Lecture 4:
Matrix functions and complex networks

Paola Boito and Federico Poloni

Università di Pisa

Pisa - Hokkaido - Roma2 Summer School
Pisa, Aug 27 - Sept 8, 2018

Complex networks

Complex networks are graphs that model physical, biological or social systems. For instance: social networks, transportation networks, food webs, protein interaction, neural networks, computer code, power distribution, epidemiology networks, telecommunications networks, web pages...

Typical features

Interesting classes of complex networks generally exhibit typical behaviors and properties. For instance:

- degree distribution,
- assortativity,
- small world property,
- spectral properties...

Graphs and networks

- As a network model, let us consider G a connected simple graph with N nodes and m edges (i.e., G is unweighted, undirected and has no loops).
- Assume we have assigned labels $1, \ldots, N$ to the nodes.
- The degree d_i of a node i is the number of nodes that are adjacent (i.e., connected by an edge) to this node.
- The adjacency matrix $A \in \mathbb{R}^{N \times N}$ associated with G is defined as:
  - $A_{ij} = 1$ if nodes i and j are adjacent,
  - $A_{ij} = 0$ otherwise.

  In particular, A is symmetric, and $A_{ii} = 0$ for $1 \leq i \leq N$.

Degree distribution

The degree distribution of a network is given by

$$p(k) = \frac{N(k)}{N},$$

where $N(k)$ is the number of nodes having degree k and N is the total number of nodes.

For instance: power-law distribution (scale-free networks)

$$p(k) = Ak^{-\gamma}, \qquad A, \gamma > 0,$$

$\rightsquigarrow$ low probability of finding high-degree nodes, high probability of finding low-degree nodes.

Barabási-Albert model: each new node is adjacent to an existent one with probability proportional to its degree (preferential attachment). This gives a scale-free network.

See e.g., [Barabási and Bonabeau, *Scale-free networks*, Scientific American, May 2003] for an easy-to-read presentation.

Assortativity

What can be said about degree-degree correlation?

- A network in which high-degree nodes tend to connect to each other is called assortative.
- A network in which high-degree nodes tend to connect to low-degree nodes is called disassortative.
- Assortativity can be quantified by the correlation coefficient of the degrees of nodes at either side of each edge (Newman 2002).
- For instance, social networks tend to be assortative, other networks (e.g., biological) are often disassortative.

Shortest path distance

- Undirected network: $d(i, j)$ is the number of links in the shortest path connecting nodes i and j.
- Directed network: $\overrightarrow{d}(i, j)$ is the number of links in the directed shortest path going from node i to node j.
- If no such path exists: distance is set to ∞.
- Note that in general $\overrightarrow{d}(i, j) \neq \overrightarrow{d}(j, i)$.

Let's consider the undirected case:

- the Wiener index is $W(G) = \frac{1}{2} \sum_i \sum_j d(i, j)$,
- the average path length is $\ell = \frac{2W(G)}{N(N-1)}$.

Small-world networks

Milgram's experiment (1967): how many "degrees of separation" from Omaha (Nebraska) to Boston (Massachusetts)? Answer: about 5.5 on average.

Small-world networks are characterized by a short average distance between nodes (or by a slowly growing diameter) $\rightsquigarrow$ small world effect (Milgram 1967, Watts 1999, Buchanan 2003).

Short average distance is generally taken to mean $\ell \approx \ln N$.

Barabási-Albert is also an example of small-world model.

Small-world networks

Examples from CONTEST (Matlab toolbox by Taylor and Higham): a Watts-Strogatz model.

`A=smallw(`N`,`k`,`p`)`

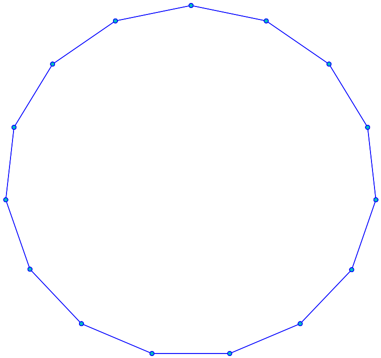
- N positive integer (number of nodes)
- k positive integer (number of connected neighbors)
- $0 \leq p \leq 1$ (probability of a shortcut)

This small-world model interpolates between a regular lattice and a "random" graph.

On the other hand, it is not scale-free (exponential decay of $p(k)$).

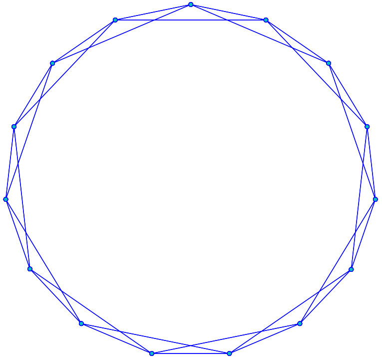
Small-world networks

Begin with a ring (e.g., a 1-ring)...



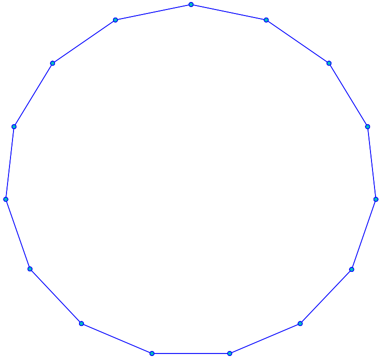
Small-world networks

Begin with a ring (e.g., a 2-ring)...



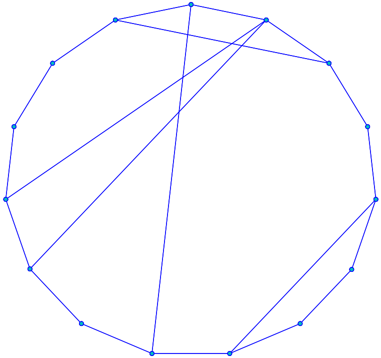
Small-world networks

Begin with a ring (e.g., a 1-ring)...



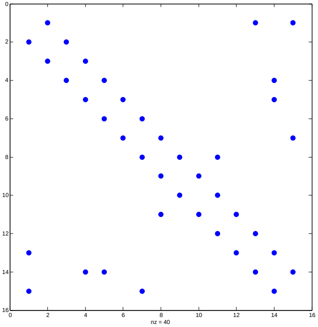
Small-world networks

Begin with a ring... and add shortcuts (with probability *p*).



Small-world networks

Here is the corresponding adjacency matrix ($N = 15$, $k = 1$, $p = 0.4$):



Spectral properties

- The spectra of the matrices related to a graph (adjacency matrix, Laplacian) give useful information on the graph itself.
- For instance, consider the eigenvalue of the Laplacian:

  $$0 = \mu_N \leq \mu_{N-1} \leq \ldots \leq \mu_1.$$

  The graph is connected iff $\mu_{N-1} > 0$.
- For an introduction, see e.g.
  P. Van Mieghem, *Graph Spectra for Complex Networks*, Cambridge University Press 2011.

Recall that the Laplacian of a graph is the $N \times N$ matrix defined as

$$L = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_N \end{bmatrix} - A.$$

Motivation and some literature

There has been a growing interest in complex networks during the last years in the applied mathematics/numerical analysis community:

- vast scientific literature,
- books
  - P. Van Mieghem, *Graph Spectra for Complex Networks*, Cambridge University Press 2011
  - M. Newman, *Networks: An Introduction*, Oxford University Press 2010
  - E. Estrada, *The Structure of Complex Networks. Theory and Applications*, Oxford University Press 2012.
- journals
  - *Journal of Complex Networks*, Oxford University Press,
  - *Network Science*, Cambridge University Press,
- review papers
- S. Strogatz, *Exploring complex networks*, Nature 410(8) 2001
- R. Albert and A.-L. Barabasi, *Statistical mechanics of complex networks*, Rev. Modern Physics 74 (2002)
- M. Newman, *The structure and function of complex networks*, SIAM rev. 45 (2003)
- E. Estrada, D. Higham, *Network Properties Revealed Through Matrix Functions*, SIREV 2010.

Complex networks

- Networks are described by graphs and associated matrices;
- Some network properties can be quantified via matrix functions;
- Relevant matrix functions can be investigated via quadrature formulas and decay properties.

Examples of interesting properties:

- importance/centrality of a vertex,
- connectivity of two given nodes,
- presence of hubs and authorities...

Plan (and more literature)

We will focus on :

- Networks and matrix functions:
  - Estrada, Higham, *Network Properties Revealed Through Matrix Functions*, SIREV 2010,
  - E. Estrada, *The Structure of Complex Networks. Theory and Applications*, Oxford University Press 2012.
- Krylov methods for approximating matrix functions (see [Higham 2008]).
- The directed case (hubs and authorities):
  - Benzi, Estrada, Klymko, *Ranking hubs and authorities using matrix functions*, LAA 2013
- Decay bounds: Benzi et al.

Graphs

Let us consider

- G a connected simple graph with N nodes and m edges (i.e., G is unweighted, undirected and has no loops),
- $A \in \mathbb{R}^{N \times N}$ the associated adjacency matrix:
  - assign labels $1, \ldots, N$ to the nodes,
  - $A_{ij} = 1$ if nodes i and j are adjacent, $A_{ij} = 0$ otherwise,
  - A is symmetric,
  - $A_{ii} = 0$, $1 \leq i \leq N$,
- $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ the eigenvalues of A.

Renumbering the nodes corresponds to a transformation $A \longleftarrow P\,A\,P^T$, where P is a suitable permutation matrix.

Powers of A

The link between complex networks and functions of matrices is based on the following property:

Let k be a positive integer. Then $A^k(i,j)$ counts the number of walks of length k in G that connect node i to node j.

Recall that a walk is an ordered list of nodes such that successive nodes in the list are connected. The nodes need not be distinct: some nodes may be revisited along the way. Compare to the notion of path, where nodes are required to be distinct.

The length of a walk is the number of edges that form the walk, that is, the number of nodes in the list minus one.

Powers of A

Proof. By induction on k:

- **For $k = 1$:** $A^1 = A$ and A_{ij} is the number of edges from v_i to v_j, that is, the number of walks of length 1 from v_i to v_j.
- **Inductive step:** assume that the property is true for k and let us prove it for $k+1$. The (i,j)-th entry of A^{k+1} is, by definition of matrix multiplication:

$$\begin{aligned} (A^{k+1})_{ij} &= (A\,A^k)_{ij} = \\ &= A_{i1}(A^k)_{1j} + A_{i2}(A^k)_{2j} + \ldots + A_{iN}(A^k)_{Nj} = \\ &= \sum_{\ell=1}^{N} A_{i\ell}(A^k)_{\ell j}. \end{aligned}$$

  By the inductive hypothesis, $A_{i1}(A^k)_{1j}$ is the number of walks of length k from v_1 to v_j times the number of walks of length 1 from v_j to v_1. So it is the number of walks of length $k+1$ from v_i to v_j such that v_1 is the second vertex. Analogously, for each ℓ we have that $A_{i\ell}(A^k)_{\ell j}$ is the number of walks of length $k+1$ from v_i to v_j such that v_ℓ is the second vertex. So the sum of all terms $A_{i\ell}(A^k)_{\ell j}$ is the number of all possible walks of length $k+1$ from v_i to v_j.

Degree and subgraph centrality

For a node i, define

- its degree $d_i := \sum_{k=1}^{N} A_{ik} = A\,\mathbb{1}$
- its subgraph centrality $c(i) = [e^A]_{ii}$.

Both d_i and $c(i)$ quantify how "well-connected" a node is:

- d_i counts the number of neighbors of node i (but does not take into account their importance),
- $c(i)$ counts the number of walks in G that begin and end at node i; each walk $\mathcal{W}$ carries a weight $\frac{1}{\text{length}(\mathcal{W})!}$, so that longer walks are penalized. In fact we have

$$
\begin{aligned}
c(i) &= \frac{1}{2}A^2(i,i) + \frac{1}{3!}A^3(i,i) + \frac{1}{4!}A^4(i,i) + \dots \\
&= A(i,i) + \frac{1}{2}A^2(i,i) + \frac{1}{3!}A^3(i,i) + \frac{1}{4!}A^4(i,i) + \dots \\
&\approx 1 + A(i,i) + \frac{1}{2}A^2(i,i) + \frac{1}{3!}A^3(i,i) + \frac{1}{4!}A^4(i,i) + \dots \\
&= \left[I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \frac{1}{4!}A^4 + \dots \right](i,i)
\end{aligned}
$$

Centrality: other definitions

Other definitions of centrality have been proposed. For instance:

- Katz (1953):

$$k(i) = \sum_{j=1}^{N} \sum_{k=0}^{\infty} \alpha^k A^k(i,j) = (((I - \alpha A)^{-1} - I)\,\mathbb{1})_i,$$

  where α is a suitably chosen parameter ($\alpha < \lambda_1^{-1}$);
- Bonacich (*eigenvector centrality*): $b(i)$ is the i-th entry of the dominant eigenvector of A, that is, the Perron-Frobenius eigenvector (compare with PageRank);
- definitions based on shortest paths (*closeness centrality, betweenness centrality*).

However, we will not use these definitions here. See [Benzi and Klymko 2014] for a comparison.

Estrada index

The Estrada index of G is

$$EE(G) := \sum_{i=1}^{N} c(i) = \sum_{i=1}^{N} [e^{A}]_{ii} = \sum_{i=1}^{N} e^{\lambda_i}$$

(see e.g., [Estrada and Higham 2008]).

Variants of the Estrada index can also be used to measure the "bipartiteness" of a graph. Define

$$EE_{even}(G) = \sum_{i=1}^{N} \cosh(\lambda_i), \qquad EE_{odd}(G) = \sum_{i=1}^{N} \sinh(\lambda_i).$$

- A graph is bipartite iff there is no closed walk of odd length.
- Given a graph G, the quantity $\frac{EE_{even}(G)}{EE(G)}$ tells us how close G is to being bipartite.

Communicability

The subgraph communicability between nodes i and j quantifies "how easily" information can be passed from i to j. It counts the number of walks that connect i and j, again with weights $\frac{1}{\text{length}(\mathcal{W})!}$:

$$C(i,j) := [e^A]_{ij}.$$

Analogously to the Estrada index, one can introduce

- the total communicability of a node:

  $$TC(i) = \sum_{j=1}^{N} C(i,j),$$

  which can be seen as another measure of centrality,
- and the total network communicability of a graph:

  $$TC(G) = \sum_{i=1}^{N} \sum_{j=1}^{N} C(i,j) = \mathbb{1}^T e^A \mathbb{1},$$

  which quantifies the global connectedness of G. See [Benzi and Klymko, Journal of Complex Networks, 2013].

Betweenness

How does the overall communicability change when a node is removed?

- Let $A - E(r)$ be the adjacency matrix for the network with node r removed.
- Then the betweenness of node r is

$$B(r) = \frac{1}{(N-1)^2 - (N-1)} \sum_{i \neq j, i \neq r, j \neq r} \frac{[e^A]_{i,j} - [e^{A-E(r)}]_{i,j}}{[e^A]_{i,j}},$$

  where $N \geq 3$.

Centrality, Estrada index, communicability and betweenness can be generalized by choosing other sets of weights (the exponential is replaced by other functions).

More generally:

Assume that $f(x)$ admits a Taylor-Maclaurin expansion

$$f(x) = f_0 + f_1 x + f_2 x^2 + f_3 x^3 + f_4 x^4 + \dots \text{ with } f_k \geq 0.$$

Then we can define f-based centrality and communicability by choosing f_k as a weight for the number of walks of length k:

$$\begin{aligned} c_f(i) &= f_0 + f_1 A + f_2 (A^2)_{ii} + f_3 (A^3)_{ii} + \dots \\ &= [f_0 I + f_1 A + f_2 A^2 + f_3 A^3 + f_4 A^4 + \dots]_{ii} = f(A)_{ii}, \end{aligned}$$

$$\begin{aligned} C_f(i,j) &= f_0 + f_1 A_{ij} + f_2 (A^2)_{ij} + f_3 (A^3)_{ij} + f_4 (A^4)_{ij} + \dots \\ &= [f_0 I + f_1 A + f_2 A^2 + f_3 A^3 + f_4 A^4 + \dots]_{ij} = f(A)_{ij}. \end{aligned}$$

A spectral point of view

Let $(\lambda_k, x_k)_{k=1,\ldots,N}$ be the eigenpairs of A.

- Since A is symmetric, we can write

$$A = \sum_{k=1}^{N} \lambda_k x_k x_k^T.$$

- Therefore we have, for f-centrality and f-communicability:

$$c(i) = \sum_{k=1}^{N} f(\lambda_k) x_k(i)^2,$$

$$C(i,j) = \sum_{k=1}^{N} f(\lambda_k) x_k(i) x_k(j).$$

Resolvent-based definitions

Define the resolvent function

$$r(x) := \left(1 - \frac{x}{N-1}\right)^{-1}.$$

Estrada and Higham have proposed the notions of

- resolvent centrality of node i:

$$c_r(i) := [r(A)]_{ii},$$

- resolvent Estrada index of G:

$$EE_r := \sum_{i=1}^{N} [r(A)]_{ii},$$

- resolvent communicability between nodes i and j:

$$C_r(i,j) := [r(A)]_{ij}.$$

Why $r(A)$?

Analogously, observe that for resolvent-based subgraph centrality and communicability we have:

$$\begin{aligned} c_r(i) &= \left[\left(I - \frac{A}{N-1}\right)^{-1}\right]_{ii} \\ &= \left[I + \frac{1}{N-1}A + \frac{1}{(N-1)^2}A^2 + \frac{1}{(N-1)^3}A^3 + \dots\right]_{ii} \\ &\approx \text{ weighted sum of closed walks based at node } i, \end{aligned}$$

$$\begin{aligned} C_r(i,j) &= \left[\left(I - \frac{A}{N-1}\right)^{-1}\right]_{ij} \\ &= \left[I + \frac{1}{N-1}A + \frac{1}{(N-1)^2}A^2 + \frac{1}{(N-1)^3}A^3 + \dots\right]_{ij} \\ &\approx \text{ weighted sum of walks joining nodes } i \text{ and } j, \end{aligned}$$

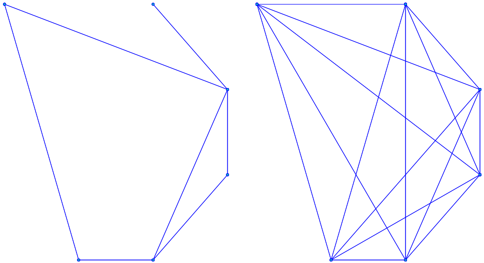
with weights $(N-1)^{-k}$.

Why $r(A)$?

Motivation for this choice of weights:

$$(N-1)^{1-k} \approx \frac{\#\text{ of walks of length } k \text{ in } G}{\#\text{ of walks of length } k \text{ in } K_N},$$

where K_N is the complete graph with N nodes.



What do we need to compute?

Some comments on the computation of matrix functions for complex networks:

- adjacency matrices typically have large size (but are sparse),
- single entries of $f(A)$ may be required (no need to compute the whole matrix),
- for some applications, the product of $f(A)$ times a vector is required,
- in many cases, results need not be very accurate (e.g., for rankings).

As a consequence:

- bounds on entries of $f(A)$ can be useful,
- approximate computations are ok.

Quadrature-based bounds

- Gauss-type quadrature rules can be used to obtain bounds on the entries of certain functions of symmetric matrices (see [Golub and Meurant 1993], [Benzi and Golub 1999] and the book [Golub and Meurant 2009]).
- Upper and lower bounds are available for the bilinear form

$$u^T f(A) v$$

  with $u, v \in \mathbb{R}^N$, $A \in \mathbb{R}^{N \times N}$ symmetric and $f(x)$ strictly completely monotonic on an interval containing the spectrum of A.
- If N is very large and f-centrality values are needed for several nodes, techniques based on low rank approximation of A can be useful; see [Fenu, Martin, Reichel, Rodriguez 2013].
- Another approach explored in [Fenu, Martin, Reichel, Rodriguez 2013] consists in combining Gauss and anti-Gauss quadrature rules.

Krylov methods

- Krylov subspace methods are used to solve many large-scale linear algebra problems, such as linear systems, eigenvalue problems, matrix equations... and, last but not least, to compute $f(A)b$.
- Recall that the k-th Krylov subspace associated with $A \in \mathbb{C}^{n\times n}$ and $b \in \mathbb{C}^n$ is defined as

$$\mathcal{K}_k(A, b) = \text{span}\{b, Ab, A^2b, \ldots, A^{k-1}b\}.$$

- Also recall that $f(A)b \in \mathcal{K}_d(A, b)$, where $d = \deg \psi_{A,b}$ and $\psi_{A,b}$ is the monic polynomial of lowest degree such that $\psi_{A,b}(A)b = 0$. Equivalently, d is the smallest integer such that $\mathcal{K}_d(A, b) = \mathcal{K}_{d+1}(A, b)$.
- As a key example of Krylov method, we outline the Arnoldi process.

Krylov methods: Arnoldi process

Let $A \in \mathbb{C}^{n\times n}$. We would like to compute its Hessenberg reduction, i.e., the factorization

$$A = QHQ^*,$$

where $Q \in \mathbb{C}^{n\times n}$ is unitary and $H \in \mathbb{C}^{n\times n}$ is upper Hessenberg.

Let $q_1, q_2, \ldots, q_n$ be the columns of Q. Equate columns in $AQ = QH$:

$$Aq_k = \sum_{i=1}^{k+1} h_{ik} q_i, \qquad k = 1, \ldots, n-1.$$

$$h_{k+1,k} q_{k+1} = Aq_k - \sum_{i=1}^{k} h_{ik} q_i =: r_k,$$

where $h_{ik} = q_i^* A q_k$ for $i = 1, \ldots, k$. If $r_k \neq 0$ we obtain

$$q_{k+1} = r_k / h_{k+1,k}$$

with $h_{k+1,k} = \|r_k\|_2$.

Krylov methods: Arnoldi process

Remark

From $Aq_k = \sum_{i=1}^{k+1} h_{ik} q_i$ it follows that

$$\text{span}\{q_1, \ldots, q_k\} = \text{span}\{q_1, Aq_1, \ldots, A^{k-1}q_1\} = \mathcal{K}_k(A, q_1),$$

that is, $q_1, \ldots, q_k$ form an orthonormal basis of $\mathcal{K}_k(A, q_1)$.

The Arnoldi process produces the factorization

$$AQ_k = Q_k H_k + h_{k+1,k} q_{k+1} e_k^T$$

where $Q_k = [q_1, \ldots, q_k]$ and $H_k = (h_{ij})$ is $k \times k$ upper Hessenberg.
Note that

$$Q_k^* A Q_k = H_k,$$

therefore H_k is the orthogonal projection of A onto $\mathcal{K}_k(A, q_1)$.

Krylov methods: Arnoldi process

Input: matrix A, normalized vector q_1.
Output: matrices Q, H of sizes $n \times d$ and $d \times d$, respectively.

for $k = 1 : n$
    $z = Aq_k$
    for $i = 1 : k$
        $h_{ik} = q_i^* z$
        $z = z - h_{ik} q_i$
    end
    $h_{k+1,k} = \|z\|_2$
    if $h_{k+1,k} = 0$, $m = k$, quit, end
    $q_{k+1} = z / h_{k+1,k}$
end

Krylov methods: Arnoldi process

- The process terminates in at most d steps, where $d = \deg \psi_{A,q_1}$.
- In the Arnoldi process, A does not have to be stored explicitly. We only need to be able to compute Aq_k. (Good for large sparse matrices!)
- The Arnoldi process typically suffers from loss of orthogonality. Reorthogonalization should be applied.
- Hermitian case: Lanczos algorithm.

Arnoldi approximation of $f(A)b$

How can we use the Arnoldi process to approximate $f(A)b$?
Take $q_1 = b/\|b\|_2$ and apply k steps of Arnoldi:

$$f(A) \approx f_k := \|b\|_2 Q_k f(H_k) e_1 = Q_k f(H_k) Q_k^* b.$$

The approximation is exact if $k = \deg \psi_{A,b}$. But in practice we may stop earlier.

- Since k is small, $f(H_k)$ can be computed explicitly.
- Caveat: in general $f(H_k)$ might not be defined (but sufficient conditions are available to ensure that it is).
- It is common practice to restart the Arnoldi process after a fixed number of steps to reduce storage.
- Convergence, error bounds and stopping criterion need to be discussed.

Some remarks on the directed case

Suppose now that G is a directed graph.

- It is still true that $[A^k]_{i,j}$ counts the number of (directed) walks from node i to node j. One can again define subgraph centrality, Estrada index etc.
- In particular: the subgraph centrality $c(i) = [e^A]_{i,i}$ can be seen as a measure of returnability [Estrada and Hatano 2009]. Directed walks that start and end at node i tell us if information will "come back" to the node.
- However, subgraph centrality may not always be a good choice in the directed case (see next example).

An example: the path graph

Consider the directed path graph $G_p = (V, E)$:

$$V = (v_1, v_2, \dots, v_N), \qquad E = ((1,2), (2,3), \dots, (N-1, N)).$$

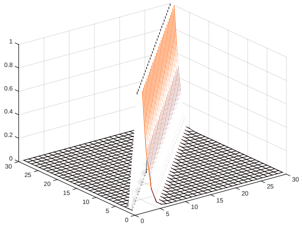
Its adjacency matrix is

$$A_p = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & 0 & 1 \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix}$$

In this case, subgraph centralities are all equal to 1 (although the first and last node are certainly special).

Part of the problem is that there is no closed walk...

A city plot of $\exp(A)$



- Diagonal entries $= 1$,
- zero lower triangular part,
- fast decay of upper triangular part.

Hubs and authorities

In a directed network there are two distinct centrality roles: "broadcasters" and "receivers". Centrality measures need to address this point.

Of course, the in-degree (number of incoming edges) and the out-degree (number of outgoing edges) provide a first, rough measure of in- and out-centrality. But more refined approaches are available.

In a directed graph/network (V, E)

- Hubs point to "important" nodes
- Authorities are these important nodes.

Good hubs point to many good authorities and good authorities are pointed to by many good hubs.

The HITS algorithm

HITS=Hypertext Induced Topics Search [Kleinberg 1999].
Each node i is assigned

- an authority weight x_i,
- and a hub weight y_i,

which are updated through successive iterations until convergence.

It turn out that HITS is essentially a power method that computes the dominant eigenvalue and eigenvector of AA^T and A^TA.

- A^TA is the authority matrix,
- AA^T is the hub matrix,
- the Perron-Frobenius theorem applies to both.

Bipartization of directed networks

For a directed graph G with adjacency matrix A, consider the symmetric matrix

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

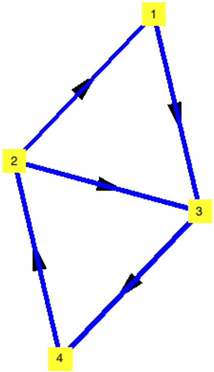
which is associated with a bipartite graph $\widetilde{G} = (\widetilde{V}, \widetilde{E})$:

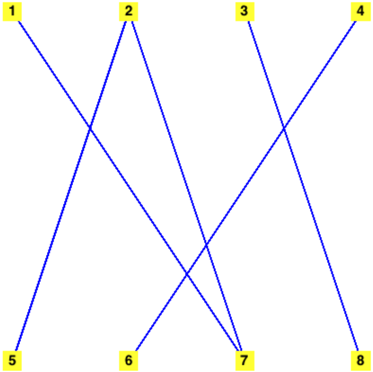
- $\widetilde{V}$ contains two copies V, V' of the vertex set of G,
- $\widetilde{E}$ contains the edges (i, j') such that $(i, j) \in E$.

It can be shown that

$$e^{\mathcal{A}} = \begin{bmatrix} \cosh\left(\sqrt{AA^T}\right) & A\left(\sqrt{A^TA}\right)^{\dagger} \sinh\left(\sqrt{A^TA}\right) \\ \sinh\left(\sqrt{A^TA}\right)\left(\sqrt{A^TA}\right)^{\dagger} A^T & \cosh\left(\sqrt{A^TA}\right) \end{bmatrix}.$$

Bipartization of a directed graph





Matrix powers and alternating walks

- Alternating walk on a directed graph (starting with out-edge):

$$v_1 \longrightarrow v_2 \longleftarrow v_3 \longrightarrow v_4 \longleftarrow v_5 \longrightarrow \dots$$

- The entry

$$[AA^TA\dots]_{ij}$$

  (k matrices being multiplied) counts the number of even alternating walks of length k from node i to node j, starting from an out-edge;

- the entry

$$[A^TAA^T\dots]_{ij}$$

  (k matrices being multiplied) counts the number of even alternating walks of length k from node i to node j, starting from an in-edge;

- so $(AA^T)^k$ and $(A^TA)^k$ count alternating walks of length $2k$.

Hub and authority centrality

If node i is a good hub, there should be many even closed walks based at i, starting with an out-edge.

Let $A = U\Sigma V^T$ be the SVD of A. We have

$$\begin{aligned}
& I + \frac{AA^T}{2!} + \frac{(AA^T)^2}{4!} + \ldots + \frac{(AA^T)^k}{(2k)!} + \ldots = \\
&= U\left(I + \frac{\Sigma^2}{2!} + \frac{\Sigma^4}{4!} + \ldots + \frac{\Sigma^{2k}}{(2k)!} + \ldots\right)U^T = \\
&= U\cosh(\Sigma)U^T = \cosh\left(\sqrt{AA^T}\right).
\end{aligned}$$

- The hub centrality of node i can be quantified by

$$[e^{\mathcal{A}}]_{ii} = [\cosh\left(\sqrt{AA^T}\right)]_{ii},$$

- and the authority centrality of node i can be quantified by

$$[e^{\mathcal{A}}]_{N+i,N+i} = [\cosh\left(\sqrt{A^TA}\right)]_{ii}.$$

Hub and authority communicability

- The hub communicability between nodes i and j is

$$[e^{\mathcal{A}}]_{ij} = [\cosh\left(\sqrt{AA^T}\right)]_{ij},$$

- the authority communicability between nodes i and j is

$$[e^{\mathcal{A}}]_{N+i,N+j} = [\cosh\left(\sqrt{A^TA}\right)]_{ij},$$

- and the off-diagonal blocks in $e^{\mathcal{A}}$ give hub-authority and authority-hub communicabilities.

See [Benzi, Estrada and Klymko, *Ranking Hubs and Authorities Using Matrix Functions*, LAA 2013].

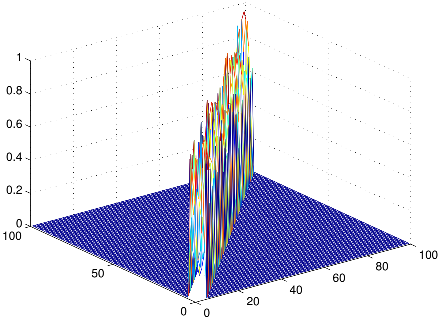
Decay of matrix functions

Functions of sparse matrices typically exhibit a decay behavior.

For instance:

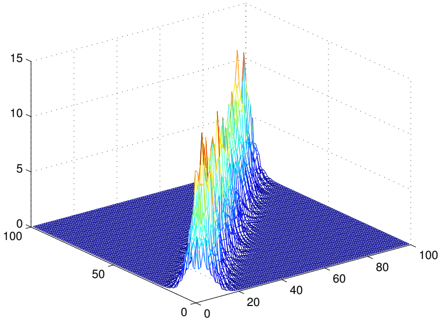
- Let $A \in \mathbb{R}^{N \times N}$ be a banded matrix and compute $B = e^A$.
- B is a full matrix. However, its entries decrease (in absolute value) away from the main diagonal (off-diagonal decay).
- Therefore, B can be approximated by a banded matrix.
- A similar behavior is observed for other matrix functions (e.g., matrix inverse).

An example



A banded matrix *A*...

An example



...and its exponential B

Decay bounds

Given

- a banded/sparse matrix $A \in \mathbb{R}^{N \times N}$,
- a function $f(x)$ such that $B = f(A)$ is well defined,

we would like to formulate *a priori* bounds on the off-diagonal decay behavior of B. Ideally

$$|B_{ij}| \leq K e^{-\alpha|i-j|},$$

with K, α positive constants (independent of the size of A).

For networks: such bounds may, for instance, help us identify nodes with low communicability.

Some history

- Demko, Moss, Smith (1984): bounds for inverse of banded spd matrices; see also Jaffard (1991), Blatov (1996) et al.
- Benzi, Golub (1999): bounds for functions of banded, symmetric matrices
- Iserles (2000): bounds for the exponential of banded matrices
- Del Buono, Lopez, Peluso (2005): bounds for functions of skew-symmetric matrices
- Benzi, Razouk (2007): extension to non-normal matrices
- Benzi, B., Razouk (2013): review on applications to electronic structure computations
- Benzi, B. (2014): extension to C^*-algebras
- Benzi, Simoncini (2015): approach based on Laplace-Stieltjes transforms.

Application of decay bounds

Theorem (Benzi and Golub, Benzi and Razouk)

Let A be a real symmetric m-banded / sparse matrix and let $f(x)$ be a smooth function on an interval containing the spectrum of A. Then we can compute constants $C > 0$ and $0 < \lambda < 1$ such that

$$|[f(A)]_{ij}| \leq C\lambda^{\frac{|i-j|}{m}}$$

and

$$|[f(A)]_{ij}| \leq C\lambda^{d(i,j)},$$

where $d(i,j)$ is the graph distance between nodes i and j.

Application of decay bounds

We can use decay results with A adjacency matrix, $f(x) = e^x$ or $r(x)$ or another suitable function, to give bounds on graph communicability. For instance, in the banded case:

- choose a threshold $\eta > 0$ (i.e., values of communicability smaller than η are considered negligible),
- find the smallest integer $\hat{m}$ such that

$$|[f(A)]_{ij}| \leq C\lambda^{\frac{|i-j|}{\hat{m}}} < \eta,$$

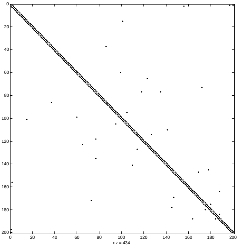
- then, for the purpose of computing communicability, we can truncate $f(A)$ to bandwidth $\hat{m}$ and ignore the entries of $f(A)$ outside the band.

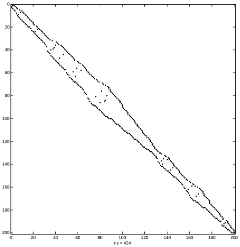
If the bandwidth of A (possibly after reordering) is independent of N, then $\hat{m}$ is also independent of N.

In this case, the number of node pairs that have non-negligible communicability grows only linearly with N.

Decay bounds and communicability

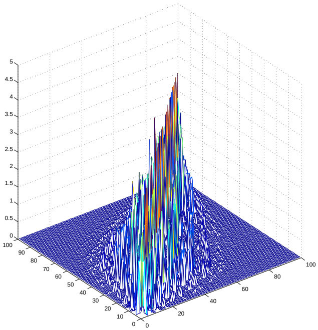
Example: take A a 200×200 small world matrix with $k = 1$ and $p = 0.1$, normalized so that $\|A\|_2 = 1$, and reorder it via reverse Cuthill-McKee.





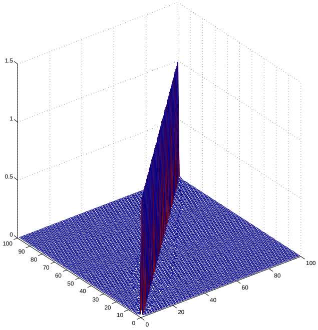
Decay bounds and communicability

This is what e^A looks like:



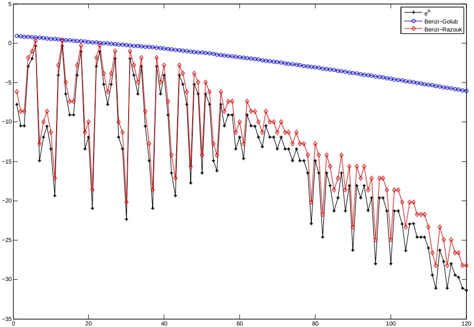
Decay bounds and communicability

And this is $r(A)$:



Decay bounds and communicability

The decay bounds give:



How to compute decay bounds?

- Let A be a Hermitian $n \times n$ matrix of bandwidth m, $f(x)$ a sufficiently regular function, $p_k(x)$ a polynomial of degree k.
- Observe that $p_k(A)$ is a banded matrix of bandwidth km.
- Then:

$$
\begin{aligned}
|[f(A)]_{ij}| = |[f(A) - p_k(A)]_{ij}| &\le \|f(A) - p_k(A)\|_2 \\
&\le \max_{x \in \sigma(A)} |f(x) - p_k(x)| \quad \text{for} \quad |i - j| > km.
\end{aligned}
$$

- Therefore we can use polynomial approximation techniques to develop decay bounds for $f(A)$.

Polynomial approximation

We need (asymptotic) upper bounds on the *k-th best approximation error*

$$E_k(f) = \inf\{\max_{-1\leq x\leq 1} |f(x) - p(x)| : p \in P_k\},$$

where $P_k \subset \mathbb{R}[x]$ is the set of polynomials with degree less or equal to k.

Bernstein's theorem

Denote by $\mathcal{E}_\chi$ the ellipse in the complex plane with foci in ± 1 and sum of semi-axes $\chi > 1$.

Bernstein's theorem

Denote by $\mathcal{E}_\chi$ the ellipse in the complex plane with foci in ± 1 and sum of semi-axes $\chi > 1$.

Theorem (Bernstein)

Let the function f be analytic in the interior of the ellipse $\mathcal{E}_\chi$ and continuous on $\mathcal{E}_\chi$. Assume that $f(z)$ is real for real z. Then

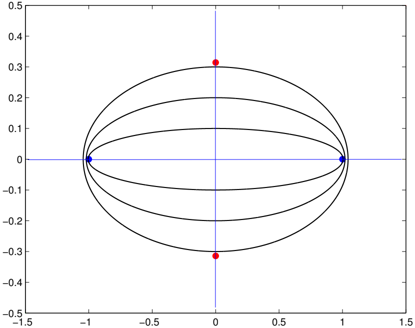
$$E_k(f) \leq \frac{2M(\chi)}{\chi^k(\chi - 1)},$$

where $M(\chi) = \max_{z \in \mathcal{E}_\chi} |f(z)|$.

So we can choose $C = \frac{2M(\chi)}{\chi - 1}$ and $\lambda = \frac{1}{\chi}$ in the decay bounds.

Choice of a Bernstein ellipse

If f has poles in $\mathbb{C}$, for instance at $\pm\frac{\pi i}{\beta}$ as below, it is analytic in $\mathcal{E}_\chi$ as long as the poles do not belong to the interior of $\mathcal{E}_\chi$.



A generalization

The previous results can be extended to the case where the matrix A has a more general sparsity pattern.

- Define the graph G associated with A, such that
  - G_n has N nodes,
  - nodes i and j are connected by and edge iff $A_{ij} \neq 0$.

  This is just the graph for which A is an adjacency matrix!
- The distance $d(i,j)$ in G is the number of edges in the shortest path connecting nodes i and j (∞ if there is no such path).

The decay bound then becomes

$$|[f(A)]_{ij}| \leq C\,\lambda^{d(i,j)}.$$

Cartesian products

Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs, with adjacency matrices A_1 and A_2.

The Cartesian product of G_1 and G_2 is a graph $\mathcal{G}$ such that:

- the vertex set of $\mathcal{G}$ is the Cartesian product $V_1 \times V_2$,
- there is an edge between (u_1, u_2) and (v_1, v_2) if
  - either $u_1 = v_1$ and $(u_2, v_2) \in E_2$,
  - or $u_2 = v_2$ and $(u_1, v_1) \in E_1$.

The adjacency matrix of $\mathcal{G}$ is

$$\mathcal{A} = A_1 \oplus A_2 = A_1 \otimes I + I \otimes A_2$$

(Kronecker sum).

Entries of $f(\mathcal{A})$ can be efficiently approximated via Krylov methods, especially for particular instances of $f(x)$ [Benzi and Simoncini 2015].

Integral bounds

Benzi and Simoncini propose new bounds for the entries of $f(A)$, where

- A is a Hermitian banded (or sparse) matrix,
- f belongs to classes of functions that can be represented as integral transforms of measures (e.g., exponential and resolvent).

Laplace-Stieltjes functions

Let f be strictly completely monotonic in $(0, +\infty)$, i.e.,

$$(-1)^k f^{(k)}(x) > 0 \quad \text{for all} \quad 0 < x < +\infty, \quad k \in \mathbb{N}.$$

Then it can be represented as

$$f(x) = \int_0^{+\infty} e^{-\tau x} d\alpha(\tau)$$

For instance, for $x > 0$:

- $\frac{1}{x} = \int_0^{+\infty} e^{-\tau x} d\alpha_1(\tau)$, with $\alpha_1(\tau) = \tau$, $\tau \geq 0$,
- $e^{-x} = \int_0^{+\infty} e^{-\tau x} d\alpha_2(\tau)$, with $\alpha_2(\tau) = 0$ for $0 \leq \tau < 1$ and $\alpha_2(\tau) = 1$ for $\tau > 1$,
- $\frac{1-e^{-x}}{x} = \int_0^{+\infty} e^{-\tau x} d\alpha_3(\tau)$, with $\alpha_3(\tau) = \tau$ for $0 \leq \tau < 1$ and $\alpha_3(\tau) = 1$ for $\tau \geq 1$.

Cauchy-Stieltjes functions

Cauchy-Stieltjes functions can be written as

$$f(z) = \int_{-\infty}^{0} \frac{d\gamma}{z - \omega}, \qquad z \in \mathbb{C} \setminus (-\infty, 0],$$

with γ a real measure.

This class includes

$$z^{-\frac{1}{2}}, \qquad \frac{e^{-t\sqrt{z}} - 1}{z}, \qquad \frac{\log(1+z)}{z}$$

Integral bounds

Let $f(x)$ be a Cauchy-Stieltjes function.

Then one can use exponential decay bounds (as seen above) in the integral definition of $f(x)$ and obtain for any banded Hermitian positive definite matrix M

$$|f(M)_{kt}| \leq \int_{-\infty}^{0} C(\omega) q(\omega)^{\frac{|k-t|}{\beta}} |d\gamma(\omega)|.$$

A similar approach can be used for Laplace-Stieltjes functions, together with bounds on $\exp(-\tau A)$...

Integral bounds

The following result by Benzi and Simoncini comes from a theorem by Hochbruck and Lubich on the error of Arnoldi approximations of exponential integrators.

Theorem

Let M be a Hermitian, β-banded positive semidefinite matrix with eigenvalues in $[0, 4\rho]$. For $k \neq t$, let $\xi = |k - t|/\beta$. Then:

1. *for $\rho\tau \geq 1$ and $\sqrt{4\rho\tau} \leq \xi \leq 2\rho\tau$,*

$$|[\exp(-\tau M)]_{kt}| \leq 10 \exp\left(-\frac{1}{5\rho\tau}\xi^2\right),$$

2. *for $\xi \geq 2\rho\tau$,*

$$|[\exp(-\tau M)]_{kt}| \leq 10 \frac{\exp(-\rho\tau)}{\rho\tau} \left(\frac{e\rho\tau}{\xi}\right)^{\xi}.$$