




SoftNet: A package for the analysis of complex networks

Caterina Fenu ^{1,†} , Lothar Reichel ^{2,†}  and Giuseppe Rodriguez ^{1,†,*} 

¹ Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale, 72, 09124 Cagliari, Italy; kate.fenu@unica.it, rodriguez@unica.it

² Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA; reichel@math.kent.edu

* Correspondence: rodriguez@unica.it; Tel.: +39-070-675-5617 (G.R.)

† These authors contributed equally to this work.

Abstract: Identifying the most important nodes according to specific centrality indices is an important issue in network analysis. Node metrics based on the computation of functions of the adjacency matrix of a network have been defined by Estrada and his collaborators in various papers. This paper describes a MATLAB toolbox for computing such centrality indices using efficient numerical algorithms based on the connection between the Lanczos method and Gauss-type quadrature rules.

Keywords: Complex network analysis, centrality measure, matrix function, Lanczos algorithm.

1. Introduction

Let G be a connected, undirected, unweighted graph with a large number of nodes n and a number of edges significantly smaller than n^2 . We assume there are no self-loops or multiple edges in G . Networks represented by such kind of graph occur in many applications, such as epidemiology, genetics, telecommunications, and energy distribution; see [7,13,14,30]. It is usual to associate to the graph G a symmetric adjacency matrix $A = [A_{ij}] \in \mathbb{R}^{n \times n}$ with entries $A_{ij} = 1$, if nodes i and j are connected by an edge, and $A_{ij} = 0$, otherwise.

It is often meaningful to extract from a large graph numerical values describing global properties of the graph, such as the ease of traveling between vertices, or the importance of a chosen node. A walk in a network is an ordered list of nodes such that successive entries of the list are connected. A well-known fact in graph theory is that the number of walks of length $m \geq 1$ starting at node i and ending at node j is given by $[A^m]_{ij}$, that is, the entry (i, j) of the m -th power of the adjacency matrix. Let us assume that the coefficients c_m in the matrix-valued function

$$f(A) = \sum_{m=0}^{\infty} c_m A^m \tag{1.1}$$

are nonnegative and decay fast enough to ensure convergence of the series. Then, the ease of traveling between the nodes i and j can be measured by $[f(A)]_{ij}$, with $i \neq j$, while the importance of node i can be quantified by $[f(A)]_{ii}$.

A common choice (see [13,14,18,19]) is to set the coefficients c_m in (1.1) to be non-increasing positive functions of m , with the aim of attributing less importance to long walks than to short ones. For example, $c_m = 1/m!$ [16] yields the matrix exponential

$$f(A) = \exp(A), \tag{1.2}$$

while setting $c_m = \alpha^m$, with $0 < \alpha < (\rho(A))^{-1}$, where $\rho(A)$ denotes the spectral radius of A , leads to the resolvent

$$f(A) = (I - \alpha A)^{-1}. \tag{1.3}$$

Let $\mathbf{e} = [1, 1, \dots, 1]^T \in \mathbb{R}^n$ and let $\mathbf{e}_i = [0, \dots, 0, 1, 0, \dots, 0]^T \in \mathbb{R}^n$ be the axis vector with the i th component equal to 1. As usual, the superscript T denotes transposition.

Citation: Fenu, C.; Reichel, L.; Rodriguez, G. SoftNet: A package for the analysis of complex networks. *Algorithms* **2022**, *1*, 0. <https://doi.org/>

Received:
Accepted:
Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Submitted to *Algorithms* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The following definitions, which are discussed in [13–15,18,19,21], are motivated by the discussion above:

- the *degree* of node i , given by $[A\mathbf{e}]_i = \mathbf{e}_i^T A\mathbf{e}$, provides a measure of the importance of node i ;
- the *f -subgraph centrality* of node i , defined by

$$[f(A)]_{ii} = \mathbf{e}_i^T f(A)\mathbf{e}_i, \quad (1.4)$$

- furnishes a more sophisticated measure of the importance of node i than its degree;
- the *f -communicability* between nodes i and j ,

$$[f(A)]_{ij} = \mathbf{e}_i^T f(A)\mathbf{e}_j, \quad (1.5)$$

- quantifies the ease of traveling between nodes i and j ;
- the *f -starting convenience* of node i , given by

$$\frac{\mathbf{e}^T f(A)\mathbf{e}_i}{\mathbf{e}^T f(A)\mathbf{e}}, \quad (1.6)$$

quantifies the ease of traveling from node i to anywhere in the network. This is the sum of the communicabilities from node i to all other nodes, scaled so that the sum of the quantity over all nodes is one.

Note that all the centrality measures (1.4)–(1.6) are of the form

$$\mathbf{u}^T f(A)\mathbf{v} \quad (1.7)$$

for specific vectors \mathbf{u} and \mathbf{v} . The purpose of this paper is to present a software package that makes it easy to compute the above defined centrality measures, whose use and methods for their computation have received considerable attention in the literature; see [1,5,6,8–22,25,27] as well as many other references. In these references many real applications are discussed.

When the adjacency matrix A is large, i.e., when the graph G has many nodes, direct evaluation of $f(A)$ generally is not feasible. Benzi and Boito [5] applied pairs of Gauss and Gauss-Radau rules to compute upper and lower bounds for selected entries of $f(A)$. This work is based on the connection between the symmetric Lanczos process, orthogonal polynomials, and Gauss-type quadrature, explored by Golub and his collaborators in many publications; see Golub and Meurant [24] for details and references. A brief review of this technique is provided in Section 2. An application of pairs of block Gauss-type quadrature rules to simultaneously determine approximate upper and lower bounds of expressions of the form (1.7) when \mathbf{u} and \mathbf{v} are “block vectors”, i.e., matrices with many rows and very few columns, is described in [21].

The main drawback of quadrature-based methods is that the computational effort is proportional to the number of desired bounds. Therefore these methods may be expensive to use when bounds for many expressions of the form (1.7) are to be evaluated. This situation arises, for instance, when we would like to determine one or a few nodes with the largest f -subgraph centrality in a large graph, because this requires the computation of upper and lower bounds for all diagonal entries of $f(A)$.

A method to produce upper and lower bounds for quantities of the form (1.7) has been proposed in [20]. It is based on that knowledge of a few leading eigenvalue-eigenvector pairs gives bounds for every entry of $f(A)$, with little computational effort in addition to computing the eigenvalue-eigenvector pairs. For example, determining the m most important nodes of a graph, with m much smaller than the number of nodes n , amounts to finding the m nodes with the largest f -subgraph centrality. It is possible to quickly evaluate bounds for all entries $[f(A)]_{ij}$ if a partial spectral factorization of A is available. Using these

bounds we can determine a set of $\ell \geq m$ nodes containing the m nodes of interest, and compute tighter bounds for the nodes in this set, if necessary, by employing Gauss-type quadrature rules. When $\ell \ll n$, the complexity of this hybrid algorithm is much smaller than computing upper and lower bounds for all entries $[f(A)]_{ii}$, $i = 1, \dots, n$, by Gauss quadrature.

In this work, we present a MATLAB package for the identification of the m most important nodes according to the centrality/communicability indices discussed above, based on two matrix functions, namely, the exponential (1.2) and the resolvent (1.3). Either the f -subgraph centrality, the f -communicability or the f -starting convenience can be computed. The computation can be performed by using one of three different methods: Gauss quadrature, partial spectral factorization, or the hybrid method; the latter two algorithm have been introduced in [20].

This paper is organized as follows. Section 2 recalls how upper and lower bounds for quantities of the form (1.7) can be determined via Gauss quadrature. Approximation via partial spectral factorization of A is discussed in Section 3 and the hybrid method is summarized in Section 4. Section 5 presents the SoftNet package as well as a graphical user interface (GUI) that simplifies its use. A brief description of the code and its use also is provided. Section 6 describes some numerical experiments and Section 7 contains concluding remarks.

2. Approximation by Gauss quadrature

Let A be a symmetric matrix of order n and suppose that we are interested in computing bounds for bilinear forms

$$F_{\mathbf{u},\mathbf{v}}(A) := \mathbf{u}^T f(A) \mathbf{v}, \quad (2.1)$$

where \mathbf{u} and \mathbf{v} are given vectors and f is a smooth function defined on an interval $[a, b] \subset \mathbb{R}$ that contains the spectrum of A . Since

$$F_{\mathbf{u},\mathbf{v}}(A) = \frac{1}{4}(F_{\mathbf{u}+\mathbf{v},\mathbf{u}+\mathbf{v}}(A) - F_{\mathbf{u}-\mathbf{v},\mathbf{u}-\mathbf{v}}(A)),$$

we can focus on the case $\mathbf{u} = \mathbf{v}$.

The matrix A has the spectral decomposition $A = Q\Lambda Q^T$. Then we can write

$$F_{\mathbf{u},\mathbf{u}}(A) = \mathbf{u}^T Q f(\Lambda) Q^T \mathbf{u} = \boldsymbol{\mu}^T f(\Lambda) \boldsymbol{\mu} = \sum_{i=1}^n f(\lambda_i) \mu_i^2 = \int_a^b f(\lambda) d\mu(\lambda), \quad (2.2)$$

i.e., we may regard $F_{\mathbf{u},\mathbf{u}}(A)$ as a Stieltjes integral; see [20,24] for further details. We approximate this integral by Gauss-type quadrature rules as follows. Let \mathbf{u} be of unit Euclidean norm. Application of k steps of the Lanczos algorithm to A with initial vector \mathbf{u} gives the $k \times k$ symmetric tridiagonal matrix T_k . It can be shown that $\mathbf{e}_1^T f(T_k) \mathbf{e}_1$ is a k -node Gauss quadrature rule \mathcal{G}_k for the Stieltjes integral (2.2). A $(k+1)$ -node Gauss-Radau quadrature formula $\hat{\mathcal{G}}_{k+1}$ with a fixed node at $\theta \geq \rho(A)$ for approximating the Stieltjes integral also can be defined. This discussion assumes that the Lanczos algorithm does not break down. Breakdown is very rare and allows the computations to be simplified.

Under the assumption that the derivatives of $f(x)$ have constant sign on the convex hull of the support of the measure, which is met by the functions (1.2) and (1.3), and the Radau node θ is suitably chosen, pairs of Gauss and Gauss-Radau rules furnish lower and upper bounds of increasing accuracy for the quadratic form (2.2). For the functions (1.2) and (1.3), and the Radau node θ chosen as described, we have

$$\mathcal{G}_k \leq \mathcal{G}_{k+1} \leq F_{\mathbf{u},\mathbf{u}}(A) \leq \hat{\mathcal{G}}_{k+2} \leq \hat{\mathcal{G}}_{k+1}.$$

For a user-chosen accuracy τ , we terminate the iterations with the Lanczos algorithm when

$$\frac{|\mathcal{G}_k - \hat{\mathcal{G}}_{k+1}|}{|\mathcal{G}_k|} \leq \tau. \quad (2.3)$$

The default value in the code is $\tau = 10^{-3}$.

The matrix functions are applied to the tridiagonal matrices by using their spectral factorization. Thus, let $T_k = W_k \Lambda_k W_k^T$ be the spectral factorization. Then $f(T_k) = W_k f(\Lambda_k) W_k^T$. When f is the exponential function, we let μ be the largest eigenvalue of T_k and evaluate

$$e^{-\mu} e^{T_k} = e^{T_k - \mu I} = W e^{\Lambda_k - \mu I} W^T \quad (2.4)$$

instead of e^{T_k} to avoid overflow.

Regarding the choice of the Radau node θ , we often may let $\theta = \|A\|_\infty$. Alternatively, we can use the MATLAB function `eigs` or the function `irbleigs` described in [2,3] to determine an estimate of the largest eigenvalue λ_1 of A .

3. Bounds via partial spectral factorization

This section recalls how to derive bounds for expressions of the form (2.1), with $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$, by using a partial spectral factorization of A . Introduce the spectral factorization

$$A = Q \Lambda Q^T,$$

where the eigenvector matrix $Q = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n] \in \mathbb{R}^{n \times n}$ is orthogonal and the eigenvalues in the diagonal matrix $\Lambda = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_n] \in \mathbb{R}^{n \times n}$ are ordered according to $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Then

$$f(A) = Q f(\Lambda) Q^T = \sum_{k=1}^n f(\lambda_k) \mathbf{q}_k \mathbf{q}_k^T,$$

so that

$$F_{\mathbf{u}, \mathbf{v}}(A) = \mathbf{u}^T f(A) \mathbf{v} = \sum_{k=1}^n f(\lambda_k) \tilde{u}_k \tilde{v}_k,$$

where $\tilde{u}_k = \mathbf{u}^T \mathbf{q}_k$ and $\tilde{v}_k = \mathbf{v}^T \mathbf{q}_k$. Let the first N eigenpairs $\{\lambda_k, \mathbf{q}_k\}_{k=1}^N$ of A be known. Then $F_{\mathbf{u}, \mathbf{v}}(A)$ can be approximated by

$$F_{\mathbf{u}, \mathbf{v}}(A) \approx F_{\mathbf{u}, \mathbf{v}}^{(N)} := \sum_{k=1}^N f(\lambda_k) \tilde{u}_k \tilde{v}_k. \quad (3.1)$$

The following results from [20] shows how upper and lower bounds for $F_{\mathbf{u}, \mathbf{v}}(A)$ can be determined with the aid of the first N eigenpairs of A .

Theorem 1. *Let the function f be nondecreasing and nonnegative on the convex hull of the spectrum of A and let $F_{\mathbf{u}, \mathbf{v}}^{(N)}$ be defined by (3.1). Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ be the N largest eigenvalues of A and let $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N$ be associated orthonormal eigenvectors. Then we have*

$$L_{\mathbf{u}, \mathbf{v}}^{(N)} \leq F_{\mathbf{u}, \mathbf{v}}(A) \leq U_{\mathbf{u}, \mathbf{v}}^{(N)}, \quad (3.2)$$

where

$$L_{\mathbf{u}, \mathbf{v}}^{(N)} := F_{\mathbf{u}, \mathbf{v}}^{(N)} - f(\lambda_N) \left(1 - \sum_{k=1}^N \tilde{u}_k^2\right)^{1/2} \left(1 - \sum_{k=1}^N \tilde{v}_k^2\right)^{1/2},$$

$$U_{\mathbf{u}, \mathbf{v}}^{(N)} := F_{\mathbf{u}, \mathbf{v}}^{(N)} + f(\lambda_N) \left(1 - \sum_{k=1}^N \tilde{u}_k^2\right)^{1/2} \left(1 - \sum_{k=1}^N \tilde{v}_k^2\right)^{1/2}.$$

When $\mathbf{u} = \mathbf{v}$, we have

$$F_{\mathbf{u},\mathbf{u}}^{(N)} \leq F_{\mathbf{u},\mathbf{u}}(A) \leq U_{\mathbf{u},\mathbf{u}}^{(N)} \quad (3.3)$$

and

$$F_{\mathbf{u},\mathbf{u}}^{(N)} \leq F_{\mathbf{u},\mathbf{u}}^{(N+1)}, \quad U_{\mathbf{u},\mathbf{u}}^{(N)} \geq U_{\mathbf{u},\mathbf{u}}^{(N+1)}, \quad 1 \leq N < n. \quad (3.4)$$

To determine which nodes have the largest f -subgraph centrality (1.4), we use the inequalities (3.3) and (3.4). The N leading eigenpairs $\{\lambda_k, \mathbf{q}_k\}_{k=1}^N$ of A and the bounds (3.2) and (3.3) can be used to determine a subset of nodes that contains the vertices with the largest value of the node metric we are considering.

Let $L_{\mathbf{u},\mathbf{v}}^{(N)}$ and $U_{\mathbf{u},\mathbf{v}}^{(N)}$ be the lower and upper bounds defined in Theorem 1. Since we seek an approximation of the centrality value for all the nodes of the network, we will either set $\mathbf{u} = \mathbf{v} = \mathbf{e}_i$, as in (1.4), or $\mathbf{u} = \mathbf{e}$ and $\mathbf{v} = \mathbf{e}_i$, as in (1.6). So, $F_{\mathbf{u},\mathbf{v}}^{(N)}$ will be a quantity depending on an index $i = 1, \dots, n$. We will write $F_i^{(N)} = F_{\mathbf{u},\mathbf{v}}^{(N)}$ to simplify the notation when we are computing either $F_{\mathbf{e}_i,\mathbf{e}_i}^{(N)}$ or $F_{\mathbf{e},\mathbf{e}_i}^{(N)}$. When approximating (1.5), we will fix a value of j and consider $F_i^{(N)} = F_{\mathbf{e}_j,\mathbf{e}_i}^{(N)}$ for $i = 1, \dots, n$.

Let $\mathcal{F}_m^{(N)}$ denote the m th largest value of the vector $(F_i^{(N)})_{i=1}^n$. The index sets

$$S_m^{(N)} = \left\{ i : U_{\mathbf{u},\mathbf{v}}^{(N)} \geq \mathcal{F}_m^{(N)} \right\}, \quad N = 1, 2, \dots, n. \quad (3.5)$$

contains the indices of the nodes that can be considered important with respect to the desired centrality index.

A computational difficulty to overcome is that we do not know in advance how the dimension N of the leading invariant subspace $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$ of A should be chosen in order to obtain useful bounds (3.2) or (3.3). We use the restarted block Lanczos method `irbleigs` described in [2,3], which computes the leading invariant subspace $\{\lambda_k, \mathbf{q}_k\}_{k=1}^\ell$ of A for a user-chosen dimension ℓ , and allows the extension of such subspace by successively increasing the value of ℓ . Using `irbleigs`, we compute more and more eigenpairs of A until N is such that

$$|S_m^{(N)}| = m, \quad (3.6)$$

where $|S|$ denotes the number of elements of the set S . This stopping criterion is referred to as the *strong convergence condition*. As shown in [20], the set $S_m^{(N)}$ contains the indices of the m nodes with the largest f -subgraph centrality.

The criterion (3.6) for choosing N is useful if the required value of N is not too large. The *weak convergence criterion* has been introduced to be used for problems for which a large value of N is required in order to satisfy (3.6), and this makes it impractical to compute the associated bounds (3.2). The weak convergence criterion is well suited also for application in the hybrid algorithm described in Section 4. This criterion is designed to stop increasing N when the values $F_{\mathbf{u},\mathbf{v}}^{(N)}$ do not increase significantly with N . Specifically, we stop increasing N when the average increment of the values in the vector $F_{\mathbf{u},\mathbf{v}}^{(N)}$ is small when the N th eigenpair $\{\lambda_N, \mathbf{q}_N\}$ is included in the bounds. The average contribution of this eigenpair to $F_{\mathbf{u},\mathbf{v}}^{(N)}$, $1 \leq i \leq n$, is

$$F_i = f(\lambda_N)[\tilde{u}_N]_i[\tilde{v}_N]_i$$

(see (3.1)), and we stop increasing N when

$$\frac{1}{n} \sum_{i=1}^n F_i < \tau \cdot \mathcal{F}_m^{(N)} \quad (3.7)$$

for a user-specified tolerance τ , whose default value in the code is 10^{-3} . Note that when this criterion is satisfied, but not (3.6), the nodes with index in $S_m^{(N)}$ and with the largest

value $F_{\mathbf{u},\mathbf{v}}^{(N)}$ are not guaranteed to be the nodes with the largest index value we are searching for. 111

Also the weak convergence criterion (3.7) may yield a set $S_m^{(N)}$ with many more than m indices. In particular, we may not be willing to compute accurate bounds for a specific node metric by applying the approach of Section 2 to all nodes with index in $S_m^{(N)}$. We therefore describe how to determine a smaller index set \mathcal{J} , which is likely to contain the indices of the m most important nodes. We discard from the set $S_m^{(N)}$ indices for which $F_{\mathbf{u},\mathbf{v}}^{(N)}$ is much smaller than $\mathcal{F}_m^{(N)}$. Thus, for a user-chosen parameter $\sigma > 0$, we include in the set \mathcal{J} all indices $i \in S_m^{(N)}$ such that 112

$$\mathcal{F}_m^{(N)} - F_{\mathbf{u},\mathbf{v}}^{(N)} < \sigma \cdot \mathcal{F}_m^{(N)}.$$

The default value for σ in the software is 10^{-3} . 113

4. The hybrid method 114

We summarize here the algorithm corresponding to the hybrid method. The first step is to compute a partial spectral factorization of the adjacency matrix A . Such a partial factorization makes it possible to determine a set of candidate nodes that contains the most important nodes according to a chosen criterion, e.g., the f -subgraph centrality. The accuracy of upper and lower bounds for the candidate nodes is then improved by a suitable application of Gauss and Gauss-Radau quadrature rules. 115
116
117
118
119
120

5. The SoftNet software package 121

The package `SoftNet` for MATLAB is available at the web page <http://bugs.unica.it/cana/software> as a compressed archive. Uncompressing it, a directory named `SoftNet` will be created; in order to use the package the user should add its name to the search path. The package `SoftNet` consists of 14 MATLAB routines for the identification of the m most important nodes in a network according to different centrality indices. The package also includes the function `irbleigs` from [2,3], and the following 5 adjacency matrices of real-world networks that can be used to test the software 122
123
124
125
126
127
128

- karate (34 nodes, 78 edges): represents the social relationships among the 34 individuals of a university karate club [31]; 129
130
- yeast (2114 nodes, 4480 edges): describes the protein interaction network for yeast [4,26,32]; 131
132
- power (4941 nodes, 13188 edges): undirected representation of the topology of the western states power grid of the United States [31,35]; 133
134
- internet (22963 nodes, 96872 edges): snapshot of the structure of the Internet at the level of autonomous systems from data for July 22, 2006 [31]; 135
136
- collaborations (40421 nodes, 351304 edges): collaboration network of scientists who posted preprints at www.arxiv.org between January 1, 1995 and March 31, 2005 [29,31]; 137
138
- facebook (63731 nodes, 1545686 edges): user-to-user links (*friendship*) from the Facebook New Orleans network, studied in [34] and available at [28]. 139
140

The package `Contest` by Taylor and Higham [33] contains different kind of synthetic networks and can be used to generate further numerical tests. We provide a convenient interface to this package. 141
142
143

Table 1 lists the 14 MATLAB routines with a description of their purpose. The first group, “Computational Routines,” includes the functions for computing different centralities (subgraph centrality (1.4), communicability (1.5), and starting convenience (1.6)) with respect to two different matrix functions, the exponential (1.2) and the resolvent (1.3). The computations can be performed with three different methods, namely the Gauss quadrature method recalled in Section 2, the low-rank approximation presented in Section 3 and the hybrid method described in Section 4. The section “Auxiliary Routines for the Graphical User Interface” lists some routines required to start and use the graphical user interface. 144
145
146
147
148
149
150
151

Table 1. Routines and GUI.

Computational Routines	
<code>commgauss</code>	Identifies the m most important nodes according to f -communicability of node i through Gauss quadrature; see Section 2.
<code>commhyb</code>	Identifies the m most important nodes according to f -communicability of node i through partial singular value decomposition (SVD) and Gauss quadrature; see Section 4.
<code>commlr</code>	Identifies the m most important nodes according to f -communicability of node i through partial SVD; see Section 3.
<code>gaussexp</code>	Computes the bilinear form $\mathbf{u}^T f(A)\mathbf{v}$, with $f(A) = \exp(A)$ through Gauss quadrature; see Section 2.
<code>gaussres</code>	Computes the bilinear form $\mathbf{u}^T f(A)\mathbf{v}$, with $f(A) = (I - \alpha A)^{-1}$ through Gauss quadrature; see Section 2.
<code>sgcengauss</code>	Identifies the m most important nodes according to f -subgraph centrality through Gauss quadrature; see Section 2.
<code>sgcenhyb</code>	Identifies the m most important nodes according to f -subgraph centrality through partial SVD and Gauss quadrature; see Section 4.
<code>sgcenlr</code>	Identifies the m most important nodes according to f -subgraph centrality through partial SVD; see Section 3.
<code>stconvgauss</code>	Identifies the m most important nodes according to f -starting convenience through Gauss quadrature; see Section 2.
<code>stconvhyb</code>	Identifies the m most important nodes according to f -starting convenience through partial SVD and Gauss quadrature; see Section 4.
<code>stconvlr</code>	Identifies the m most important nodes according to f -starting convenience through partial SVD; see Section 3.
Auxiliary Routines for the Graphical User Interface	
<code>vipnodes</code>	Starts the graphical user interface.
<code>compute</code>	Performs the computations according to the chosen parameters.
<code>choose_contest</code>	Allow selection of a synthetic network from the Contest Package [33].
<code>initialize_gui</code>	Initializes the default settings for the graphical user interface.

The computational routines are totally independent from the graphical user interface and can be used by the user from the MATLAB command line. For example, the command

```
[vip, vipsgc] = sgcenlr(A, 'exp', 10);
```

identifies the 10 most important nodes according to the subgraph centrality when the low-rank approximation is used for the computation. `vip` and `vipsgc` are vectors containing the indices of the nodes that are candidates to being the most important nodes and the values of their subgraph centrality, respectively.

Identifying the 5 most important nodes of a network whose adjacency matrix is A with respect to the starting convenience can be done by the following lines of code

```
func = 'exp'; % the function to be used
nnodes = 5; % the number of nodes to be identified
theta = eigs(double(A), 1, 'LA'); % estimation of the largest eigenvalue
opts = struct('gausstolq', 1e-5, 'gaussmaxn', 150, 'gaussmu', theta, 'show', 1)
[vip, vipsgc, info, iters, allstconv] = stconvgauss(A, func, nnodes, opts);
```

The third line computes the largest eigenvalue, since its estimation is needed for the computation of the Gauss-Radau rule. The struct `opts` is initialized on the fifth line, where the tolerance (2.3) for the convergence of Gauss quadrature is chosen, as well as the

maximum number of iterations, and the value μ used for the spectrum shift (2.4). Setting the show variable to 1 displays a waitbar during the computations. 169
170

The output values are: 171

- vip: indices for the most important nodes; 172
- vipsgc: values of starting convenience for the identified nodes; 173
- info: a vector containing a flag that indicates convergence and shows the number of matrix-vector products; 174
175
- iters: the number of iterations performed for each node; 176
- allstconv: the values of the starting convenience for each node. 177

Table 2 reports a subset of the options used for tuning the performance of the package; all the options have a default value. Refer to the second column of the table and to the description of the algorithms in [20] for their meaning. The available options are described in the various functions. 178
179
180
181

Table 2. Problem definition and options.

Problem Definition	
func	function to be used ((1.2) or (1.3))
nnodes	number of nodes to be identified
Fields for the opts variable	
gausstolq	tolerance for Gauss quadrature
gaussmaxn	maximum number of iterations
gaussmu	approximation of the largest eigenvalue for the spectrum shift
alpha	constant for the resolvent
show	if $\neq 0$ shows some information during the computation
sgcmaxva	maximum number of stored eigenvectors
sgctoll	tolerance for weak convergence
sgcshift	if $\neq 0$ applies spectrum shift

All the functions can be used interactively with the vipnodes graphical user interface, located in the main directory of the package. The GUI starts by typing the command vipnodes in the MATLAB Command Window; see Figure 1. 182
183
184

The GUI consists of one input panel, on the left, and an output area, on the right. The former allows the user to set different parameters to perform the computations, the latter shows some information about the loaded network and the results, once the computations are done. A drop-down menu at the top of the window allows the user to perform different tasks as follows: 185
186
187
188
189

- File. This menu allows the user to load a network in three different ways: load it from a mat file, extract it from the workspace, and create it by the Contest package [33], if the latter is installed. 190
191
192
- Export. This menu allows the user to export the results as a mat file, as a text file, or export them to variables in the workspace. 193
194
- Reset. Reset options and computed results or just the results. 195
- Stop. Interrupt the computations if they take too long time. 196
- Previous results. Display a table with results of the previous computation. 197

The first step to complete in order to carry out the computations is to load an adjacency matrix through the "File" menu on the top left of the main window. Once this task is done, general information about the network are shown, namely the number of nodes, the number of edges and, if the network contains self-loops, the number of removed edges. The parameters are set to their default values, and can be modified by the user. By pressing the "Find nodes" button the computations start. If the user chooses to show the animation (this possibility is not given if computation via Gauss quadrature is selected), a new window "Animation" will appear. It contains a spy plot of the adjacency matrix associated to the 198
199
200
201
202
203
204
205

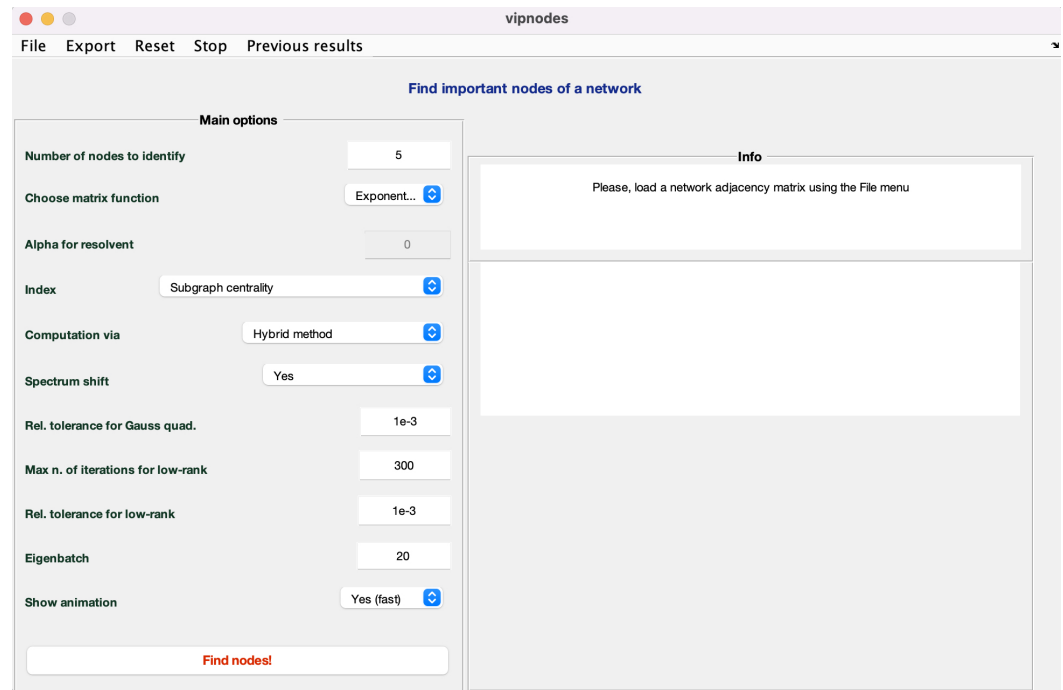


Figure 1. The graphical user interface (GUI).

network with the number of non-zero elements, that is, the number of edges, shown on the bottom of the figure. Below, the spectrum of A is drawn and the graph is updated once a new set of eigenvalues is computed. On the right, an animation with the lower and upper bounds computed in correspondence of each new eigenpair added to the sum (3.2) is shown. If either the strong or weak convergence criteria are satisfied, then the candidate nodes are highlighted with red circles. The title of the last graph reports the number of used eigenpair and the cardinality of the set $S_m^{(N)}$ defined in (3.5).

Figure 2 shows a typical Animation window. In this case, the computations aim to identify the 5 most important nodes with respect to the subgraph centrality of the Power network included in the package. The computations were carried out by the low-rank method, with the strong convergence condition. The number of computed eigenpairs is the minimal integer N such that the cardinality of $S_5^{(N)}$ is 5.

Figure 3 shows the same window after identifying the same number of nodes as in Figure 2 by the hybrid method. In this case, the cardinality of the set $S_5^{(N)}$ is larger than 5, and the final computation to identify the 5 most important nodes is performed by Gauss quadrature.

Once the computations are done, the main window shows the following information:

- the method used to perform the computation (low-rank with strong convergence, hybrid method or Gauss quadrature);
- the line of code that has to be written on the command window to perform the same computation without using the graphical user interface;
- whether the strong or weak convergence criteria are satisfied (if one of them was selected);
- the number of used eigenpairs (if either the low-rank or hybrid methods have been used for the computations);
- the number of VIP nodes identified (if either the low-rank or hybrid methods have been used);
- the elapsed time;
- a table with the index of the identified nodes and the value of the corresponding centrality index.

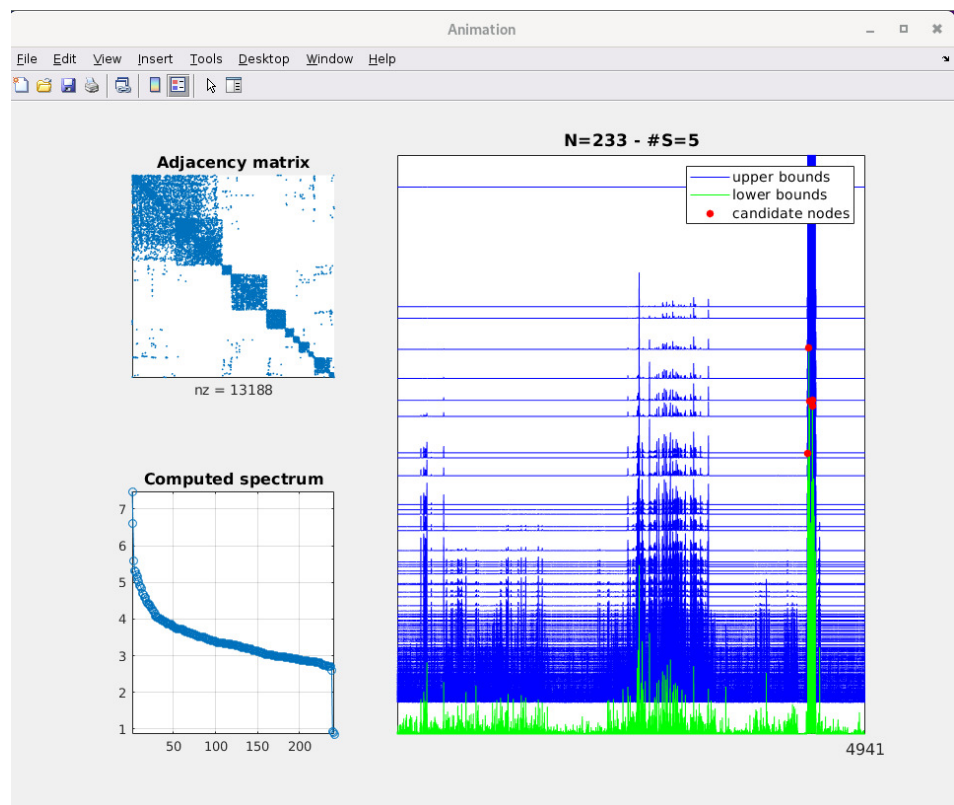


Figure 2. The animation window after the identification of the 5 most important nodes for the Power network by the low-rank approximation method.

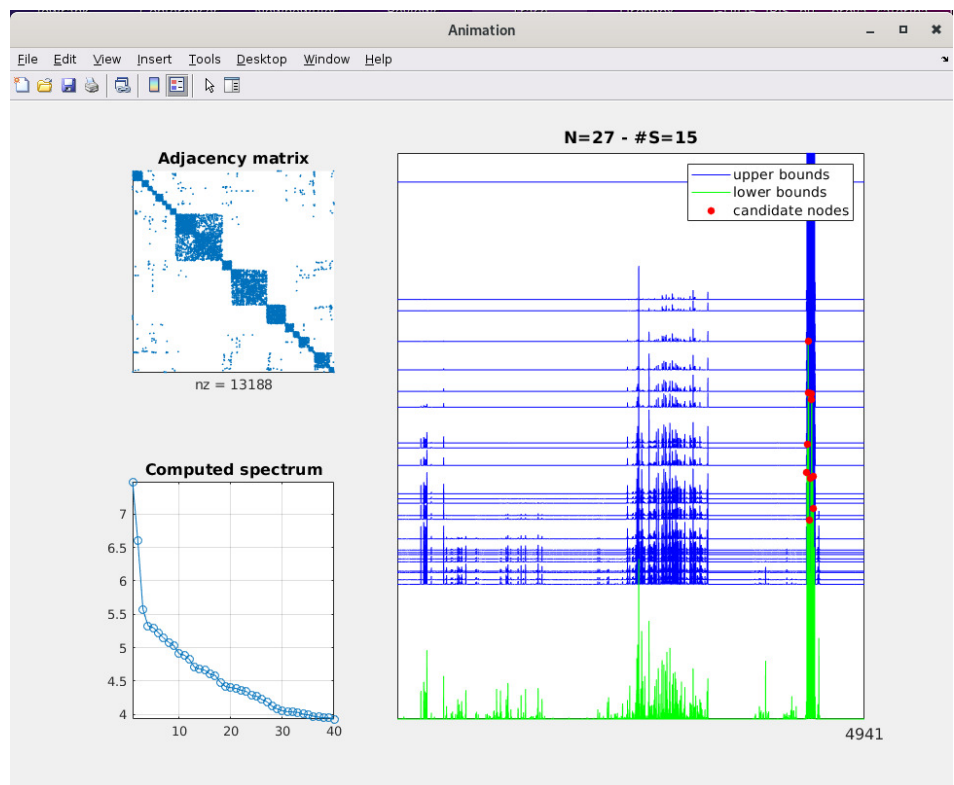


Figure 3. The animation window after the identification of the 5 most important nodes for the Power network by the hybrid method.

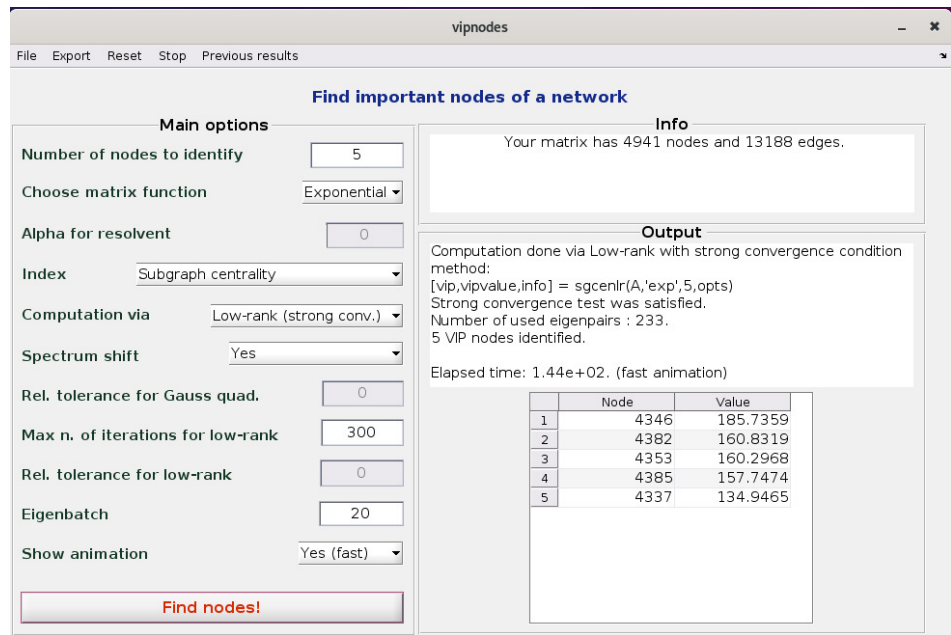


Figure 4. Main window after the identification of the 5 most important nodes for the Power network by the low-rank approximation method.

Figure 4 shows the main window once the computations related to Figure 2 have been carried out. Figure 5 shows the same window after the hybrid method has been used.

Note that the lists of nodes produced by the two methods is the same, but the values of the centrality index are slightly different. This happens because the value of the subgraph centrality computed by the low-rank approximation is estimated as an average of the lower and upper bounds computed by the method, while the value computed by Gauss quadrature is more accurate.

6. Numerical experiments

This section provides some numerical experiments to explore the performance of the centrality indices used in the software, namely the f -subgraph centrality and the f -starting convenience. In particular, we compare them to the following well-known centrality indices:

- degree: the number of edges adjacent to a node;
- betweenness: the number of shortest paths that pass through the node;
- closeness: the reciprocal of the sum of the length of the shortest paths between a node and all other nodes in the graph;
- eigenvector: a score is assigned to each node taking into account connections with nodes that have high scores;
- pagerank: a variant of the eigenvector centrality.

The computation of the centrality indices listed above has been done by the centrality function included in Matlab. An example of its usage it is the following:

```
centr = 'betweenness'; % the centrality to be used
nodes = 5; % the number of nodes to be identified
G = graph(A); % converts the adjacency matrix A in to the graph G
values = centrality(G,centr); % computes all the centralities of graph G
[~, node_ind] = sort(values,'descend'); % sorts all the centralities
disp(node_ind(1:nodes)); % displays the index for the nodes with the largest centrality
```

The string `centr` can be set to `degree`, `betweenness`, `closeness`, `eigenvector` and `pagerank`.

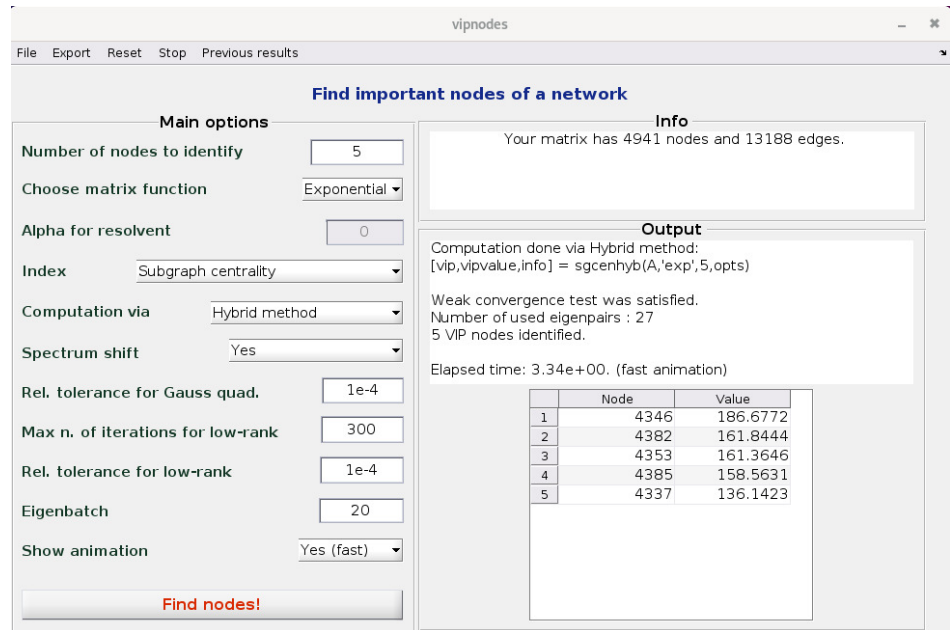


Figure 5. Main window after the identification of the 5 most important nodes for the Power network by the hybrid method.

Table 3. Ranking of the 5 most important nodes for the karate network identified by the centrality function of Matlab, the f -subgraph centrality, and f -starting convenience.

degree	betw	clos	eigvec	pagerank	exp-sgc	res-sgc	exp-stc	res-stc
34	1	1	34	34	34	34	34	34
1	34	3	1	1	1	1	1	1
33	33	34	3	33	33	33	3	33
3	3	32	33	3	3	3	33	3
2	32	9	2	2	2	2	2	2

The first network we analyze is the famous Zachary's karate club network [31]. The most important nodes of the network are node 1 and node 34, which stand for the instructor and the club president, respectively.

Each column of Table 3 reports the ranking of the 5 most important nodes obtained by using the centrality indices listed above, namely degree, betweenness, closeness, eigenvector, and pagerank centralities, compared with the ranking obtained by the exp -subgraph centrality, the res -subgraph centrality, the exp -starting convenience, and the res -starting convenience. The value used for α in (1.3) is $0.95 \cdot (\rho(A))^{-1}$. We remark that for this example, the ranking is not very sensitive to the choice of the parameter α .

It is worth noting that all the centrality indices correctly identify nodes 1 and 34 as the most important ones. The list of the 5 most important nodes contains the same indices except for the betweenness centrality, which includes in the list node 32, and the closeness centrality, which determines that node 32 and 9 are among the 5 most important nodes.

The second example we are going to consider is the Facebook network included in the package. The graph has 63731 nodes and 1545686 edges. Neither the exponential nor the resolvent of the adjacency matrix A can be evaluated in a straightforward manner due to the large size of the matrix. We therefore apply the hybrid algorithm described in Section 4 to find the 10 most important nodes in the network.

Table 4 reports in each column the ranking of the 10 most important nodes according to the centrality indices described above and computed by the centrality function of Matlab. Table 5 reports the ranking of the 10 most important nodes obtained by the exp -subgraph,

Table 4. Ranking of the 10 most important nodes for the facebook network identified by the centrality function of Matlab.

degree	betweenness	closeness	eigenvector	pagerank
2332	554	2332	9904	554
471	471	471	2322	471
554	2332	23	5170	2332
2322	23	554	5157	23
451	451	1463	2362	451
23	280	207	3943	2208
2208	1463	280	7765	1463
9904	207	451	133	423
1463	84	1996	2332	280
3943	1996	2805	1902	207

Table 5. Ranking of the 10 most important nodes for the facebook network identified by the f -subgraph centrality and f -starting convenience.

sgcen_exp	sgcen_res(.95)	sgcen_res(.1)	stconv_exp	stconv_res(.95)	stconv_res(.1)
9904	9904	2332	9904	9904	2332
2322	2322	471	2322	2322	471
5170	5170	451	5170	5170	451
5157	3943	2208	5157	3943	2208
2362	2362	9904	2362	2362	9904
3943	7765	3943	3943	7765	3943
7765	5157	133	7765	5157	133
133	2332	423	133	2332	423
2332	1902	7765	2332	1902	7765
1902	133	14253	1902	133	14253

the res -subgraph, the exp -starting convenience and the res -starting convenience. The value used for α in (1.3) is $0.95 \cdot (\rho(A))^{-1}$ and $0.1 \cdot (\rho(A))^{-1}$.

It can be seen that the considered centrality indices generally produce different rankings. This confirms that they measure different features of the nodes in a network. It is remarkable to observe that in this example the exp -subgraph centrality and the exp -starting convenience produce the same list than the eigenvector centrality.

7. Discussion

This article introduces the SoftNet toolbox written in MATLAB, designed to compute the most important nodes of a network by some centrality indices based on the computation of matrix functions. The methods used to perform the computation were introduced in [20]. The use of the toolbox is illustrated by examples.

Author Contributions: All authors equally contributed to the research that led to this paper. All authors have read and agreed to the published version of the manuscript.

Funding: C.F. and G.R. were partially supported by the Regione Autonoma della Sardegna research project “Algorithms and Models for Imaging Science (AMIS)” [RASSR57257] and the INdAM-GNCS research project “Tecniche numeriche per l’analisi delle reti complesse e lo studio dei problemi inversi”. C.F. gratefully acknowledges Regione Autonoma della Sardegna for the financial support provided under the Operational Programme P.O.R. Sardegna F.S.E. (European Social Fund 2014-2020 - Axis III Education and Formation, Objective 10.5, Line of Activity 10.5.12).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in the numerical experiments is contained in the software package presented in the paper. Data sources are acknowledged in the text.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GUI graphical user interface

References

- M. Aprahamian, D. J. Higham, and N. J. Higham, *Matching exponential-based and resolvent-based centrality measures*, *J. Complex Networks*, 4 (2016), pp 157–176.
- J. Baglama, D. Calvetti, and L. Reichel, *IRBL: An implicitly restarted block Lanczos method for large-scale Hermitian eigenproblems*, *SIAM J. Sci. Comput.* **2003**, 24, 1650–1677.
- J. Baglama, D. Calvetti, and L. Reichel, *Algorithm 827: irbleigs: A MATLAB program for computing a few eigenpairs of a large sparse Hermitian matrix*, *ACM Trans. Math. Soft.* **2003**, 29, 337–348.
- V. Batagelj and A. Mrvar *Pajek datasets 2006*, Available online: <http://vlado.fmf.uni-lj.si/pub/networks/data/> (accessed on 28 June 2022).
- M. Benzi and P. Boito, *Quadrature rule-based bounds for functions of adjacency matrices*, *Linear Algebra and its Appl.* **2010**, 433, 637–652.
- M. Benzi and C. Klymko, *Total communicability as a centrality measure*, *J. Complex Networks*, 1 (2013), pp. 124–149.
- D. A. Bini, G. M. Del Corso, and F. Romani, *Evaluating scientific products by means of citation-based models: a first analysis and validation*, *Electron. Trans. Numer. Anal.* **2008**, 33, 1–16.
- D. A. Bini, G. M. Del Corso, and F. Romani, *A combined approach for evaluating papers, authors and scientific journals*, *J. Comput. Appl. Math.*, 234 (2010), pp. 3104–3121.
- P. F. Bonacich, *Power and centrality: A family of measures*, *Am. J. Sociol.*, 92 (1987), pp. 1170–1182.
- J. J. Crofts, E. Estrada, D. J. Higham, and A. Taylor, *Mapping directed networks*, *Electron. Trans. Numer. Anal.*, 37 (2010), pp. 337–350.
- J. J. Crofts and D. J. Higham, *Googling the brain: Discovering hierarchical and asymmetric network structures, with applications in neuroscience*, *Internet Mathematics*, 7 (2011), pp. 233–254.
- G. M. Del Corso and F. Romani, *Versatile weighting strategies for a citation-based research evaluation model*, *Bull. Belg. Math. Soc. Simon Stevin*, 16 (2009), pp. 723–743.
- E. Estrada, *The Structure of Complex Networks*, Oxford University Press: Oxford, 2012.
- E. Estrada and N. Hatano. *Statistical-mechanical approach to subgraph centrality in complex networks*, *Chemical Physics Letters*, 439 (2007), pp. 247–251.
- E. Estrada and N. Hatano, *Communicability in complex networks*, *Phys. Rev. E* **2008**, 77, 036111 (12 pages).
- E. Estrada, N. Hatano, and M. Benzi, *The physics of communicability in complex networks*, *Phys. Rep.* **2012**, 514, 89–119.
- E. Estrada and D. J. Higham, *Network properties revealed through matrix functions*, *SIAM Rev.* **2010**, 52, pp. 696–714.
- E. Estrada and J. A. Rodríguez-Velázquez, *Subgraph centrality in complex networks*, *Phys. Rev. E* **2005**, 71, 056103.
- E. Estrada and J. A. Rodríguez-Velázquez, *Subgraph centrality and clustering in complex hyper-networks*, *Physica A* **2006**, 364, 581–594.
- C. Fenu, D. Martin, L. Reichel, and G. Rodriguez, *Network analysis via partial spectral factorization and Gauss quadrature*, *SIAM J. Sci. Comput.* **2013**, 35, A2046–A2068.
- C. Fenu, D. Martin, L. Reichel, and G. Rodriguez, *Block Gauss and anti-Gauss quadrature with application to networks*, *SIAM J. Matrix Anal. Appl.* **2013**, 34, 1655–1684.
- D. F. Gleich, *PageRank beyond the web*. *SIAM Review*, 57 (2015), pp. 321–363.
- G. H. Golub and G. Meurant, *Matrices, moments and quadrature* in *Numerical Analysis*, eds. D. F. Griffiths and G. A. Watson, Longman: Essex, England, 1994; pp. 105–156.
- G. H. Golub and G. Meurant, *Matrices, Moments and Quadrature with Applications*, Princeton University Press: Princeton, 2010.
- V. E. Henson and G. Sanders, *Locally supported eigenvectors and matrices associated with connected and unweighted power-law graphs*, *Electron. Trans. Numer. Anal.*, 39 (2012), pp. 353–378.
- H. Jeong, S. Mason, A.-L. Barabási, and Z. N. Oltvai, *Lethality and centrality of protein networks*, *Nature* **2001**, 411, 41–42.
- A. N. Langville and C. D. Meyer, *Google’s Pagerank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, 2006.
- The Max Plank Institute for Software Systems web site, Available online: <http://socialnetworks.mpi-sws.org/data-wosn2009.html> (accessed on 28 June 2022).
- M. E. J. Newman, *The structure of scientific collaboration networks*, *Proc. Natl. Acad. Sci. USA* **2001**, 98, 404–409.
- M. E. J. Newman, *Networks: An Introduction*, Oxford University Press: Oxford, 2010.
- M. E. J. Newman’s web page, Available online: <http://www-personal.umich.edu/~mejn/netdata/> (accessed on 28 June 2022).
- S. Sun, L. Ling, N. Zhang, G. Li, and R. Chen, *Topological structure analysis of the protein-protein interaction network in budding yeast*, *Nucleic Acids Res.* **2003**, 31, 2443–2450.

-
33. A. Taylor and D. J. Higham, *CONTEST: A controllable test matrix toolbox for MATLAB*, *ACM Trans. Math. Softw.* **2009**, 35, 26:1–26:17. 365
 34. B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, *On the evolution of user interaction in Facebook*, In Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks (WOSN'09), Barcelona, Spain, August 2009. 366
 35. D. J. Watts and S. H. Strogatz, *Collective dynamics of 'small-world' networks*, *Nature* **1998**, 393, 440–442. 368