Sparse Linear Algebra Issues Arising in the Analysis of Complex Networks

Michele Benzi

Department of Mathematics and Computer Science
Emory University
Atlanta, GA 30322, USA

Acknowledgments

- Christine Klymko (Emory)
- Ernesto Estrada (Strathclyde, UK)
- Support: NSF DMS-1115692
- Papers and preprints available on web page

Outline

- Complex networks: motivation and background
- Quantitative analysis of networks
- Subgraph centrality and communicability
- Algorithms and numerical experiments
- Some open challenges

Complex networks: motivation and background

- Complex networks provide models for physical, biological, engineered or social systems (e.g., molecular structure, gene and protein interaction, food webs, transportation networks, power grids, social networks,...).
- Graph analysis provides quantitative tools for the study of complex networks.
- Techniques from spectral graph theory, linear and multilinear algebra, probability, approximation theory, etc. play a major role.

Network science today is a vast multidisciplinary field. Important early work was done by social scientists: sociologists, anthropologists, experimental psychologists, economists and even bibliometrists. More recently, physicists, computer scientists and applied mathematicians have made major contributions.

Complex networks: motivation and background

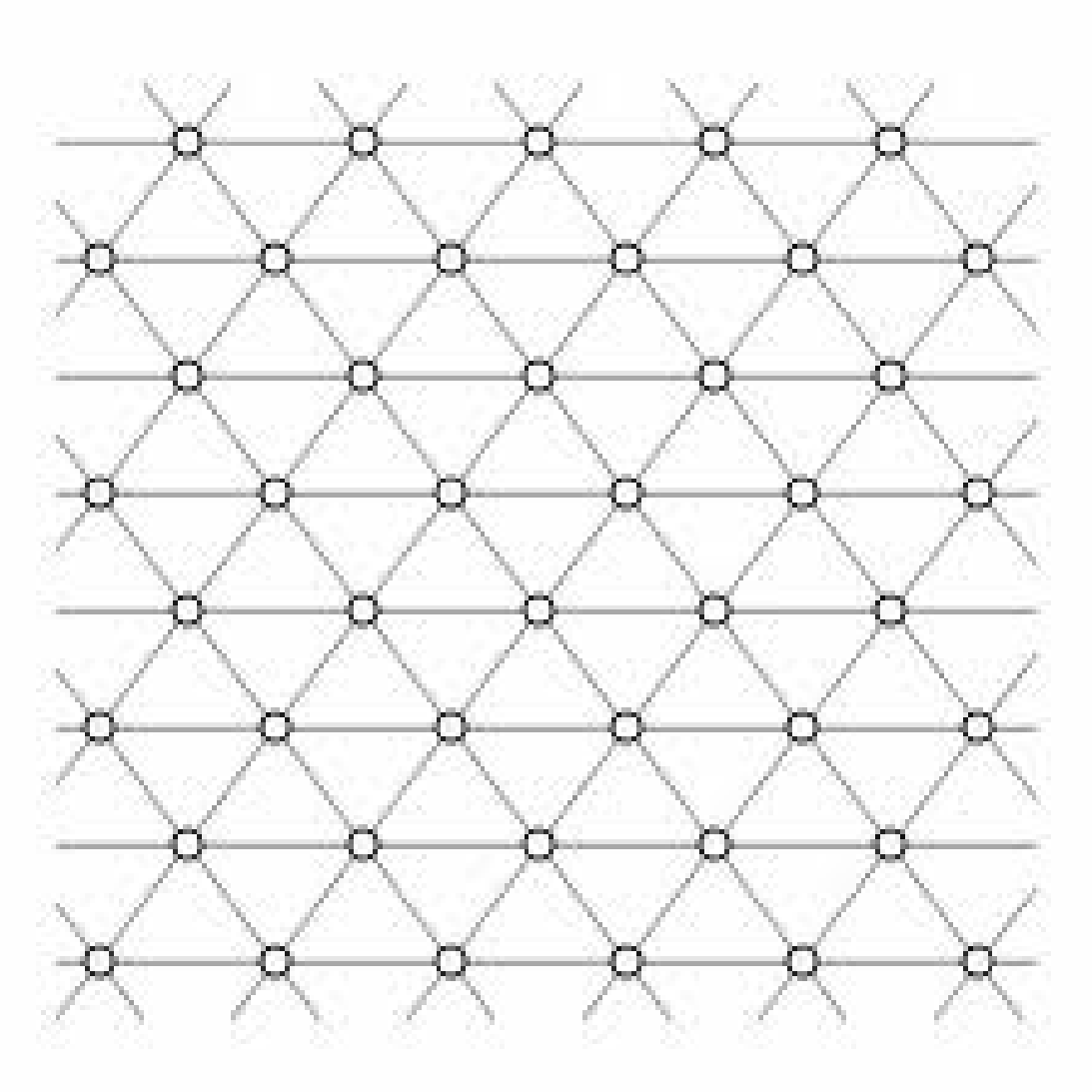
But what exactly is a complex network?

Unfortunately, no precise definition exists.

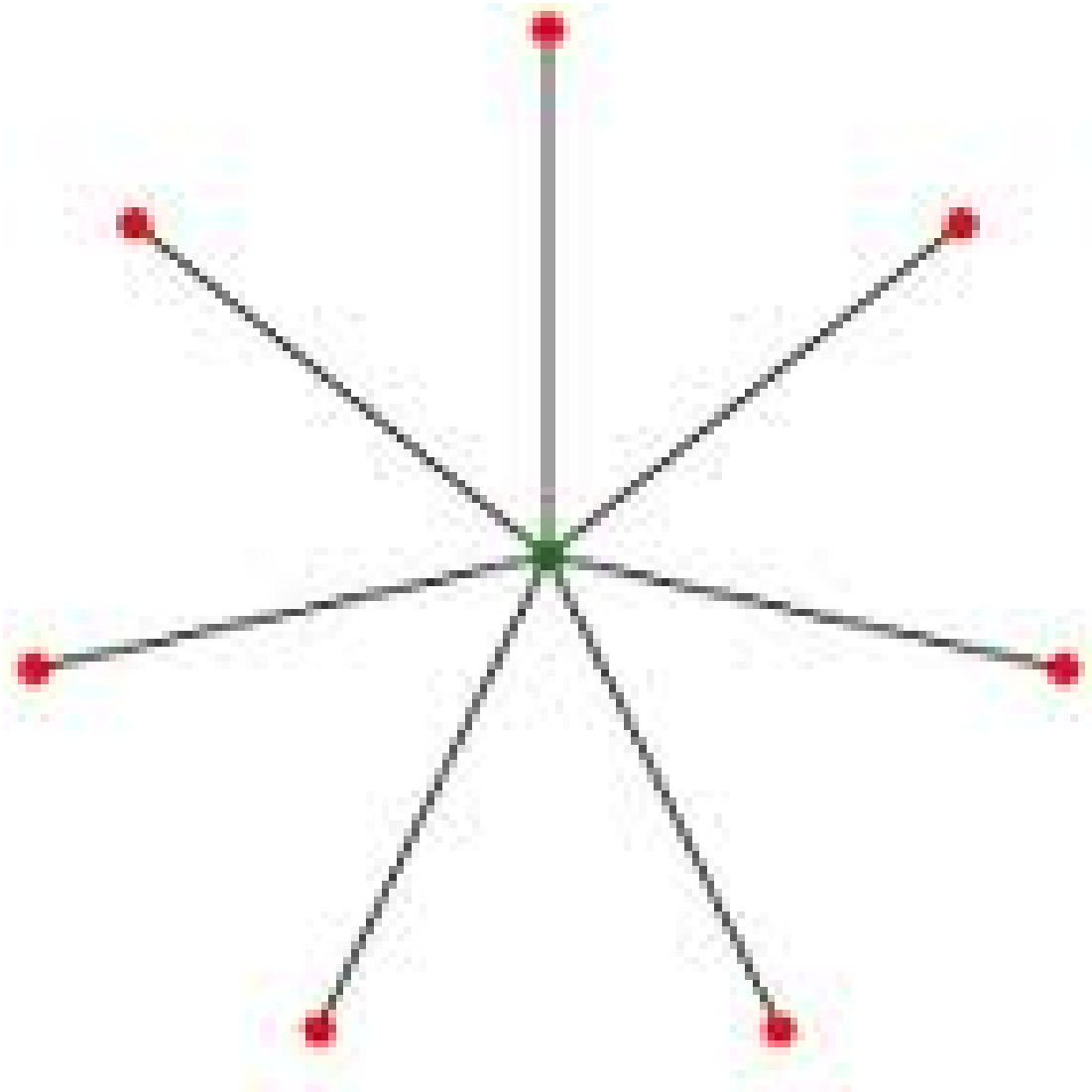
It is easier to say which graphs are not complex networks. Regular lattices are not considered complex networks, and neither are completely random graphs such as the Gilbert or Erdös–Rényi models.

Random graphs are, however, considered by many to be useful as toy models for complex networks.

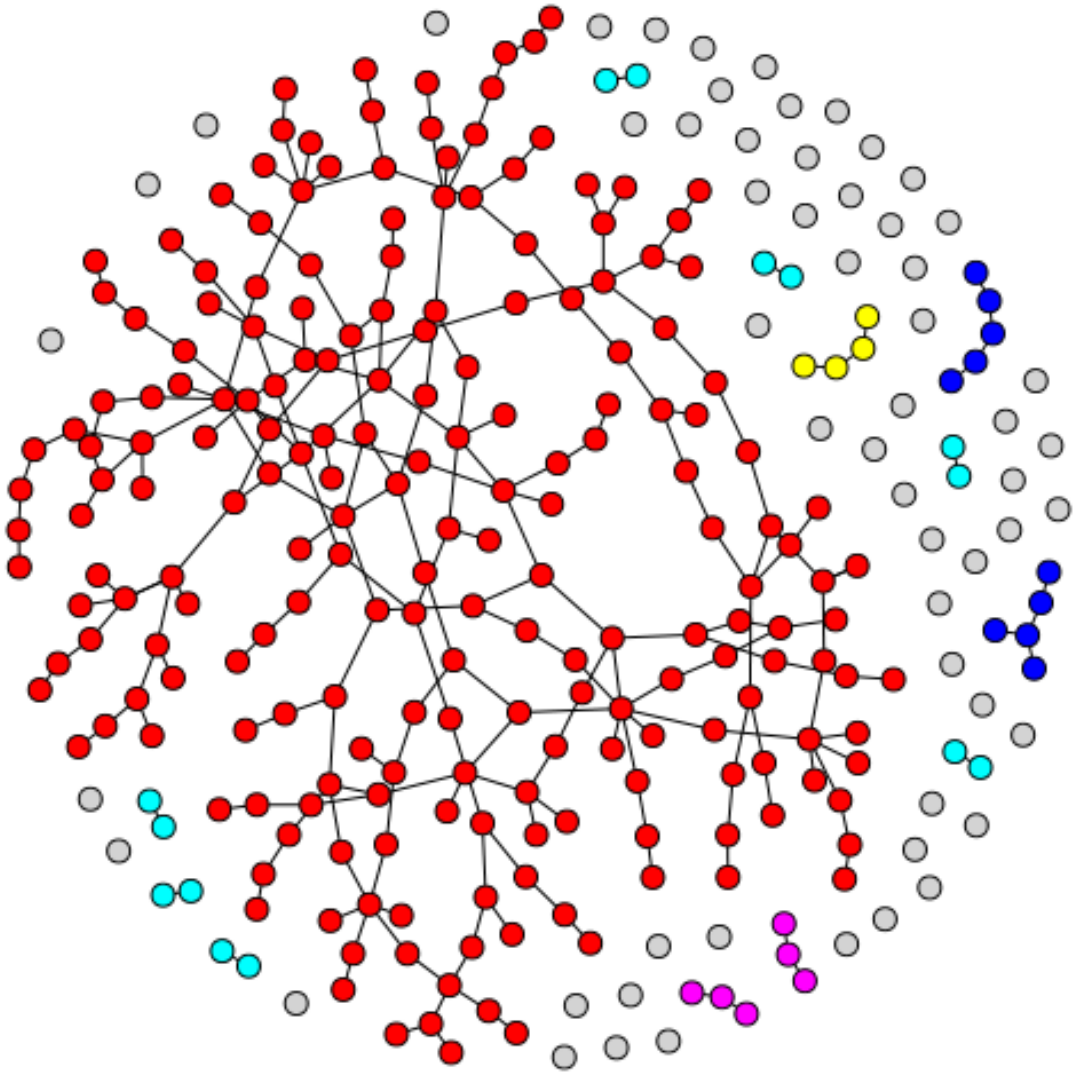
Regular lattice: not a complex network!



Star graph: also not a complex network!



Erdös–Rényi graph: also not a complex network!



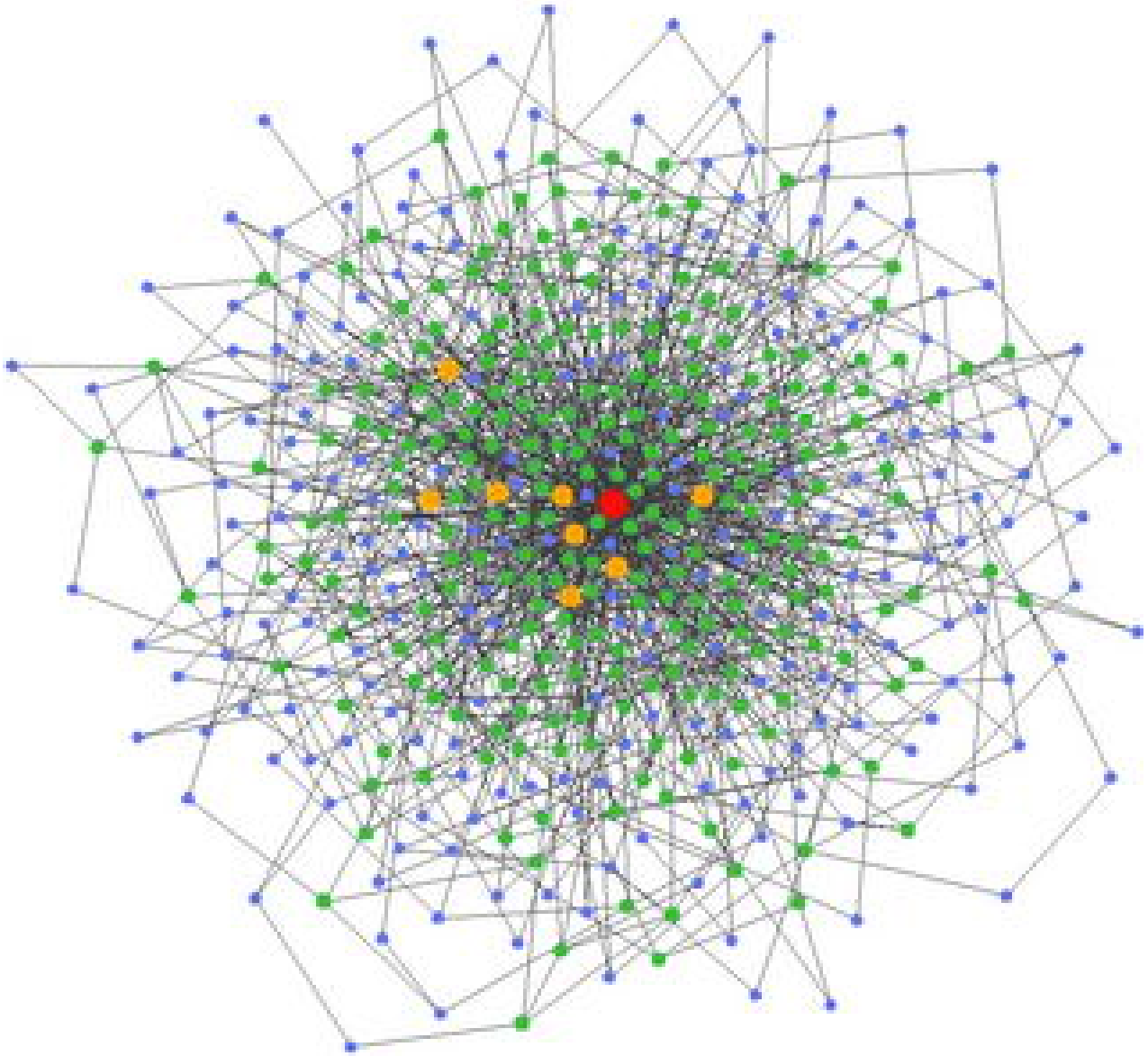
Complex networks: motivation and background

Some of the attributes typical of many real-world complex networks are:

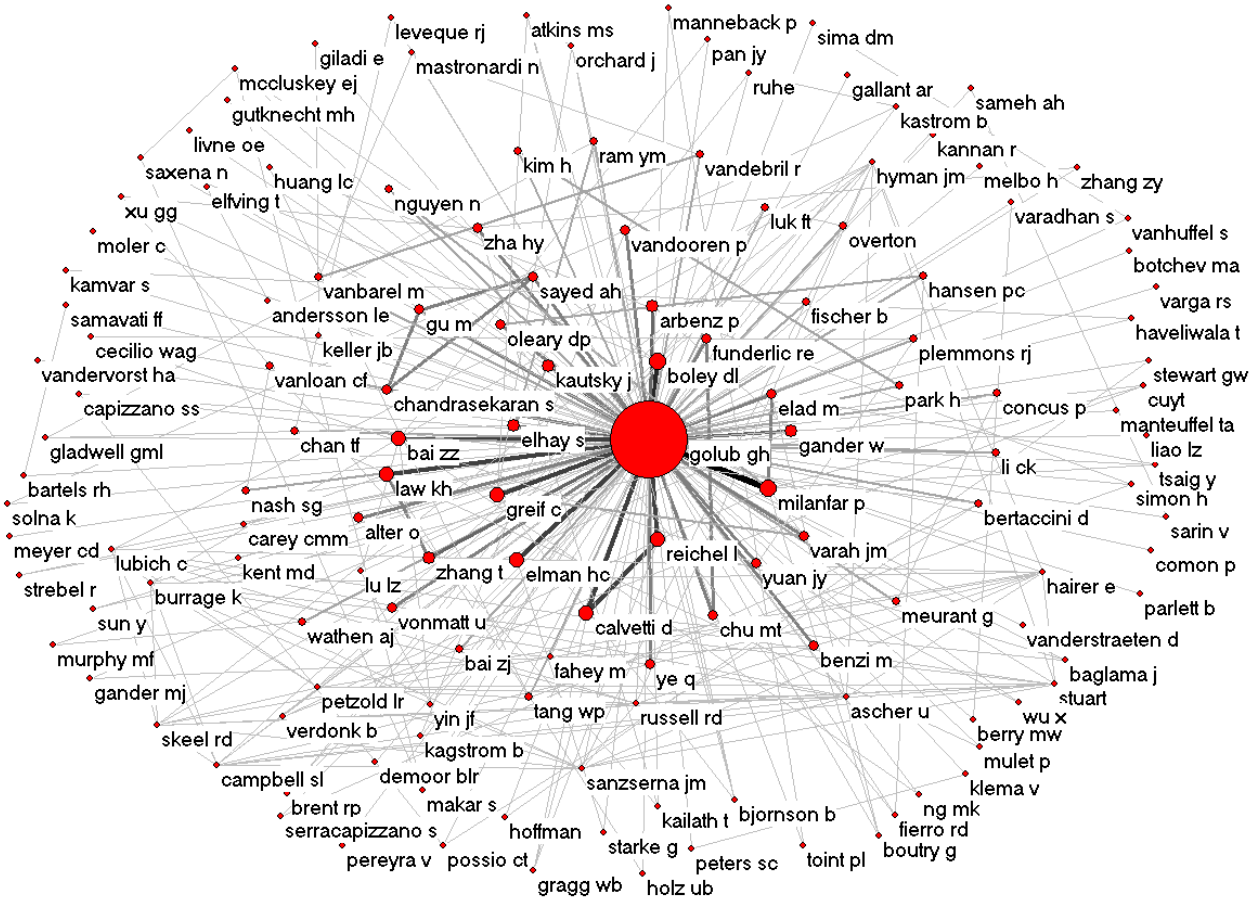
- Scale-free (power law degree distribution)
- Small-world (small graph diameter)
- Highly clustered (many triangles, hubs...)
- Hierarchical
- Rich in 'motifs'
- Self-similar

Caveat: there are important examples of real-world complex networks lacking one or more of these attributes.

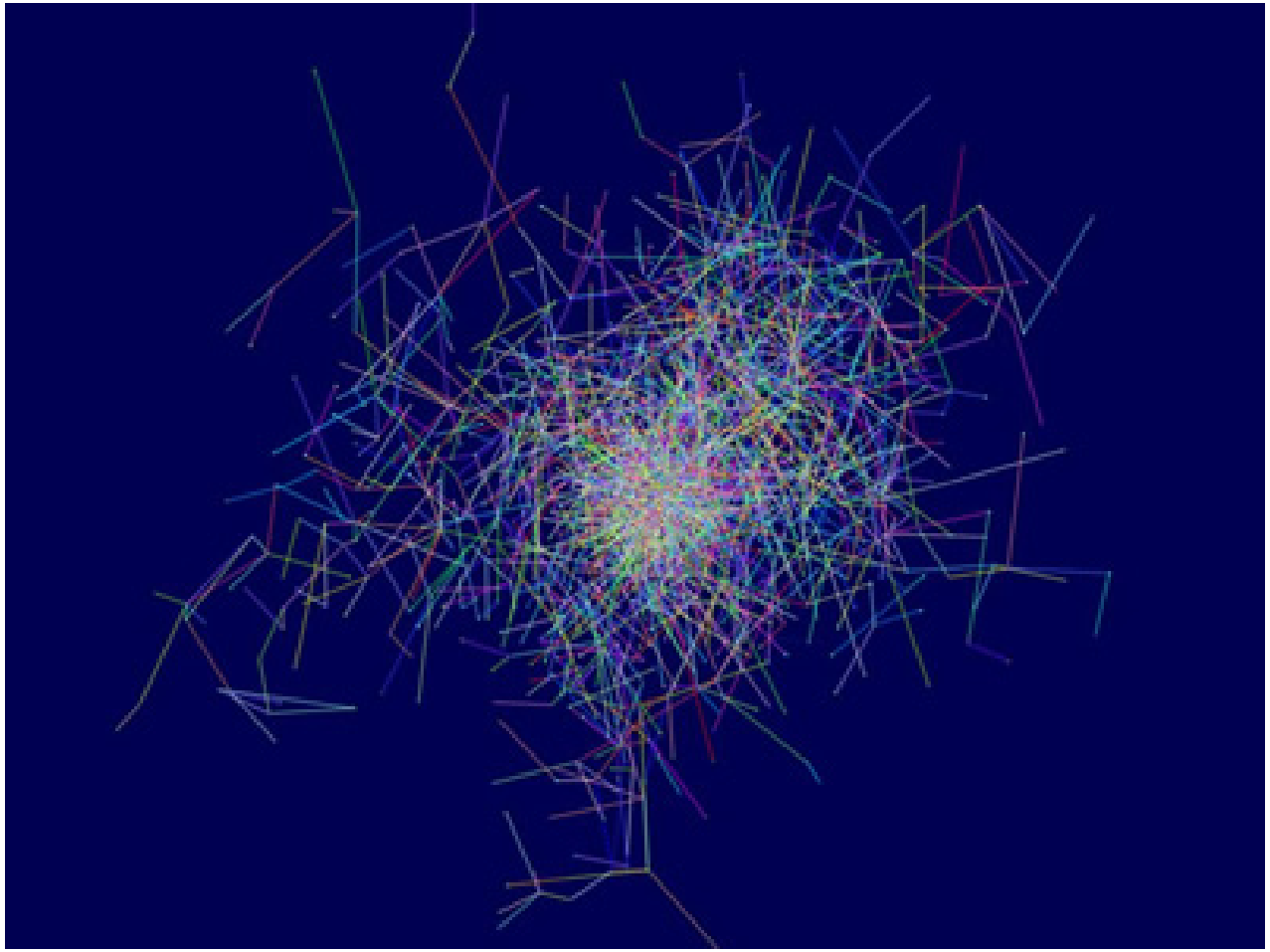
Example of complex network: B–A model



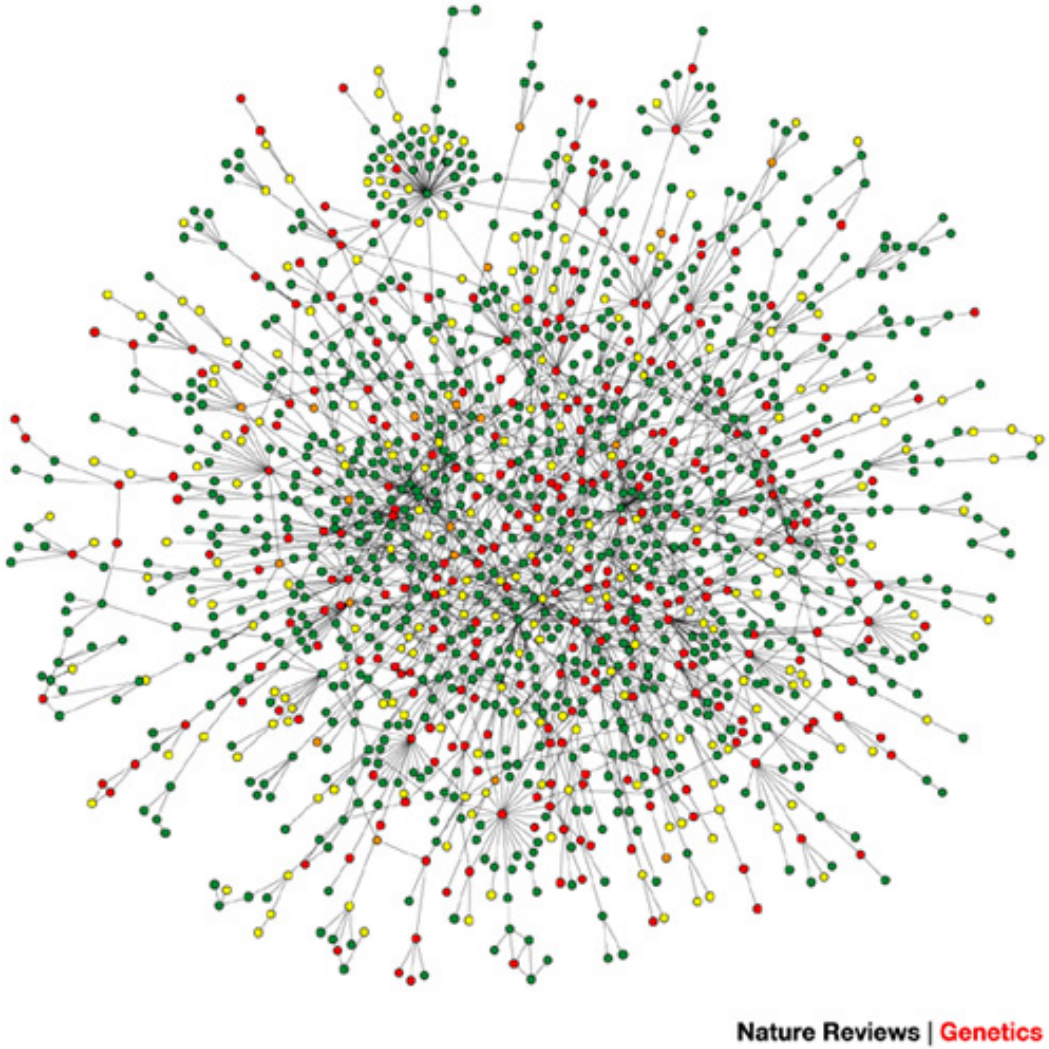
Example of complex network: Golub collaboration graph



Example of complex network: Erdös collaboration graph

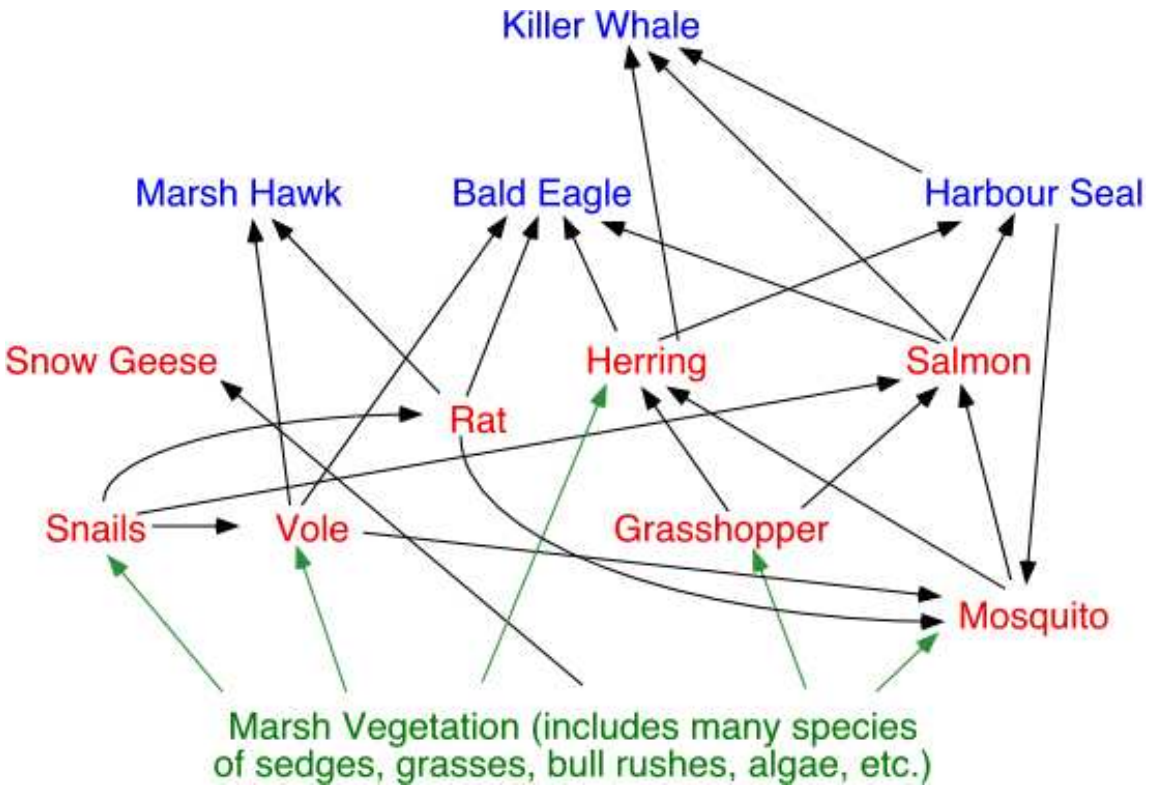


Example of complex network: PPI network of *Saccharomyces cerevisiae* (beer yeast)



Example of complex network: the Internet

Example of (directed) complex network: a food web

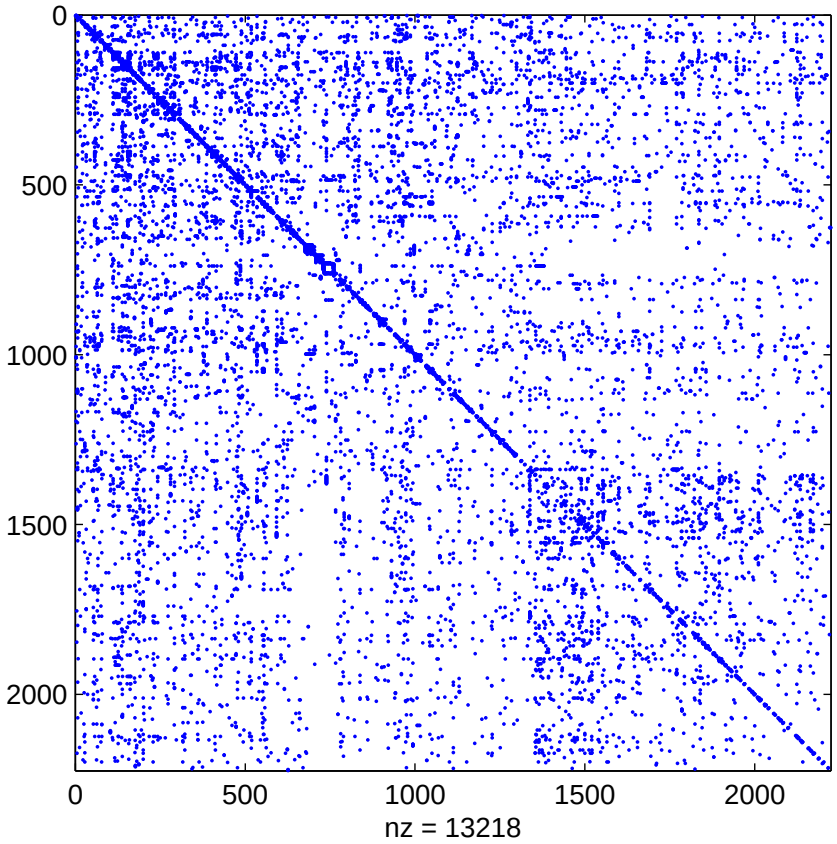


Complex networks: motivation and background

Notation:

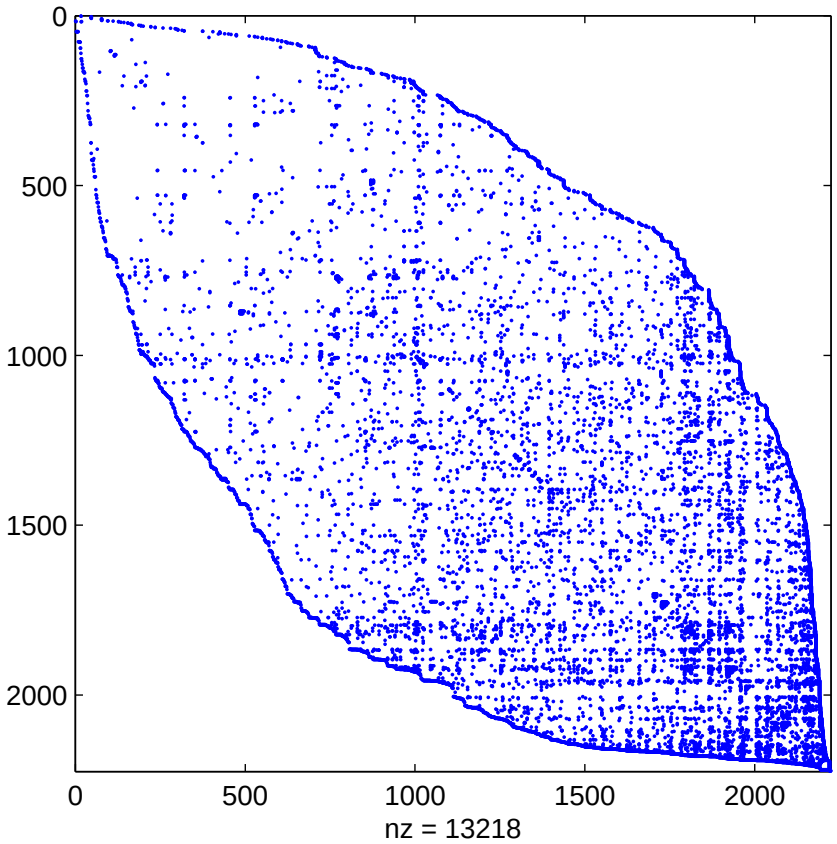
- $G = (V, E)$ is a connected simple graph with $N = |V|$ nodes and $m = |E|$ edges (i.e., G is unweighted, undirected and has no loops)
- $A \in \mathbb{R}^{N \times N}$ is the associated adjacency matrix:
  - $A_{ij} = 1$ if nodes i and j are adjacent, $A_{ij} = 0$ otherwise
  - A is symmetric
  - $A_{ii} = 0$, $1 \leq i \leq N$
  - $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ eigenvalues of A
- For a node i, define its degree $d_i := \sum_{k=1}^{N} A_{ik}$ (number of immediate neighbors in G)
- Graph Laplacian: $L := D - A$ where $D = \mathrm{diag}\,(d_1, \ldots, d_N)$

Example: PPI network of *Saccharomyces cerevisiae*



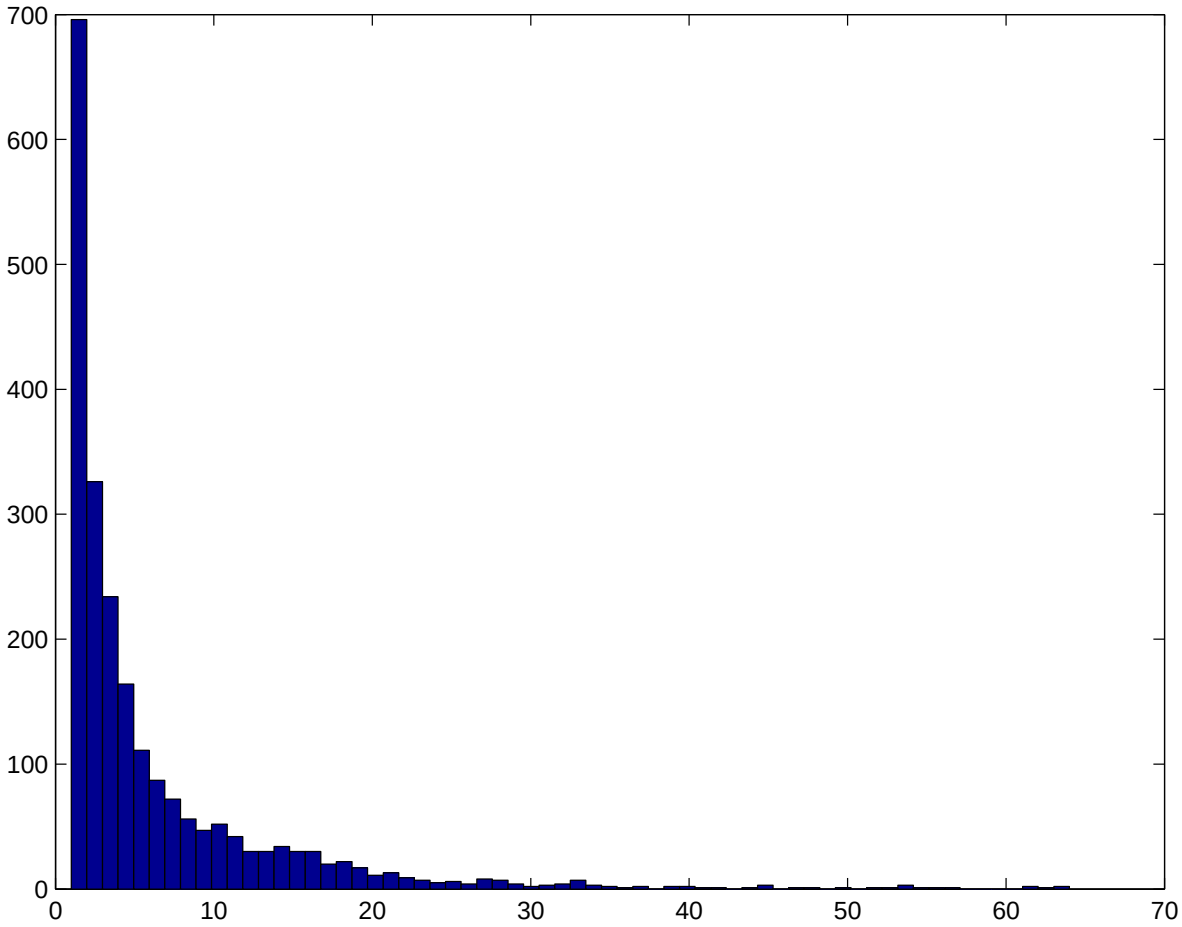
Adjacency matrix, $N = 2224$, $m = 6609$.

Example: PPI network of *Saccharomyces cerevisiae*



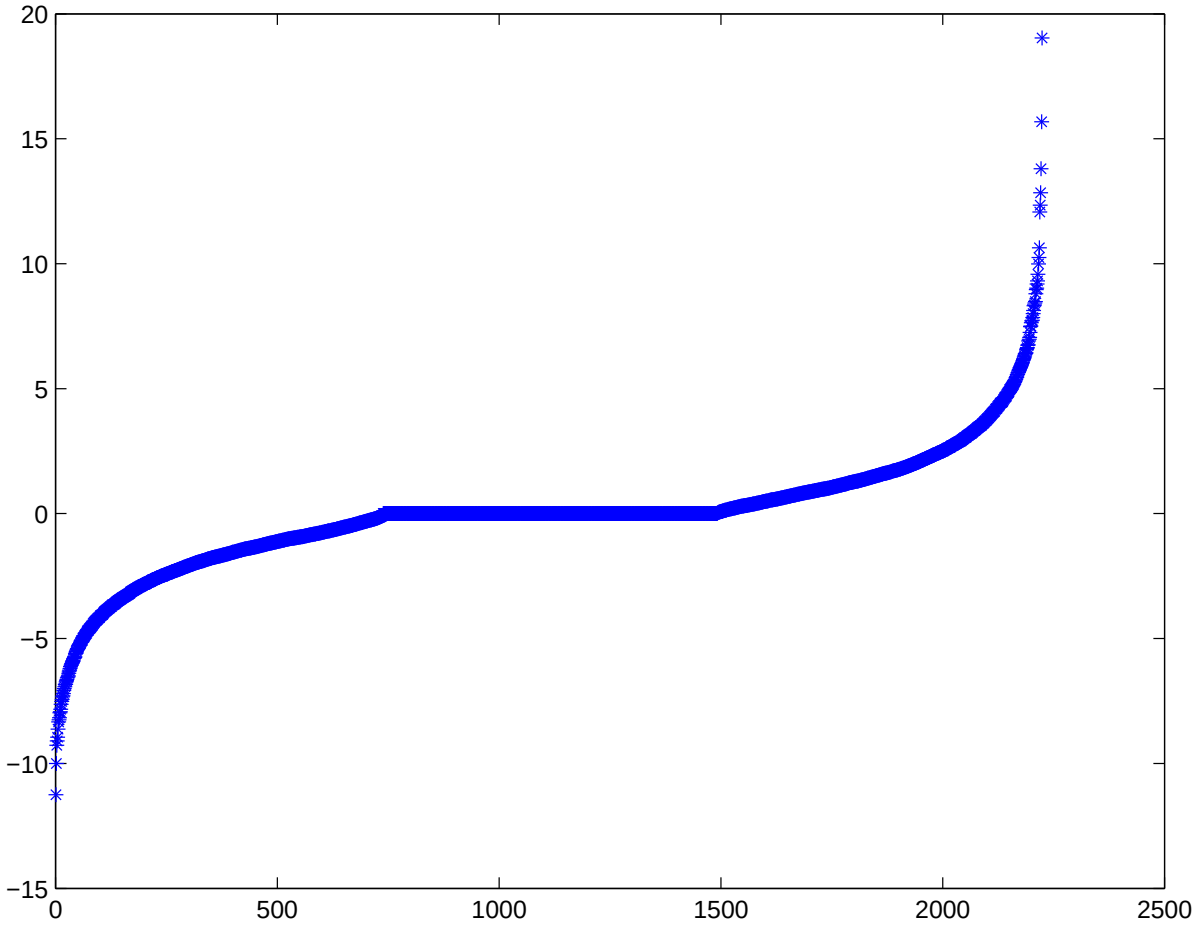
Same, reordered with Reverse Cuthill–McKee

Example: PPI network of *Saccharomyces cerevisiae*



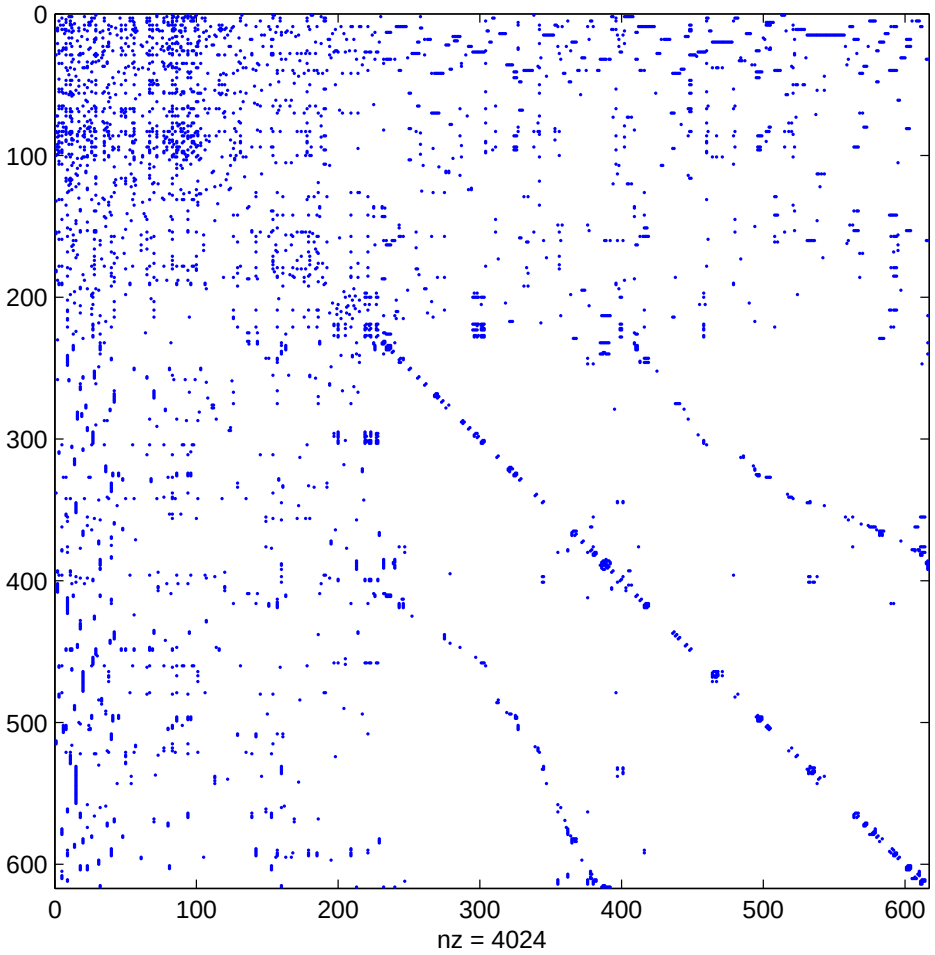
Degree distribution ($d_{\min} = 1$, $d_{\max} = 64$)

Example: PPI network of *Saccharomyces cerevisiae*



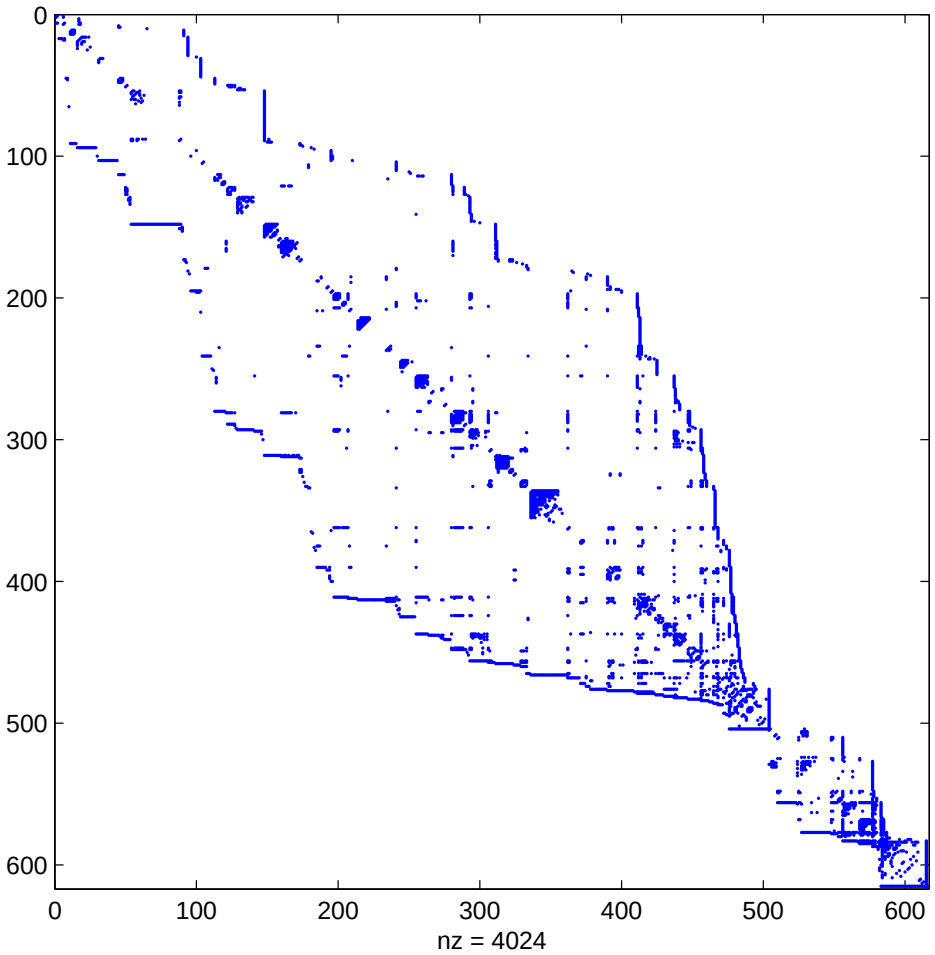
Eigenvalues

Example: Intravenous drug users network



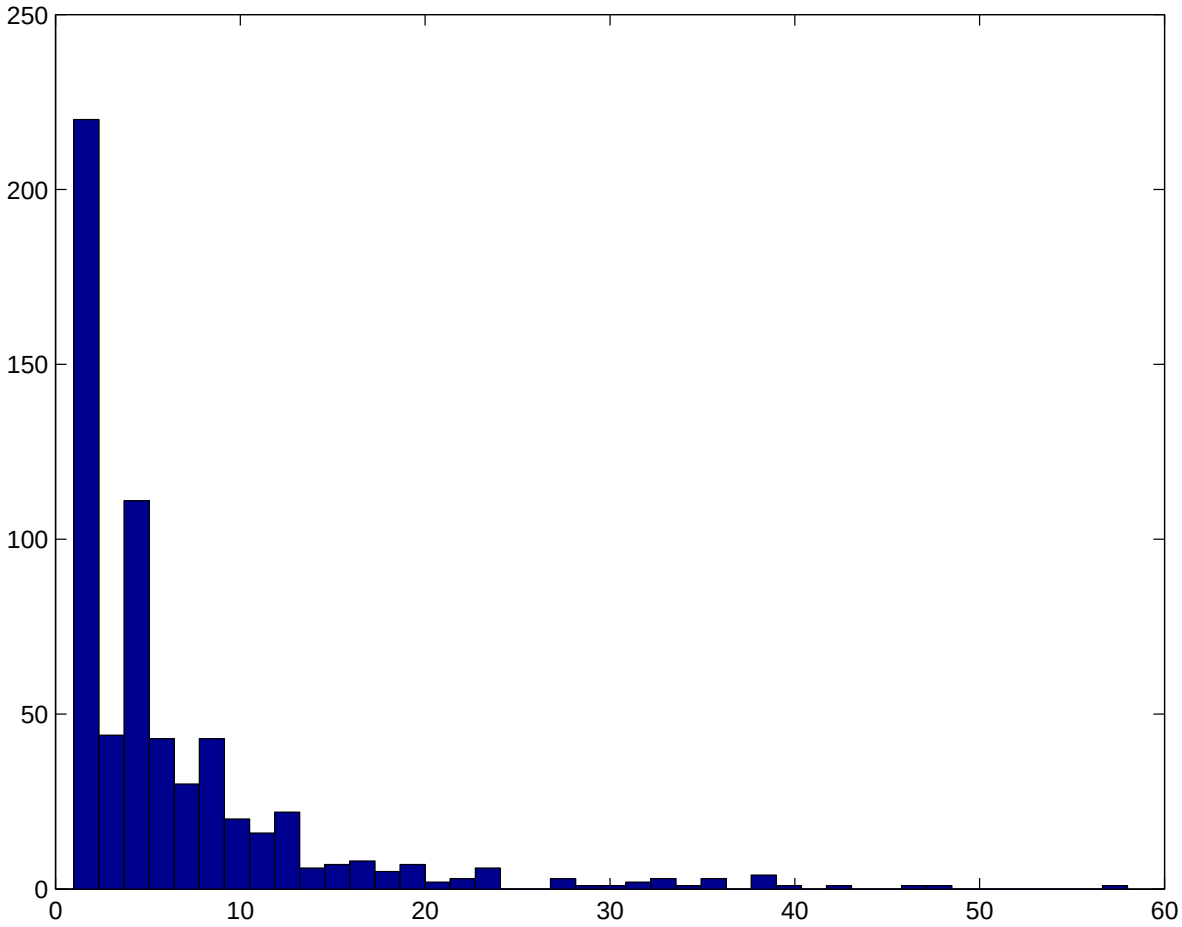
Adjacency matrix, $N = 616$, $m = 2012$.

Example: Intravenous drug users network



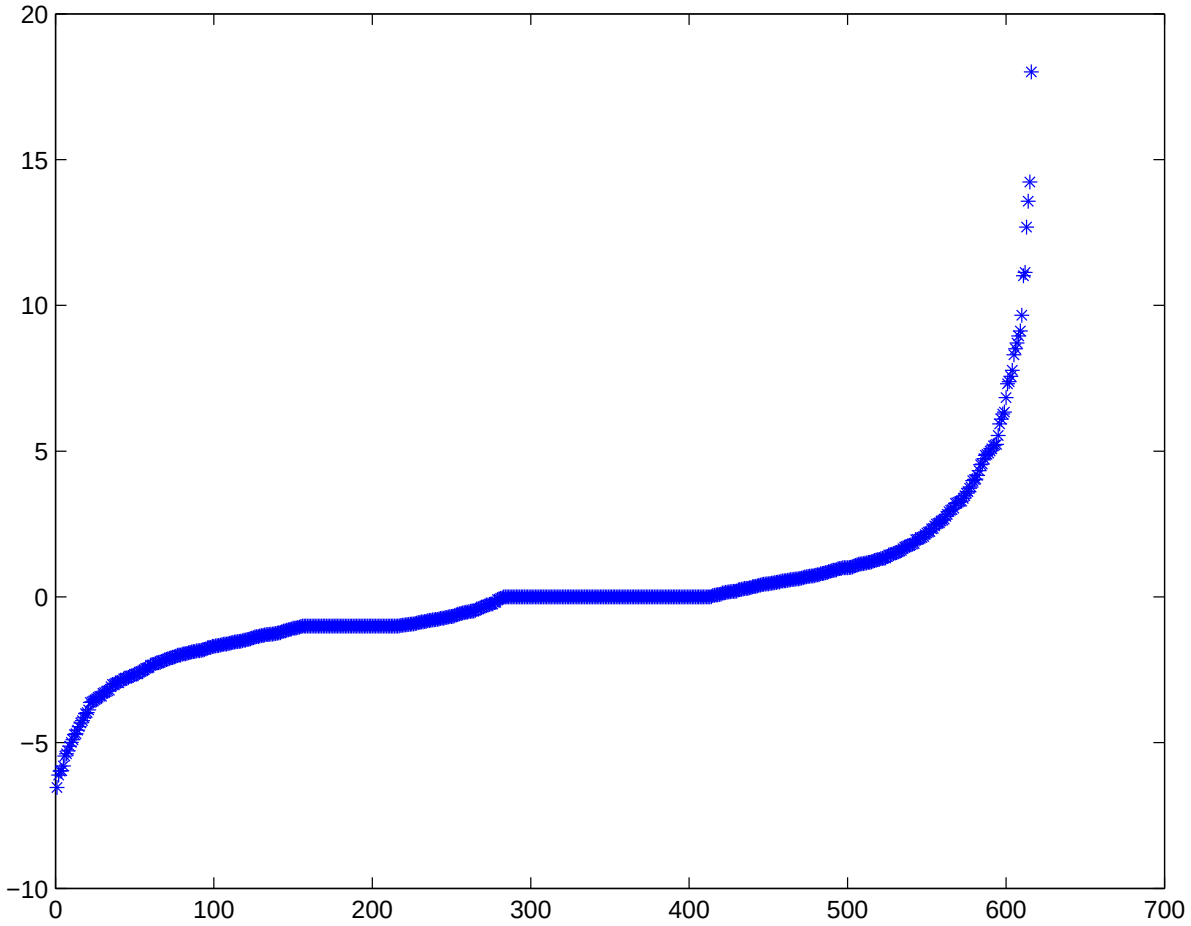
Same, reordered with Reverse Cuthill–McKee

Example: Intravenous drug users network



Degree distribution ($d_{\text{min}} = 1$, $d_{\text{max}} = 58$)

Example: Intravenous drug users network



Eigenvalues

Quantitative analysis of networks

Basic questions about networks include centrality, communicability and community detection issues:

- Which are the most "important" nodes?
  - Network connectivity (vulnerability)
  - Lethality in PPI networks
  - Author centrality in collaboration networks
- How do "disturbances" spread in a network?
  - Spreading of epidemics, rumors, fads,...
  - Routing of messages; returnability
- How to identify "community structures" in a network?
  - Clustering, transitivity
  - Partitioning

Subgraph centrality

There are dozens of different definitions of centrality for nodes in a graph. The simplest is degree centrality, which is just the degree d_i of node i. This does not take into account the "importance" of the nodes a given nodes is connected to—only their number.

Subgraph centrality (Estrada & Rodríguez-Velásquez, 2005) measures the centrality of a node by taking into account the number of subgraphs the node "participates" in.

This is done by counting, for all $k = 1, 2, \ldots$ the number of closed walks in G starting and ending at node i, with longer walks being penalized (given a smaller weight).

Subgraph centrality

Recall that

- $(A^k)_{ii} = \#$ of closed walks of length k based at node i,
- $(A^k)_{ij} = \#$ of walks of length k that connect nodes i and j.

Using $1/k!$ as weights leads to the notion of subgraph centrality:

$$\begin{aligned} SC(i) &= \left[I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots \right]_{ii} \\ &= [e^A]_{ii} \end{aligned}$$

Note: the additional term of order $k = 0$ does not alter the ranking of nodes in terms of subgraph centrality.

Subgraph centrality

Of course different weights can be used, leading to different matrix functions, such as the resolvent (Katz, 1953):

$$(I - cA)^{-1} = I + cA + c^2A^2 + \cdots, \quad 0 < c < 1/\lambda_N.$$

Subgraph centrality has been used successfully in various settings, especially proteomics and neuroscience.

Also, in the case of a directed network one can use the solution vectors of the linear systems

$$(I - cA)\mathbf{x} = \mathbf{1} \quad \text{and} \quad (I - cA^T)\mathbf{y} = \mathbf{1}$$

to rank hubs and authorities.

Subgraph centrality

Other matrix functions of interest are $\cosh(A)$ and $\sinh(A)$, which correspond to considering only walks of even and odd length, respectively.

Because in a bipartite graph $\sinh(A) = 0$, the quantity

$$\langle B(G) \rangle := \frac{\mathrm{Tr}\left(\cosh(A)\right)}{\mathrm{Tr}\left(\mathrm{e}^{A}\right)}$$

provides a measure of how "close" a graph is to being bipartite.

Hyperbolic matrix functions are also used to define the notion of returnability in digraphs (Estrada & Hatano, 2009).

Communicability

Communicability measures how "easy" it is to send a message from node i to node j in a graph. It is defined as (Estrada & Hatano, 2008):

$$\begin{aligned} C(i,j) &= [\mathrm{e}^A]_{ij} \\ &= \left[I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots \right]_{ij} \\ &\approx \text{weighted sum of walks joining nodes } i \text{ and } j. \end{aligned}$$

As before, other power series expansions (weights) can be used, leading to different functions of A.

For a large graph, computing all communicabilities $C(i,j)$ $(i \neq j)$ is prohibitively expensive. Instead, averages are often used—as in statistical physics.

Communicability

The average communicability of a node is defined as

$$\langle C(i) \rangle := \frac{1}{N-1} \sum_{j \neq i} C(i,j) = \frac{1}{N-1} \sum_{j \neq i} [e^A]_{ij}.$$

Communicability functions can be used to study the spreading of diseases (or rumors) and to identify bottlenecks in networks.

Recently, community detection algorithms based on various communicability measures have been developed (Fortunato, 2010; Ma, Gao & Yong, 2010; Estrada, 2011).

Algorithms

We saw that many problems in graph analysis lead to the computation of selected entries of $f(A)$, where A is the adjacency matrix and $f(x)$ is an analytic function.

Typically, the diagonal entries of $f(A)$ are wanted, and for large graphs some global averages over subsets of entries of $f(A)$, often expressed in terms of traces.

High accuracy is not always required. For example, in the case of centralities it is important to be able to identify just the top-ranked nodes. For the communicabilities, bounds are often sufficient.

Algorithms

All of these computations can be formulated as evaluation of bilinear forms $\mathbf{u}^T f(A)\,\mathbf{v}$ for suitably chosen vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ and for suitable functions $f(x)$:

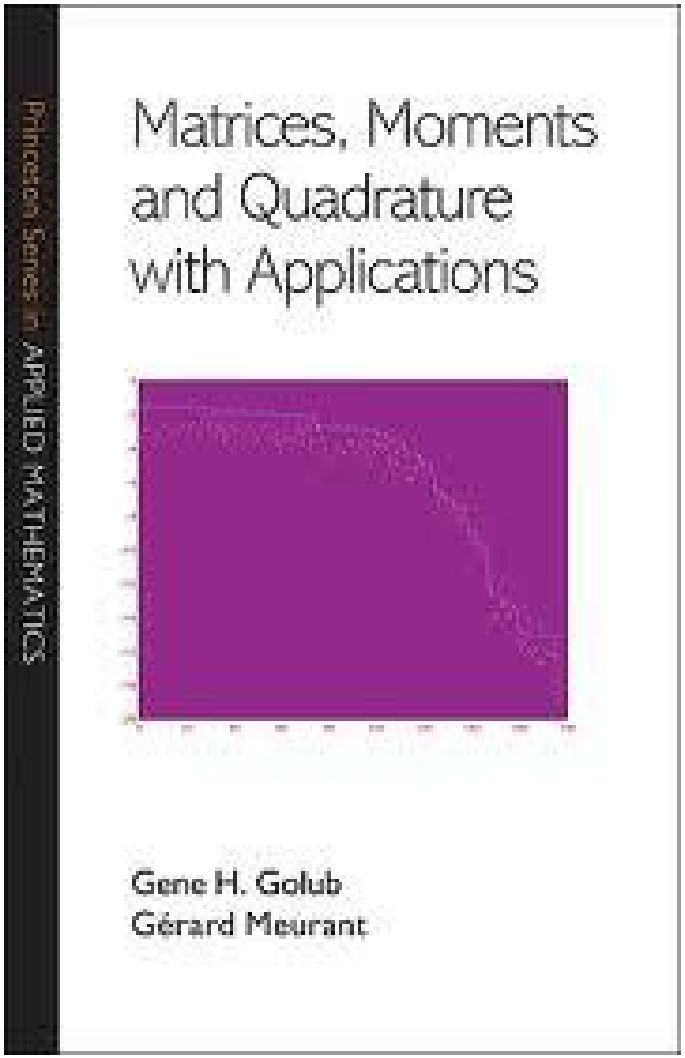
- Subgraph centrality: $SC(i) = \mathbf{e}_i^T \exp(A)\,\mathbf{e}_i$
- Communicability: $C(i,j) = \mathbf{e}_i^T \exp(A)\,\mathbf{e}_j$
- Average communicability:

$$\langle C(i) \rangle = \frac{1}{N-1}\left[\mathbf{e}^T \exp(A)\,\mathbf{e}_i - \mathbf{e}_i^T \exp(A)\,\mathbf{e}_i\right]$$

  where $\mathbf{e}$ is the vector of all ones.

How do we compute these quantities?

Algorithms


Princeton Series in APPLIED MATHEMATICS
Matrices, Moments and Quadrature with Applications
Gene H. Golub
Gérard Meurant

Algorithms

Consider the spectral decompositions

$$A = Q\Lambda Q^T, \qquad f(A) = Qf(\Lambda)Q^T.$$

For $u, v \in \mathbb{R}^N$ we have

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

where $p = Q^T u$ and $q = Q^T v$. Rewrite this as a Riemann-Stieltjes integral:

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

Algorithms

The general Gauss-type quadrature rule gives in this case:

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

where the nodes $\{z_k\}$ are prescribed.

- Gauss: $M = 0$,
- Gauss–Radau: $M = 1$, $z_1 = a$ or $z_2 = b$,
- Gauss–Lobatto: $M = 2$, $z_1 = a$ and $z_2 = b$.

The evaluation of these quadrature rules is reduced to

- computation of orthogonal polynomials via three-term recurrence,
- or, equivalently, computation of entries and spectral information of the corresponding tridiagonal matrix (Lanczos).

Algorithms

The tridiagonal matrix J_n corresponds to the three-term recurrence relationship satisfied by the set of polynomials orthonormal with respect to $d\mu$.

$$J_n = \begin{pmatrix} \omega_1 & \gamma_1 & & & \\ \gamma_1 & \omega_2 & \gamma_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \gamma_{n-2} & \omega_{n-1} & \gamma_{n-1} \\ & & & \gamma_{n-1} & \omega_n \end{pmatrix}$$

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

The quadrature rule is computed with the Golub–Welsch QR algorithm.

Algorithms

Consider the case $u = v = e_i$ (corresp. to the (i, i) entry of $f(A)$).

The entries of J_n are computed using the symmetric Lanczos algorithm:

$$\gamma_j x_j = r_j = (A - \omega_j I)x_{j-1} - \gamma_{j-1} x_{j-2}, \qquad j = 1, 2, \ldots$$
$$\omega_j = x_{j-1}^T A x_{j-1},$$
$$\gamma_j = \|r_j\|$$

with initial vectors $x_{-1} = 0$ and $x_0 = e_i$.

Each additional Lanczos step amounts to adding another node to the Gauss-type quadrature rule, resulting in tighter and tighter bounds (implementation based on MMQ Matlab toolbox by G. Meurant).

Algorithms

For $f(A) = \mathrm{e}^A$ and $f(A) = (I - cA)^{-1}$ we obtain:

- bounds on $[f(A)]_{ii}$ from symmetric Lanczos,
- bounds on $[f(A)]_{ii} + [f(A)]_{ij}$ from unsymmetric Lanczos,
- lower bounds from the Gauss and the Gauss-Radau rules,
- upper bounds from the Gauss-Radau and Gauss-Lobatto rules.

In computations we often use the simple (Gerschgorin) estimates

$$\lambda_{\min} \approx - \max_{1 \le i \le N} \{\deg(i)\}, \qquad \lambda_{\max} \approx \max_{1 \le i \le N} \{\deg(i)\}.$$

Using more accurate estimates of the extreme eigenvalues of A generally leads to improved results, especially in scale-free case.

Summary of results

Numerical experiments were performed on a number of networks, both synthetic (generated with the CONTEST Toolbox for Matlab, Taylor & Higham 2010) and from real-world applications, including PPI and transportation networks.

- A few (3-5) Lanczos steps per estimate are usually enough
- Independent of N for graphs with bounded max degree
- Almost as good for power-law degree distributions
- Very efficient approach for computing global averages

Summary of results (cont.)

Example: Range dependent network from CONTEST, $N = 100$.

- MMQ approximations of the Estrada index $EE(G) = 425.0661$

# it	1	2	3	4	5
Gauss	348.9706	416.3091	424.4671	425.0413	425.0655
Radau (lower)	378.9460	420.6532	424.8102	425.0570	425.0659
Radau (upper)	652.8555	437.7018	425.6054	425.0828	425.0664
Lobatto	2117.9233	531.1509	430.3970	425.2707	425.0718

- MMQ approximations of $[e^A]_{1,5} = 0.396425$

# it	1	2	3	4	5
Radau (lower)	-2.37728	0.213316	0.388791	0.396141	0.396431
Radau (upper)	4.35461	0.595155	0.404905	0.396626	0.396420

Summary of results (cont.)

Bounds for the Estrada index of small world matrices of increasing size. Watts–Strogatz model with parameters $(4, 0.1)$; five Lanczos iterations.

N	Gauss	Gauss-Radau (l)	Gauss-Radau (u)	Gauss-Lobatto
1000	1.64e5	1.63e5	1.64e5	1.66e5
2500	4.09e5	4.10e5	4.11e5	4.12e5
5000	8.18e5	8.18e5	8.20e5	8.33e5
7500	1.23e6	1.22e6	1.22e6	1.25e6
10000	1.63e6	1.63e6	1.64e6	1.66e6

Five iterations suffice to reach a small relative error, independent of N.

Summary of results (cont.)

Experiments were performed on scale-free networks (Barabási–Albert model) of increasing size, to see the effect of increasing maximum degree.

In this case, the Gerschgorin estimate $\lambda_{\max} \approx d_{\max}$ gives poor results. Indeed, whereas the max degree increases with N, the spread of the eigenvalues tend to grow very slowly.

A better strategy is to estimate $\lambda_{\min}$, $\lambda_{\max}$ using Lanczos.

The number of Lanczos steps per estimate is then approximately independent of N, for a fixed accuracy. Typically, 9-11 Lanczos steps suffice to achieve relative errors around 10^{-6}.

Some challenges

- Relate convergence rates of Gaussian quadrature rules to network properties
- Can one quickly locate subsets of nodes with the largest centralities $SC(i)$?
- For truly massive networks, current approaches are too expensive
  - Fast estimates for $\mathrm{Tr}\,(f(A))$: randomized algorithms? (Avron & Toledo)
  - Low-rank approximations?
  - Exploitation of locality?
- Parallelization is tricky
- Major issue in this area: how to evaluate and compare measures?

Basic references

Some classic early references:

- L. Katz, *A new status index derived from sociometric analysis*, Psychometrika, 18 (1953), pp. 39–43.
- A. Rapoport, *Mathematical models of social interaction*, in Handbook of Mathematical Psychology, vol. 2, pp. 493–579. Wiley, New York, 1963.
- D. J. de Solla Price, *Networks of scientific papers*, Science, 149 (1965), pp. 510–515.
- S. Milgram, *The small world problem*, Psychology Today, 2 (1967), pp. 60–67.
- J. Travers and S. Milgram, *An experimental study of the small world problem*, Sociometry, 32 (1969), pp. 425–443.

Basic references (cont.)

The field exploded in the late 1990s due to several breakthroughs by physicists, applied mathematicians and computer scientists. This helped make the field more quantitative and mathematically sophisticated. Landmark papers include

- D. J. Watts and S. H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature, 393 (1998), pp. 440–442.
- L.-A. Barabási and R. Albert, *Emergence of scaling in random networks*, Science, 386 (1999), pp. 509–512.
- M. E. J. Newman, *Models of the small world*, J. Stat. Phys., 101 (2000), pp. 819–841.
- J. Kleinberg, *Navigation in a small world*, Nature, 406 (2000), p. 845.
- R. Albert and L.-A. Barabási, *Statistical mechanics of complex networks*, Rev. Modern Phys., 74 (2002), pp. 47–97.

Basic references (cont.)

Recent systematic accounts include

- U. Brandes and T. Erlebach (Eds.), *Network Analysis. Methodological Foundations*, Springer, LNCS 3418, 2005.
- F. Chung and L. Lu, *Complex Graphs and Networks*, American Mathematical Society, 2006.
- A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks*, Cambridge University Press, 2008.
- M. E. J. Newman, *Networks. An Introduction*, Oxford University Press, 2010.
- P. Van Mieghem, *Graph Spectra for Complex Networks*, Cambridge University Press, 2011.
- Popular accounts by Barabási (2002) and by Watts (2003).