Applications of Matrix Functions to Network Analysis and Quantum Chemistry
Part I: Complex Networks

Michele Benzi

Department of Mathematics and Computer Science

Emory University

Atlanta, Georgia, USA

SSMF2013, Lille, France

Outline

Outline

Plan and acknowledgements

- Lecture I: Background on Complex Networks
- Lecture II: Algorithms
- Lecture III: Applications to Quantum Chemistry

Many thanks to Paola Boito, Ernesto Estrada, Christine Klymko, and Nader Razouk. I would also like to gratefully acknowledge the early influence of Gene Golub.

Gene Howard Golub (1932–2007)

Ghent, Belgium, September 2006 (courtesy of Gérard Meurant)

Outline

Complex networks

- 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.

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.
- E. Estrada, *The Structure of Complex Networks*, Oxford University Press, 2011.

Basic references (cont.)

Two new interdisciplinary journals in 2013:

- Journal of Complex Networks (Oxford University Press)
- Network Science (Cambridge University Press)

`http://www.oxfordjournals.org/`

`http://journals.cambridge.org/`

Complex networks: what are they?

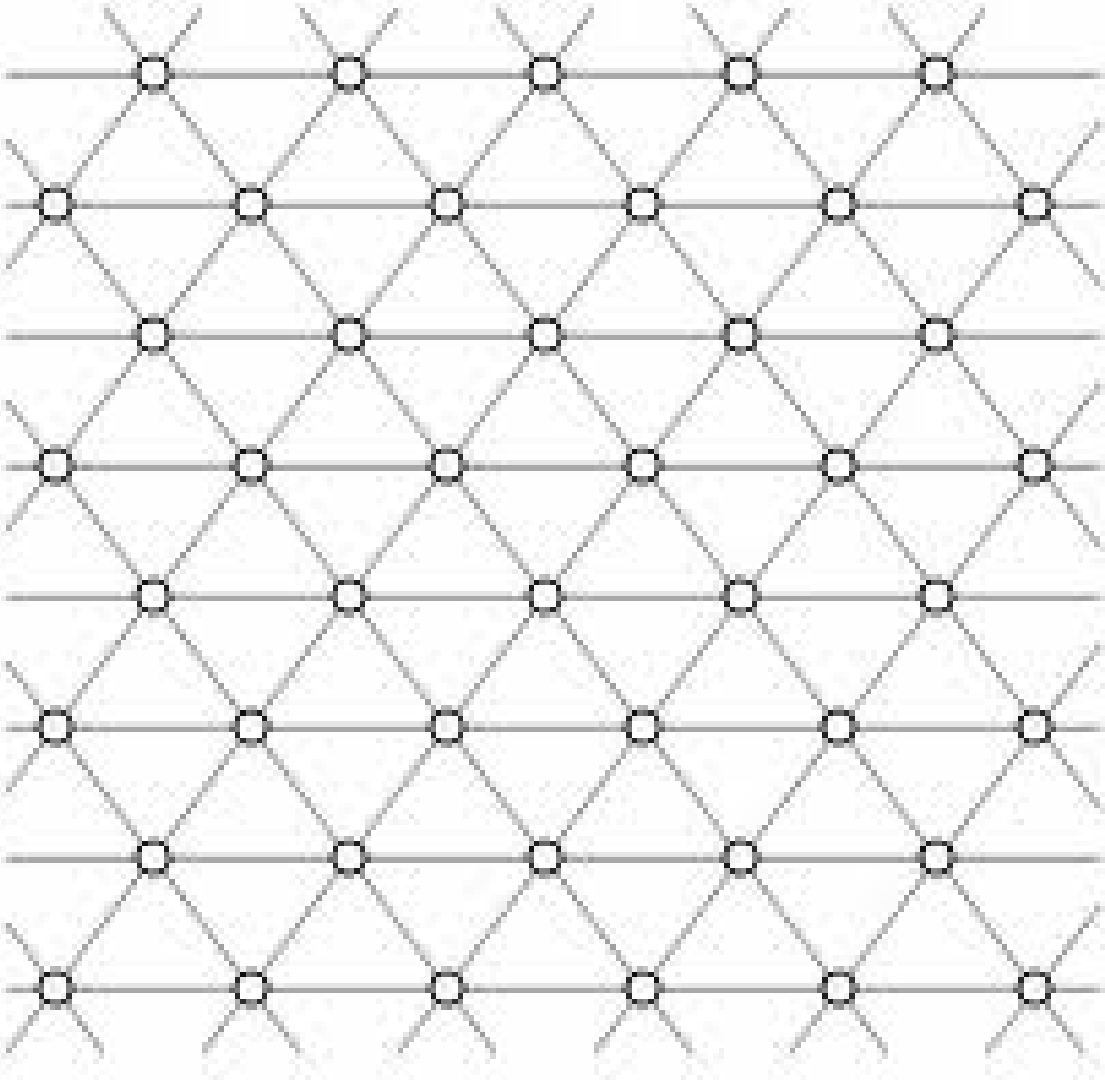
But what exactly is a complex network?

Unfortunately, no precise definition exists.

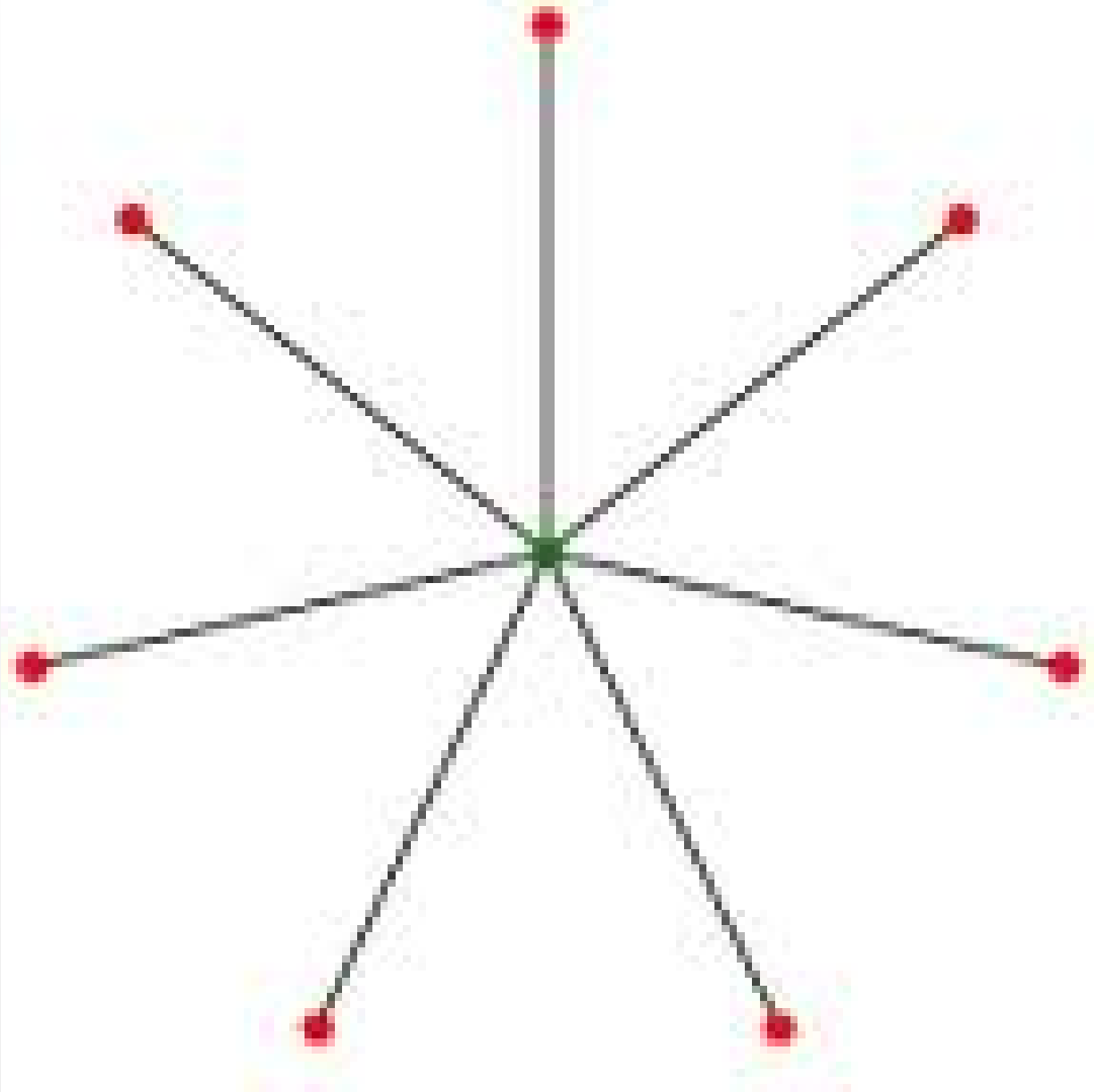
It is easier to say which graphs are not complex networks. Regular lattices are not considered complex networks, and neither are completely random graphs such as the Erdös–Rényi model.

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

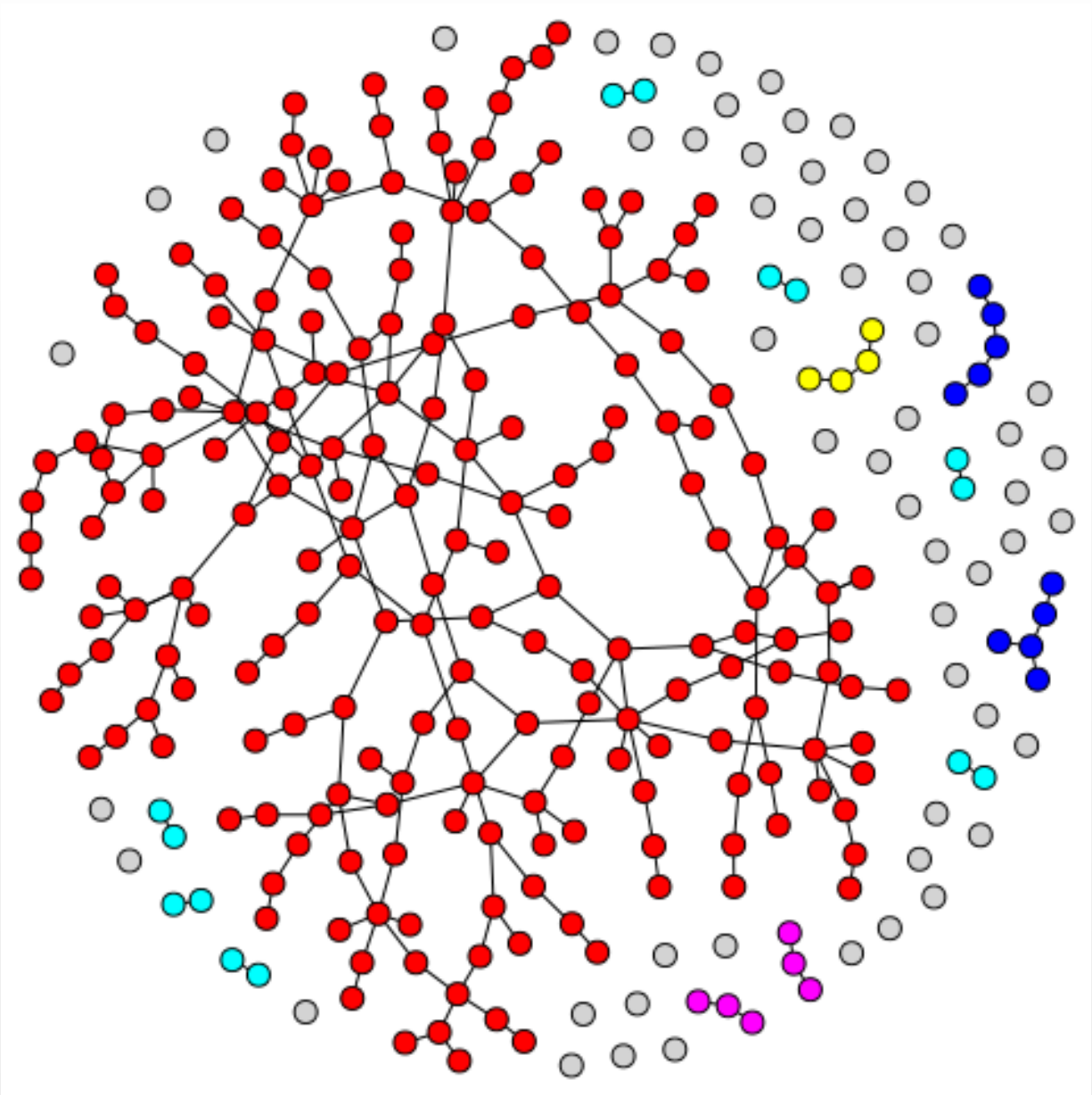
Regular lattice: not a complex network!



Star graph: also not a complex network!



Erdös–Rényi graph: also not a complex network!



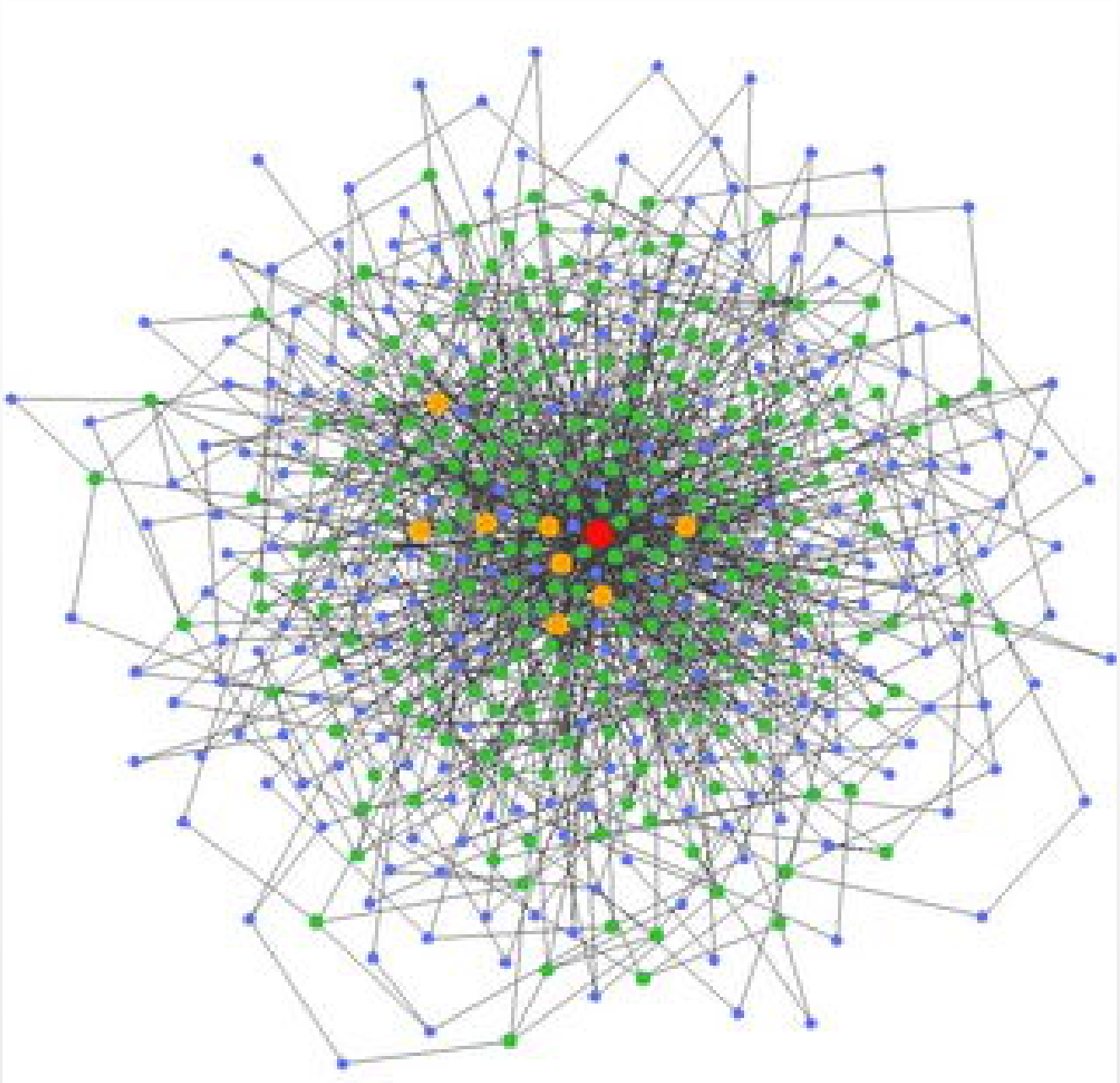
Some features of complex networks

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

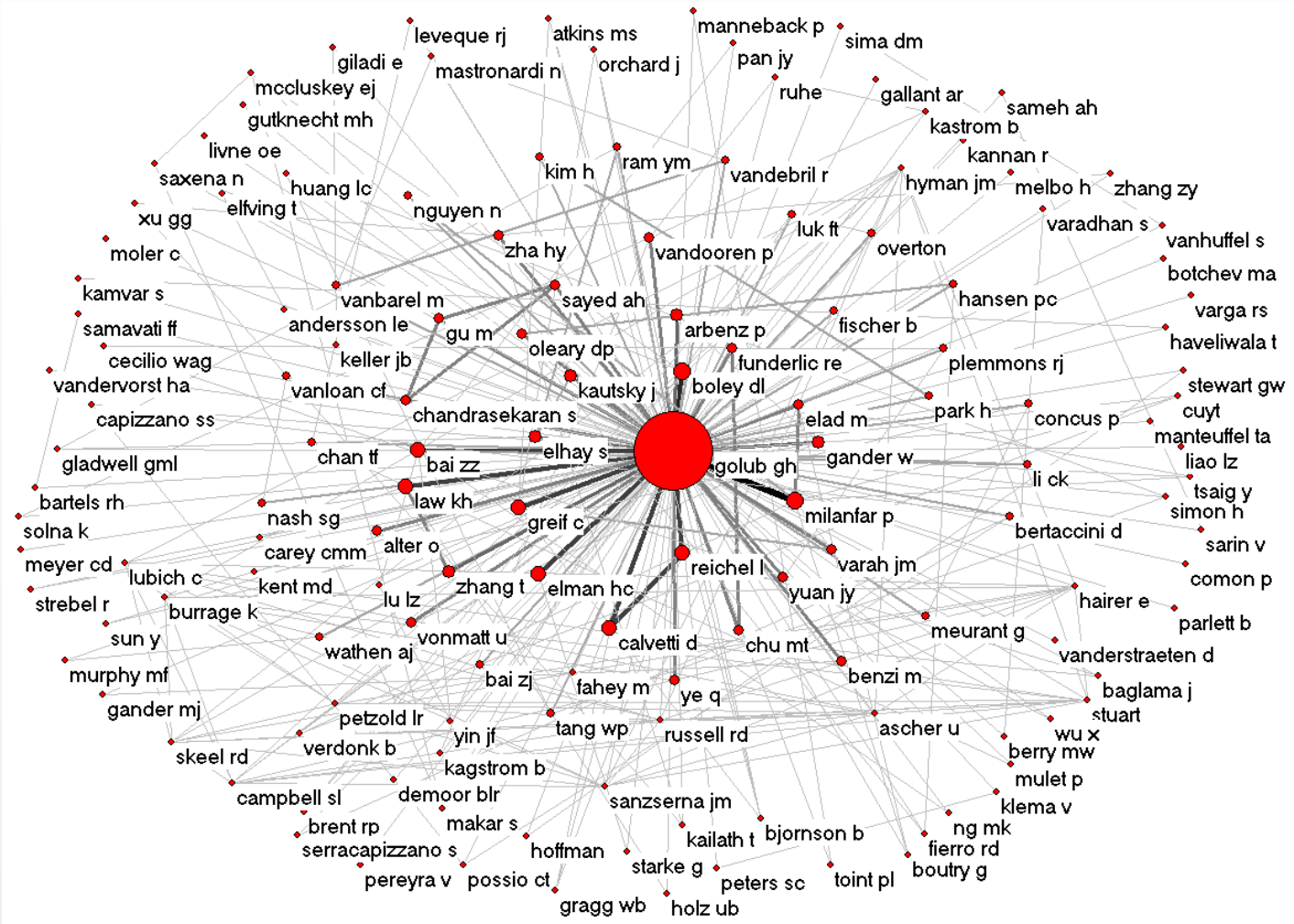
- Scale-free (power law degree distribution)
- Small-world (small graph diameter)
- Highly clustered (many triangles, hubs...)
- Hierarchical
- Rich in 'motifs'
- Self-similar

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

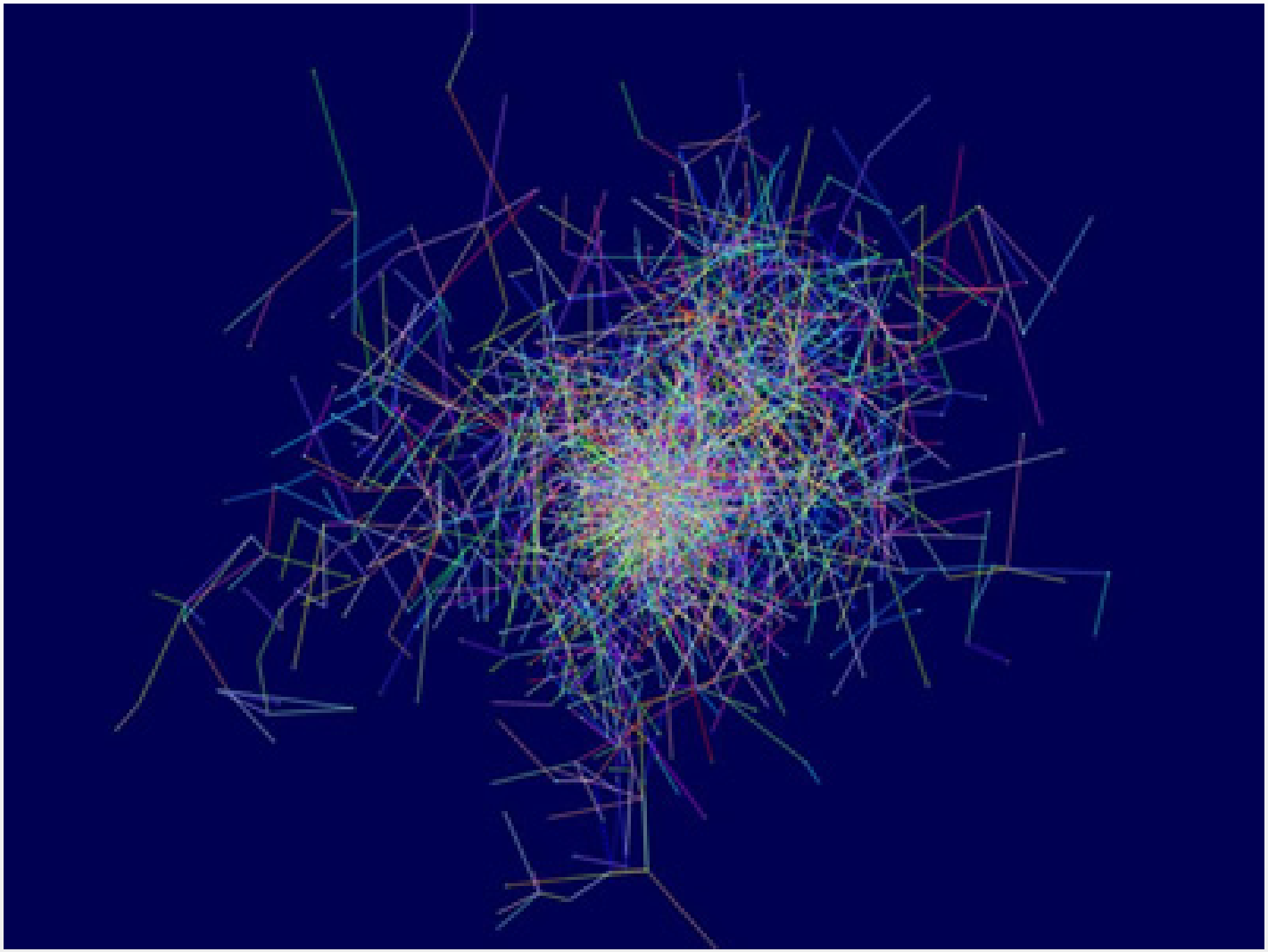
Example of complex network: B–A model



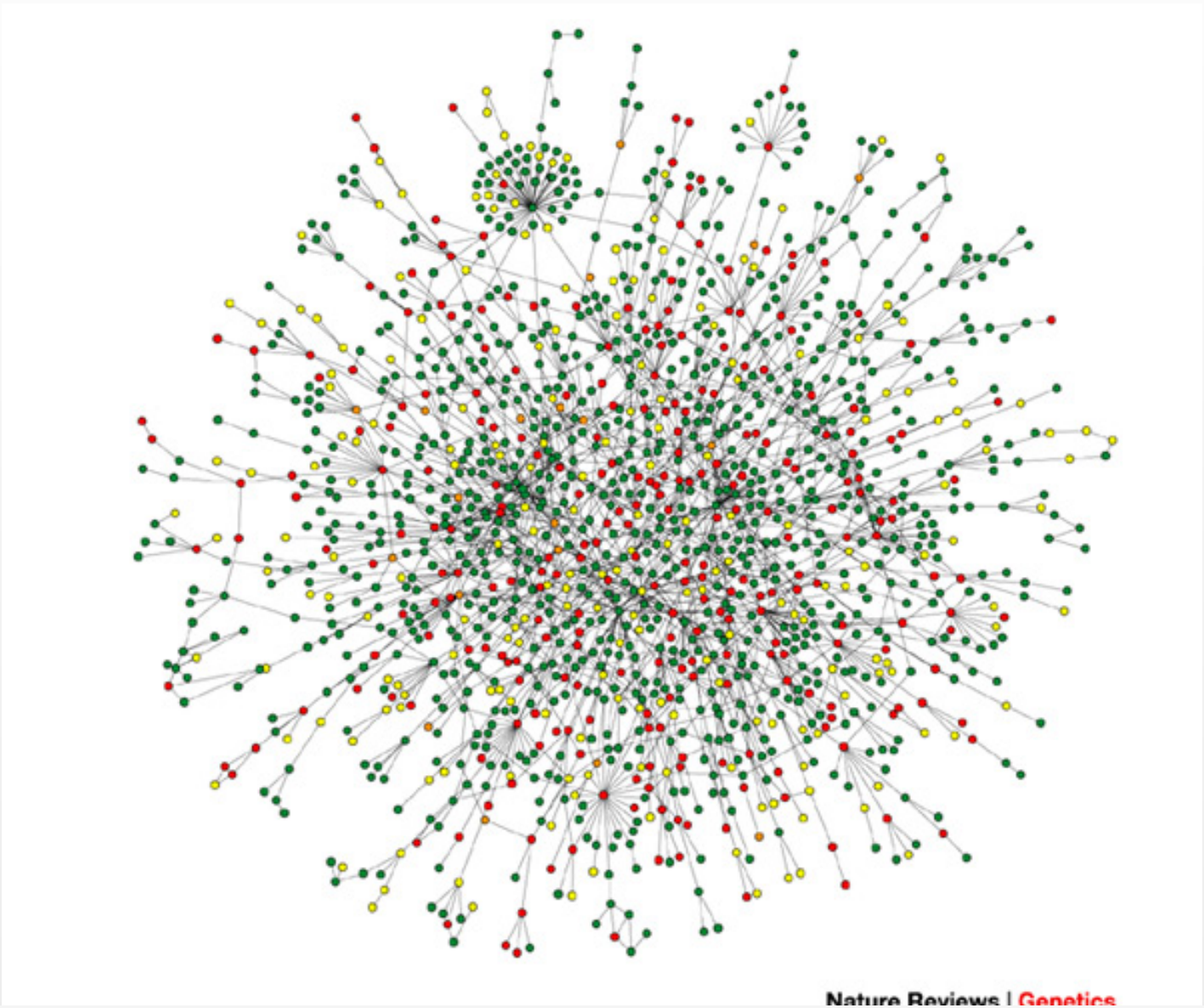
Example of complex network: Golub collaboration graph



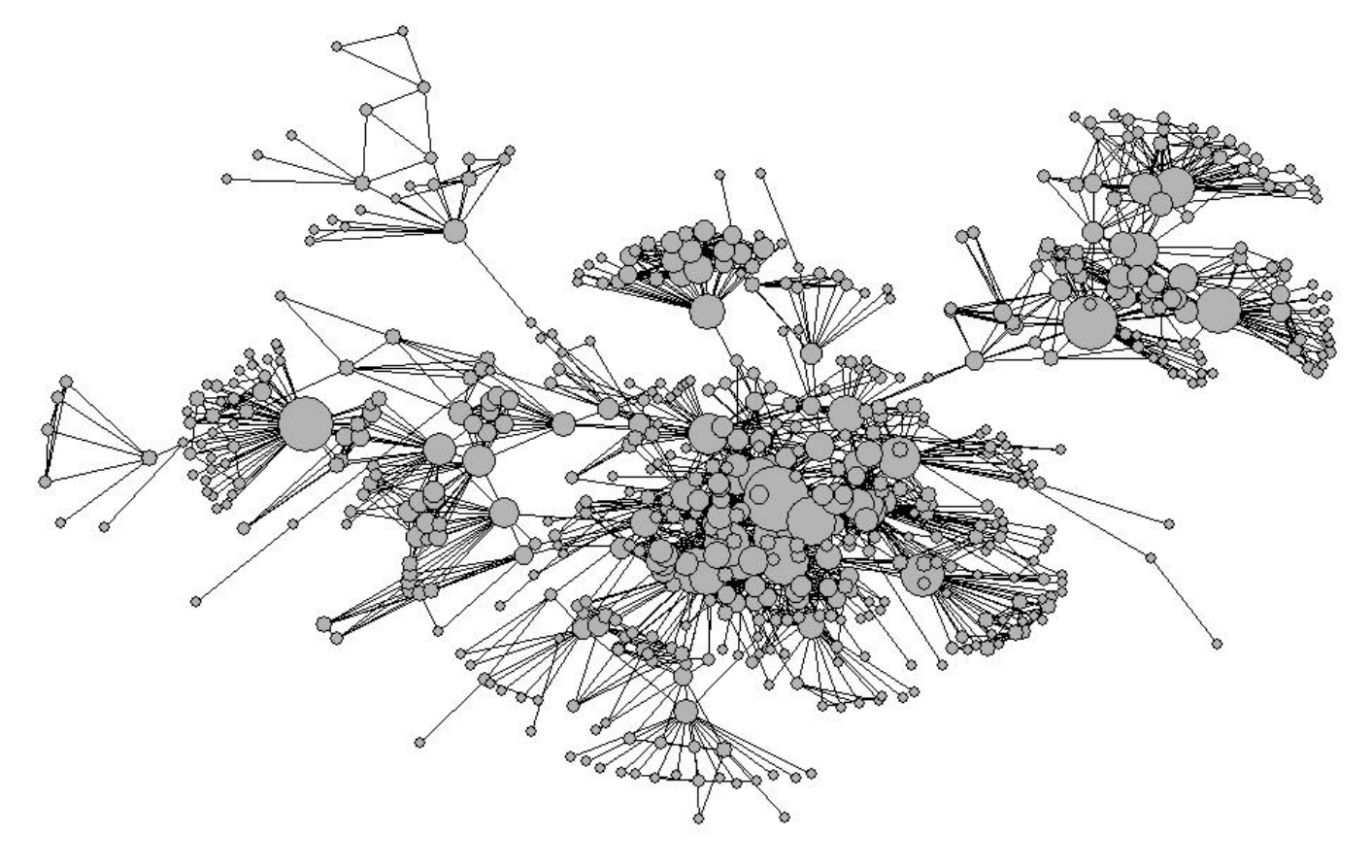
Example of complex network: Erdös collaboration graph



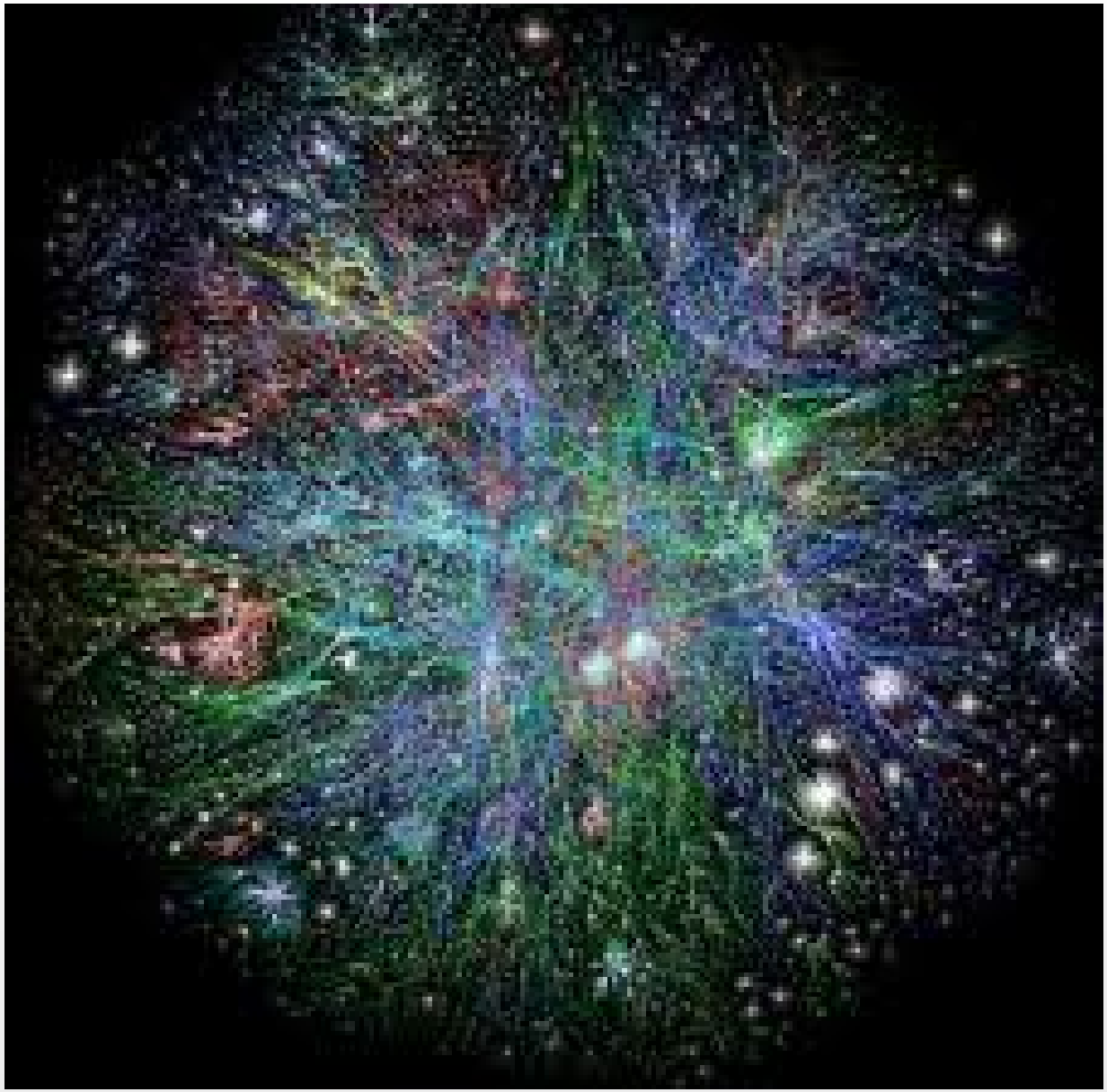
Example of complex network: PPI network of *Saccharomyces cerevisiae* (beer yeast)



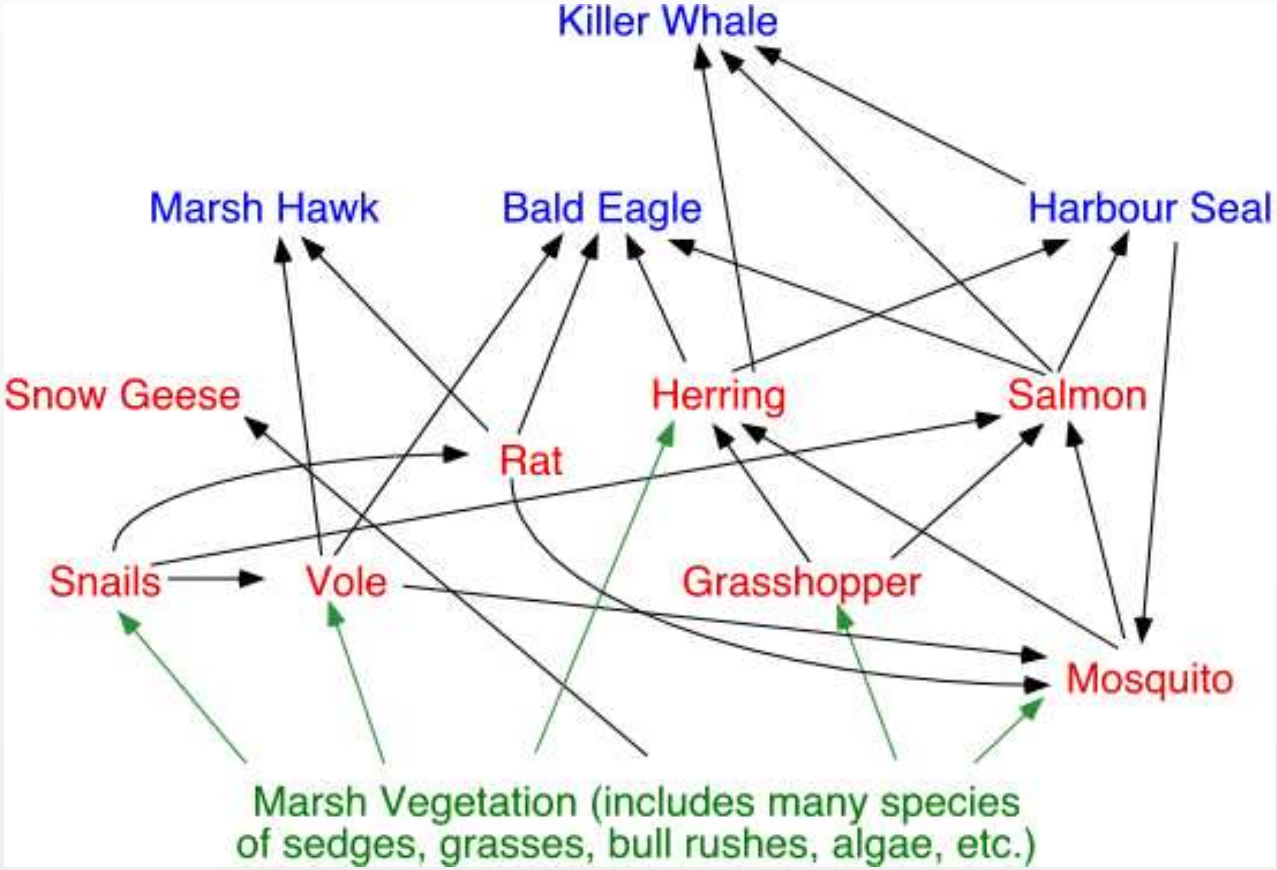
Example of complex network: social network of injecting drug users in Colorado Springs, CO



Example of complex network: the Internet



Example of (directed) complex network: a food web



Network analysis

Basic questions about network structure 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, beliefs, rumors, fads,...
  - Routing of messages; returnability
- How to identify "community structures" in a network?
  - Clustering, transitivity
  - Partitioning

Formal definitions

Real-world networks are usually modelled by means of graphs.

A graph $G = (V, E)$ consists of a (finite) set $V = \{v_1, v_2, \ldots, v_N\}$ of nodes (or vertices) and a set E of edges (or links), which are pairs $\{v_i, v_j\}$ with $v_i, v_j \in V$.

The graph G is directed if the edges $\{v_i, v_j\} \in E$ are ordered pairs $(= (v_i, v_j) \in V \times V)$, otherwise G is undirected. A directed graph is often referred to as a digraph.

A loop in G is an edge from a node to itself. Loops are often ignored or excluded.

A graph G is weighted if numerical values are associated with its edges. If all the edges are given the same value 1, we say that the graph is unweighted.

A simple graph is an unweighted graph without multiple edges or loops.

Formal definitions (cont.)

A walk of length k in G is a set of nodes $v_{i_1}, v_{i_2}, \ldots v_{i_k}, v_{i_{k+1}}$ such that for all $1 \leq j \leq k$, there is an edge between v_{i_j} and $v_{i_{j+1}}$.

A closed walk is a walk where $v_{i_1} = v_{i_{k+1}}$.

A path is a walk with no repeated nodes.

A cycle is a path with an edge between the first and last node. In other words, a cycle is a closed path.

A triangle in G is a cycle of length 3.

Formal definitions (cont.)

The geodetic distance $d(v_i, v_j)$ between two nodes is the length of the shortest path connecting v_i and v_j. We let $d(v_i, v_j) = \infty$ if no such path exists.

The diameter of a graph $G = (V, E)$ is defined as

$$\text{diam}(G) := \max_{v_i, v_j \in V} d(v_i, v_j)\,.$$

A graph G is connected if for every pair of nodes v_i and v_j there is a path in G that starts at v_i and ends at v_j; i.e., $\text{diam}(G) < \infty$.

These definitions apply to both undirected and directed graphs, though in the latter case the orientation of the edges must be taken into account.

Formal definitions (cont.)

To every unweighted graph $G = (V, E)$ we associate its adjacency matrix $A = [a_{ij}] \in \mathbb{R}^{N \times N}$, with

$$a_{ij} = \begin{cases} 1, & \text{if } (v_i, v_j) \in E, \\ 0, & \text{else.} \end{cases}$$

Any renumbering of the graph nodes results in a symmetric permutation $A \longrightarrow PAP^T$ of the adjacency matrix of the graph.

If G is an undirected graph, A is symmetric with zeros along the main diagonal (A is "hollow"). In this case, the eigenvalues of A are all real. We label the eigenvalues of A in non-increasing order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. Note that A is always indefinite if $E \neq \emptyset$.

If G is connected, then λ_1 is simple and satisfies $\lambda_1 > |\lambda_i|$ for $2 \leq i \leq N$ (this follows from the Perron–Frobenius Theorem).

In particular, the spectral radius $\rho(A)$ is given by λ_1.

Formal definitions (cont.)

If G is undirected, the degree d_i of node v_i is the number of edges incident to v_i in G. In other words, d_i is the number of "immediate neighbors" of v_i in G. A regular graph is a graph where every node has the same degree d.

Note that in terms of the adjacency matrix, $d_i = \sum_{j=1}^{N} a_{ij}$.

For a directed graph, we define the in-degree of node v_i as the number d_i^{in} of edges ending in v_i, and the out-degree of v_i as as the number d_i^{out} of edges originating at v_i.

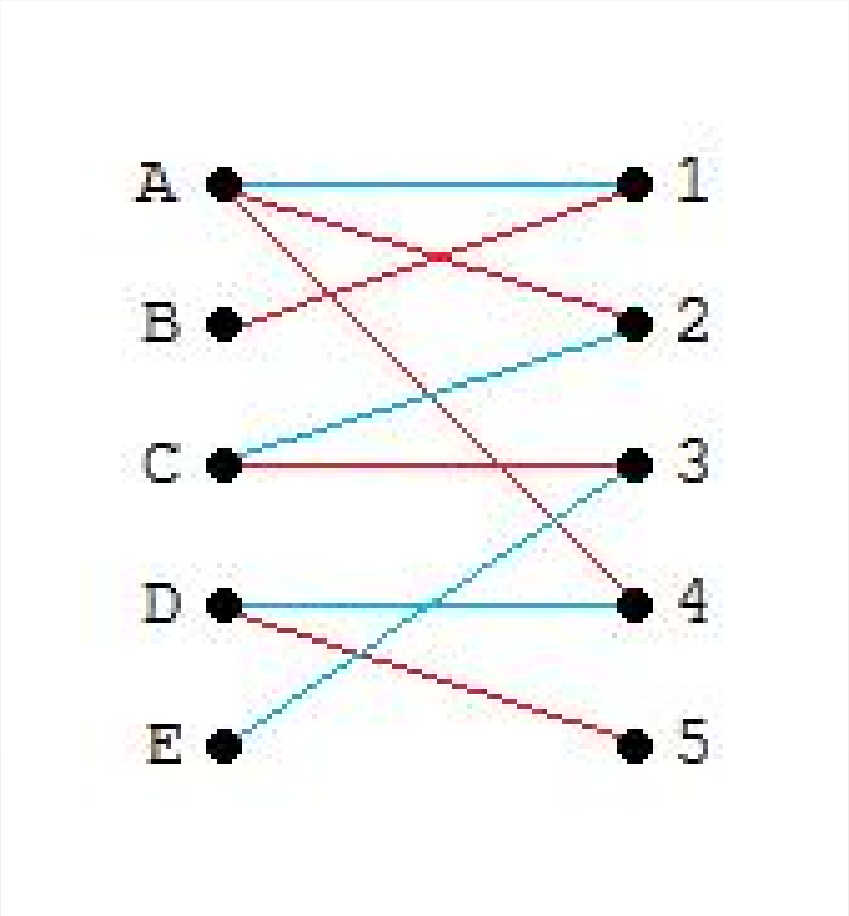
In terms of the (nonsymmetric) adjacency matrix,

$$d_i^{in} = \sum_{i=1}^{N} a_{ij}, \quad d_i^{out} = \sum_{j=1}^{N} a_{ij}.$$

Hence, the column sums of A give the in-degrees and the row sums give the out-degrees.

Formal definitions (cont.)

An undirected graph $G = (V, E)$ is bipartite if there are $V_1, V_2 \subset V$, with $V_1, V_2 \neq \emptyset$, $V = V_1 \cup V_2$, $V_1 \cap V_2 = \emptyset$ such that edges can exist only between nodes belonging to different subsets V_1, V_2. In other terms, a graph is bipartite if it does not contain any odd-length cycles.



Formal definitions (cont.)

Let $G = (V, E)$ be bipartite with $|V_1| = m$, $V_2 = n$, $m + n = N$.

Then there exists a numbering of the nodes of G such that the adjacency matrix of G is of the form

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

with $B \in \mathbb{R}^{m \times n}$.

Note: The nonzero eigenvalues of A are of the form $\pm\sigma_i$, where σ_i denote the singular values of B.

Bipartite graphs can also be used to give an alternative representation of directed graphs.

Formal definitions (cont.)

Indeed, given a digraph $G = (V, E)$ with N nodes we can make a copy $V' = \{v'_1, \ldots, v'_N\}$ of $V = \{v_1, \ldots, v_N\}$ and define a new, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ wth $2N$ nodes, where $\mathcal{V} := V \cup V'$ and

$$\mathcal{E} = \{(v_i, v'_j) \,|\, (v_i, v_j) \in E\}.$$

If A is the adjacency matrix of the original digraph G, the adjacency matrix of the corresponding bipartite graph $\mathcal{G}$ is given by

$$\mathcal{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \in \mathbb{R}^{2N \times 2N}.$$

As before, the nonzero eigenvalues of $\mathcal{A}$ come in opposite pairs, $\pm\sigma_i(A)$.

We will make use of this bipartite representation of digraphs in Lecture II.

Formal definitions (cont.)

Let now G be an undirected graph with adjacency matrix A.

Let $D = \text{diag}(d_1, \ldots, d_N)$ where d_i is the degree of node v_i in G. The matrix

$$L := D - A$$

is called the graph Laplacian of the graph G.

It is easy to see that L is symmetric positive semidefinite. By construction, $L\mathbf{1} = \mathbf{0}$ where $\mathbf{1}$ denotes the vector of all 1's. Hence, L is singular. Indeed, L is a singular M-matrix.

Furthermore, the dimension of the null space $\mathcal{N}(L)$ is equal to the number of connected components of G.

If G is connected, the null space is 1-dimensional: $\mathcal{N}(L) = \text{span}\{\mathbf{1}\}$.

Formal definitions (cont.)

Assume G is connected. The eigenvector associated with the smallest nonzero eigenvalue of L is called the Fiedler vector of the graph G. The Fiedler vector has been widely used in graph partitioning problems.

If the graph G is regular, then $D = dI_N - A$, therefore the eigenvalues of the Laplacian are just

$$\lambda_i(L) = d - \lambda_i(A), \quad 1 \leq i \leq N\,.$$

If G is not a regular graph, there is no simple relationship between the eigenvalues of L and those of A.

Also useful is the notion of normalized Laplacian:

$$\widehat{L} := I_N - D^{-1/2} A D^{-1/2}\,.$$

In the case of directed graphs, there are several distinct notions of graph Laplacian in the literature.

Complex networks

Complex graphs arising in real-world applications tend to be highly irregular and exhibit a nontrivial topology—in particular, they are far from being completely "random".

Complex networks are very often

- Scale-free, meaning that their degree distribution tends to follow a power law: $p(d) =$ number of nodes of degree $d \approx c \cdot d^{-\alpha}$, $\alpha > 0$. Frequently, $2 < \alpha \leq 3$. This implies sparsity but also the existence of several highly connected nodes (hubs).
- Small-world, meaning that the diameter grows very slowly with the number N of nodes; e.g.,

$$\operatorname{diam}(G) = O(\log N), \quad N \to \infty.$$

- Highly clustered, i.e., they contain a very large proportion of triangles.

The last property is especially frequent in social networks.

Complex networks (cont.)

A clustering coefficient measures the degree to which the nodes in a network tend to cluster together. For a node v_i with degree d_i, it is defined as

$$CC(i) = \frac{2\Delta_i}{d_i(d_i - 1)}$$

where Δ_i is the number of triangles in G having node v_i as one of its vertices.

The clustering coefficient of a graph G is defined as the sum of the clustering coefficients over all the nodes of degree ≥ 2.

Many real world small-world networks, and particularly social networks, tend to have high clustering coefficient.

This is not the case for random networks. For example, Erdös–Renyi (ER) graphs. are small-world graphs but have very low clustering coefficients. Also, the degree distribution in ER graphs falls off exponentially (does not follow a power law).

Complex networks (cont.)

The number of triangles in G that a node participates in is given by

$$\Delta_i = \frac{1}{2}[A^3]_{ii},$$

while the total number of triangles in G is given by

$$\Delta(G) = \frac{1}{6}\mathrm{Tr}\,(A^3).$$

Here $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, i.e., the sum of the diagonal entries. Computing clustering coefficients for a graph G reduces to computing diagonal entries of A^3.

We note for many networks, A^3 is a rather dense matrix. For example, for the PPI network of beer yeast the percentage of nonzero entries in A^3 is about 19%, compared to 0.27% for A. This fact is related to the small-world property.

Complex networks (cont.)

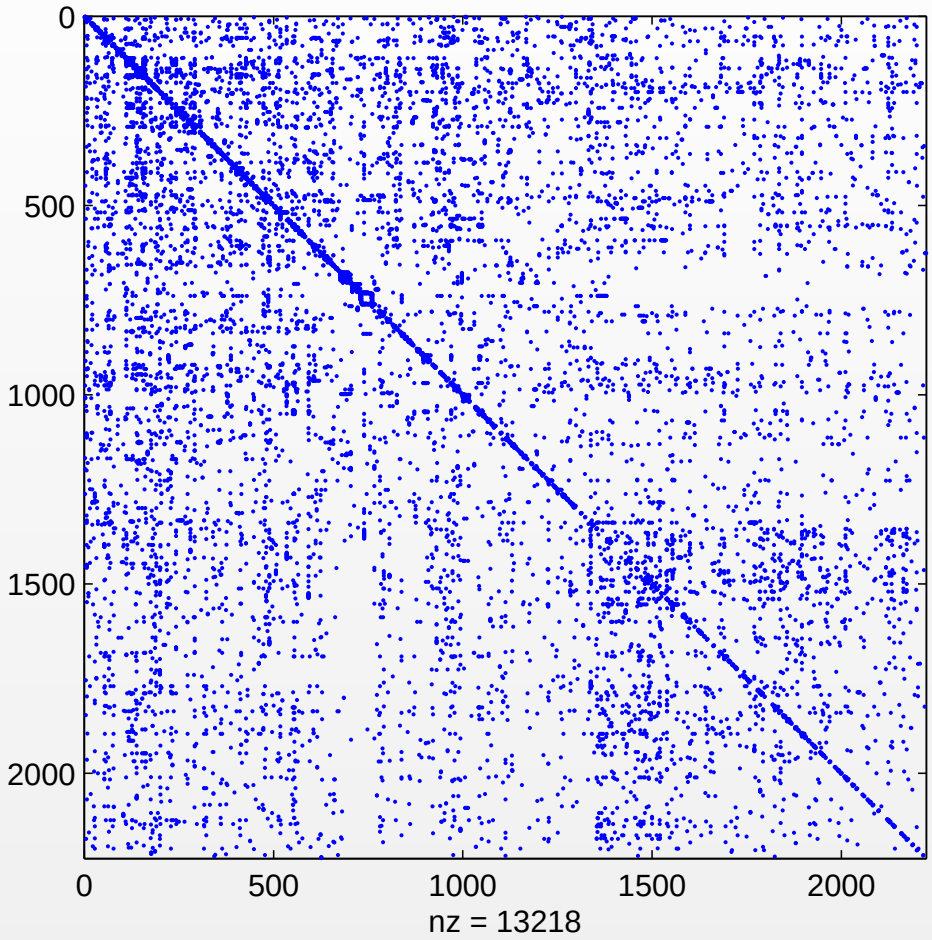
Summarizing: completely random graphs (like Erdös–Renyi graphs) are not scale-free and have low clustering coefficients. This makes them ill-suited as models of real-world complex networks.

The Watts–Strogatz (WS) model starts with a regular graph (say, a ring), which is then "perturbed" by rewiring some of the links between nodes in a randomized fashion. The WS model interpolates between a regular and a random graph model. With this technique, one can obtain small-world graphs with high clustering coefficients; the degree distribution, however, is rather homogeneous (i.e., WS graphs are not scale-free).

The Barabási–Albert (BA) model uses a preferential attachment, or rich get richer, mechanism to evolve a given initial graph. The resulting networks are small-world, scale-free, and have high clustering coefficients.

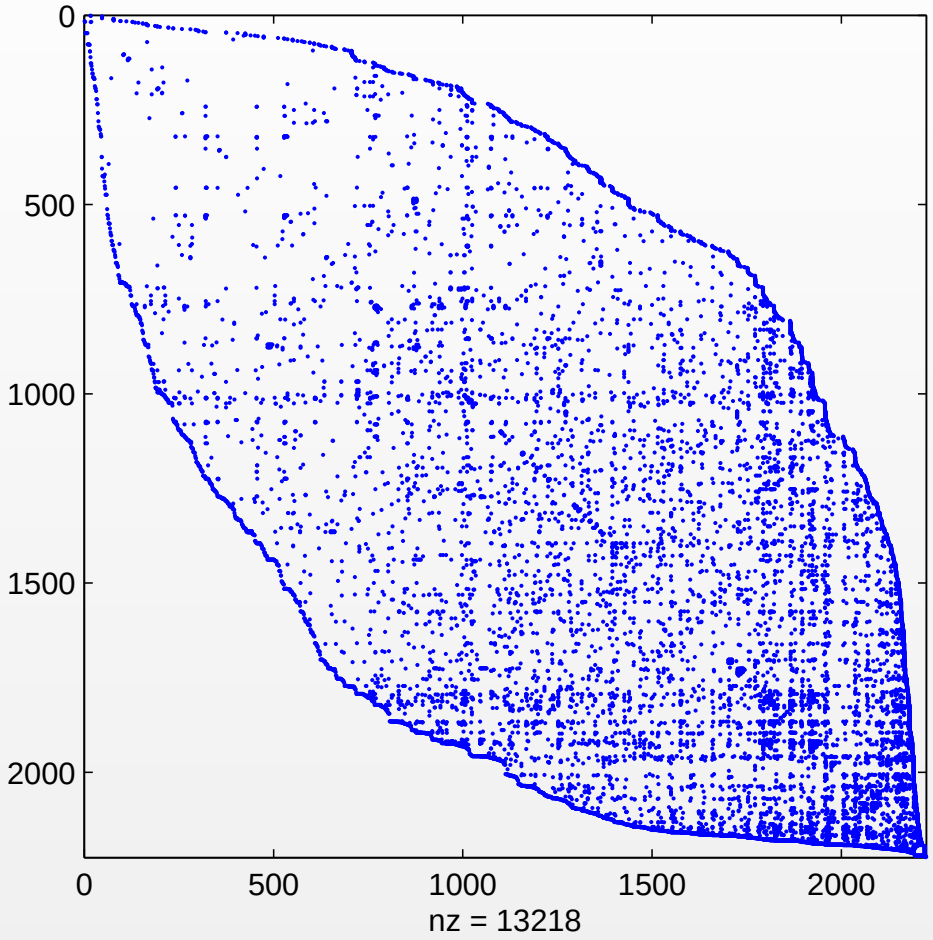
The study of generative models for constructing complex graphs with prescribed properties is still undergoing intensive development.

Example: PPI network of *Saccharomyces cerevisiae*



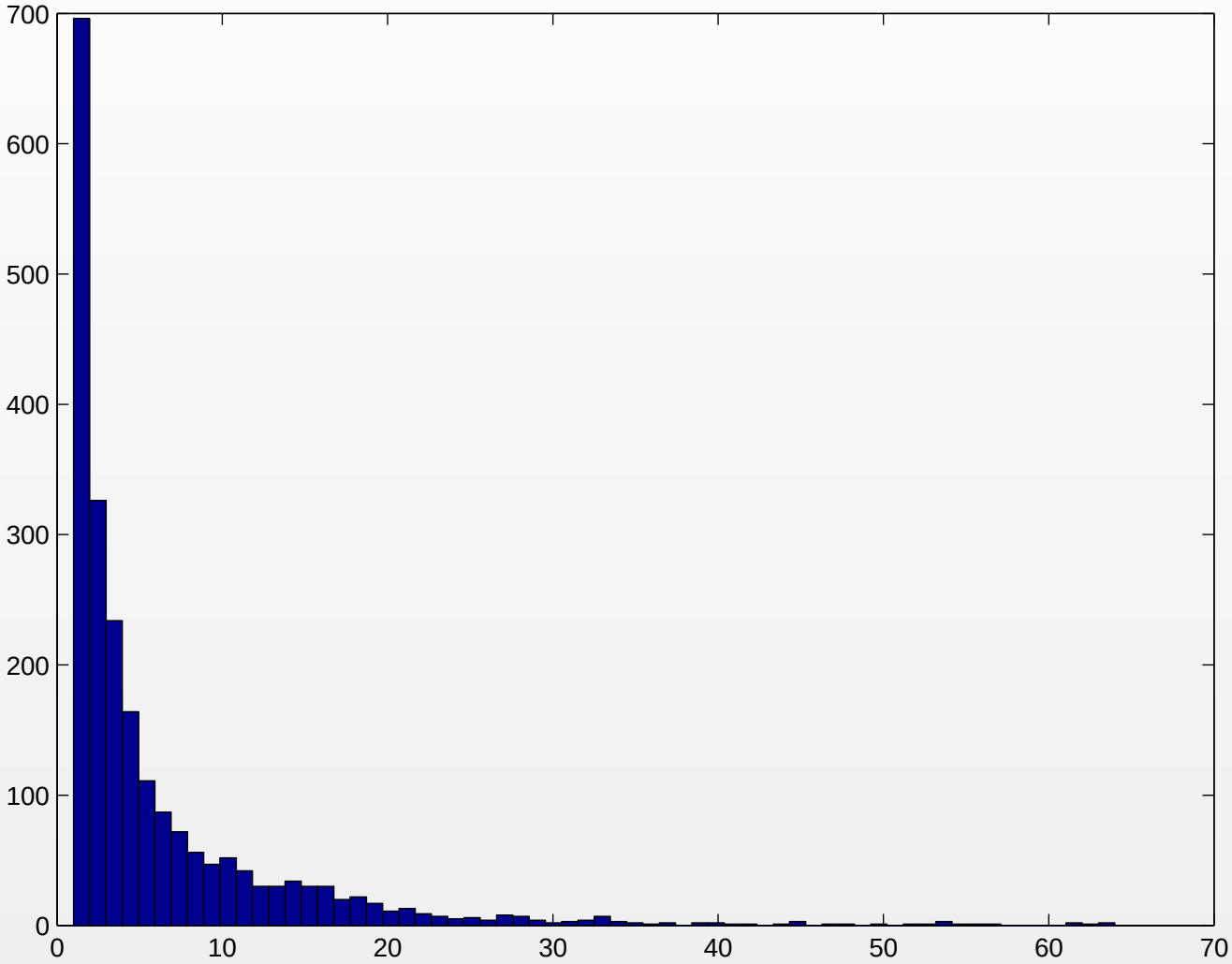
Adjacency matrix, $N = 2224$, $m = 6609$.

Example: PPI network of *Saccharomyces cerevisiae*



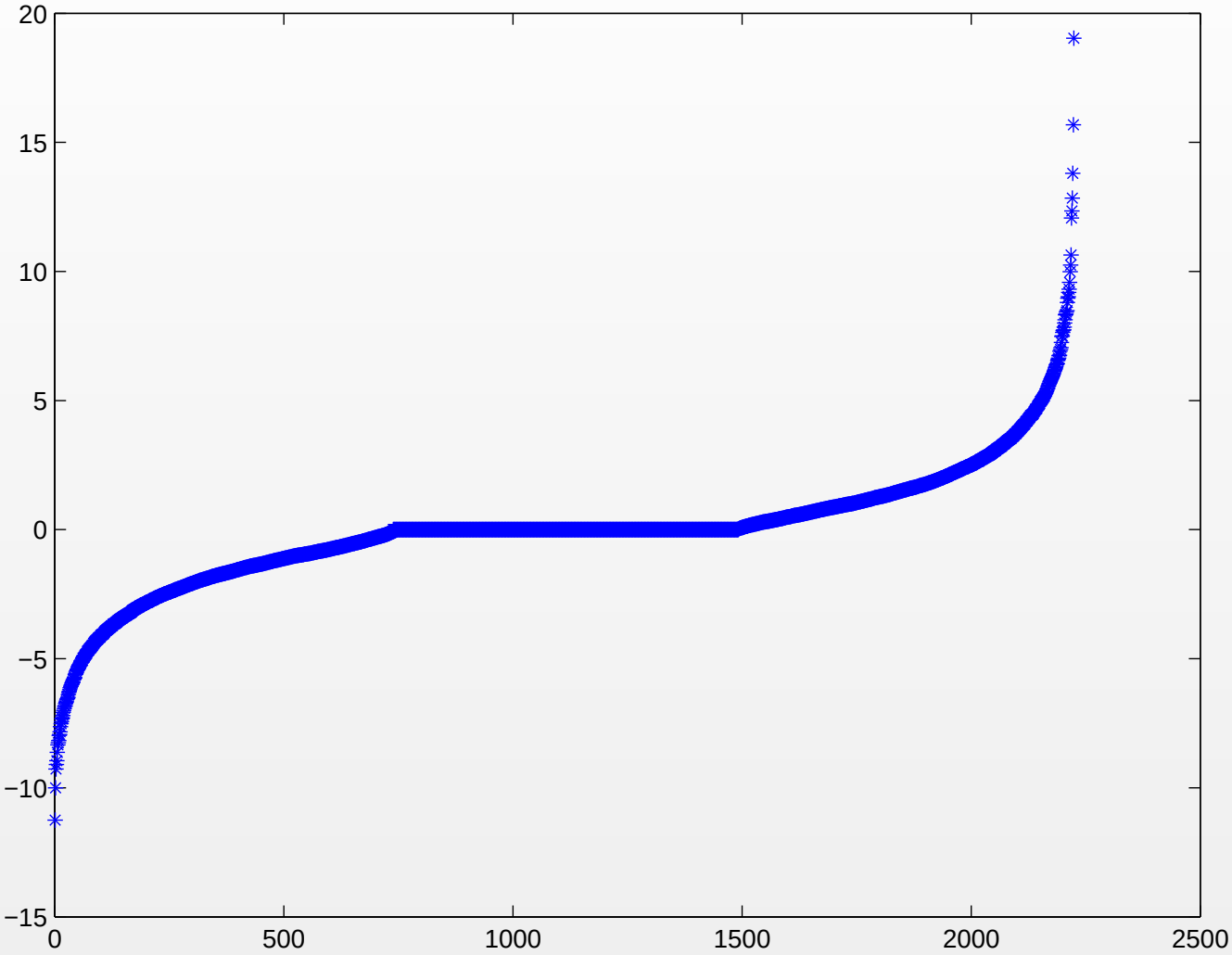
Same, reordered with Reverse Cuthill–McKee

Example: PPI network of *Saccharomyces cerevisiae*



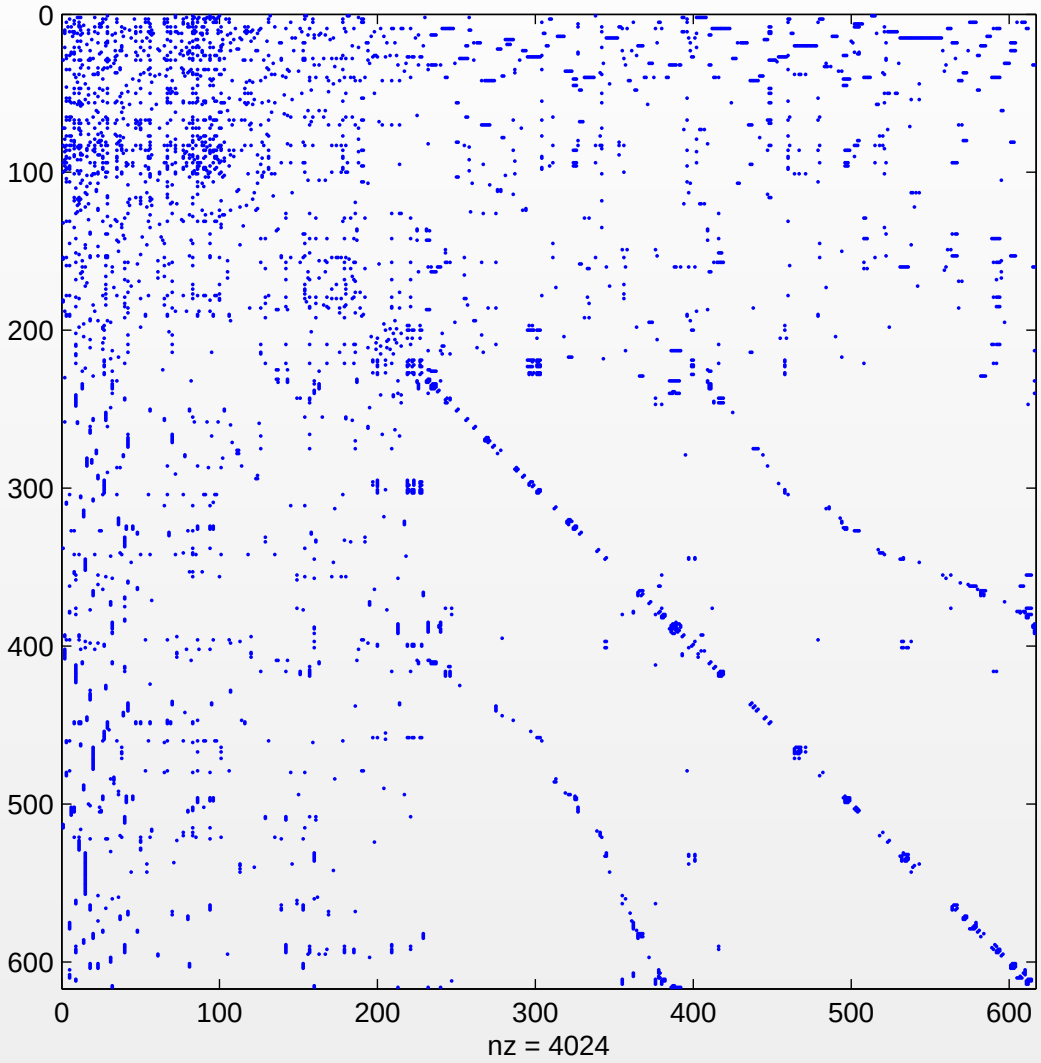
Degree distribution ($d_{\text{min}} = 1$, $d_{\text{max}} = 64$)

Example: PPI network of *Saccharomyces cerevisiae*



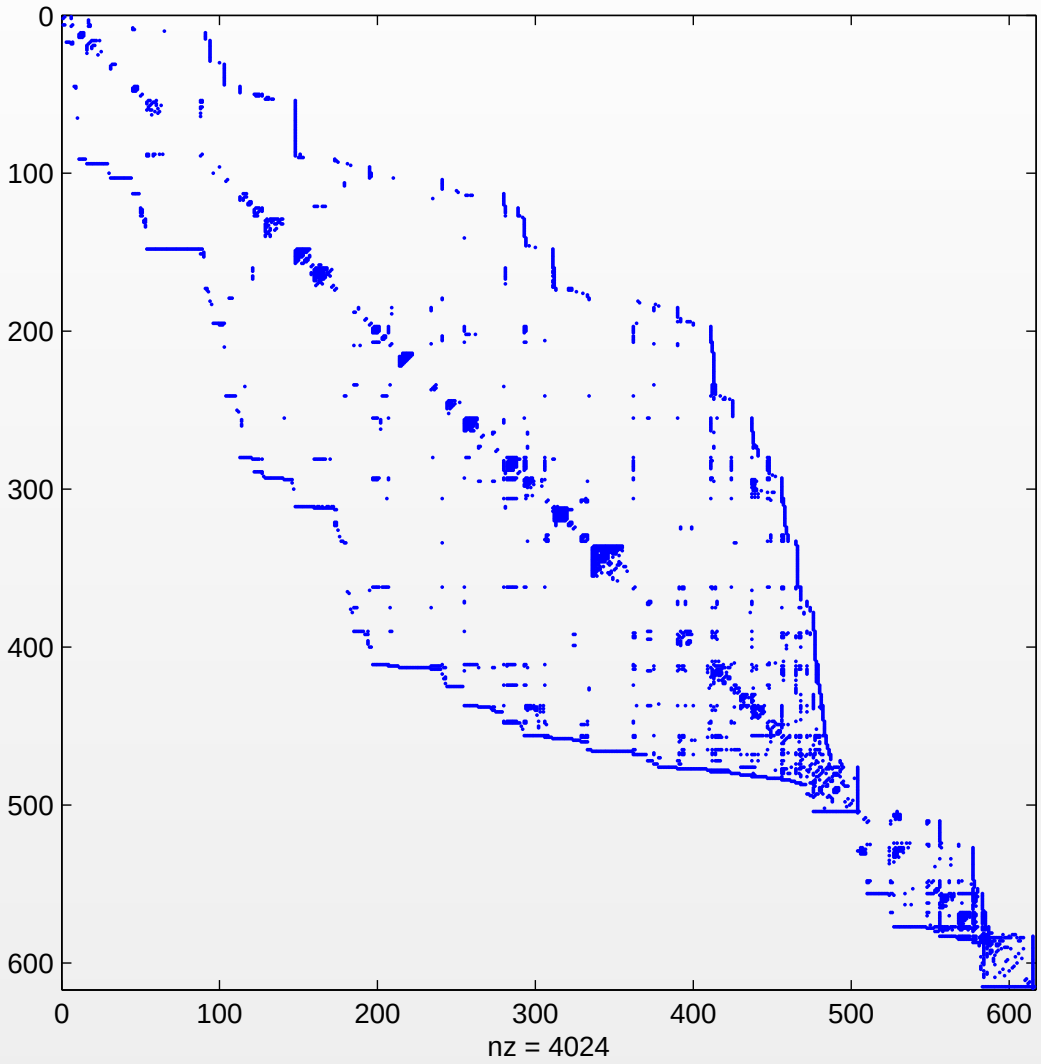
Eigenvalues

Example: Intravenous drug users network



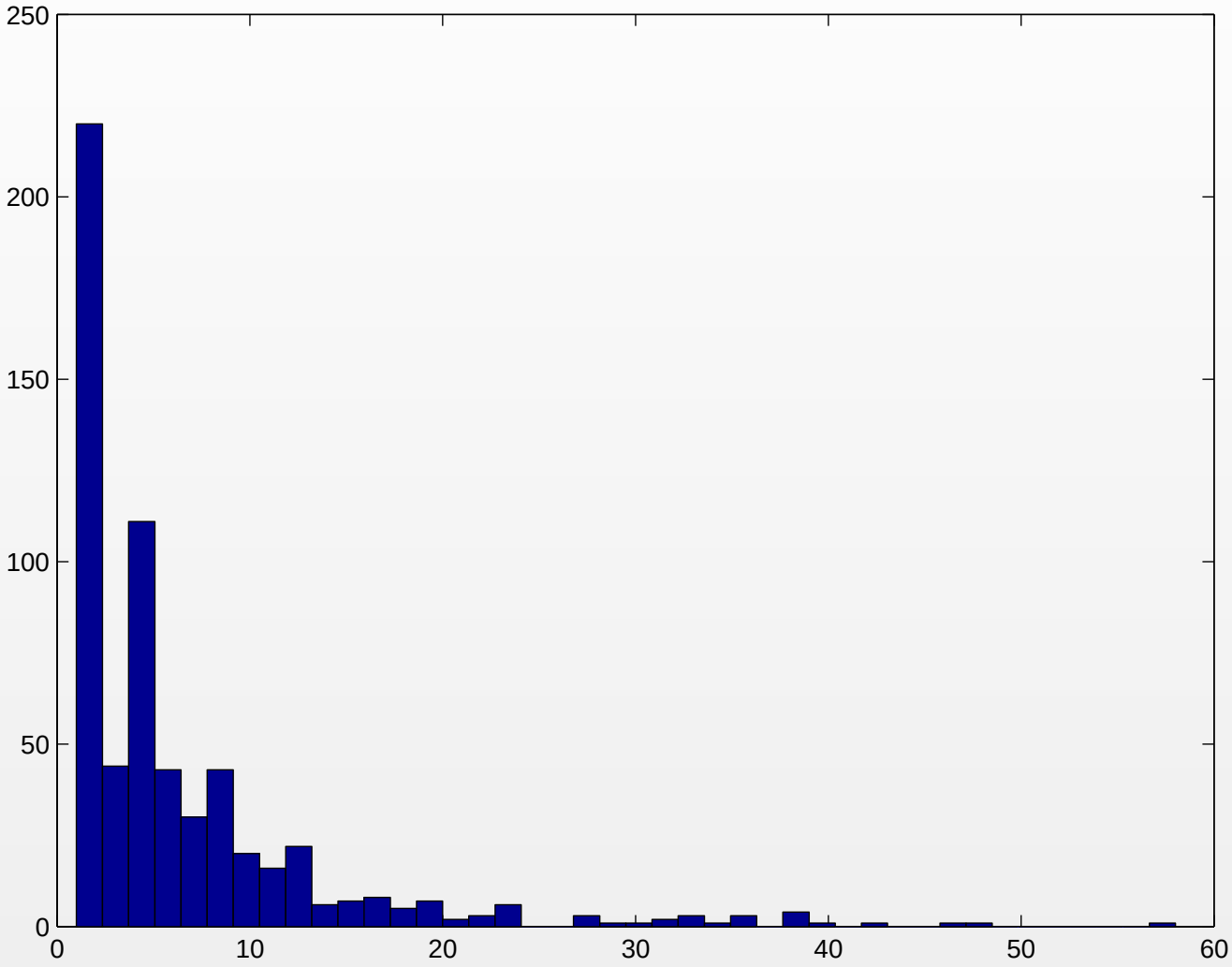
Adjacency matrix, $N = 616$, $m = 2012$.

Example: Intravenous drug users network



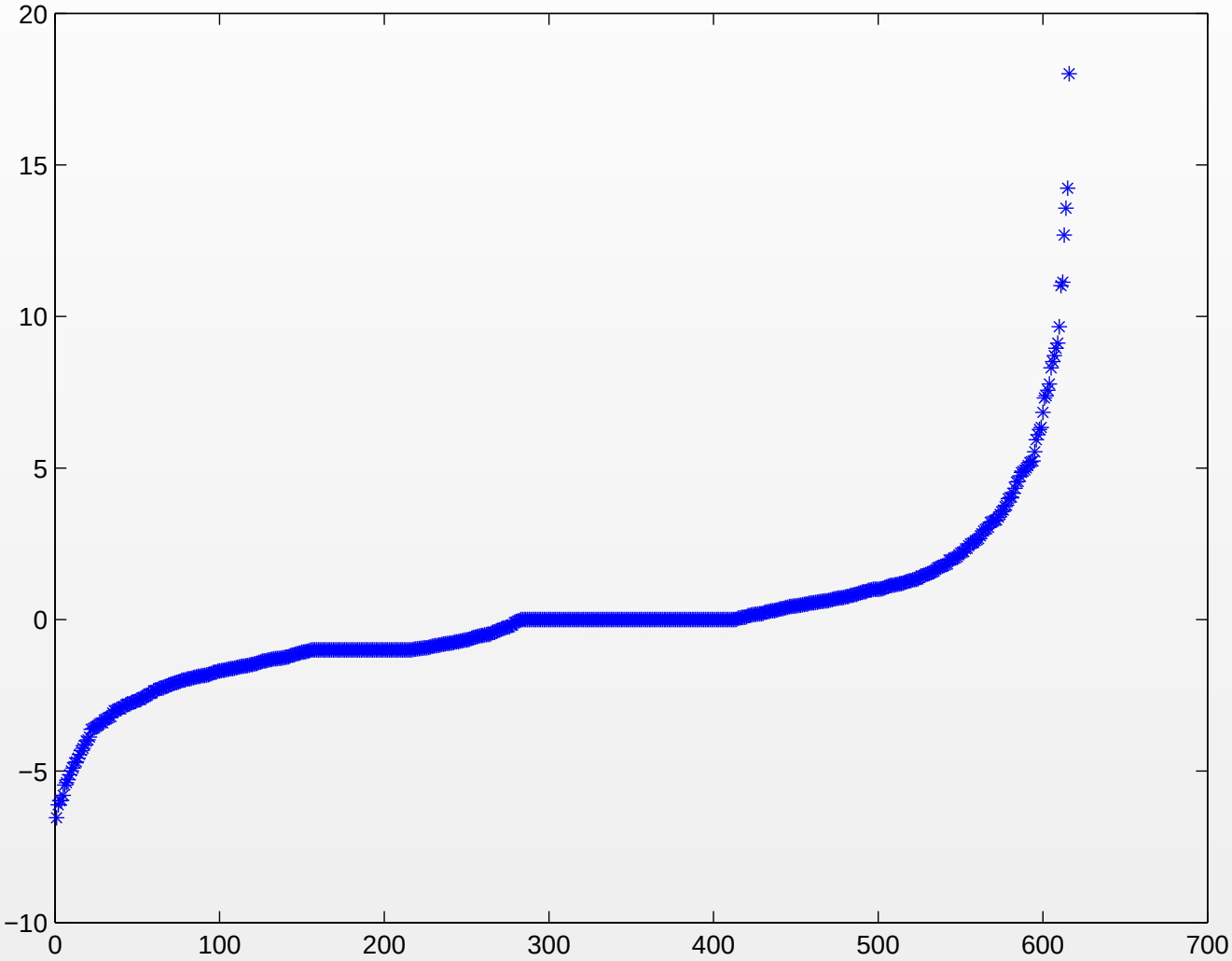
Same, reordered with Reverse Cuthill–McKee

Example: Intravenous drug users network



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

Example: Intravenous drug users network



Eigenvalues

Outline

Centrality

The centrality of a node is a measure of the importance of that node within the network.

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 v_i. This does not take into account the "importance" of the nodes a given nodes is connected to—only their number.

A popular notion of centrality is betweenness centrality (Freeman, 1977), defined for any node $v_i \in V$ as

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

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

$$\delta_{jk}(i) := \frac{\#\text{ of shortest paths between } v_j, v_k \text{ containing } v_i}{\#\text{ of shortest paths between } v_j, v_k} \,.$$

Centrality (cont.)

Betweenness centrality assumes that all communication in the network takes place via shortest paths, but this is often not the case.

This observation has motivated a number of alternative definitions of centrality, which aim at taking into account the global structure of the network and the fact that all walks between pairs of nodes should be considered, not just shortest paths.

We mention here that for directed graphs it is sometimes necessary to consider both hubs and authorities.

A crude measure of authoritativeness and "hubbiness" of a node v_i is provided by the in-degree d_i^{in} and by the out-degree d_i^{out} of the node, respectively. We will come back to this topic later on.

Spectral centrality measures

Eigenvector centrality (Bonacich, 1987) uses the entries of the dominant eigenvector (that is, the eigenvector $\mathbf{x}$ corresponding to the spectral radius of the adjacency matrix) to rank the nodes in the network in order of importance: the larger an entry x_i is, the more important node v_i is considered to be. By the Perron–Frobenius Theorem, this eigenvector is unique provided the network is connected.

The underlying idea is that "a node is important if it is linked to many important nodes." This *circular definition* corresponds to the fixed-point iteration

$$\mathbf{x}^{k+1} = A\mathbf{x}^k, \quad k = 0, 1, \ldots$$

which converges, under mild conditions, to the dominant eigenvector of A as $k \to \infty$. The rate of convergence depends on the spectral gap $\gamma = \lambda_1 - \lambda_2$. The larger γ, the faster the convergence.

Spectral centrality measures (cont.)

Google's PageRank algorithm (Brin and Page, 1998) is a variant of eigenvector centrality, applied to the (directed) graph representing web pages (documents), with hyperlinks between pages playing the role of directed edges.

Since the WWW graph is *not* connected, some tweaks are needed to have a unique PageRank eigenvector. The entries of this vector have a probabilistic interpretation in terms of random walks on the web graph, and their sizes determine the ranking of web pages. The PageRank vector is the unique stationary probability distribution of the Markov chain corresponding to the random walk on the Web graph.

An alternative approach (HITS), proposed by J. Kleinberg in 1999, uses the dominant left and right singular vectors of the (nonsymmetric) adjacency matrix of the graph in order to obtain both hub and authority scores. We will return to HITS in Lecture II.

Subgraph centrality

We now turn to centrality measures that are based on matrix functions.

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 v_i, with longer walks being penalized (given a smaller weight).

Subgraph centrality (cont.)

It is a well-known fact that

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

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

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

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

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

Subgraph centrality (cont.)

It is often convenient to normalize the subgraph centrality of a node by the sum

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

of all the subgraph graph centralities. The quantity $EE(G)$ is known as the Estrada index of the graph G. It is analogous to the partition function Z in statistical physics, and it plays an important role in the statistical mechanics of complex networks.

Note that the normalized subgraph centralities define a probability distribution $P(i) := SC(i)/EE(G)$ on the set V of nodes.

Katz centrality

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

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

Katz's original idea was to use the row sums of this matrix as centrality scores. These are just the components of the solution of the linear system $(I - cA)\mathbf{x} = \mathbf{1}$.

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. These are the row and column sums of the matrix resolvent $(I - cA)^{-1}$, respectively.

Other centrality measures

Similarly, one can use the row and column sums of other matrix functions, such as e^A, to rank nodes. This can be done without computing any entries of e^A, and is much faster than computing the diagonal entries of e^A (more on this later).

Systematic comparisons of various centrality measures can be found in:

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

M. Benzi and C. Klymko, *Total communicability as a centrality measure*, J. Complex Networks, in press (2013).

Comparing centrality measures

Suppose A is the adjacency matrix of an undirected, connected network. Let $\lambda_1 > \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_N$ be the eigenvalues of A, and let $\mathbf{x}_k$ be a normalized eigenvector corresponding to λ_k. Then

$$\mathrm{e}^A = \sum_{k=1}^{N} \mathrm{e}^{\lambda_k} \mathbf{x}_k \mathbf{x}_k^T$$

and therefore

$$SC(i) = \sum_{k=1}^{N} \mathrm{e}^{\lambda_k} x_{ki}^2 \,,$$

where x_{ki} denotes the ith entry of $\mathbf{x}_k$.

Analogous spectral formulas hold for other centrality measures based on matrix functions (e.g., Katz).

Comparing centrality measures (cont.)

It is clear that when the spectral gap is sufficiently large ($\lambda_1 \gg \lambda_2$), the centrality ranking is essentially determined by the dominant eigenvector $\mathbf{x}_1$, since the contributions to the centrality scores from the remaining eigenvalues/vectors becomes negligible, in relative terms.

Hence, in this case all these spectral centrality measures tend to agree, especially on the top-ranked nodes, and they all reduce to eigenvector centrality. Degree centrality is also known to be strongly correlated with eigenvector centrality in this case.

In contrast, the various measures can give significantly different results when the spectral gap is small or moderate. In this case, going beyond the dominant eigenvector can result in significant improvements.

More matrix functions

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

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

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

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, Linear Algebra Appl., 2009).

Outline

Communicability

Communicability measures how "easy" it is to send a message from node v_i to node v_j in a graph (Estrada & Hatano, Phys. Rev. E, 2008). It is defined as

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

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 (cont.)

The average communicability of a node is defined as

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

The total communicability centrality of a node

$$CC(i) := \sum_{j=1}^{N} C(i,j) = [\mathrm{e}^A \mathbf{1}]_i = \mathbf{e}_i^T \mathrm{e}^A \mathbf{1}$$

(Benzi and Klymko, 2013) attributes a high score to nodes that have high communicability with every node in the network, and can be used as a cheaper alternative to subgraph centrality. Note that if e^A is replaced with $(I - cA)^{-1}$, one recovers the definition of Katz centrality.

Communicability (cont.)

The row sum vector $\mathrm{e}^A\mathbf{1}$, and more generally of $f(A)\mathbf{1}$, can be efficiently computed using Krylov-based methods. The computation of $\overline{C(i)} = CC(i) - SC(i)$ is easy once $SC(i)$ has been estimated (see next lecture).

The (normalized) total network communicability

$$C(G) := \frac{1}{N}\sum_{i,j=1}^{N} C(i,j) = \frac{1}{N}\mathbf{1}^T\mathrm{e}^A\mathbf{1}$$

(Benzi and Klymko, 2013) provides a global measure of the flow of information in a network, and can be used to compare different network designs. It can be computed efficiently using Krylov-based methods even for large networks.

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

Communicability (cont.)

Estrada and collaborators have also introduced the scaled matrix exponential

$$\mathrm{e}^{\beta A} = I + \beta A + \frac{\beta^2}{2!}A^2 + \frac{\beta^3}{3!}A^3 + \cdots = \sum_{k=0}^{\infty} \frac{\beta^k}{k!}A^k.$$

Here β is a parameter that plays the role of an inverse temperature:

$$\beta = (\kappa_B T)^{-1},$$

where κ_B is the Boltzmann constant and T the absolute temperature.

Varying β allows for studying properties of networks subject to "external disturbances" such as increased tension in a social network, financial distress in the banking system, etc. Note that as the temperature increases ($\beta \to 0$) the communicability between nodes decreases ($\mathrm{e}^{\beta A} \to I$).

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

Communicability (cont.)

Communicability and centrality measures can also be defined using the negative graph Laplacian $-L = -D + A$ instead of the adjacency matrix A.

When the graph is regular (the nodes have all the same degree d, hence $D = dI$) nothing is gained by using $-L$ instead of A, owing to the obvious identity

$$\mathrm{e}^{-L} = \mathrm{e}^{-dI+A} = \mathrm{e}^{-d}\mathrm{e}^{A}.$$

Complex networks, however, have highly skewed degree distributions, and using e^{-L} instead of e^{A} can lead to substantial differences. Note that the interpretation in terms of weighted walks is no longer valid in this case.

Laplacian-based centrality and communicability measures are especially useful in the study of dynamic processes on graphs.

Outline

Differential equations

A fundamental problem in Network Science is the study of diffusion on graphs. In the absence of sources, it can be modeled by the simple first-order initial value problem

$$\frac{d\mathbf{u}}{dt} = -L\mathbf{u}, \quad t > 0; \quad \mathbf{u}(0) = \mathbf{u}_0,$$

where $\mathbf{u_0} \in \mathbb{R}^N$ is prescribed; for example, $\mathbf{u_0} = \mathbf{e}_i$ for some i.

The solution is given explicitly by $\mathbf{u}(t) = \mathrm{e}^{-tL}\mathbf{u}_0$, for all $t \geq 0$. The matrix function e^{-tL} is known as the heat kernel.

If G is connected then, regardless of the choice of $\mathbf{u_0}$, the solution $\mathbf{u}(t)$ tends to the uniform steady-state distribution $\mathbf{u}^\infty = c\mathbf{1}$ as $t \to \infty$, since

$$\mathbf{u}^\infty = \lim_{t\to\infty} \mathbf{u}(t) = \lim_{t\to\infty} \mathrm{e}^{-tL}\mathbf{u}_0 = P\mathbf{u}_0\,,$$

where P is the orthogonal projector onto $\mathcal{N}(L) = \mathrm{span}\{\mathbf{1}\}$. Note that $c = \mathbf{1}^T\mathbf{u}_0/N$, hence $\mathbf{u}^\infty = \mathbf{0}$ if and only if $\mathbf{u}_0 \perp \mathrm{span}\{\mathbf{1}\}$.

Differential equations (cont.)

Indeed, if the spectral decomposition of L is $L = \sum_{k=1}^{N} \mu_i \mathbf{q}_i \mathbf{q}_i^T$ with eigenvalues $0 = \mu_1 < \mu_2 \leq \cdots \leq \mu_N$ and eigenvectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N$, $\mathbf{q_1} = \mathbf{1}/\sqrt{N}$, then

$$\mathrm{e}^{-tL} = \sum_{k=1}^{N} \mathrm{e}^{-t\mu_i} \mathbf{q}_i \mathbf{q}_i^T = \mathbf{q}_1 \mathbf{q}_1^T + \sum_{k=2}^{N} \mathrm{e}^{-t\mu_i} \mathbf{q}_i \mathbf{q}_i^T \to \mathbf{q}_1 \mathbf{q}_1^T = P, \quad t \to \infty.$$

Note that, in contrast with the situation for regular graphs (like 2D or 3D grids), in a complex network the steady-state is often reached rather quickly, owing to the small-world (small diameter) property.

Suppose that the system is initially in the state described by $\mathbf{u}_0 = \mathbf{e}_j$ (the jth standard basis vector). Then the "temperature" of node i at time t is given by $[\mathrm{e}^{-tL}]_{ij}$, for all $i = 1 : N$. This fact provides a useful "physical" interpretation of time-dependent, Laplacian-based communicability.

Differential equations (cont.)

Also of interest in Network Science is the Schrödinger equation on a graph:

$$i\frac{d\boldsymbol{\psi}}{dt} = L\boldsymbol{\psi}, \quad t \in \mathbb{R}, \quad \boldsymbol{\psi}(0) = \boldsymbol{\psi}_0,$$

where $\boldsymbol{\psi}_0 \in \mathbb{R}^N$ is a prescribed initial state with $\|\boldsymbol{\psi}_0\| = 1$.

The solution is given esplicitly by $\boldsymbol{\psi}(t) = \mathrm{e}^{-itL}\boldsymbol{\psi}_0$, for all $t \in \mathbb{R}$; note that since itL is skew-Hermitian, the propagator $U(t) = \mathrm{e}^{-itL}$ is a unitary matrix, which guarantees that the solution has unit norm for all t:

$$\|\boldsymbol{\psi}(t)\| = \|U(t)\boldsymbol{\psi}_0\| = \|\boldsymbol{\psi}_0\| = 1, \quad \forall t \in \mathbb{R}.$$

The amplitudes $|[U(t)]_{ij}|^2$ express the probability to find the system in state $\mathbf{e}_i$ at time t, given that the initial state was $\mathbf{e}_j$.

The Schrödinger equation has recently been proposed as a tool to analyze properties of complex graphs in P. Suau, E. R. Hancock and F. Escolano, *Graph characteristics from the Schrödinger operator*, LNCS 7877 (2013), pp. 172–181.

Summary

- Matrix functions can be used to analyze the structural properties of networks, and to describe dynamical processes taking place on them;
- The matrix resolvent and the matrix exponential are especially important in Network Science;
- In many applications, one needs to compute both expressions of the form $\mathbf{u} = f(A)\mathbf{v}$ and scalar quantities like $\mathbf{u}^T f(A)\mathbf{v}$, including individual entries $[f(A)]_{ij} = \mathbf{e}_i^T f(A)\mathbf{e}_j$;
- The diagonal entries of $f(A)$ are especially valuable;
- The study of dynamic processes on networks often involves the exponential of the graph Laplacian and related matrix functions.