Computation of centrality and communicability indices using generalized matrix functions

Francesca Arrigo, Michele Benzi, and Caterina Fenu

Structured Matrix Days, Limoges
May 9–10, 2016

Overview

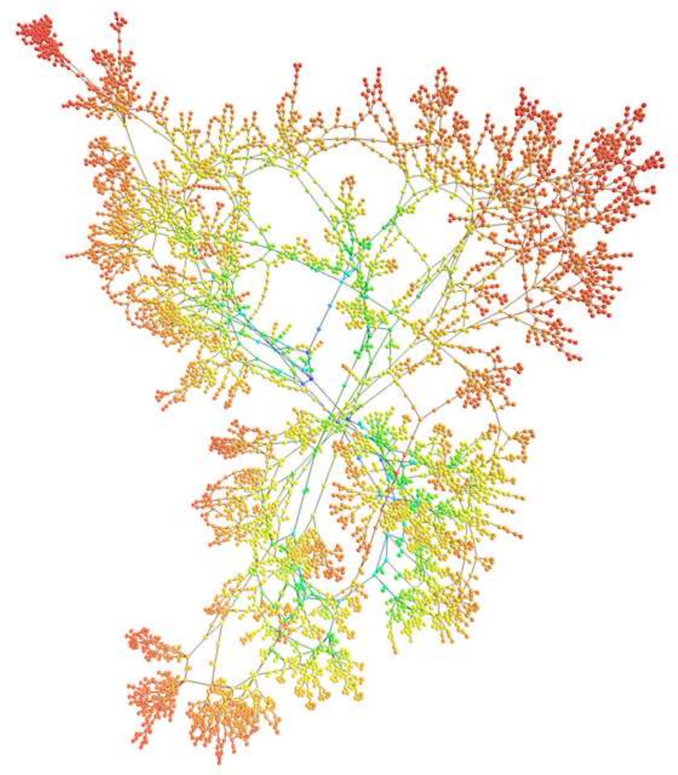
Complex Networks

A complex network is a sparse graph that exhibits non-trivial topological features.

A complex network is a sparse graph that exhibits non-trivial topological features.

They are used to model interactions of various types:

- social networks: collaboration, friendship,...;
- biological networks: PPI, food webs,...;
- technological networks: www, internet,...;
- transportation network: road maps, air routes,...
- Facebook, Twitter,...



In a nutshell

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a complex network with $n = |\mathcal{V}|$ nodes and $m = |\mathcal{E}|$ edges.
The graph is said to be undirected if $(i, j) \in \mathcal{E} \Leftrightarrow (j, i) \in \mathcal{E}$, directed otherwise.

Any graph $\mathcal{G}$ can be represented using an adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ defined as

$$a_{ij} = \begin{cases} \omega_{ij} & \text{if } (i, j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

where $\omega_{ij} \in \mathbb{R}_{>0}$ are weights for the elements in $\mathcal{E}$.

In a nutshell

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a complex network with $n = |\mathcal{V}|$ nodes and $m = |\mathcal{E}|$ edges.
The graph is said to be undirected if $(i, j) \in \mathcal{E} \Leftrightarrow (j, i) \in \mathcal{E}$, directed otherwise.

Any graph $\mathcal{G}$ can be represented using an adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ defined as

$$a_{ij} = \begin{cases} \omega_{ij} & \text{if } (i, j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

where $\omega_{ij} \in \mathbb{R}_{>0}$ are weights for the elements in $\mathcal{E}$.

Suppose $\mathcal{G}$ unweighted, with no multiple edges nor self loops, and sparse.
Thus, A is binary, sparse, and $a_{ii} = 0$ for all $i = 1, 2, \ldots, n$.

A walk of length k is a sequence of $k+1$ nodes $i_1, i_2, \ldots, i_{k+1} \in \mathcal{V}$ such that $(i_l, i_{l+1}) \in \mathcal{E}$ for all $1 \leq l \leq k$.

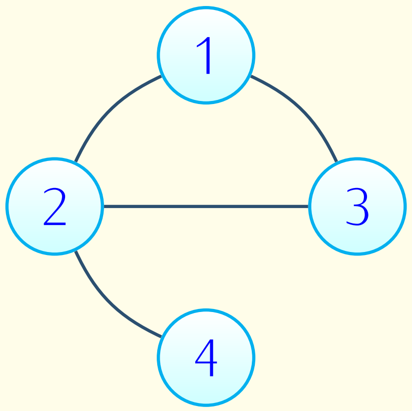
$$i_1 — i_2 — i_3 — \cdots — i_{k+1}.$$

The walk is said to be closed if $i_1 = i_{k+1}$.

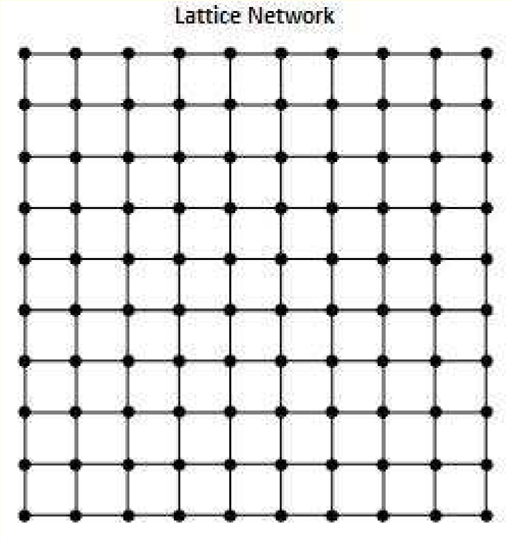
The quantities $\left(A^k\right)_{ii}$, $\left(A^k\right)_{ij}$ count closed (resp., open) walks of length k.

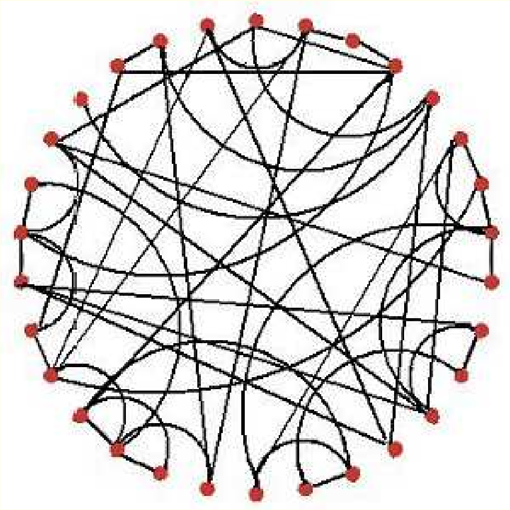
Example: the degree of node i is defined as

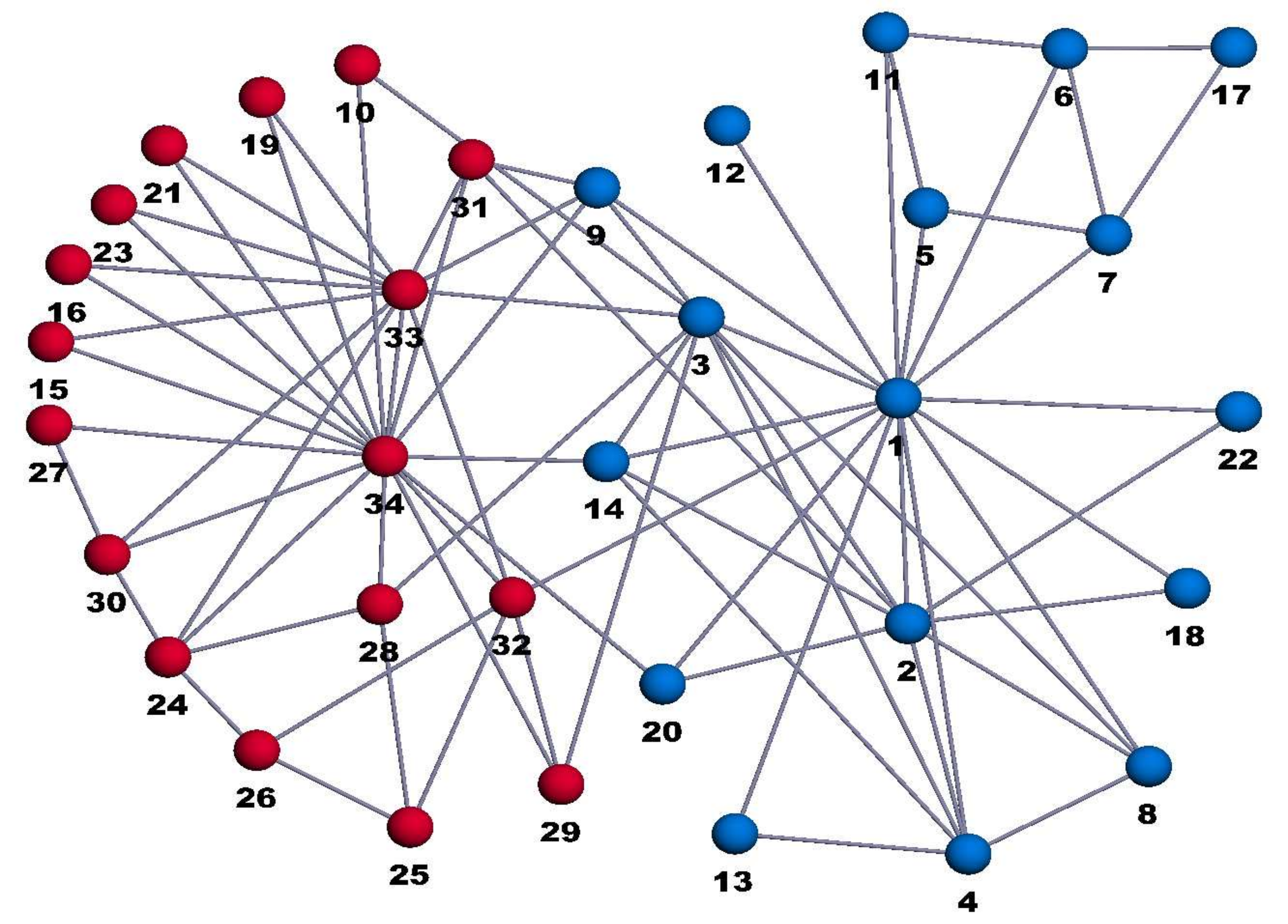
$$d_i = |\{j \in \mathcal{V} : (i,j) \in \mathcal{E}\}| = \left(A^2\right)_{ii}.$$



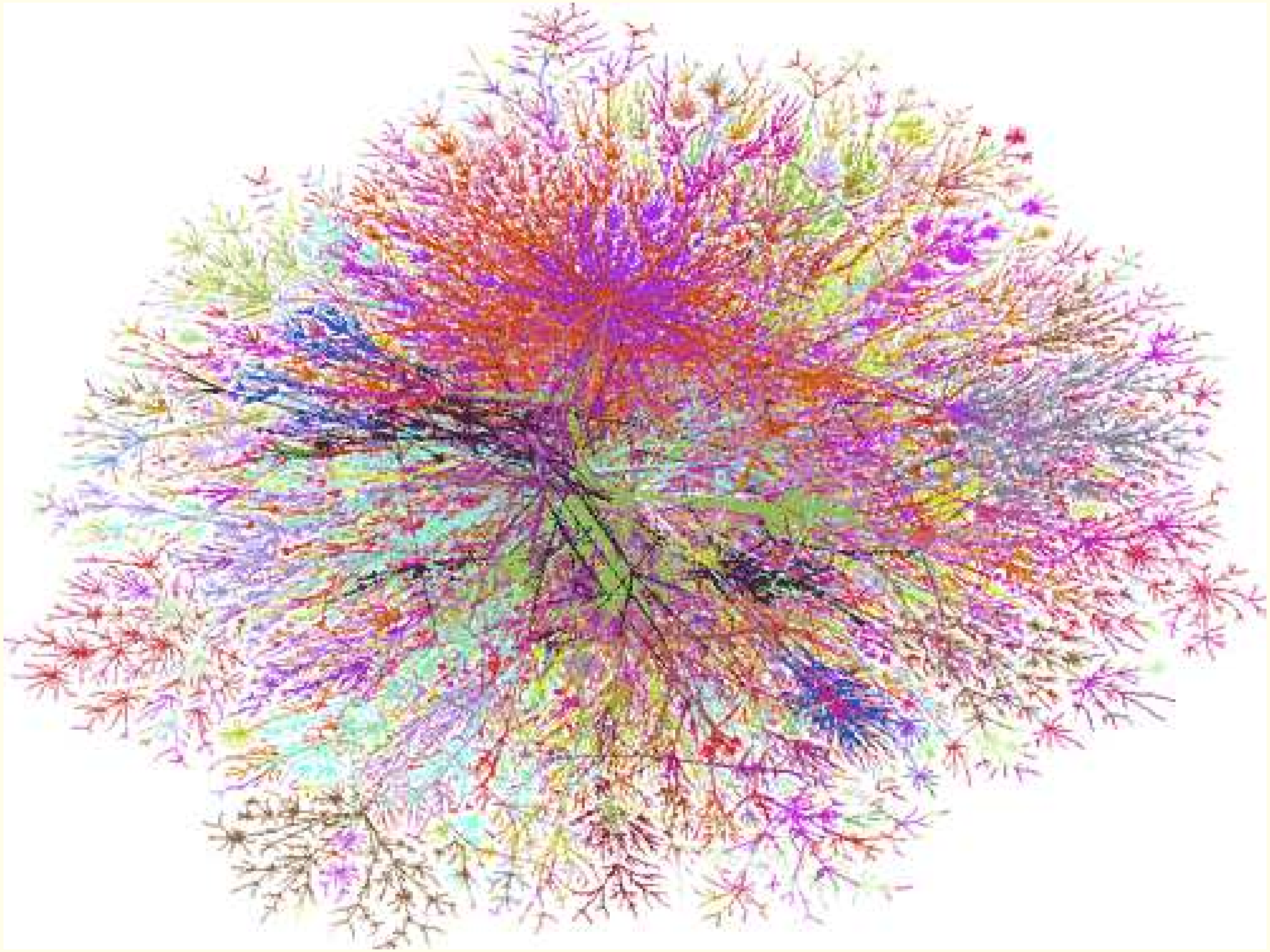
A complex network is not a regular, nor a completely random graph.





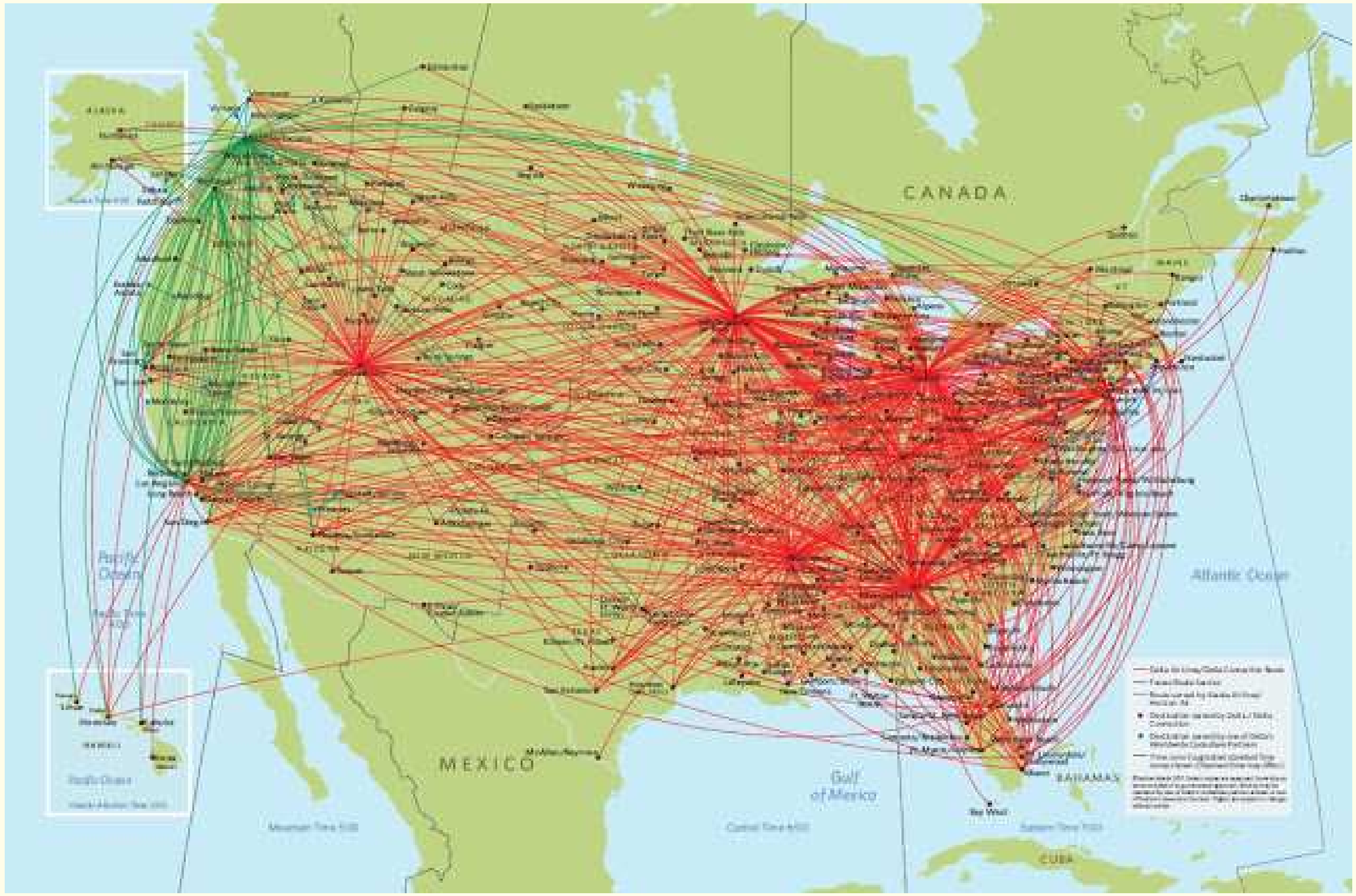

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

Structure?



Structure?

Complex Networks
In a nutshell
Structure?
A hint of (generalized) matrix functions
Centrality and Communicability
Let's focus on the case $A^T \neq A$
Approximating centrality and communicability indices



- High degree of *transitivity*: the ratio of the number of triangles to the number of open paths of length two is very high.

$$\overline{C} = \frac{1}{n} \sum_{j=1}^{n} \frac{2(A^3)_{jj}}{d_j(d_j - 1)}$$

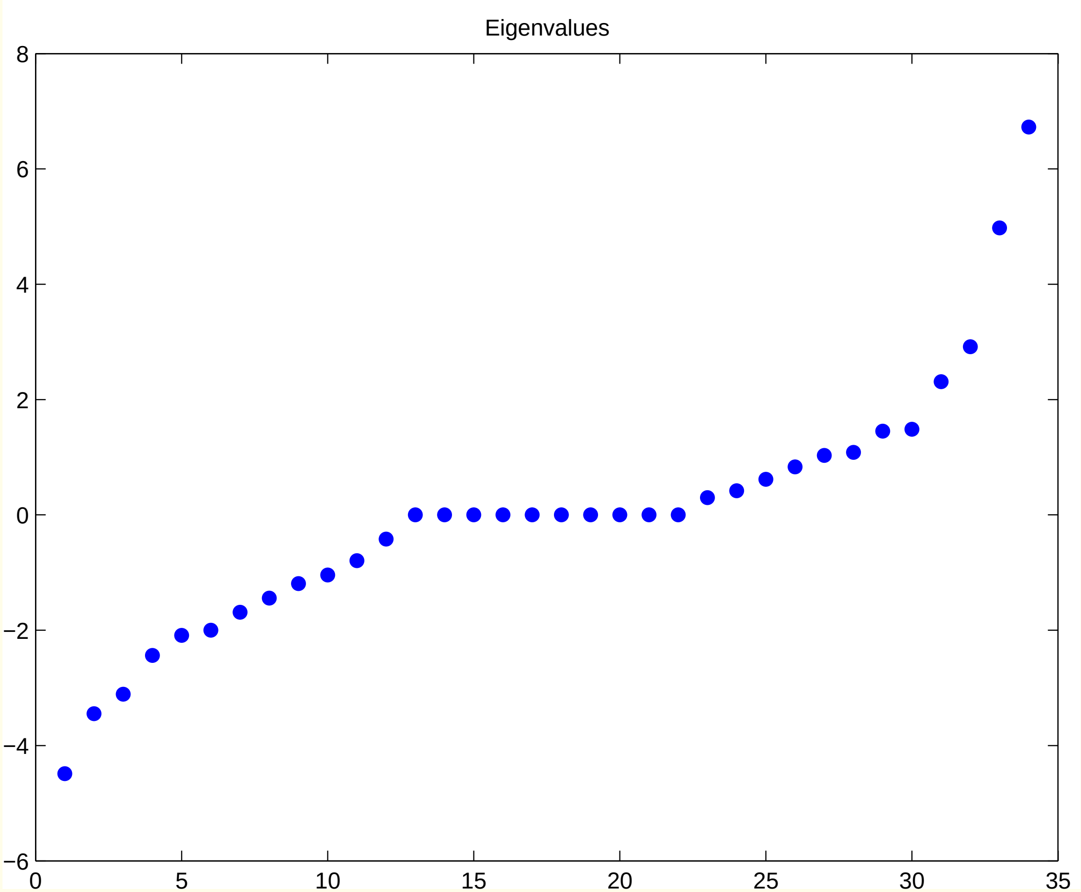
- Small *average path length*: on average it takes a small number of steps to go from one node to another.

$$\overline{\ell} = \frac{1}{n(n-1)} \sum_{i,j} \mathrm{d}(i,j)$$

- Small *diameter*: the largest distance between two nodes is proportional to $O(\log n)$.

$$\mathrm{diam}(\mathcal{G}) = \max_{i,j \in \mathcal{V}} \mathrm{d}(i,j)$$

The eigenvalues usually behave like this:



This is how the *degrees* of the nodes distribute:

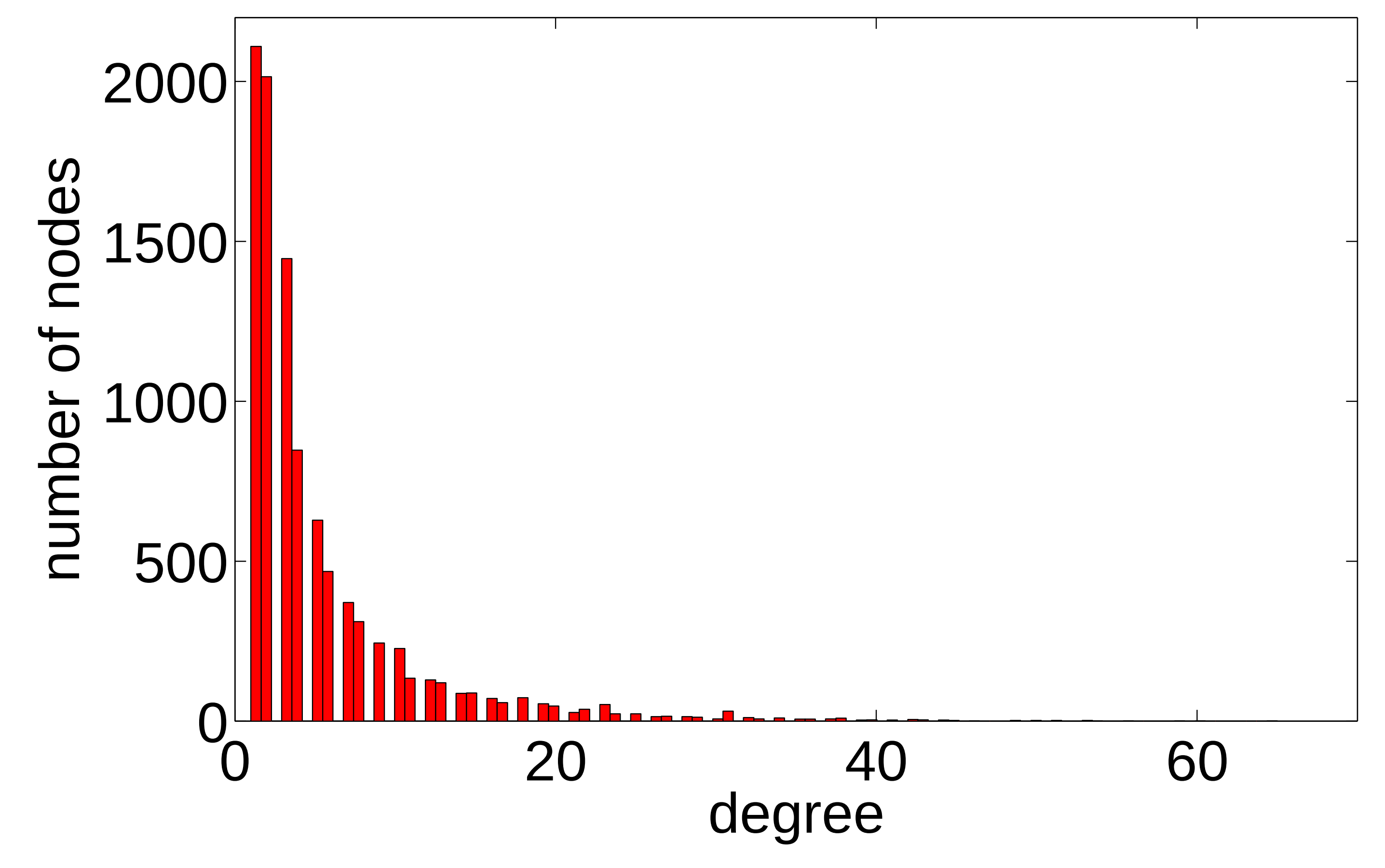


Figure courtesy of Prof. Benzi and Dr. Klymko!

A hint of (generalized) matrix functions

Let $\{\lambda_1, \lambda_2, \dots, \lambda_s\}$ be the set of distinct eigenvalues of A and let n_i denote the *index* of the ith eigenvalue, i.e., the size of the largest Jordan block associated with λ_i.

Let $Z^{-1}AZ = J = \mathrm{diag}(J_1, J_2, \dots, J_s)$ be the *Jordan canonical form* of the matrix, where Z is nonsingular. Each diagonal block $J_1, J_2, \dots, J_s$ is block diagonal and has the form

$$J_i = \mathrm{diag}(J_i^{(1)}, J_i^{(2)}, \dots, J_i^{(g_i)}),$$

where g_i is the geometric multiplicity of the ith distinct eigenvalue,

$$J_i^{(j)} = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix} \in \mathbb{C}^{\nu_i^{(j)} \times \nu_i^{(j)}},$$

and $\sum_{i=1}^{s} \sum_{j=1}^{g_i} \nu_i^{(j)} = n$.

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

Matrix functions

Let f be *defined on the spectrum of* $A \in \mathbb{C}^{n \times n}$, i.e., let the values $f^{(j)}(\lambda_i)$ exist for all $j = 0, \ldots, n_i - 1$ and for all $i = 1, \ldots, s$, where $f^{(j)}$ is the jth derivative of the function and $f^{(0)} = f$.

Def: Then the matrix function $f(A)$ is defined as

$$f(A) := Zf(J)Z^{-1} = Z\mathrm{diag}(f(J_1^{(1)}), \ldots, f(J_1^{(g_1)}), f(J_2^{(1)}), \ldots, f(J_s^{(g_s)}))Z^{-1},$$

where

$$f(J_i^{(j)}) := \begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{f^{(\nu_i^{(j)}-1)}(\lambda_i)}{(\nu_i^{(j)}-1)!} \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & f'(\lambda_i) \\ & & & f(\lambda_i) \end{pmatrix}.$$

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

Let $A \in \mathbb{C}^{m \times n}$ be a rank-r matrix. Then, the *Singular Value Decomposition* (*SVD*) of A is

$$A = U \Sigma V^*$$

where $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal and its entries $\Sigma_{ii} = \sigma_i$ are ordered as

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_q = 0,$$

where $q = \min\{m, n\}$. The matrices $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m] \in \mathbb{C}^{m \times m}$ and $V = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] \in \mathbb{C}^{n \times n}$ are unitary.

Remark: the matrix Σ is uniquely determined, while U and V are not.

Consider now the matrices $U_r = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r] \in \mathbb{C}^{m \times r}$ and $V_r = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r] \in \mathbb{C}^{n \times r}$, and let $\Sigma_r = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r) \in \mathbb{R}^{r \times r}$.

Then a *Compact SVD* (*CSVD*) of the matrix A is

$$A = U_r \Sigma_r V_r^* = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^* .$$

Generalized matrix functions

Linear and Multilinear Algebra, 1973, Vol. 1, pp. 163–171

Printed in Great Britain

On Generalized Matrix Functions

J. B. HAWKINS† and A. BEN-ISRAEL‡§
Department of Engineering Science, Northwestern University, Evanston, Illinois 60201

(Received August 25, 1972)

Def: Let $A \in \mathbb{C}^{m \times n}$ be a rank-r matrix and let $A = U_r \Sigma_r V_r^*$ be its CSVD. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a scalar function such that $f(\sigma_i)$ is well defined for all $i = 1, 2, \ldots, r$. The generalized matrix function $f^\diamond : \mathbb{C}^{m \times n} \longrightarrow \mathbb{C}^{m \times n}$ induced by f is defined as

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

where $f(\Sigma_r) = \mathrm{diag}(f(\sigma_1), f(\sigma_2), \ldots, f(\sigma_r))$.

Proposition: Let $A = U_r \Sigma_r V_r^* \in \mathbb{C}^{m \times n}$ be a matrix of rank r. Let $f : \mathbb{R} \to \mathbb{R}$ be a scalar function and let $f^\diamond : \mathbb{C}^{m \times n} \to \mathbb{C}^{m \times n}$ be the induced generalized matrix function, assumed to be defined at A. Then the following properties hold true.

(i) $[f^\diamond(A)]^* = f^\diamond(A^*)$;

(ii) let $X \in \mathbb{C}^{m \times m}$ and $Y \in \mathbb{C}^{n \times n}$ be two unitary matrices, then $f^\diamond(XAY) = X[f^\diamond(A)]Y$;

(iii) Let $P \in \mathbb{R}^{r \times r}$ be a permutation matrix and let $U_r = \mathscr{U}_r P$, $V_r = \mathscr{V}_r P$, and $\Sigma_r = P^T \mathscr{D}_r P$. Then $f^\diamond(A) = \mathscr{U}_r f(\mathscr{D}_r) \mathscr{V}_r^*$.

(iv) if $A = \mathrm{diag}(A_{11}, A_{22}, \ldots, A_{kk})$, then $f^\diamond(A) = \mathrm{diag}(f^\diamond(A_{11}), f^\diamond(A_{22}), \ldots, f^\diamond(A_{kk}))$;

(v) $f^\diamond(I_k \otimes A) = I_k \otimes f^\diamond(A)$, where I_k is the $k \times k$ identity matrix and $\otimes$ is the Kronecker product;

(vi) $f^\diamond(A \otimes I_k) = f^\diamond(A) \otimes I_k$.

Proposition [composite functions]: Let $A = U_r \Sigma_r V_r^* \in \mathbb{C}^{m \times n}$ be a rank-r matrix and let $\{\sigma_i : 1 \leq i \leq r\}$ be its singular values. Assume that $h : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are two scalar functions such that $h(\sigma_i) \neq 0$ and $g(h(\sigma_i))$ exist for all $i = 1, 2, \ldots, r$. Let $g^\diamond : \mathbb{C}^{m \times n} \to \mathbb{C}^{m \times n}$ and $h^\diamond : \mathbb{C}^{m \times n} \to \mathbb{C}^{m \times n}$ be the induced generalized matrix functions. Moreover, let $f : \mathbb{R} \to \mathbb{R}$ be the composite function $f = g \circ h$. Then the induced matrix function $f^\diamond : \mathbb{C}^{m \times n} \to \mathbb{C}^{m \times n}$ satisfies

$$f^\diamond(A) = g^\diamond(h^\diamond(A)).$$

Proposition: Let $A \in \mathbb{C}^{m \times n}$ be a rank-r matrix and let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be two scalar functions such that $f^\diamond(A)$ and $g(AA^*)$ are defined. Then

$$g(AA^*) f^\diamond(A) = f^\diamond(A) g(A^*A).$$

Centrality and Communicability

A node centrality measure is a function

$$C : \mathcal{V} \longrightarrow \mathbb{R}_{\geq 0}$$

which is invariant under graph isomorphism[a] and that assigns to each node in the graph a nonnegative score that quantifies its importance within the network.

[a] Two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ with associated adjacency matrices A_1 and A_2 are isomorphic if there exists a permutation matrix P such that $A_1 = PA_2P^T$.

Centrality measure	Undirected graph
degree	$d_i = (A\mathbf{1})_i$
eigenvector	$\mathbf{q}_1(i)$
f-subgraph	$\mathbf{e}_i^T f(A)\mathbf{e}_i$

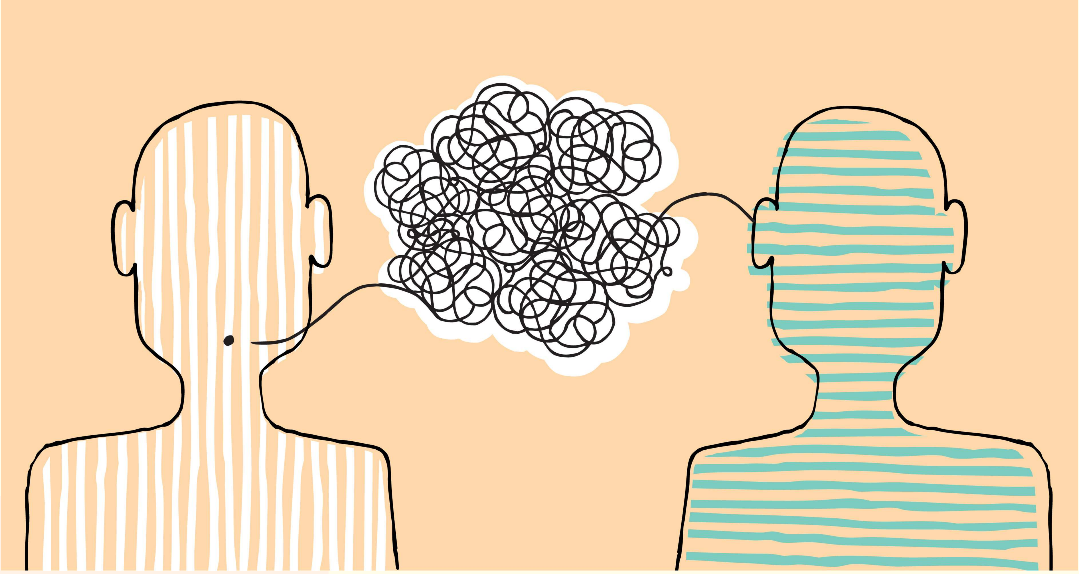
where

- $\mathbf{q}_1$ is the eigenvector associated with the leading eigenvaule λ_1 of A.
- $\mathbf{1}$ is the vector of all ones,
- $\mathbf{e}_i$ is the ith vector of the standard basis or $\mathbb{R}^n$,
- $f : \mathbb{R} \to \mathbb{R}$ is a scalar function (usually $f(x) = e^x$ and $f(x) = (1 - \alpha x)^{-1}$, where $\alpha \in (0, \lambda_1^{-1})$).

Communicability: a measure of the amount of information two nodes in the network exchange if we allow information to flow along walks of up to infinite length:

$$\mathbf{e}_i^T \left(e^A\right) \mathbf{e}_j = \sum_{k=0}^{\infty} \frac{\left(A^k\right)_{ij}}{k!}.$$

E. Estrada and N. Hatano, *"Communicability in complex networks"*, Phys. Rev. E, 77 (2008), 036111.



Total node communicability: a *centrality index* that describes the overall ability of a node to exchange information with the rest of the network:

$$TC(i) = \mathbf{e}_i^T e^A \mathbf{1} = \sum_{k=0}^{\infty} \sum_{j=1}^{n} \frac{\left(A^k\right)_{ij}}{k!}$$

M. Benzi and C. Klymko, *"Total communicability as a centrality measure"*, J. Complex Networks 1(2) (2013), pp. 124–149.

Total network communicability: a *global index* that describes the overall ability of the network to exchange information:

$$TC(A) = \mathbf{1}^T e^A \mathbf{1} = \sum_{k=0}^{\infty} \sum_{i,j=1}^{n} \frac{\left(A^k\right)_{ij}}{k!}$$

Let's focus on the case $A^T \neq A$

Twofold nature

Nodes play two different roles in digraphs:

Broadcast (hubs)

Receive (authorities)

Moreover, these two concepts are related through a recursive relationship:

good hubs are nodes that point to many good authorities.

good authorities are nodes that are pointed to by many good hubs.

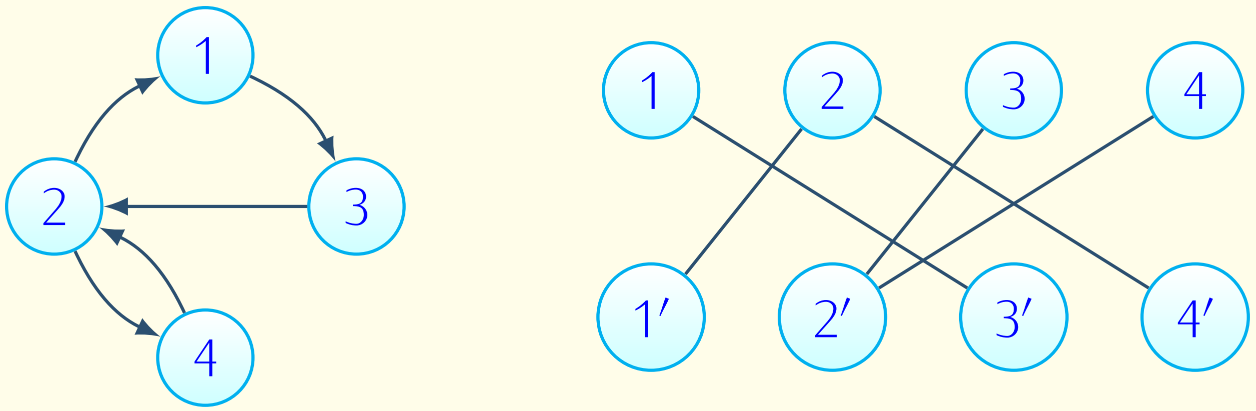
Figures from: http://www.bookkeeping-essentials.com/sensitive-information.html, http://www.carolinamailbox.com/

Digraphs as bigraphs

Let $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ be a graph such that

$$\begin{aligned} \mathscr{V} &= \mathcal{V} \cup \mathcal{V}' \\ \mathcal{V}' &= \{1' = n+1, 2' = n+2, \dots, n' = 2n\} \\ &= \text{set of copies of the nodes} \\ \mathscr{E} &= \{(i, j') : (i, j) \in \mathcal{E}\}. \end{aligned}$$

The associated adjacency matrix is $\mathscr{A} = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}$.



Centrality measure	Directed graph hub	Directed graph authority
degree	$d_{out}(i) = (A\mathbf{1})_i$	$d_{in}(i) = (A^T\mathbf{1})_i$
eigenvector	$\mathbf{x}_1(i)$	$\mathbf{y}_1(i)$
f–subgraph	$\mathbf{e}_i^T f(\mathscr{A})\mathbf{e}_i$	$\mathbf{e}_{i'}^T f(\mathscr{A})\mathbf{e}_{i'}$
f–total communicability	$[f(A)\mathbf{1}]_i$	$\left[f(A^T)\mathbf{1}\right]_i$
HITS	$\mathbf{u}_1(i)$	$\mathbf{v}_1(i)$

where

- $(\sigma_1, \mathbf{u}_1, \mathbf{v}_1)$ is the leading singular triplet of A.
- $\mathbf{x}_1, \mathbf{y}_1$ are the left and right e–vectors associated with the leading e–value of A.

Suppose that $f(z) = \sum_{k=0}^{\infty} a_k z^k$.

Then, within the radius of convergence:

$$f(\mathscr{A}) = \begin{pmatrix} f_{even}(\sqrt{AA^T}) & f^{\diamond}_{odd}(A) \\ f^{\diamond}_{odd}(A^T) & f_{even}(\sqrt{A^T A}) \end{pmatrix},$$

where

$$f(z) = f_{even}(z) + f_{odd}(z) = \sum_{k=0}^{\infty} a_{2k} z^{2k} + \sum_{k=0}^{\infty} a_{2k+1} z^{2k+1}.$$

Interpretation of the entries:

- the off-diagonal entries of the diagonal blocks describe the communicabilities between two nodes, both playing the role of hubs (resp., authorities).
- the off-diagonal entries of the off-diagonal blocks describe the communicabilities between two nodes, when one is playing the role of hub and the other that of authority.

Let $f(x) = (1-\alpha x)^{-1}$, where $\alpha \in (0, \sigma_1^{-1})$:

$$(I - \alpha\mathscr{A})^{-1} = \begin{pmatrix} (I-\alpha^2 AA^T)^{-1} & h^\diamond(A) \\ h^\diamond(A^T) & (I-\alpha^2 A^T A)^{-1} \end{pmatrix},$$

where $h(t) = \frac{\alpha t}{1-(\alpha t)^2}$.

In the off-diagonal blocks we have:

- *resolvent-based communicability* between node i (hub) and node j (authority): $[h^\diamond(A)]_{ij}$.
- *resolvent-based communicability* between node i (authority) and node j (hub): $[h^\diamond(A^T)]_{ij}$.

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

The matrix exponential of the adjacency matrix $\mathscr{A}$ is:

$$e^{\mathscr{A}} = \begin{pmatrix} \cosh(\sqrt{AA^T}) & \sinh^{\diamond}(A) \\ \sinh^{\diamond}(A^T) & \cosh(\sqrt{A^T A}) \end{pmatrix}.$$

Generalized total communicability: a hub (resp., authority) centrality measure that accounts for the overall ability of the node to exchange information with all the other nodes in the network, when they are acting as authorities (resp., hub).

$$T_h C = \mathbf{e}_i^T \sinh^{\diamond}(A)\mathbf{1} \qquad T_a C = \mathbf{e}_i^T \sinh^{\diamond}(A^T)\mathbf{1}.$$

F.A. and M. Benzi, *"Edge modification criteria for enhancing the communicability of digraphs"*, SIAM J. Matrix Anal. & Appl., 37(1), pp. 443–468 (2016).

Goal: we wanted to approximate bilinear forms involving generalized matrix functions, i.e., quantities like

$$\mathbf{z}^T f^{\diamond}(A)\mathbf{w},$$

where $A \in \mathbb{R}^{m \times n}$, $\mathbf{z} \in \mathbb{R}^m$, and $\mathbf{w} \in \mathbb{R}^n$.

Goal: we wanted to approximate bilinear forms involving generalized matrix functions, i.e., quantities like

$$\mathbf{z}^T f^{\diamond}(A)\mathbf{w},$$

where $A \in \mathbb{R}^{m\times n}$, $\mathbf{z} \in \mathbb{R}^m$, and $\mathbf{w} \in \mathbb{R}^n$.

Idea: Use the *Golub–Kahan bidiagonalization algorithm* to obtain approximations to the quantities of interest.

Generalized matrix functions have appeared in several framework, without being recognized as such:

- When computing $e^{A}\mathbf{b}$ for a skew symmetric matrix A, which is orthogonally similar to a matrix to $\mathscr{A}$ with B bidiagonal.

  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.

- They arise when *filter factors* are used to regularize discrete ill-posed problems, see, e.g.,

  M. Hanke, J. Nagy, and R. Plemmons, *Preconditioned iterative regularization for ill-posed problems*, in L. Reichel, A. Ruttan, and R. S. Varga, Eds., *Numerical Linear Algebra. Proceedings of the Conference in Numerical Linear Algebra and Scientific Computation, Kent, Ohio, USA, March 13–14, 1992*, de Gruyter, Berlin and New York, 1993, pp. 141–163.

- When approximating directed bipartite communities in digraphs: $f(A) = \cosh(\sqrt{AA^*}) - \sinh^{\diamond}(A)$.

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

- Applications in computer vision, structure from motion, photometric stereo and optical flow, finance, control system, computation of solution of Hamiltonian differential systems, etc...

Approximating centrality and communicability indices

We propose three different approaches to approximate the bilinear forms of interest:

(1) The first approach exploits a description of $\mathbf{z}^T f^\diamond(A)\mathbf{w}$ as a bilinear form that involves standard matrix functions of a tridiagonal matrix.

We propose three different approaches to approximate the bilinear forms of interest:

(1) The first approach exploits a description of $\mathbf{z}^T f^\diamond(A)\mathbf{w}$ as a bilinear form that involves standard matrix functions of a tridiagonal matrix.

Theorem: Let $A \in \mathbb{C}^{m\times n}$ be a rank-r matrix and let $f : \mathbb{R} \to \mathbb{R}$ be a scalar function. Let $f^\diamond : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ be the induced generalized matrix function. Then

$$f^\diamond(A) = f(\sqrt{AA^*})(\sqrt{AA^*})^\dagger A = A(\sqrt{A^*A})^\dagger f(\sqrt{A^*A}).$$

We propose three different approaches to approximate the bilinear forms of interest:

(1) The first approach exploits a description of $\mathbf{z}^T f^\diamond(A)\mathbf{w}$ as a bilinear form that involves standard matrix functions of a tridiagonal matrix.

Theorem: Let $A \in \mathbb{C}^{m\times n}$ be a rank-r matrix and let $f : \mathbb{R} \to \mathbb{R}$ be a scalar function. Let $f^\diamond : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ be the induced generalized matrix function. Then

$$f^\diamond(A) = f(\sqrt{AA^*})(\sqrt{AA^*})^\dagger A = A(\sqrt{A^*A})^\dagger f(\sqrt{A^*A}).$$

(2) The second approach works directly with the generalized matrix function and the Moore–Penrose pseudo-inverse of a bidiagonal matrix.

We propose three different approaches to approximate the bilinear forms of interest:

(1) The first approach exploits a description of $\mathbf{z}^T f^\diamond(A)\mathbf{w}$ as a bilinear form that involves standard matrix functions of a tridiagonal matrix.

Theorem: Let $A \in \mathbb{C}^{m\times n}$ be a rank-r matrix and let $f : \mathbb{R} \to \mathbb{R}$ be a scalar function. Let $f^\diamond : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ be the induced generalized matrix function. Then

$$f^\diamond(A) = f(\sqrt{AA^*})(\sqrt{AA^*})^\dagger A = A(\sqrt{A^*A})^\dagger f(\sqrt{A^*A}).$$

(2) The second approach works directly with the generalized matrix function and the Moore–Penrose pseudo-inverse of a bidiagonal matrix.

(3) The third approach first approximates the action of a generalized matrix function on a vector and then derives the approximation for the bilinear form of interest.

After ℓ steps, we have the decompositions

$$AQ_\ell = P_\ell B_\ell, \qquad A^T P_\ell = Q_\ell B_\ell^T + \gamma_\ell \mathbf{q}_\ell \mathbf{e}_\ell^T,$$

where

- $\mathbf{q}_\ell$ is the Lanczos vector computed at iteration $\ell + 1$,
- $Q_\ell = [\mathbf{q}_0, \mathbf{q}_1, \ldots, \mathbf{q}_{\ell-1}] \in \mathbb{R}^{n\times\ell}$ and $P_\ell = [\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{\ell-1}] \in \mathbb{R}^{m\times\ell}$ have orthonormal columns;
- $$B_\ell = \begin{pmatrix} \omega_1 & \gamma_1 & & \\ & \ddots & \ddots & \\ & & \omega_{\ell-1} & \gamma_{\ell-1} \\ & & & \omega_\ell \end{pmatrix} \in \mathbb{R}^{\ell\times\ell},$$
- $\mathbf{q}_0 = \mathbf{w}$ is the normalized starting vector.

Let $\widetilde{\mathbf{z}} = A^T\mathbf{z}$ and $g(t) = (\sqrt{t})^{-1} f(\sqrt{t})$; then we can write

$$\mathbf{z}^T f^{\diamond}(A)\mathbf{w} = \widetilde{\mathbf{z}}^T g(A^T A)\mathbf{w}.^{(1)}$$

(1) Remark: Unless $\widetilde{\mathbf{z}} = \mathbf{w}$, one has to use the *polarization identity*.

Let $\widetilde{\mathbf{z}} = A^T\mathbf{z}$ and $g(t) = (\sqrt{t})^{-1} f(\sqrt{t})$; then we can write

$$\mathbf{z}^T f^\diamond(A)\mathbf{w} = \widetilde{\mathbf{z}}^T g(A^T A)\mathbf{w}.^{(1)}$$

We can rewrite $\mathbf{w}^T g(A^T A)\mathbf{w}$ as a Riemann–Stieltjes integral:

$$\mathbf{w}^T g(A^T A)\mathbf{w} = \mathbf{w}^T V_r g(\Sigma_r^2) V_r^T \mathbf{w} = \sum_{i=1}^{r} \frac{f(\sigma_i)}{\sigma_i} (\mathbf{v}_i^T \mathbf{w})^2 = \int_{\sigma_r^2}^{\sigma_1^2} g(t)\, \mathrm{d}\alpha(t),$$

where

$$\alpha(t) = \begin{cases} 0, & \text{if } t < \sigma_r^2 \\ \sum_{i=j+1}^{r} (\mathbf{v}_i^T \mathbf{w})^2, & \text{if } \sigma_{j+1}^2 \leq t < \sigma_j^2 \\ \sum_{i=1}^{r} (\mathbf{v}_i^T \mathbf{w})^2, & \text{if } t \geq \sigma_r^2. \end{cases}$$

(1) Remark: Unless $\widetilde{\mathbf{z}} = \mathbf{w}$, one has to use the *polarization identity*.

Combining the equations leads to

$$A^T A Q_\ell = Q_\ell B_\ell^T B_\ell + \gamma_\ell \omega_\ell \mathbf{q}_\ell \mathbf{e}_\ell^T.$$

The matrix

$$T_\ell = B_\ell^T B_\ell$$

is symmetric and tridiagonal and coincides (in exact arithmetic) with the matrix obtained when the Lanczos algorithm is applied to $A^T A$.

Combining the equations leads to

$$A^T A Q_\ell = Q_\ell B_\ell^T B_\ell + \gamma_\ell \omega_\ell \mathbf{q}_\ell \mathbf{e}_\ell^T.$$

The matrix

$$T_\ell = B_\ell^T B_\ell$$

is symmetric and tridiagonal and coincides (in exact arithmetic) with the matrix obtained when the Lanczos algorithm is applied to $A^T A$.

We approximate $\mathbf{w}^T g(A^T A)\mathbf{w}$ using an ℓ-point Gauss quadrature rule:

$$\mathcal{G}_\ell := \mathbf{e}_1^T g(T_\ell)\mathbf{e}_1 = \mathbf{e}_1^T (\sqrt{T_\ell})^\dagger f(\sqrt{T_\ell})\mathbf{e}_1.$$

We compute the ℓ-point Gauss quadrature rule in terms of the generalized matrix function of the bidiagonal matrix B_ℓ.

Remark: unless $\mathbf{z} = A\mathbf{w}$ or $\mathbf{w} = A^T\mathbf{z}$, one has to use the polarization identity to estimate the bilinear forms of interest.

We compute the ℓ-point Gauss quadrature rule in terms of the generalized matrix function of the bidiagonal matrix B_ℓ.

Remark: unless $\mathbf{z} = A\mathbf{w}$ or $\mathbf{w} = A^T\mathbf{z}$, one has to use the polarization identity to estimate the bilinear forms of interest.

Proposition: Let be $A \in \mathbb{R}^{m\times n}$ and let $B_\ell \in \mathbb{R}^{\ell\times\ell}$ be the bidiagonal matrix computed at step ℓ of the Golub–Kahan bidiagonalization algorithm. Then, the ℓ-point Gauss quadrature rule $\mathcal{G}_\ell$ is given by

$$\mathcal{G}_\ell = \mathbf{e}_1^T B_\ell^\dagger f^\diamond(B_\ell)\mathbf{e}_1, \quad \text{if } \tilde{\mathbf{z}} = \mathbf{w},$$

or

$$\mathcal{G}_\ell = \mathbf{e}_1^T f^\diamond(B_\ell) B_\ell^\dagger \mathbf{e}_1, \quad \text{if } \mathbf{z} = \tilde{\mathbf{w}}.$$

After $\ell = r = \text{rank}(A)$ steps, we obtain the matrices P_r, B_r, and Q_r such that $A = P_r B_r Q_r^T$. The SVD of the bidiagonal matrix is $B_r = \mathcal{U}_r \Sigma_r \mathcal{V}_r^T$ and $\text{rank}(P_r B_r Q_r^T) = \text{rank}(B_r) = r$.

Thus

$$\mathbf{z}^T f^\diamond(A)\mathbf{w} = \mathbf{z}^T (P_r \mathcal{U}_r) f(\Sigma_r)(Q_r \mathcal{V}_r)^T \mathbf{w} = \widehat{\mathbf{z}}^T f^\diamond(B_r)\mathbf{e}_1,$$

where $\widehat{\mathbf{z}} = P_r^T \mathbf{z}$ and $Q_r^T \mathbf{w} = \mathbf{e}_1$.

Third approach

After $\ell = r = \text{rank}(A)$ steps, we obtain the matrices P_r, B_r, and Q_r such that $A = P_r B_r Q_r^T$. The SVD of the bidiagonal matrix is $B_r = \mathcal{U}_r \Sigma_r \mathcal{V}_r^T$ and $\text{rank}(P_r B_r Q_r^T) = \text{rank}(B_r) = r$.

Thus

$$\mathbf{z}^T f^\diamond(A)\mathbf{w} = \mathbf{z}^T (P_r \mathcal{U}_r) f(\Sigma_r)(Q_r \mathcal{V}_r)^T \mathbf{w} = \widehat{\mathbf{z}}^T f^\diamond(B_r)\mathbf{e}_1,$$

where $\widehat{\mathbf{z}} = P_r^T \mathbf{z}$ and $Q_r^T \mathbf{w} = \mathbf{e}_1$.

Assume now that $\ell < r$. We can then truncate the bidiagonalization process and approximate $f^\diamond(A)\mathbf{w}$ as

$$f^\diamond(A)\mathbf{w} \approx P_\ell f^\diamond(B_\ell)\mathbf{e}_1$$

and then obtain the approximation to the bilinear form of interest as

$$\mathbf{z}^T f^\diamond(A)\mathbf{w} \approx \mathbf{z}^T P_\ell f^\diamond(B_\ell)\mathbf{e}_1.$$

- Network: **ITWiki**. $n = 49728$, $m = 941425$, and $r > 200$.
- Computation of the *generalized total communicability* of a node i (broadcaster):

$$[\sinh^{\diamond}(A)\mathbf{1}]_i.$$

Stopping criterion:

$$\mathcal{R}_\ell = \left| \frac{x^{(\ell+1)} - x^{(\ell)}}{x^{(\ell)}} \right| \leq \text{tol},$$

where $x^{(\ell)}$ is the approximation at step ℓ obtained from the method under consideration.

Relative error:

$$\mathcal{E}_\ell = \frac{|x^{(\ell)} - \mathbf{e}_i^T \sinh^{\diamond}(A)\mathbf{1}|}{|\mathbf{e}_i^T \sinh^{\diamond}(A)\mathbf{1}|}.$$

ITwiki: $\mathbf{e}_i^T \sinh^{\diamond}(A)\mathbf{1}$ and tol $= 10^{-6}$.

	First approach		Second approach		Third approach	
	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$
1	5	3.88e-08	5	2.90e-08	6	8.02e-09
2	10	4.72e-05	9	4.68e-05	7	1.27e-08
3	5	3.20e-08	5	3.17e-08	6	7.01e-09
4	7	2.31e-05	9	2.33e-05	8	4.31e-09
5	8	4.20e-05	20	5.77e-05	8	5.91e-09
6	9	2.19e-04	24	2.13e-04	8	2.70e-08
7	6	4.26e-07	6	5.85e-07	7	3.15e-09
8	14	1.91e-04	29	2.24e-04	8	3.38e-09
9	5	8.57e-08	5	9.31e-08	6	5.07e-09
10	9	9.36e-06	8	1.12e-05	8	3.22e-10

- Network: **Roget**. $n = 1022$, $m = 5075$, and $r = 984$.
- Computation of the *resolvent-based communicability* between node i (broadcaster) and node j (receiver):

$$[h^\diamond(A)]_{ij},$$

where $h(t) = \alpha t(1 - (\alpha t)^2)^{-1}$ and $\alpha \in (0, \sigma_1^{-1})$.

Stopping criterion:

$$\mathcal{R}_\ell = \left| \frac{x^{(\ell+1)} - x^{(\ell)}}{x^{(\ell)}} \right| \leq \text{tol},$$

where $x^{(\ell)}$ is the approximation at step ℓ obtained from the method under consideration.

Relative error:

$$\mathcal{E}_\ell = \frac{|x^{(\ell)} - \mathbf{e}_i^T h^\diamond(A)\mathbf{e}_j|}{|\mathbf{e}_i^T h^\diamond(A)\mathbf{e}_j|}.$$

Roget: $\mathbf{e}_i^T h^\diamond(A)\mathbf{e}_j$, $h(t) = \frac{\alpha t}{1-(\alpha t)^2}$, $\alpha = \frac{1}{8\sigma_1}$, and tol $= 10^{-4}$.

	First approach		Second approach		Third approach	
	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$
1	619	2.44e+01	619	2.44e+01	6	1.76e-06
2	175	5.66e-02	175	5.66e-02	4	1.99e-08
3	160	1.12e+01	160	1.12e+01	5	6.54e-09
4	548	3.24e+00	548	3.24e+00	5	3.78e-08
5	204	5.62e-02	204	5.62e-02	4	1.41e-08
6	669	2.31e-01	669	2.31e-01	5	2.27e-08
7	69	2.64e-02	69	2.64e-02	4	6.60e-09
8	666	5.80e-01	666	5.80e-01	5	1.68e-08
9	124	4.19e-02	124	4.19e-02	4	8.40e-08
10	125	6.47e+01	125	6.47e+01	5	2.67e-08

Example 2

Roget: $\mathbf{e}_i^T h^{\diamond}(A)\mathbf{e}_j$, $h(t) = \frac{\alpha t}{1-(\alpha t)^2}$, $\alpha = \frac{1}{2\sigma_1}$, and tol $= 10^{-4}$.

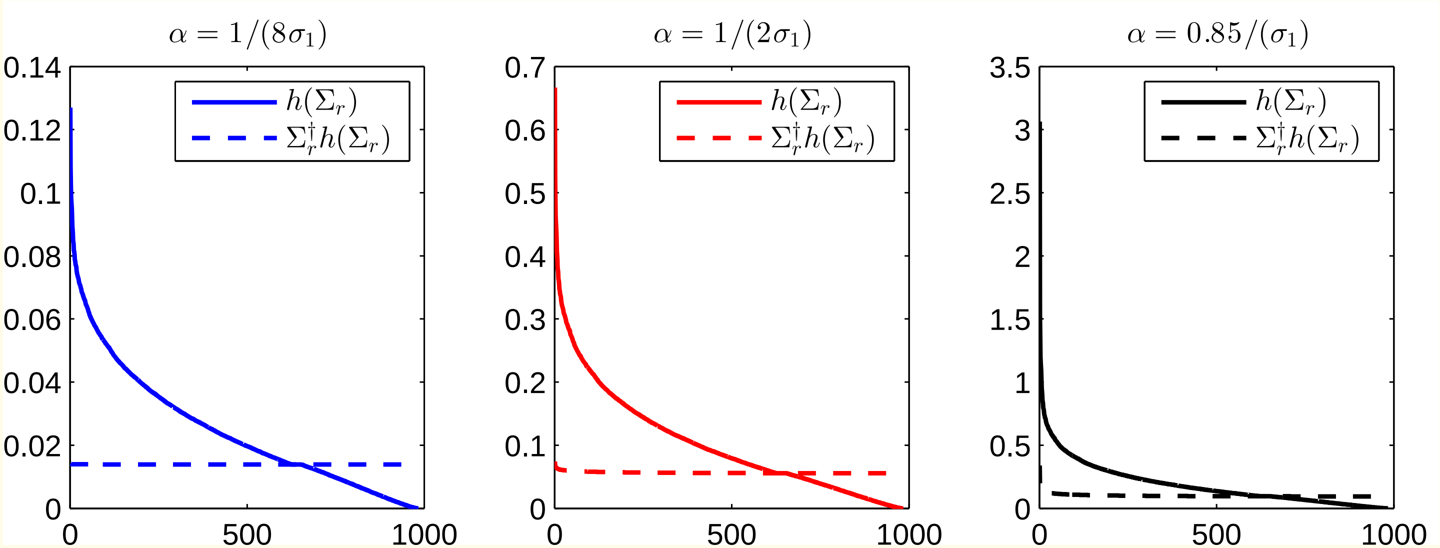
	First approach		Second approach		Third approach	
	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$
1	861	9.32e-03	861	9.32e-03	7	5.97e-05
2	312	2.04e+01	312	2.04e+01	8	2.48e-06
3	128	1.52e-01	128	1.52e-01	7	2.14e-06
4	241	4.16e-02	241	4.16e-02	7	4.59e-07
5	51	6.85e-02	51	6.85e-02	6	3.15e-08
6	125	1.86e-01	125	1.86e-01	7	5.85e-07
7	145	3.25e-02	145	3.25e-02	7	1.49e-07
8	36	2.67e-02	36	2.67e-02	6	1.28e-06
9	118	1.01e-01	118	1.01e-01	7	7.66e-07
10	342	4.52e-02	342	4.52e-02	7	1.45e-06

Roget: $\mathbf{e}_i^T h^\diamond(A)\mathbf{e}_j$, $h(t) = \frac{\alpha t}{1-(\alpha t)^2}$, $\alpha = \frac{0.85}{\sigma_1}$, and tol $= 10^{-4}$.

	First approach		Second approach		Third approach	
	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$
1	31	2.05e-02	31	2.05e-02	9	5.63e-07
2	73	2.72e-02	73	2.72e-02	9	3.38e-07
3	231	2.48e-02	231	2.48e-02	11	3.47e-06
4	151	1.90e-02	151	1.90e-02	10	1.94e-06
5	166	1.39e-01	166	1.39e-01	10	2.46e-06
6	43	3.64e-02	43	3.64e-02	9	1.49e-06
7	212	1.41e-01	212	1.41e-01	10	1.25e-05
8	84	1.45e-02	84	1.45e-02	10	1.64e-06
9	62	1.35e-03	62	1.35e-03	8	1.45e-05
10	82	3.18e-02	82	3.18e-02	9	2.58e-05

Example 2

Diagonal entries of $h(\Sigma_r)$ and $\Sigma_r^\dagger h(\Sigma_r)$ for $h(t) = \frac{\alpha t}{1-(\alpha t)^2}$, when $\alpha = \frac{1}{8\sigma_1}, \frac{1}{2\sigma_1}, \frac{0.85}{\sigma_1}$.



When dealing with bilinear forms that involve submatrices of $f(\mathscr{A})$, it is natural to compare these approaches with the use of Gauss quadrature-based bounds and estimates based on the Lanczos process (**mmq**).

Pros and cons:

(+) it can provide lower and upper bounds on the quantities being computed, if bounds on the singular values of A are available.

(=) the cost per step is comparable for the two methods.

(-) it requires working with vectors of length $2n$ instead of n.

(-) not applicable to generalized matrix functions that do not arise as submatrices of standard matrix functions.

G. Meurant, *MMQ toolbox for MATLAB*, `http://pagesperso-orange.fr/gerard.meurant/`.

Roget: $\mathbf{e}_i^T \sinh^\diamond(A)\mathbf{1}$ and tol $= 10^{-6}$.

	First approach		Second approach		Third approach		mmq	
	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$
1	10	4.69e-06	10	4.69e-06	9	1.09e-08	11	2.46e-09
2	6	4.48e-06	6	4.48e-06	9	3.71e-08	10	3.38e-08
3	9	4.27e-06	9	4.27e-06	9	5.07e-10	10	3.35e-08
4	5	1.82e-06	5	1.82e-06	8	2.36e-07	10	4.34e-08
5	13	4.25e-06	13	4.25e-06	7	1.29e-08	10	3.47e-09
6	10	1.20e-05	10	1.20e-05	9	4.34e-08	10	1.44e-08
7	4	1.96e-06	4	1.96e-06	9	4.63e-09	9	5.29e-10
8	17	1.02e-05	17	1.02e-05	9	8.20e-08	11	3.53e-09
9	6	1.26e-05	6	1.26e-05	10	2.13e-09	10	4.86e-08
10	13	3.62e-06	13	3.62e-06	9	7.45e-08	11	2.29e-09

G. Meurant, *MMQ toolbox for MATLAB*, `http://pagesperso-orange.fr/gerard.meurant/`.

- We've seen some basic things about complex networks. I hope I've convinced you that there is some structure here as well!
- Some new results on generalized matrix functions have been presented.
- We've described three different approximation approaches to estimate bilinear forms that involve generalized matrix functions (more about block techniques in the paper).

Conclusions

- We've seen some basic things about complex networks. I hope I've convinced you that there is some structure here as well!
- Some new results on generalized matrix functions have been presented.
- We've described three different approximation approaches to estimate bilinear forms that involve generalized matrix functions (more about block techniques in the paper).

Future work: we are just started!!

Conclusions

- We've seen some basic things about complex networks. I hope I've convinced you that there is some structure here as well!
- Some new results on generalized matrix functions have been presented.
- We've described three different approximation approaches to estimate bilinear forms that involve generalized matrix functions (more about block techniques in the paper).

Future work: we are just started!!

Further reading:

V. Noferini. *A Daleckii-Krein formula for the Frèchet derivative of a generalized matrix function* (2016).

F. Andersson, M. Carlsson, & K. M. Perfekt. *Operator-Lipschitz estimates for the singular value functional calculus*, Proceedings of the American Mathematical Society (2015).

Thank you

Questions?

Bibliography

[1] F. A. & M. Benzi, *"Edge modification criteria for enhancing the communicability of digraphs"*, SIAM J. Matrix Anal. & Appl., 37(1), pp. 443–468 (2016).

[2] F. A., M. Benzi, and C. Fenu, *"Computation of generalized matrix functions"*, SIAM J. Matrix Anal. & Appl. (in press).

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

[4] M. Benzi and C. Klymko, *Total communicability as a centrality measure*, J. Complex Networks 1(2) (2013), pp. 124–149.

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

[6] J. B. Hawkins & A. Ben–Israel, *On generalized matrix functions*, Linear and Multilinear Algebra 1(2) (1973), pp. 163–171.

[7] N. J. Higham, *Function of Matrices. Theory and Computation*, SIAM, Philadelphia, PA, 2008.

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

Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$. Then the *Riemann–Stieltjes integral* is denoted by

$$I = \int_a^b f(x)\, dg(x)$$

and it is defined as follows.
Consider a partition $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ of $[a, b]$ and let $\delta(P) = \max_{i=0:n-1} |x_i - x_{i+1}|$ be its norm.
For each subinterval consider a point $c_i \in [x_i, x_{i+1}]$, $\forall i = 0, \ldots, n-1$ and define

$$S(P, f, g) = \sum_{i=0}^{n-1} f(c_i) \left[g(x_{i+1}) - g(x_i) \right].$$

Then

$$I = \int_a^b f(x)\, dg(x) = \lim_{\delta(P) \to 0} S(P, f, g).$$

Block approaches

It is known that for this kind of problem, block algorithms are generally more efficient than the separate computation of each individual entry (or column) of $Z^T f^\diamond(A) W$.

When dealing with blocks Z and W with a number of columns $k > 1$, no information about the sign of the quadrature error can be obtained from the remainder formula for the Gauss quadrature rules. Therefore, we focus on the computation of approximations for the quantities of interest, rather than on bounds.

There is no equivalent to the polarization identity and thus we work directly with the blocks $\widetilde{Z}$ and W, $\widetilde{Z} \neq W$ (the case when Z and $\widetilde{W}$ are the initial blocks is similar).

Block approaches

As already pointed out, if we let $g(t) = (\sqrt{t})^{-1} f(\sqrt{t})$, it holds that

$$Z^T f^{\diamond}(A) W = \widetilde{Z}^T g(A^T A) W = Z^T g(AA^T) \widetilde{W},$$

where $\widetilde{Z} = A^T Z$ and $\widetilde{W} = AW$.
Let $\widetilde{Z}_0 \Delta_0^T \in \mathbb{R}^{n \times k}$ and $W_0 \Gamma_0^T \in \mathbb{R}^{m \times k}$ have all zero entries. Assume moreover that $\widetilde{Z}_1 = \widetilde{Z}$ and $W_1 = W$ satisfy $\widetilde{Z}_1^T W_1 = I_k$.
After ℓ steps, the nonsymmetric block Lanczos algorithm applied to the matrix $X = A^T A$ with initial blocks $\widetilde{Z}_1$ and W_1 yields the decompositions

$$X \left[\widetilde{Z}_1, \dots, \widetilde{Z}_\ell\right] = \left[\widetilde{Z}_1, \dots, \widetilde{Z}_\ell\right] J_\ell + \widetilde{Z}_{\ell+1} \Gamma_\ell \mathbf{E}_\ell^T,$$
$$X [W_1, \dots, W_\ell] = [W_1, \dots, W_\ell] J_\ell^T + W_{\ell+1} \Delta_\ell \mathbf{E}_\ell^T,$$

where J_ℓ is the matrix

$$J_\ell = \begin{pmatrix} \Omega_1 & \Delta_1^T & & & \\ \Gamma_1 & \Omega_2 & \Delta_2^T & & \\ & \ddots & \ddots & \ddots & \\ & & \Gamma_{\ell-2} & \Omega_{\ell-1} & \Delta_{\ell-1}^T \\ & & & \Gamma_{\ell-1} & \Omega_\ell \end{pmatrix} \in \mathbb{R}^{k\ell \times k\ell},$$

and $\mathbf{E}_i$, for $i = 1, 2, \dots, \ell$ are $k \times (k\ell)$ block matrices which contain $k \times k$ zero blocks everywhere, except for the ith block, which coincides with the identity matrix I_k.

First approach

We remark that if $\widetilde{Z} = W$, the use of the symmetric block Lanczos algorithm is preferable. In this case, the matrix J_ℓ is symmetric and the decompositions can be written as

$$X\,[W_1, \ldots, W_\ell] = [W_1, \ldots, W_\ell]\, J_\ell + W_{\ell+1}\Gamma_\ell \mathsf{E}_\ell^T.$$

The ℓ-block nonsymmetric Gauss quadrature rule $\mathcal{G}_\ell$ can then be expressed as

$$\mathcal{G}_\ell = \mathsf{E}_1^T g(J_\ell)\mathsf{E}_1.$$

The $(\ell+1)$-block anti-Gauss quadrature rule $\mathcal{H}_{\ell+1}$ is defined as the $(\ell+1)$-block quadrature rule such that

$$(\mathcal{I} - \mathcal{H}_{\ell+1})\, p = -(\mathcal{I} - \mathcal{G}_\ell)\, p, \qquad p \in \mathbb{P}^{2\ell+1},$$

where $\mathcal{I}p := \sum_{i=1}^{r} p(\sigma_i^2)\alpha_i\beta_i^T + \sum_{i=r+1}^{n} p(0)\alpha_i\beta_i^T = (\widetilde{Z}^T V)p(\Sigma^T\Sigma)(W^T V)^T$, with $\widetilde{Z}^T V = [\alpha_1, \alpha_2, \ldots, \alpha_n]$, $W^T V = [\beta_1, \beta_2, \ldots, \beta_n] \in \mathbb{R}^{k\times n}$, and $\mathbb{P}^{2\ell+1}$ is the set of polynomials of degree at most $2\ell+1$.
The $(\ell+1)$-block nonsymmetric anti-Gauss rule can be computed as

$$\mathcal{H}_{\ell+1} = \mathrm{E}_1^T g(\widetilde{J}_{\ell+1})\mathrm{E}_1,$$

where

$$\widetilde{J}_{\ell+1} = \begin{pmatrix} \Omega_1 & \Delta_1^T & & & \\ \Gamma_1 & \Omega_2 & \Delta_2^T & & \\ & \ddots & \ddots & \ddots & \\ & & \Gamma_{\ell-1} & \Omega_\ell & \sqrt{2}\Delta_\ell^T \\ & & & \sqrt{2}\Gamma_\ell & \Omega_{\ell+1} \end{pmatrix} \in \mathbb{R}^{k(\ell+1)\times k(\ell+1)}.$$

First approach

A pair of scalar Gauss and anti-Gauss rules provides *estimates* of upper and lower bounds on the bilinear form of interest. If we express $g(t)$ in terms of orthonormal polynomials, the coefficients in the expansion are $k \times k$ matrices.

Remark: to obtain good entrywise approximations it is necessary that the norm of the coefficients decays rapidly as ℓ increases. This condition is satisfied if $g(t)$ is analytic in a simply connected domain D enclosing the spectrum of $A^T A$, as long as the boundary ∂D is not close to the spectrum.

Then, the arithmetic mean

$$F_\ell = \frac{1}{2}\left(\mathcal{G}_\ell + \mathcal{H}_{\ell+1}\right)$$

can be used as an approximation of the matrix-valued expression $Z^T f^\diamond(A) W$.

Second approach

Assume that $W \in \mathbb{R}^{n\times k}$ is s.t. $W^T W = I_k$ and that the matrices $\Gamma_0 \in \mathbb{R}^{k\times k}$ and $P_0 \in \mathbb{R}^{m\times k}$ are zero matrices.
The first ℓ steps of the block Golub–Kahan algorithm with starting block $Q_1 = W$ are:

$$\begin{aligned} R_j &= AQ_j - P_{j-1}\Gamma_{j-1}^T, \\ P_j\Omega_j &= R_j, \\ S_j &= A^T P_j - Q_j\Omega_j^T, \\ Q_{j+1}\Gamma_j &= S_j, \end{aligned} \qquad j = 1, \ldots, \ell,$$

where $P_j\Omega_j = R_j$ and $Q_{j+1}\Gamma_j = S_j$ are QR factorizations of R_j and S_j, respectively.

After ℓ steps, the recursions yield the decompositions

$$A[Q_1, \ldots, Q_\ell] = [P_1, \ldots, P_\ell]\, B_\ell,$$
$$A^T[P_1, \ldots, P_\ell] = [Q_1, \ldots, Q_\ell]\, B_\ell^T + Q_{\ell+1}\Gamma_\ell \mathbf{E}_\ell^T,$$

where now

$$B_\ell = \begin{pmatrix} \Omega_1 & \Gamma_1^T & & & \\ & \Omega_2 & \Gamma_2^T & & \\ & & \ddots & \ddots & \\ & & & \Omega_{\ell-1} & \Gamma_{\ell-1}^T \\ & & & & \Omega_\ell \end{pmatrix} \in \mathbb{R}^{k\ell \times k\ell}.$$

Following the same reasoning as in the third approach, when $k\ell < r = \mathrm{rank}(A)$, we can approximate the quantities of interest as

$$Z^T f^\diamond(A) W \approx Z^T [P_1, \ldots, P_\ell] f^\diamond(B_\ell) \mathbf{E}_1 = F_\ell.$$