically $O(n)$ or less, and experimental evidence suggests that for many types of networks the number of iterations required to achieve a given accuracy exhibits a weak dependence on n.

The standard Golub-Meurant approach may become expensive when n is large and f-centrality values are needed for several nodes. For instance, suppose one needs to determine a few nodes with the largest f-subgraph centrality in a large graph: this task would require the computation of upper and lower bounds for all the diagonal entries of $f(A)$. In such cases, one may apply techniques based on low rank approximation, as suggested in [71] for undirected graphs. For $k \ll n$, one can estimate the k largest eigenvalues $\lambda_1, \dots, \lambda_k$ (now numbered in non-increasing order) and the associated eigenvectors $\mathbf{q}_1, \dots, \mathbf{q}_k$ using the block Lanczos algorithm [83]. This partial spectral factorization yields the approximations $A \approx \sum_{i=1}^{k} \lambda_i \mathbf{q}_i \mathbf{q}_i^T$ and $f(A) \approx \sum_{i=1}^{k} f(\lambda_i)\mathbf{q}_i \mathbf{q}_i^T$. Using the latter approximation, one can select a (small) set of nodes that contains the nodes of interest for the original problem, and then apply quadrature rules to refine the bounds. Note that eigenvector centrality is equivalent to the rank-one approximation $A \approx \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T$. When the spectral gap $\lambda_1 - \lambda_2$ is large, the contribution of the dominant eigenvector to the

exponential-based centrality measures is very significant, and the rankings obtained with these measures tend to be similar to those given by eigenvector centrality. This is not the case when the gap is small; see the experiments reported in subsection 8.3.

In their work, Golub and Meurant also explain how to extend their Gauss-type quadrature approach to the block case. The goal here is approximation of expressions of the form $W^T f(A)W$, where the matrix $W \in \mathbb{R}^{n\times k}$, with $k \ll n$, has orthonormal columns. This block version, however, does not yield bounds on the sought quantities. The extension of Laurie's anti-Gauss quadrature rules to matrix-valued measures and their use in combination with block Gauss rules has been proposed in [72] to compute upper/lower bounds on entries of $f(A)$, both in the symmetric and nonsymmetric case. These techniques can also be applied to cases where the function f is not s.c.m., provided that suitable conditions on the decay of its Taylor coefficients are verified. Further generalizations are provided in [119] and in [4].

We also mention the use of partial SVD in combination with Gauss quadrature rules for the analysis of directed networks. As pointed out in Section 4.5, functions such as the hyperbolic sine and cosine of $\sqrt{A^T A}$ and $\sqrt{AA^T}$ yield centrality and communicability information on the associated directed network. For large n, however, explicity computing the SVD of A is impractical and low-rank approximation techniques may help reduce the computational burden. The approach proposed in [14] first computes a partial singular value decomposition, then determines a subset of "most important" nodes (or nodes between which it is easier to travel), and finally applies quadrature rules to rank nodes in this subset.

When A is nonsymmetric (as in the case of directed networks), it is difficult to extend the approach based on quadrature rules. At present, the best available approach to approximate selected entries $[f(A)]_{ij}$ for large A is perhaps the use of an Arnoldi-based method for the computation of $f(A)\mathbf{e}_j$, followed by extraction of the sought components.

Further developments of quadrature-based techniques for approximation of matrix functions rely on generalized averaged Gauss rules and their block generalizations: see [121, 122].

As mentioned at the end of Section 6.7, a quadrature-based approach in a similar vein as Golub-Meurant (with Golub-Kahan bidiagonalization replacing the Lanczos process) can also be applied to bound or estimate selected entries of generalized matrix functions; see [10]. For functions that are oscillatory, techniques based on Chebyshev polynomial approximation offer a valid alternative [13].

Finally, methods based on extrapolation for the estimation of bilinear forms involving matrix functions have been developed in [74, 75].

8 | SOME EXAMPLES

We illustrate and compare the behavior of different measures of centrality on four examples: a small undirected graph, a small directed graph and two large real-life networks.

8.1 | An example for an undirected graph

Our first example is a network of 15th century Florentine families [33]. Each node corresponds to a family and edges denote marriages between families. The graph is depicted in Figure 1 .

We compute degree, eigenvector, subgraph, total communicability (the last two with parameter $\beta = 1$), total and Katz centrality (with the parameter α chosen approximately at the midpoint of the allowed range). The rankings found by the various centrality measures are collected in Table 2 . While all rankings agree on the preminence of the Medici family, the other families take different positions depending on the chosen measure, although there is a general consensus on which families belong in the top part of the ranking and which ones are instead at the bottom.

8.2 | An example for a directed graph

Let us consider the directed graph G on four nodes with adjacency matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Below we show the digraph G and its associated undirected bipartite graph $\mathcal{G}$.

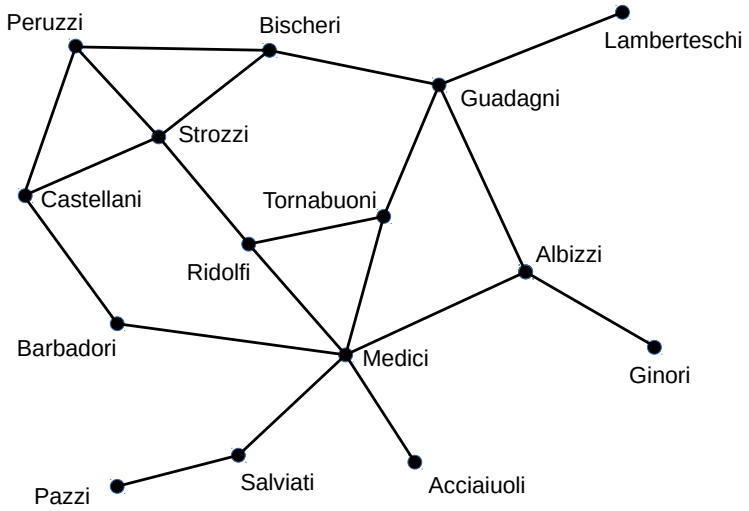

FIGURE 1 Network of 15th century Florentine families (adapted from [33]).

TABLE 2 Rankings for different measures of centrality applied to the Florentine families network.

rank	degree	eigenvector	subgraph	total comm.	Katz
1st	Medici	Medici	Medici	Medici	Medici
2nd	Guadagni	Strozzi	Strozzi	Strozzi	Strozzi
3rd	Strozzi	Ridolfi	Peruzzi	Ridolfi	Guadagni
4th	Peruzzi	Tornabuoni	Tornabuoni	Tornabuoni	Ridolfi
5th	Bischeri	Guadagni	Guadagni	Guadagni	Tornabuoni
6th	Castellani	Bischeri	Ridolfi	Bischeri	Bischeri
7th	Ridolfi	Peruzzi	Bischeri	Peruzzi	Albizzi
8th	Tornabuoni	Castellani	Castellani	Albizzi	Peruzzi
9th	Albizzi	Albizzi	Albizzi	Castellani	Castellani
10th	Barbadori	Barbadori	Barbadori	Barbadori	Barbadori
11th	Salviati	Salviati	Acciaiuoli	Salviati	Salviati
12th	Lamberteschi	Acciaiuoli	Salviati	Acciaiuoli	Acciaiuoli
13th	Ginori	Barbadori	Lamberteschi	Pazzi	Pazzi
14th	Acciaiuoli	Lamberteschi	Ginori	Lamberteschi	Lamberteschi
15th	Pazzi	Ginori	Pazzi	Ginori	Ginori

The matrix A is irreducible (G is strongly connected) with spectral radius $\rho(A) = 1.8393$ (to four decimal digits). If D is the matrix of out-degrees, the matrix $P = A^T D^{-1}$ is column stochastic and its dominant eigenvector, which is the PageRank vector of the network, is given by

$$\mathbf{p} = \begin{bmatrix} 0.1905 & 0.3810 & 0.2857 & 0.1429 \end{bmatrix}^T ,$$

hence the corresponding ranking of the nodes is 2 (top), 3, 1, and 4 (bottom).

The same ranking can be obtained from the diagonal entries of e^A:

$$\mathrm{diag}(e^A) = \begin{bmatrix} 1.8414 & 2.6377 & 2.0287 & 1.2323 \end{bmatrix}^T ,$$

and from the diagonal entries of the Katz resolvent for $\alpha = 0.5$ (note that $\frac{1}{\lambda_1} = 0.5437$):

$$\mathrm{diag}((I - 0.5\,A)^{-1}) = \begin{bmatrix} 3.3333 & 5.3333 & 4.0000 & 2.0000 \end{bmatrix}^T .$$

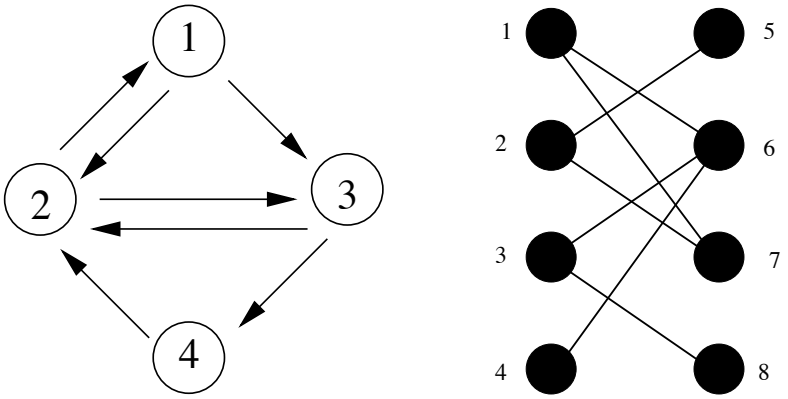

FIGURE 2 A small directed graph (left) and the associated undirected bipartite graph (right).

Hence, on this graph several of the most important centrality measures give consistent results. This is often observed for very small networks.

Things are more interesting if we look at hub and authority scores, i.e., if we distinguish between the role of a node as a broadcaster and as a receiver of information. The simplest such measure is given by the out- and in-degree vectors:

$$\mathbf{d}_{\text{out}} = \begin{bmatrix} 2 & 2 & 2 & 1 \end{bmatrix}^T, \quad \mathbf{d}_{\text{in}} = \begin{bmatrix} 1 & 3 & 2 & 1 \end{bmatrix}^T.$$

We see that out-degree results in a tie of nodes 1, 2 and 3 for first place as hub, with node 4 ranked last, while the in-degree scores give node 2 as the top authority, followed by node 3, with nodes 1 and 4 tied for last place.

The right and left dominant eigenvectors of A (normalized in the ℓ^2-norm) are given by

$$\mathbf{x}_1 = \begin{bmatrix} 0.5774 & 0.5774 & 0.4846 & 0.3139 \end{bmatrix}^T, \quad \mathbf{y}_1 = \begin{bmatrix} 0.3659 & 0.6729 & 0.5648 & 0.3071 \end{bmatrix}^T.$$

Hence, eigenvector centrality gives the hub ranking 1-2 (tied), 3, 4 and the authority ranking 2, 3, 1, 4 (which is the same ranking obtained with the metrics that do not distinguish between hubs and authorities). Note that neither out-degree, nor (right) eigenvector centrality are able to distinguish node 1 from node 2. Similar results are obtained using total communicability, whether exponential or resolvent-based:

$$\mathrm{e}^A\mathbf{1} = \begin{bmatrix} 6.9901 & 6.9901 & 6.1938 & 4.1112 \end{bmatrix}^T, \quad \mathbf{1}^T\mathrm{e}^A = \begin{bmatrix} 4.7203 & 8.4636 & 6.9901 & 4.1112 \end{bmatrix}$$

and

$$(I - 0.5\,A)^{-1}\mathbf{1} = \begin{bmatrix} 14 & 14 & 12 & 8 \end{bmatrix}^T, \quad \mathbf{1}^T(I - 0.5\,A)^{-1} = \begin{bmatrix} 9.3333 & 16.6667 & 14 & 8 \end{bmatrix}.$$

On the other hand, the SVD-based centrality measures yield somewhat different rankings. The HITS ranking is the one provided by the dominant left and right singular vectors:

$$\mathbf{u}_1 = \begin{bmatrix} 0.6555 & 0.3351 & 0.5422 & 0.4051 \end{bmatrix}^T, \quad \mathbf{v}_1 = \begin{bmatrix} 0.1685 & 0.8058 & 0.4980 & 0.2726 \end{bmatrix}^T.$$

The corresponding hub and authority rankings are given by 1 (top), 3, 4, 2 (bottom) and 2 (top), 3, 4, 1 (bottom), respectively; there are no ties. Note that in spite of some differences, the importance of node 1 as a hub and of node 2 as an authority is revealed by all hub/authority measures. This is further confirmed by looking at the diagonal entries of the exponential of the (symmetric) adjacency matrix $\mathcal{A}$ of the bipartite graph $\mathcal{G}$ associated with G. Indeed, the first four diagonal entries of $\mathrm{e}^{\mathcal{A}}$ are given by

$$\begin{bmatrix} 2.3319 & 2.2289 & 2.2812 & 1.6414 \end{bmatrix}$$

while the last four entries are given by

$$\begin{bmatrix} 1.5906 & 3.0209 & 2.2796 & 1.5922 \end{bmatrix}.$$

Note also the strong agreement of all the centrality measures considered about the bottom ranking of node 4.

We end by commenting that in most applications, network scientists do not rely on a single centrality measure, but often calculate several such measures; the most important nodes are those that score high for a majority of the measures considered. Hence, consensus between different centrality measures is a desirable outcome. In this sense we can safely conclude that for the small digraph G of this example, the top hub node is node 1 and that the top authority node is node 2.

8.3 | Examples of real-world undirected networks

In the first test of this section we use the scientific collaboration network `ca-CondMat` taken from the SuiteSparse Matrix Collection [43]. The associated (symmetric) adjacency matrix has size 23 133×23 133, with 186 936 nonzero entries. The largest eigenvalue is 37.95, the spectral gap is 7.31. Clearly in this case it is crucial to apply numerical methods that exploit the sparsity of the adjacency matrix. We determine:

- degree centrality, computed as $A \cdot \mathbf{1}$,
- eigenvector centrality,
- subgraph centrality, i.e, the vector of diagonal entries of $e^{\beta A}$ computed via quadrature rules [104], with $\beta = 0.5$,
- total communicability $e^{\beta A} \cdot \mathbf{1}$ (with $\beta = 0.5$) computed via Lanczos' method [87],
- Katz centrality $(I - \alpha A)^{-1} \cdot \mathbf{1}$, with $\alpha \approx 0.5/\lambda_1$.

Table 3 shows, for each centrality vector, the indices of the ten top-ranked nodes and the time (in seconds) required for the computation. Note that node 5013 is ranked first in three cases, fourth and fifth in the other two rankings. In general, several nodes feature in most of the rankings, although at different places.

TABLE 3 Rankings and timings (in seconds) for different measures of centrality applied to the `ca-CondMat` network.

	degree	eigenvector	subgraph	total comm.	Katz
1st	15576	5013	5013	5013	11302
2nd	11302	21052	21052	21052	15756
3rd	16897	18746	18746	11302	21052
4th	21052	11302	11302	18746	5013
5th	5013	9872	9872	9872	16897
6th	21833	13768	13768	13768	3352
7th	21068	5500	5500	5500	21833
8th	3352	20667	17245	20667	21068
9th	9886	17245	20667	17245	9886
10th	9876	4081	4081	4081	9876
time	1.39e−3	8.50e−2	178.02	0.102	2.20e−2

For this experiment we have used MATLAB 2018a running under Ubuntu 18.04 on a laptop equipped with a quad-core Intel i3-7130U processor at 2.70 GHz and 4GB RAM. Eigenvector centrality was computed using the `eigs` function in MATLAB (with default setting of the parameters). The subgraph centralities are computed using the Gauss rule (which provides lower bounds) computed using the Lanczos algorithm; the iteration was stopped when the difference between subsequent estimates was below 10^{-6} or when a maximum of 10 steps (quadrature nodes) was reached. The total communicabilities were computed using Güttel's `funm_kryl` code with a stopping accuracy set to 10^{-6}, see [87]. Finally, Katz centrality was obtained with MINRES using MATLAB's built-in function `minres`; convergence was attained in just 9 iterations (the system is very well-conditioned for the chosen value of α). Note that eigenvector, subgraph, and total communicability centrality return very similar (albeit not identical) top-10 lists.

On this example, most ranking measures can be computed very fast. The exception is subgraph centrality, requiring the calculation of all the 23 133 diagonal elements of $e^{\beta A}$, an average of 0.0077 seconds per node. Since all of the centralities can in principle be computed independently of one another, the cost of subgraph centrality could be reduced in a parallel environment. However, given the strong agreement with total communicability (observed in many other cases as well), the latter appears to be preferable for large networks.

As a second example, we show centrality results for the US road network `usroads-48`, also taken from [43]. This is an undirected network with 126 146 nodes. The associated adjacency matrix has 323 900 nonzero entries. The two largest eigenvalues

are $\lambda_1 = 3.9107$ and $\lambda_2 = 3.8403$; therefore the spectral gap $\lambda_1 - \lambda_2 = 0.0705$ is considerably smaller than the spectral gap in the previous example. Table 4 shows rankings and timings for the computation of degree, eigenvector, subgraph, total communicability and Katz centrality. The choice of algorithms and parameters is the same as above; here MINRES converges after 10 iterations with relative residual 4.7e−7.

TABLE 4 Rankings and timings (in seconds) for different measures of centrality applied to the `usroads-48` network.

	degree	eigenvector	subgraph	total comm.	Katz
1st	58080	44182	58080	44182	44182
2nd	1452	44067	44182	55471	55471
3rd	9241	44154	55471	102130	102130
4th	19694	44087	43734	43734	43734
5th	25970	44323	102130	44906	44906
6th	31409	44255	44906	71416	71416
7th	41855	44356	45671	25970	25970
8th	42726	44035	71416	45671	58080
9th	43734	44133	120961	44067	45671
10th	44182	44294	1452	19694	19694
time	2.50e−3	1.10	1.80e+3	0.202	9.93e−2

For this network the agreement among the rankings is noticeably less strong that in the previous example. This behavior is arguably related to the small size of the spectral gap and is consistent with the findings in [25]. Indeed, when the gap is small the non-dominant eigenpairs give a substantial contribution to subgraph, total communicability and Katz centrality, resulting in different rankings from the one obtained with eigenvector centrality.

9 | CONCLUSIONS AND PERSPECTIVES

In this paper we have given a broad overview of the use of matrix functions in network analysis. These include centrality and communicability metrics, and the description of diffusive processes on graphs, including discrete and continuous time random walks. Functions of matrices arise naturally in the statistical mechanics approach to the analysis of complex networks. While most of the literature concerns static networks, matrix functions can also be used in the context of temporal networks.

Overall, the application of matrix functions as a modeling tool in network science is by now well-established. Computational aspects have also been vigorously developed in recent years, in particular Krylov subspace methods have been shown to be quite effective for network science applications. These methods are able to handle both symmetric and unsymmetric problems (corresponding to undirected and directed graphs, respectively) and convergence is often fast. Another recent research direction has concerned the updating or downdating of a matrix function when the network undergoes a low-rank change.

Future work is likely to focus on the development of matrix and tensor functions arising in the study of temporal networks, and on the extension of existing techniques to more sophisticated network models such as multiplex and other multilayer networks. As an example of recent work in this area we refer to [133].

References

[1] M. Akbarzadeh and E. Estrada, *Communicability geometry captures traffic flow in cities*, Nature Human Behaviour, 2 (2018), pp. 645–652.

[2] A. Ali, F. Althukar, and E. Estrada, *Gaussianization of the spectra of graphs and networks. Theory and applications*, J. Math. Anal. Appl., 470 (2018), pp. 876–897.

[3] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin, *Non-backtracking random walks mix faster*, Commun. Contemp. Math., 9 (2007), pp. 585–603.

[4] H. Alqahtani and L. Reichel, *Generalized block anti-Gauss quadrature rules*, Numer. Math., 143 (2019), pp. 605–648.

[5] F. Andersson, M. Carlsson, and K.-M. Perfekt, *Operator-Lipschitz estimates for the singular value functional calculus*, Proc. Amer. Math. Soc., 144 (2016), pp. 1867–1875.

[6] M. Aprahamian, D. J. Higham, and D. J. Higham, *Matching exponential-based and resolvent-based centrality measures*, J. Complex Networks 4 (2016), pp. 157–176.

[7] O. Arizmendi, J. F. Hidalgo, and O. Juarez-Romero, *Energy of a vertex*, Linear Algebra Appl., 557 (2018), pp. 464–495.

[8] F. Arrigo and M. Benzi, *Updating and downdating techniques for optimizing network communicability*, SIAM J. Sci. Comput., 38 (2016), pp. B25–B49.

[9] F. Arrigo and M. Benzi, *Edge modification criteria for enhancing the communicability of digraphs*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 443–468.

[10] F. Arrigo, M. Benzi, and C. Fenu, *Computation of generalized matrix functions*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 836–860.

[11] F. Arrigo, P. Grindrod, D. J. Higham, V. Noferini, *Non-backtracking walk centrality for directed networks*, J. Complex Networks 6.1 (2017): 54–78.

[12] F. Arrigo, P. Grindrod, D. J. Higham, V. Noferini, *On the exponential generating function for non-backtracking walks*, Linear Algebra Appl. 556 (2018), pp. 381–399

[13] J. Aurentz, A. Austin, M. Benzi, and V. Kalantzis, *Stable computation of generalized matrix functions via polynomial interpolation*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 210–234.

[14] J. Baglama, C. Fenu, L. Reichel, and G. Rodriguez, *Analysis of directed networks via partial singular value decomposition and Gauss quadrature*, Linear Algebra Appl., 456 (2014), pp. 93–121.

[15] A. L. Barabasi, *Linked: The New Science of Networks*, Perseus, Cambridge, UK, 2002.

[16] P. Bartesaghi, M. Benzi, G. P. Clemente, R. Grassi, and E. Estrada, *Risk-dependent centrality in economic and financial networks*, arXiv:1907.07908v3, 25 November 2019.

[17] B. Beckermann, D. Kressner, and M. Schweitzer, *Low-rank updates of matrix functions*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 539–565.

[18] M. Benzi, *A note on walk entropies in graphs*, Linear Algebra Appl., 445 (2014), pp. 395-399.

[19] M. Benzi, D. Bertaccini, and F. Durastante, *Fractional diffusion on directed networks*, arXiv: 1912.07288, 16 December 2019.

[20] M. Benzi and P. Boito, *Quadrature rule-based bounds for functions of adjacency matrices*, Linear Algebra Appl., 433 (2010), pp. 637–652.

[21] M. Benzi, P. Boito, and N. Razouk, *Decay properties of spectral projectors with applications to electronic structure*, SIAM Rev., 55 (2013), pp. 3–64.

[22] M. Benzi, E. Estrada, and C. Klymko, *Ranking hubs and authorities using matrix functions*, Linear Algebra Appl., 438 (2013), pp. 2447–2474.

[23] M. Benzi and R. Huang, *Some matrix properties preserved by generalized matrix functions*, Special Matrices, 7 (2019), pp. 27–37.

[24] M. Benzi and C. Klymko, *Total communicability as a centrality measure*, J. Complex Networks, 1 (2013), pp. 124–149.

[25] M. Benzi and C. Klymko, *On the limiting behavior of parameter-dependent network centrality measures*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 686–706.

[26] G. Berkolaiko and P. Kuchment, *Introduction to Quantum Graphs*, American Mathematical Society, Providence, RI, 2013.

[27] M. Berljafa, S. Elsworth, and S. Güttel, *RKToolbox*, http://guettel.com/rktoolbox/index.html

[28] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Classics in Applied Mathematics, 9, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1994.

[29] R. F. Betzel, S. Gu, J. D. Medaglia, F. Pasqualetti and D. S. Bassett, *Optimally controlling the human connectome: the role of network topology*, Scientific Reports, 6 (2016), Article 30770.

[30] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Complex networks: Structure and dynamics*, Phys. Rep., 424 (2006), pp. 175–308.

[31] P. Bonacich, *Power and centrality: A family of measures*, American J. Sociology, 92 (1987), pp. 1170–1182.

[32] U. Brandes and T. Erlebach (Eds.), *Network Analysis: Methodological Foundations*, Lecture Notes in Computer Science 3418, Springer, Berlin, 2005.

[33] R. Breiger and P. Pattison, *Cumulated social roles: The duality of persons and their algebras*, Social Networks, 8 (1986), pp. 215–256.

[34] S. Brin and L. Page, *The anatomy of a large-scale hypertextual Web search engine*, Computer Networks and ISDN Systems, 33(1998), pp. 107–117.

[35] P. Chebotarev, *The walk distances in graphs*, Discr. Appl. Math., 160 (2012), pp. 1484–1500.

[36] I. Chen, M. Benzi, H. H. Chang, and V. S. Hertzberg, *Dynamic communicability and epidemic spread: A case study on an empirical dynamic contact network*, J. Complex Networks, 5 (2017), pp. 274–302.

[37] F. Chung, *The heat kernel as the pagerank of a graph*, Proc. Nat. Acad. Sci., 104 (2007), pp. 19735–19740.

[38] F. Chung and O. Simpson, *Computing heat kernel pagerank and a local clustering algorithm*, European J. Combin., 68 (2018), pp. 96–119.

[39] J. J. Crofts, E. Estrada, D. J. Higham, and A. Taylor, *Mapping directed networks*, Electron. Trans. Numer. Anal., 37 (2010), pp. 337–350.

[40] J. J. Crofts and D. J. Higham, *A weighted communicability measure applied to complex brain networks*, J. Royal Soc. Interface, 6 (2009), pp. 411–414.

[41] J. J. Crofts, D. J. Higham, R. Bosnell, S. Jbabdi, P. M. Matthews, T. E. J. Behrens, and H. Johansen-Berg, *Network analysis detects changes in the contralesional hemisphere following stroke*, Neuroimage 54 (2011), pp. 161–169.

[42] D. Cvetković, P. Rowlinson, and S. Simić, *Eigenspaces of Graphs*, Cambridge University Press, 1997.

[43] T. Davis, Y. Hu, S. Kolodziej, SuiteSparse Matrix Collection, https://sparse.tamu.edu

[44] N. Del Buono, L. Lopez, and R. Peluso, *Computation of the exponential of large sparse skew-symmetric matrices*, SIAM J. Sci. Comput., 27 (2005), pp. 278–293.

[45] N. Del Buono, L. Lopez, and T. Politi, *Computation of functions of Hamiltonian and skew-symmetric matrices*, Math. Comp. Simul., 79 (2008), pp. 1284–1297.

[46] M. A. de Reus and M. P. van den Heuvel, *Simulated rich club lesioning in brain networks: a scaffold for communication and integration?*, Front. Hum. Neurosc., 21 August 2014, https://doi.org/10.3389/fnhum.2014.00647.

[47] R. Diestel, *Graph Theory*, Springer-Verlag, Berlin, 2000.

[48] E. Estrada, *Community detection based on network communicability*, Chaos, 21 (2011), Article 016103.

[49] E. Estrada, *Complex networks in the Euclidean space of communicability distances*, Phys. Rev. E, 85 (2012), Article 066122.

[50] E. Estrada, *Path Laplacian matrices: Introduction and application to the analysis of consensus in networks*, Linear Algebra Appl., 436 (2012), pp. 3373–3391.

[51] E. Estrada, *The communicability distance in graphs*, Linear Algebra Appl., 436 (2012), pp. 4317–4328.

[52] E. Estrada, *The Structure of Complex Networks: Theory and Applications*, Oxford University Press, 2012.

[53] E. Estrada, *Communicability in temporal networks*, Phys. Rev. E, 88 (2013), Article 042811.

[54] E. Estrada, *Communicability geometry of multiplexes*, New J. Phys., 21 (2019), Article 015004.

[55] E. Estrada, A. A. Almhomaidhi, F. Al-Thukair, *Exploring the "Middle Earth" of network spectra*, Chaos, 27 (2017), Article 023109.

[56] E. Estrada and F. Arrigo, *Predicting triadic closure in networks using communicability distance functions*, SIAM J. Appl. Math., 75 (2015), pp. 1725–1744.

[57] E. Estrada and M. Benzi, *What is the meaning of the graph energy after all?*, Discr. Appl. Math., 230 (2017), pp. 71–77.

[58] E. Estrada and J. Goméz-Gardenes, *Network bipartivity and the transportation efficiency of European passenger airlines*, Physica D: Nonlinear Phenomena, 323 (2016), pp. 20–26.

[59] E. Estrada, E. Hameed, N. Hatano, and M. Langer, *Path Laplacian operators and superdiffusive processes on graphs. I. One-dimensional case*, Linear Algebra Appl., 523 (2017), pp. 307–334.

[60] E. Estrada, E. Hameed, M. Langer, and A. Puchalska, *Path Laplacian operators and superdiffusive processes on graphs. II. Two-dimensional lattice*, Linear Algebra Appl., 555 (2018), pp. 373–397.

[61] E. Estrada and N. Hatano, *Statistical-mechanical approach to subgraph centrality in complex networks*, Chem. Phys. Lett., 439 (2007), pp. 247–251.

[62] E. Estrada and N. Hatano, *Communicability in complex networks*, Phys. Rev. E, 77 (2008), Article 036111.

[63] E. Estrada and N. Hatano, *Communicability graph and community structures in complex networks*, Appl. Math. Comput., 214 (2009), pp. 500–511.

[64] E. Estrada and N. Hatano, *Communicability angle and the spatial efficiency of networks*, SIAM Rev., 58 (2016), p. 692–715.

[65] E. Estrada, N. Hatano, and M. Benzi, *The physics of communicability in complex networks*, Phys. Rep., 514 (2012), pp. 89–119.

[66] E. Estrada, D. J. Higham, *Network properties revealed through matrix functions*, SIAM Rev., 52 (2010), pp. 696–714.

[67] E. Estrada and J. A. Rodríguez-Velázquez, *Subgraph centrality in complex networks*, Phys. Rev. E, 71 (2005), Article 056103.

[68] E. Estrada and J. A. Rodríguez-Velázquez, *Spectral measures of bipartivity in complex networks*, Phys. Rev. E, 72 (2005), Article 046105.

[69] E. Estrada, E. Vargas-Estrada, and H. Ando, *Communicability angles reveal the critical links in network consensus dynamics*, Phys. Rev. E, 92 (2015), Article 052809.

[70] A. Farahat, T. LoFaro, J. Miller, G. Rae, and L. A.Ward, *Authority rankings from HITS, PageRank, and SALSA: existence, uniqueness, and effect of initialization*, SIAM J. Sci. Comput., 27 (2006), pp. 1181–1201.

[71] C. Fenu, D. Martin, D., L. Reichel, and G. Rodriguez, *Network analysis via partial spectral factorization and Gauss quadrature*, SIAM J. Sci. Comput., 35 (2013), pp. A2046–A2068.

[72] C. Fenu, D. Martin, D., L. Reichel, and G. Rodriguez, *Block Gauss and anti-Gauss quadrature with application to networks*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 1655–1684.

[73] M. Fiedler, *Algebraic connectivity of graphs*, Czech. Math. J., 23 (1973), pp. 298–305.

[74] P. Fika and M. Mitrouli, *Estimation of the bilinear form $y^* f(A)x$ for Hermitian matrices*, Linear Algebra Appl., 502 (2016), pp. 140–158.

[75] P. Fika and M. Mitrouli, *Aitken's method for estimating bilinear forms arising in applications*, Calcolo., 54 (2017), pp. 455–470.

[76] V. Fioriti and M. Chinnici, *Identifying sparse and dense sub-graphs in large graphs with a fast algorithm*, Europhysics Letters, 108 (2014), Article 50006.

[77] L. Freeman, *A set of measures of centrality based on betweenness*, Sociometry, 40 (1977), 35–41.

[78] L. Freeman, *Centrality in networks. I: Conceptual clarification*, Social Networks, 1 (1979), pp. 215–239.

[79] A. Frommer and V. Simoncini, *Matrix functions*, in *Model Order Reduction: Theory, Research Aspects and Applications*, W. Schilders and H. A. Van der Vorst (Eds.), Springer-Verlag, Berlin 2008.

[80] M. Girvan and M. E. Newman, *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 7821–7826.

[81] E. Gladilin, *Graph-theoretical model of global human interactome reveals enhanced long-range communicability in cancer networks*, PLoS ONE 12 (2017), Article e0170953. https://doi.org/10.1371/journal.pone.0170953

[82] G. H. Golub and G. Meurant, *Matrices, moments and quadrature*, in: Numerical Analysis, 1993, D.F. Griffiths, G.A. Watson (Eds.), Pitman Research Notes in Mathematics, vol. 303, Essex, England, 1994, pp. 105–156.

[83] G. H. Golub, and G. Meurant, *Matrices, Moments and Quadrature with Applications*, Princeton University Press, Princeton, NJ, 2010.

[84] G. H. Golub, and C. F. Van Loan, *Matrix Computations. Fourth Edition*, Johns Hopkins University Press, Baltimore and London, 2013.

[85] M. M. Gravert, R. J. Dolan, and T. E. J. Behrens, *A map of abstract relational knowledge in the human hippocampal-entorhinal cortex*, eLife (2017), Article 2017;6:e17086. DOI: 10.7554/eLife.17086.

[86] P. Grindrod, M. C. Parsons, D. J. Higham, E. Estrada, *Communicability across evolving networks*, Phys. Rev. E, 83 (2011), Article 046120.

[87] S. Güttel, `funm_kryl`, A Restart Code for the Evaluation of Matrix Functions, http://www.mathe.tu-freiberg.de/~guettels/funm_kryl/

[88] S. Güttel, *Rational Krylov approximation of matrix functions: Numerical methods and optimal pole selection*, GAMM-Mitteilungen 36 (2013), pp. 8-31.

[89] J. B. Hawkins and A. Ben-Israel, *On generalized matrix functions*, Linear and Multilinear Algebra, 1 (1973), pp. 163–171.

[90] N. J. Higham, *Functions of Matrices: Theory and Computation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008.

[91] M. Hochbruck and C. Lubich, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal. 34 (1997), pp. 1911–1925.

[92] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bulletin (New Series) of the American Mathematical Society, 43 (2006), pp. 439–561.

[93] R. A. Horn and C. R. Johnson, *Matrix Analysis. Second Edition*, Cambridge University Press, 2013.

[94] E. Horton, K. Kloster, and B. D. Sullivan, *Subgraph centrality and walk-regularity*, Linear Algebra Appl., 570 (2019), pp. 225–244.

[95] L. Katz, *A new status index derived from sociometric analysis*, Psychometrica, 18 (1953), 39–43.

[96] J. Kempe, *Quantum random walks – an introductory overview*, Contemporary Physics, 44 (2003), pp. 307–327.

[97] J. Kleinberg, *Authoritative sources in a hyperlinked environment*, J. ACM, 46 (1999), pp. 604–632.

[98] K. Kloster and D. F. Gleich, *A nearly-sublinear method for approximating a column of the matrix exponential for matrices from large, sparse networks*, in: *Algorithms and Models for the Web Graph*, Lecture Notes in Computer Science, Springer, 8305 (2013), Springer, Cham, pp. 68–79.

[99] K. Kloster, D. Král', and B. D. Sullivan, *Walk entropy and walk-regularity*, Linear Algebra Appl., 546 (2018), pp. 115–121.

[100] I. T. Koponen and M. Nousiainen, *Concept networks in learning: finding key concepts in learners' representations of the interlinked structure of scientific knowledge*, J. Complex Networks, 2 (2014), pp. 187–202.

[101] A. N. Langville and C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ, 2006.

[102] X. Li, Y. Shi, and I. Gutman, *Graph Energy*, Springer, New York/Heidelberg/Dordrecht/London, 2012.

[103] J. Liesen and Z. Strakoš, *Krylov Subspace Methods. Principles and Analysis*, Oxford University Press, 2013.

[104] G. Meurant, *MMQ Toolbox for MATLAB*, http://gerard.meurant.pagesperso-orange.fr/

[105] Y. Miao, L. Qi, and Y. Wei, *Generalized tensor function via the singular value decomposition based on the T-product*, Linear Algebra Appl., 590 (2020), pp. 258–303.

[106] M. Mitrouli and P. Roupa, *Vector estimates for $f(A)\mathbf{b}$ via extrapolation*, Electron. Trans. Numer. Anal., 47 (2017), pp. 179–196.

[107] D. Mugnolo, *Semigroup Methods for Evolution Equations on Networks*, Springer, Cham, 2014.

[108] L. Mussone and R. Notari, *A comparative analysis of underground and bus transit networks through graph theory*, Environment and Planning B: Urban Analytics and City Science, published online October 11, 2019, Research Article https://doi.org/10.1177/2399808319879460.

[109] E. Nathan and G. Sanders, *Personalized ranking in dynamic graphs using nonbacktracking walks*, in J. M. F. Rodrigues et al. (Eds), *International Conference on Computational Science*, Lecture Notes in Computer Science 11537, Springer, Cham, 2019, pp. 276–289.

[110] Z. Nedev, *Algebraic characterization of strongly connected graphs*, Int. J. Pure Appl. Math., 81 (2012), pp. 1–8.

[111] M. Newman, *The structure and function of complex networks*, SIAM Rev., 45 (2003), pp. 167–256.

[112] M. Newman, *Networks. Second Edition*, Oxford University Press, 2018.

[113] V. Nikiforov, *Beyond graph energy: norms of graphs and energies*, Linear Algebra Appl., 506 (2015), pp. 82–138.

[114] V. Noferini, *A formula for the Fréchet derivative of a generalized matrix function*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 434–457.

[115] R. Olfati-Saber, J. A. Fax, and R. M. Murray, *Consensus and cooperation in networked multi-agent systems*, Proc. IEEE, 95 (2007), pp. 215–233.

[116] M. Pereda and E. Estrada, *Visualization and machine learning analysis of complex networks in hyperspherical space*, Pattern Recognition, 86 (2019), pp. 320–331.

[117] P. P. Petrushev and V. A. Popov, *Rational approximation of real functions*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1987.

[118] S. Pozza and F. Tudisco, *On the stability of network indices defined by means of matrix functions*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 1521–1546.

[119] M. Pranić and L. Reichel, *Generalized anti-Gauss quadrature rules*, J. Comput. Appl. Math., 284 (2015), pp. 235–243.

[120] A. Rapoport, *Mathematical models of social interaction*, in Handbook of Mathematical Psychology, vol. 2, pp. 493–579. Wiley, New York, 1963.

[121] L. Reichel, M. M. Spalević, and T. Tang, *Generalized averaged Gauss quadrature rules for the approximation of matrix functionals*, BIT Numerical Mathematics, 56 (2016):, pp. 1045–1067.

[122] L. Reichel, G. Rodriguez, and T. Tang, *New block quadrature rules for the approximation of matrix functions*, Linear Algebra Appl., 502 (2016), pp. 299–326.

[123] A. P. Riascos and J. L. Mateos, *Fractional dynamics on networks. Emergence of anomalous diffusion and Lévy flights.*, Phys. Rev. E, 90 (2014), Article 032809.

[124] A. P. Riascos and J. L. Mateos, *Random Walks on Weighted Networks: Exploring Local and Non-Local Navigation Strategies*, arXiv:1901.05609v1 [cond-mat.stat-mech], 17 January 2019.

[125] Y. Saad, *Iterative Methods for Large Linear Systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003.

[126] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Classics in Applied Mathematics, 66, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2011.

[127] E. Santiago, J. X. Velasco-Hernandez, and M. Romero-Salcedo, *A descriptive study of fracture networks in rocks using complex network metrics*, Computers & Geosciences, 88 (2016), pp. 97–114.

[128] J. R. Seeley, *The net of reciprocal influence: A problem in treating sociometric data*, Canadian J. of Psychology, 3 (1949), pp. 234–240.

[129] T. C. Silva, S. R. Stancato Souza, and B. Miranda Tabak, *Monitoring vulnerability and impact diffusion in financial networks*, J. Econ. Dyn. Contr., 76 (2017), pp. 109–135.

[130] G. Silver, M. Akbarzadeh, and E. Estrada, *Tuned communicability metrics in networks. The case of alternative routes for urban traffic*, Chaos, Solitons & Fractals, 116 (2018), pp. 402–413.

[131] A. Streiwieser, *Molecular Orbital Theory for Organic Chemists*, John Wiley and Sons, New York, 1961.

[132] P. Suau, E.R. Hancock, and F. Escolano, *Graph characteristics from the Schrödinger operator*, in *Graph-based Representations in Pattern Recognition*, Lecture Notes in Computer Science, 7877, Springer, Cham, 2013, pp. 172–181.

[133] D. Taylor, M. A. Porter, P. J. Mucha, *Tunable Eigenvector-Based Centralities for Multiplex and Temporal Networks*, arXiv preprint arXiv:1904.02059 (2019).

[134] Ll. N. Trefethen, *Approximation Theory and Approximation Practice*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013.

[135] L. W. Veergossen, M. T. Schram, J. J. de Jong, C. D. Stehouwer, N. C. Schaper, R. M. Henry, C. J. van der Kallen, P. C. Dagnelie, M. P. van Boxtel, S. J. Eussen, W. H. Backes, J. F. Jansen, *White matter connectivity abnormalities in prediabetes and Type 2 diabetes: The Maastricht study*, Diabetes Care, October 2019, Article dc190762.

[136] D. A. Walker and A. Tordesillas, *Topological evolution in dense granular materials: a complex network perspective*, Int. J. Solids Struct., 47 (2010), pp.624–639.

[137] H. Wang and C. Ye, *Error bounds for the Krylov subspace methods for the computation of the matrix exponential*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 155–187.

[138] Y. Xu, *A spectral method to detect community structure based on the communicability modularity*, Physica A: Statistical Mechanics and Applications, 537 (2020), Article 122751.

[139] H. Yang, I. King, and M. R. Lyu, *DiffusionRank: A possible penicillin for Web spamming*, SIGIR 2007 Proceedings, ACM 978-1-59593-597-7/07/0007. Association for Computing Machinery, 2007.

[140] W. Zhao, Z. Yin, and T. Fan et al., *Research on influence spread of scientific research team based on scientific factor quantification of big data*, Int. J. Distr. Sensor Networks, published online April 9, 2019, Research Article https://doi.org/10.1177/1550147719842158