

# Centrality and Communicability Measures in Complex Networks: Mathematical and Computational Aspects, I

Michele Benzi

Department of Mathematics and Computer Science  
Emory University

Atlanta, Georgia, USA

International Summer School on Complex Networks

Bertinoro, Italy  
14-18 July, 2014

- 1 Complex networks: motivation and background
- 2 Network structure: centrality and communicability measures
- 3 Bibliography

# Acknowledgements

- Thanks to the organizers
- Collaborators:
  - ▶ Paola Boito (Limoges)
  - ▶ Ernesto Estrada (Strathclyde)
  - ▶ Christine Klymko (Livermore)

- 1 Complex networks: motivation and background
- 2 Network structure: centrality and communicability measures
- 3 Bibliography

# Complex networks: motivation and background

Networks, in particular **complex networks**, provide models for a wide variety of physical, biological, engineered or social systems.

For example: molecular structure, gene and protein interaction, anatomical and metabolic networks, food webs, transportation networks, power grids, financial and trade networks, social networks, the Internet, the WWW, Facebook, Twitter ...

**Network Science** is the study of networks, both as mathematical structures and as concrete, real world objects. It is a **growing multidisciplinary field**, with important contributions not just from mathematicians, computer scientists and physicists but also from social scientists, biologists, public health researchers and even from scholars in the humanities.

## Complex networks: motivation and background (cont.)

The field has its origins in the work of psychologists, sociologists, economists, anthropologists and statisticians dating back to the late 1940s and early 1950s. In the last 15 years or so, physicists, computer scientists and mathematicians have entered the scene and made the field more mathematically sophisticated.

Basic tools for the analysis of networks include [graph theory](#), [linear algebra](#), [probability](#), [numerical analysis](#), and of course [algorithms and data structures](#) from discrete mathematics. More advanced techniques include [statistical mechanics](#) and [multilinear algebra](#).

The study of [dynamical processes](#) often leads to [differential equations](#) posed on networks.

Some classic early references:

- J. R. Seeley, *The net of reciprocal influence: A problem in treating sociometric data*, Canadian J. Psychology, 3 (1949), pp. 234–240.
- 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.

## Two pioneers of network science



Anatol Rapoport (1911-2007) and Stanley Milgram (1933-1984)

## Basic references (cont.)

The field exploded in the late 1990s due to several breakthroughs by [physicists](#), [applied mathematicians](#) and [computer scientists](#).

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.

# Examples of complex networks

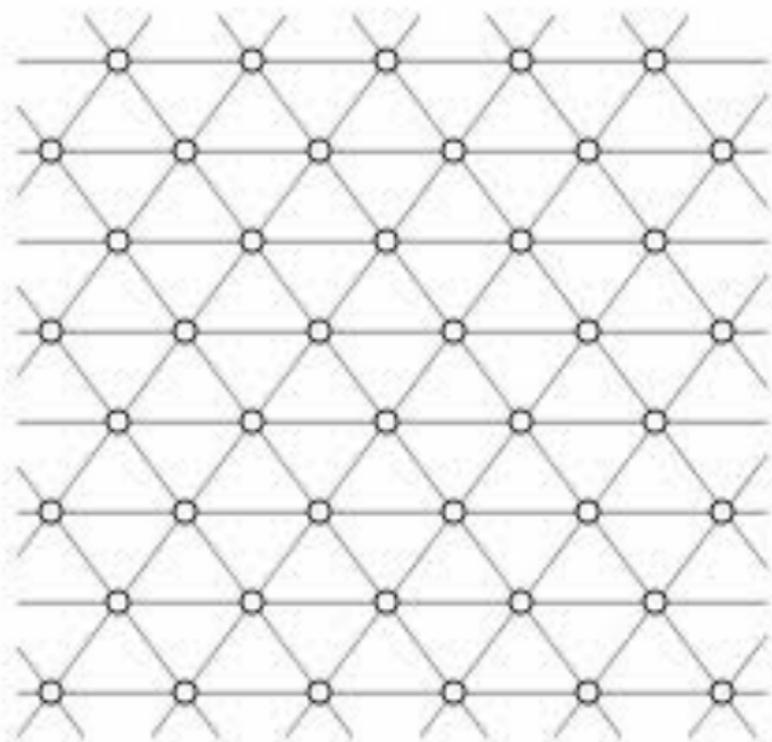
But what exactly **is** a complex network?

Unfortunately, **no precise definition** exists, although there is some ongoing work on characterizing (and quantifying) the degree of complexity in a network.

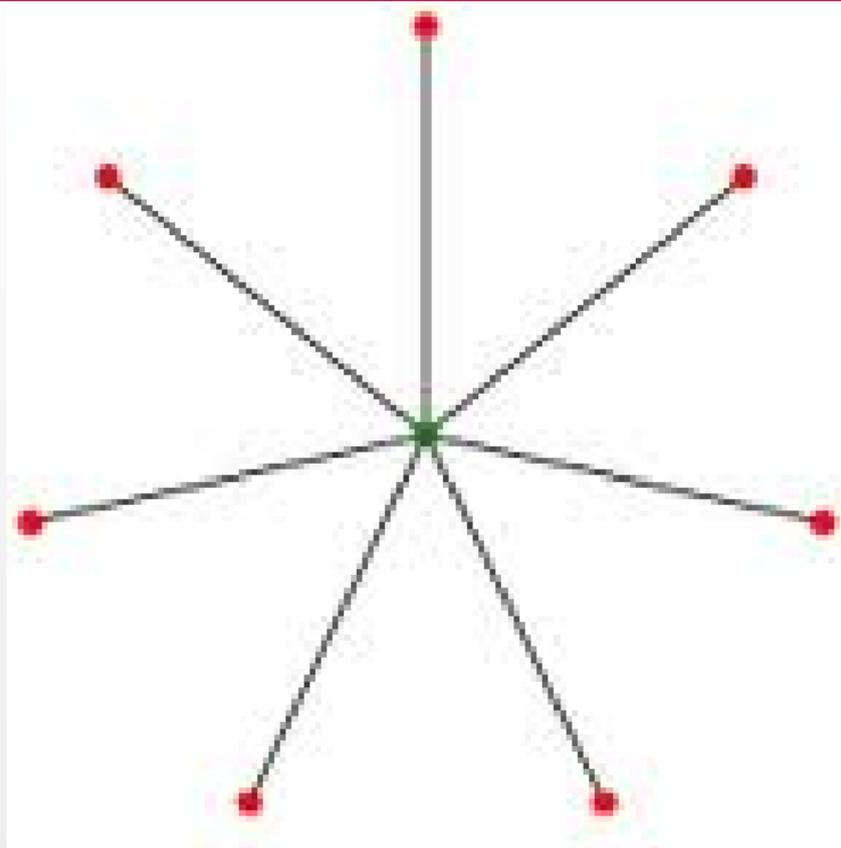
It is easy to tell which graphs are **not** complex networks. Regular lattices are not considered complex networks, and neither are completely random graphs such as the **Erdős–Rényi** model.

Random graphs, however, are useful as **null models** against which to compare (possible examples of) complex networks.

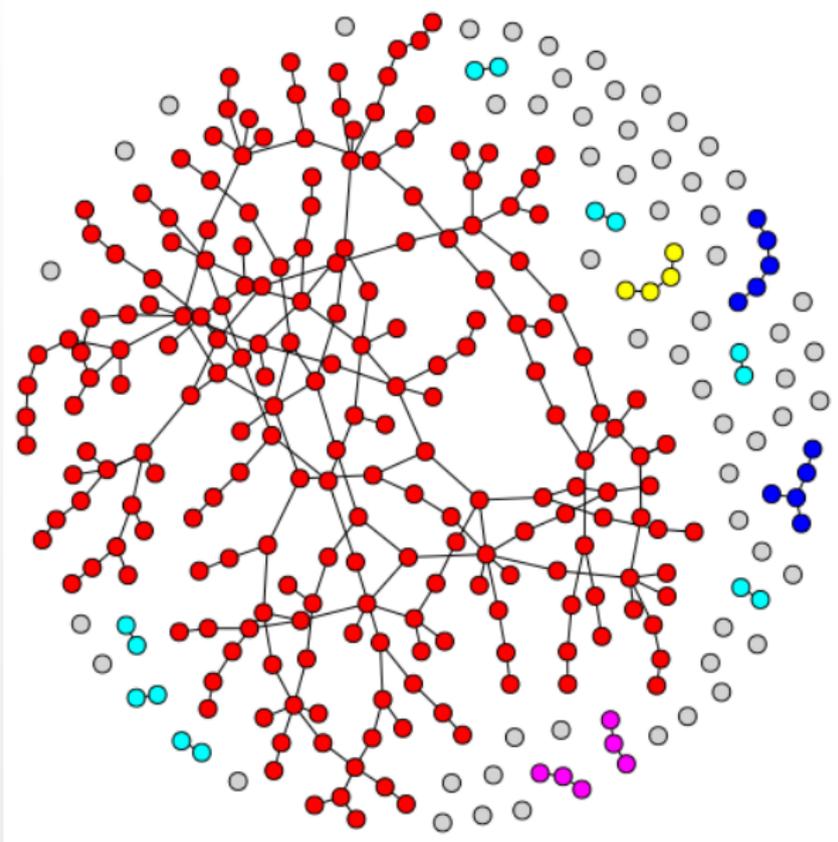
## Regular lattice: not a complex network!



## Star graph: not a complex network!



# Erdős–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": the degree distribution follows a power law (Pareto's curve)
- "Small-world":
  - ▶ Small graph diameter, short average distance between nodes
  - ▶ High clustering coefficient: many triangles, hubs, ...
- Hierarchical structure
- Rich in "motifs"
- Self-similar (as in fractals)

Briefly stated: complex networks exhibit a **non-trivial topology**.

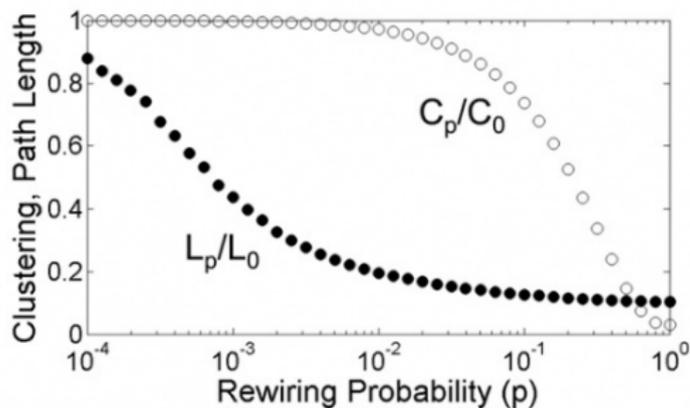
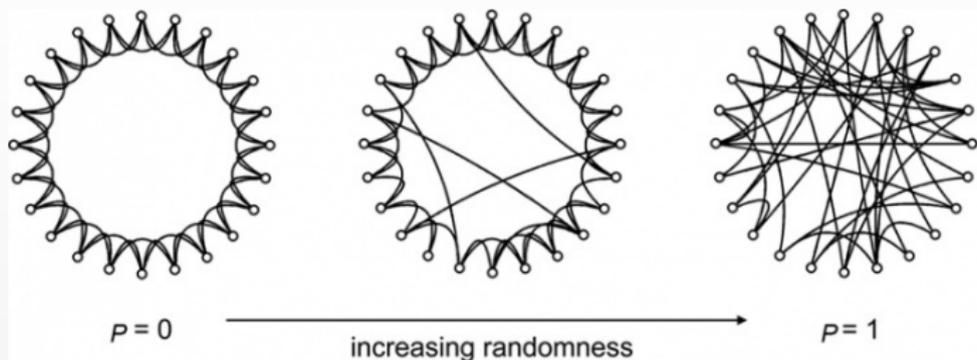
**Caveat:** there are important examples of real-world complex networks lacking one or more of these attributes.

What a small world...!

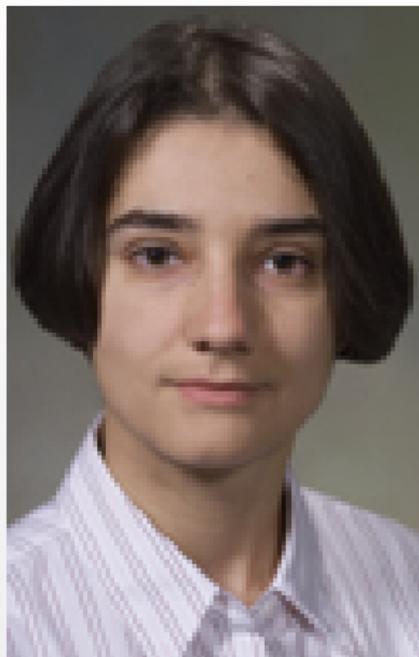


Duncan Watts and Steven Strogatz

# The Watts–Strogatz random rewire model



The rich always get richer!



Albert-László Barabási and Réka Albert

# The Barabási–Albert model

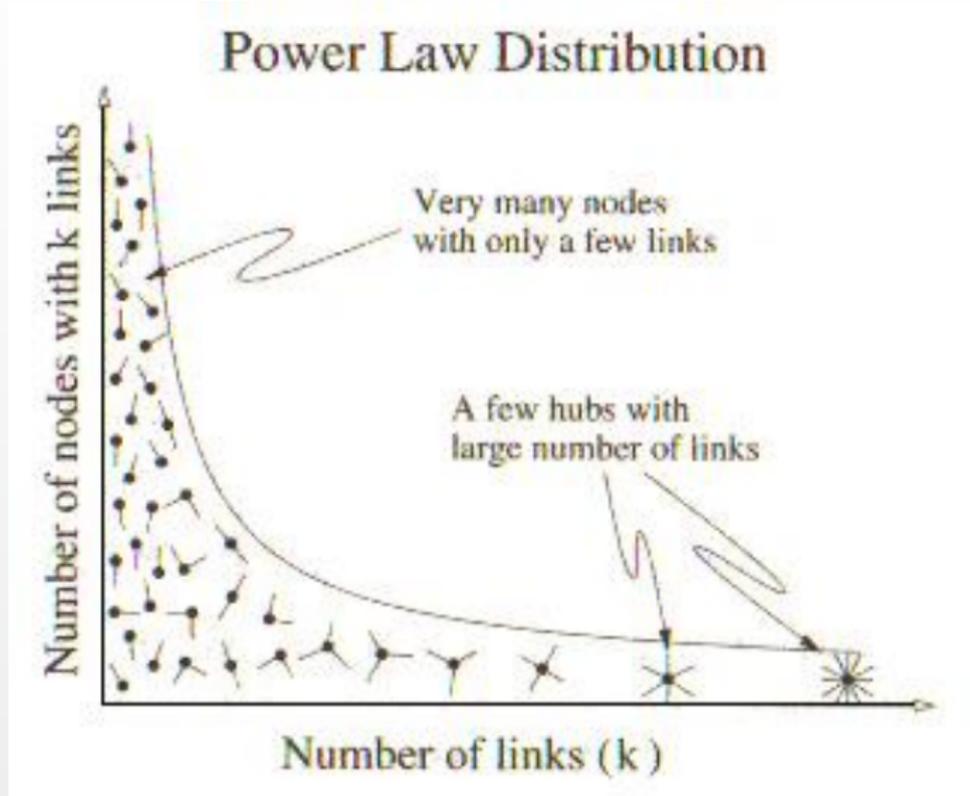
The Barabási–Albert model is based on the notion of **preferential attachment**: starting from a given sparse graph, new nodes are added (one at a time) to the network, joining them to existing nodes with a **probability proportional to the degree** (the number of adjacent nodes) of such nodes. Hence, nodes that are rich of connections to other nodes have a high probability of attracting new “neighbors”, while “poor” nodes tend to remain such.

One can show that the Barabási–Albert model leads to a **highly skewed** degree distribution given by a power law of the form

$$p(k) \propto k^{-\gamma},$$

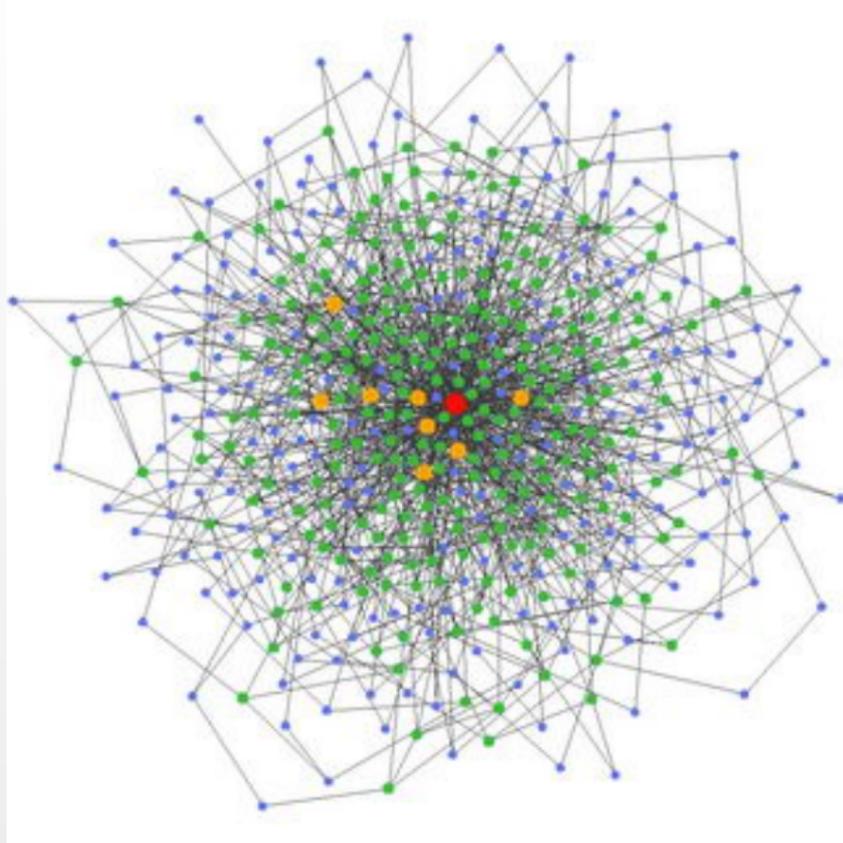
where  $p(k)$  denotes the fraction of nodes of degree  $k$  and  $\gamma$  is a constant, typically with  $2 \leq \gamma \leq 3$ .

# The Barabási–Albert model (cont.)



Degree distribution according to a power law

## The Barabási–Albert model (cont.)



## The Barabási–Albert model (cont.)

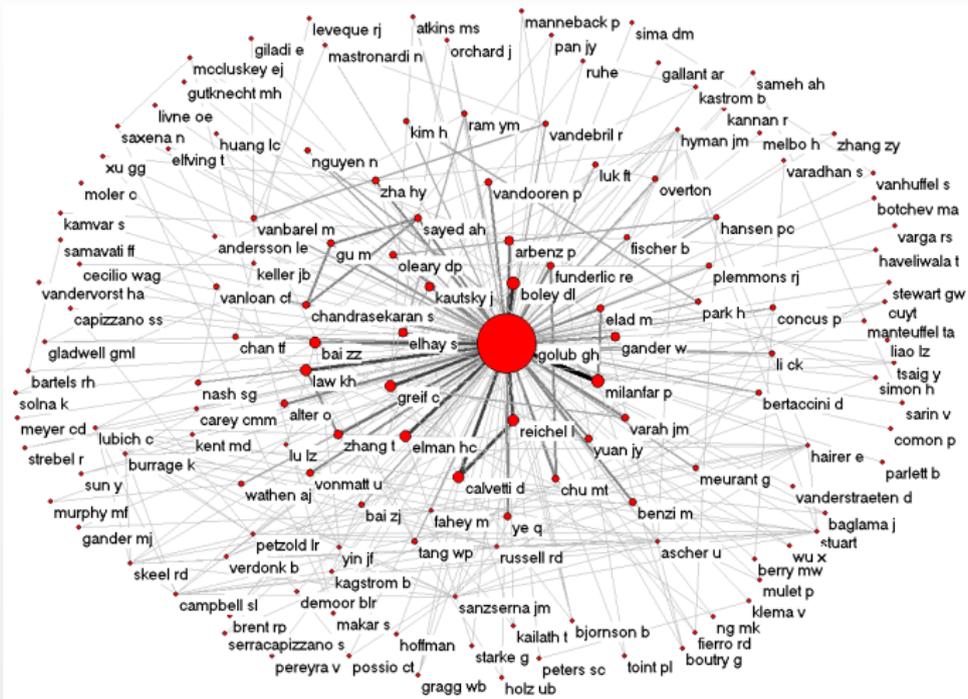
The Barabási–Albert model produces “small world” networks, in which the diameter grows slowly with the number of nodes; the growth is generally logarithmic or even log-logarithmic (in practice, the diameter remains nearly constant).

Examples: [Facebook](#); [MathSciNet](#) collaboration graph.

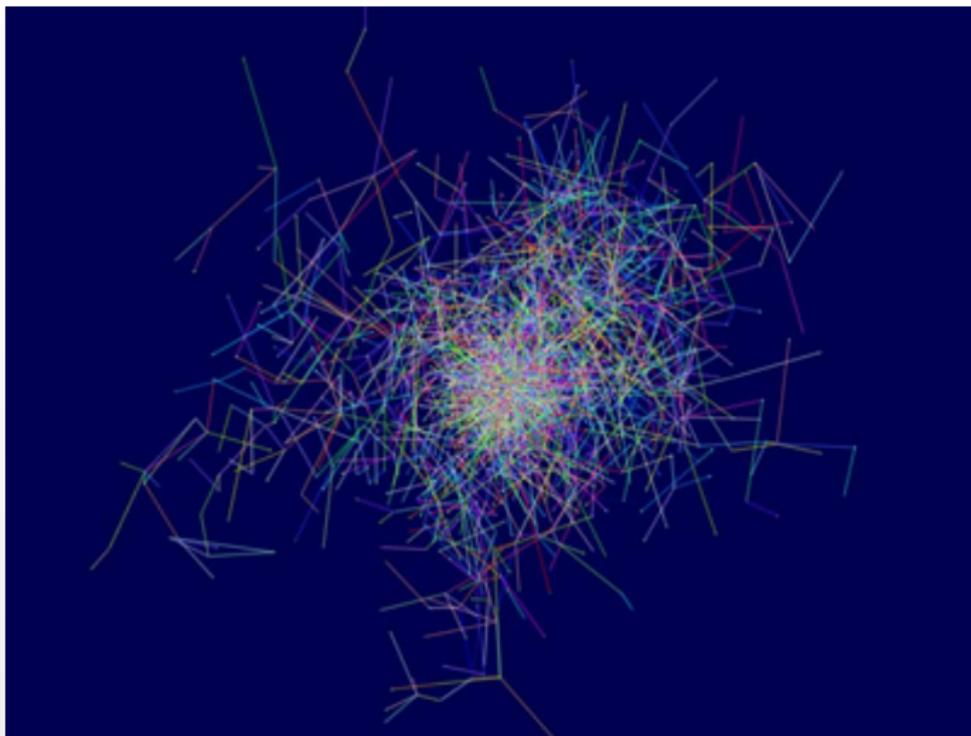
Unlike the Barabási–Albert model, the Erdős–Rényi and Watts–Strogatz models exhibit degree distributions that decay exponentially from the mean  $\langle k \rangle$ : hence, such networks are not “scale-free”.

An overwhelming number of real-world complex networks appear to follow a power law degree distribution, at least approximately.

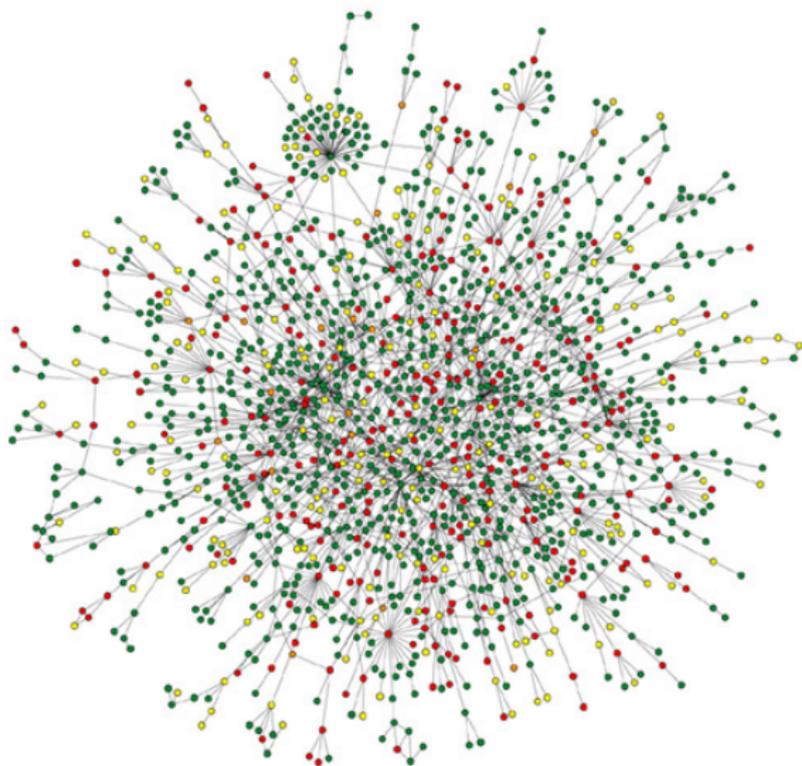
# Example of complex network: Golub collaboration graph



## Example of complex network: Erdős 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

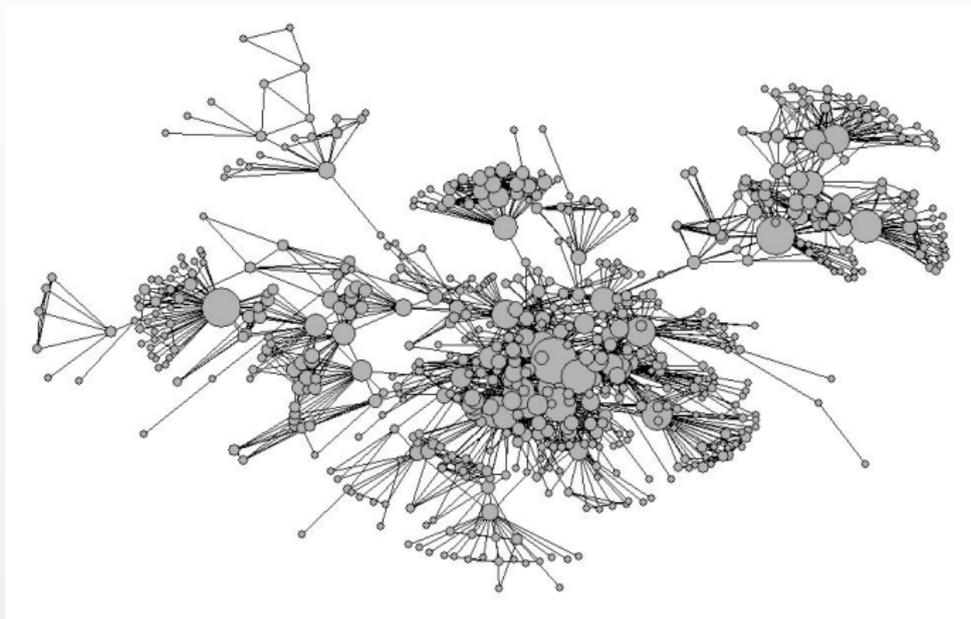
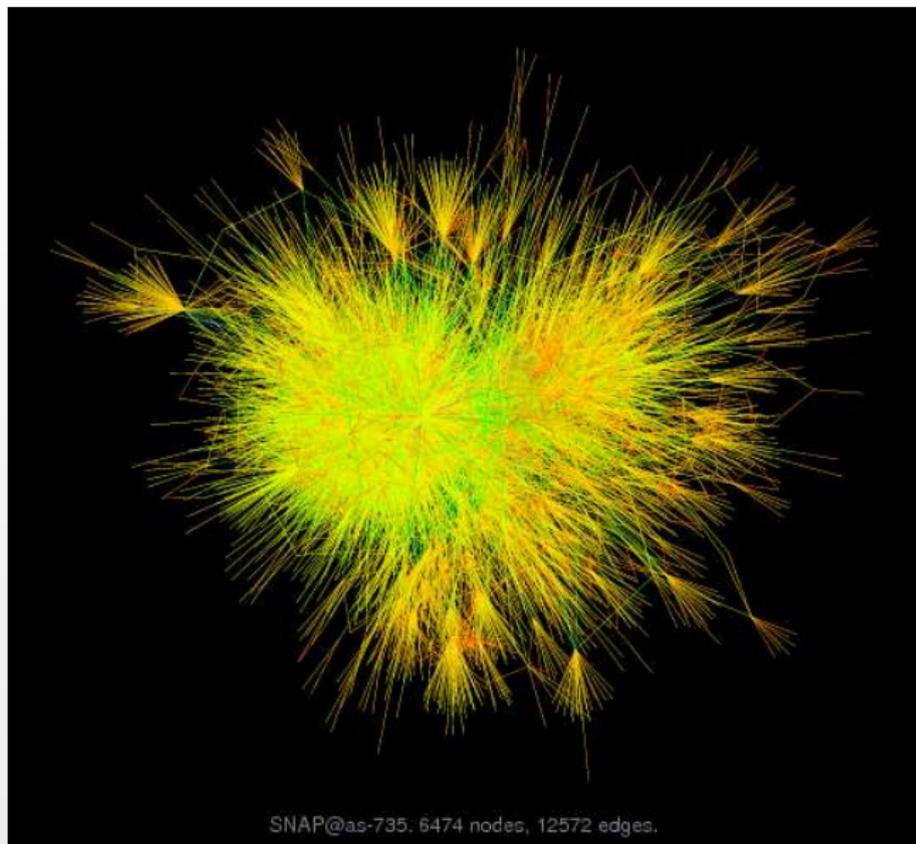
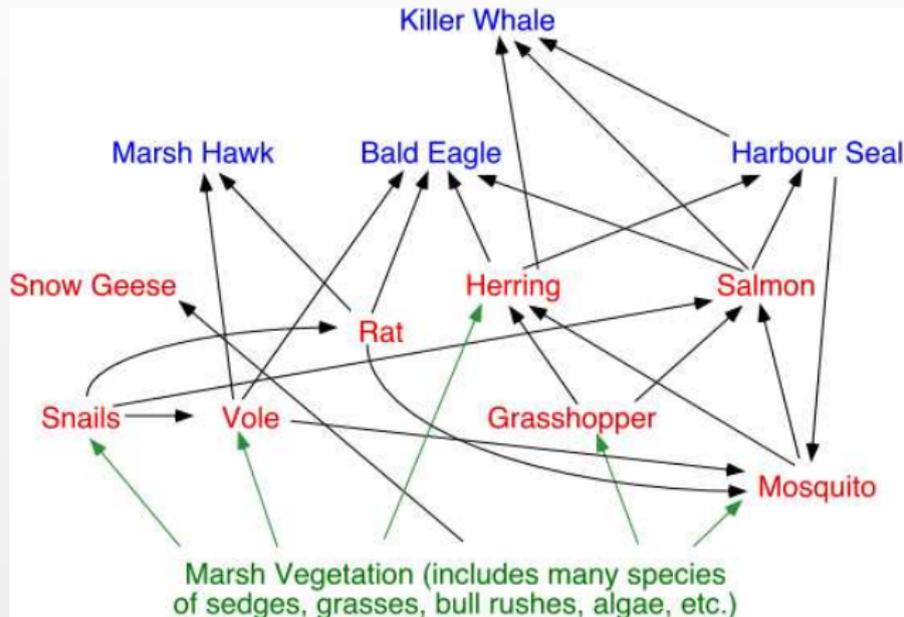


Figure courtesy of Ernesto Estrada.

## Example of complex network: a chunk of the Internet



# Example of (directed) complex network: a food web



# Network analysis

Basic questions about network structure include **centrality**, **robustness**, **communicability** and **community detection** issues:

- Which are the most “important” nodes?
  - ▶ Network connectivity and robustness/vulnerability
  - ▶ Identification of influential individuals in social networks
  - ▶ Essential proteins in PPI networks (lethality)
  - ▶ Identification of keystone species in ecosystems
  - ▶ Author centrality in collaboration networks
  - ▶ Ranking of documents/web pages on a given topic
- How do “disturbances” spread in a network?
  - ▶ Spreading of epidemics, beliefs, rumors, fads,...
  - ▶ Routing of messages; bottlenecks, returnability
- How to identify “community structures” in a network?
  - ▶ Clustering, triadic closure (transitivity)
  - ▶ Partitioning

# Formal definitions

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

A **graph**  $G = (V, E)$  consists of a (finite) set  $V = \{v_1, v_2, \dots, v_N\}$  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  $v_{i_1}, v_{i_2}, \dots, v_{i_k}, v_{i_{k+1}}$  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  $A \rightarrow PAP^T$  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:

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ . Note that  $A$  is always indefinite if  $E \neq \emptyset$ .

If  $G$  is connected, then  $\lambda_1$  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  $\lambda_1$ .

## 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 the number  $d_i^{out}$  of edges originating at  $v_i$ .

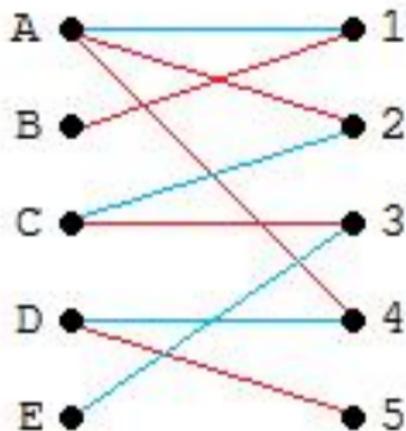
In terms of the (nonsymmetric) adjacency matrix,

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

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  $\sigma_i$  denote the **singular values** of  $B$ . [▶ More](#)

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  $V' = \{v'_1, \dots, v'_N\}$  of  $V = \{v_1, \dots, v_N\}$  and define a new, undirected graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with  $2N$  nodes, where  $\mathcal{V} := V \cup V'$  and

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

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 Part II.

# Typical features of 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 either highly regular, or completely “random”.

Complex networks are very often

- **Scale-free**, meaning that their degree distribution tends to follow a **power law**:  $p(k) = \text{number of nodes of degree } k \approx c \cdot k^{-\gamma}$ ,  $\gamma > 0$ . Frequently,  $2 \leq \gamma \leq 3$ . 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.,

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

- **Highly clustered**, i.e., they contain a very large proportion of **triangles** (unlike random graphs).

# Clustering

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  $\Delta_i$  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 average of the clustering coefficients over all the nodes of degree  $\geq 2$ .

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, Erdős–Rényi (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).

## Clustering (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

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

Hence, computing clustering coefficients for a graph  $G$  requires estimating  $\text{Tr}(A^3)$ , which for very large networks can be a challenging task.

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.

# Summary of complex networks characteristics

Summarizing: completely random graphs (like [Erdős–Rényi](#) 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.

# Graph spectra

The adjacency matrix  $A$  of an **undirected** network is always symmetric, hence its eigenvalues are all real. The **eigenvalue distribution** reflects **global properties** of  $G$ ; as we shall see, the **eigenvectors** also carry important information about the network structure.

The **spectrum** of a network is the spectrum of the corresponding adjacency matrix  $A$ .

An entire field of mathematics (**Spectral Graph Theory**) is devoted to the study of the eigenvalues and eigenvectors of graphs. Much work has been done in characterizing the spectra (eigenvalue distributions) of random graphs and of certain classes of complex graphs (e.g., **scale-free graphs**).

The adjacency matrix of a **directed** network, on the other hand, is typically nonsymmetric and will have complex (non-real) eigenvalues in general. In this case, the **singular values** of  $A$  are often useful.

P. Van Mieghem, *Graph Spectra for Complex Networks*, Cambridge University Press, 2011.

## Graph spectra (cont.)

Recall the classical

**Perron–Frobenius Theorem:** Let  $A \in \mathbb{R}^{N \times N}$  be a matrix with nonnegative entries. If  $A$  is irreducible, the **spectral radius** of  $A$ , defined as

$$\rho(A) = \max \{ |\lambda| : \lambda \in \sigma(A) \}$$

is a simple eigenvalue of  $A$ . Moreover, there exists a unique positive eigenvector  $\mathbf{x}$  associated with this eigenvalue:

$$A\mathbf{x} = \rho(A)\mathbf{x}, \quad \mathbf{x} = (x_i), \quad x_i > 0 \quad \forall i = 1 : N.$$

Note that the theorem applies, in particular, to the adjacency matrix of a connected network. This theorem is of fundamental importance, and it provides the basis for the notion of **eigenvector centrality** and for the celebrated **Google's PageRank algorithm**.

A. Langville and C. Meyer, *Google's PageRank and Beyond*, Princeton University Press, 2006.

# Perron & Frobenius



Oskar Perron (1880-1975) and Georg Ferdinand Frobenius (1849-1917)

Geršgorin's Theorem [▶ More](#) yields a simple bound for  $\lambda_{\max}(A)$ :

$$\lambda_{\max}(A) \leq d_{\max} := \max_{1 \leq i \leq N} d_i.$$

The bound is attained if and only if  $G$  is regular. If  $G$  is also bipartite, then we have  $\lambda_{\min}(A) = -d_{\max}$ .

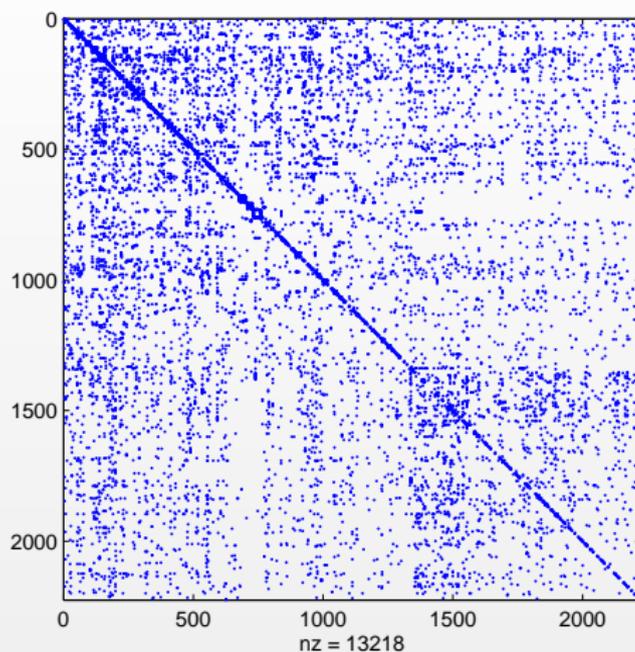
For [directed graphs](#), the Geršgorin bound becomes

$$\lambda_{\max}(A) \leq \min\{d_{\max}^{in}, d_{\max}^{out}\}.$$

A large body of work exists on the distribution of the eigenvalues of various types of graphs, such as Erdős-Rényi graphs and graphs with power law degree distributions. For example, it can be shown that for power law graphs the largest eigenvalue grows approximately like  $\sqrt{d_{\max}}$  as  $N \rightarrow \infty$ .

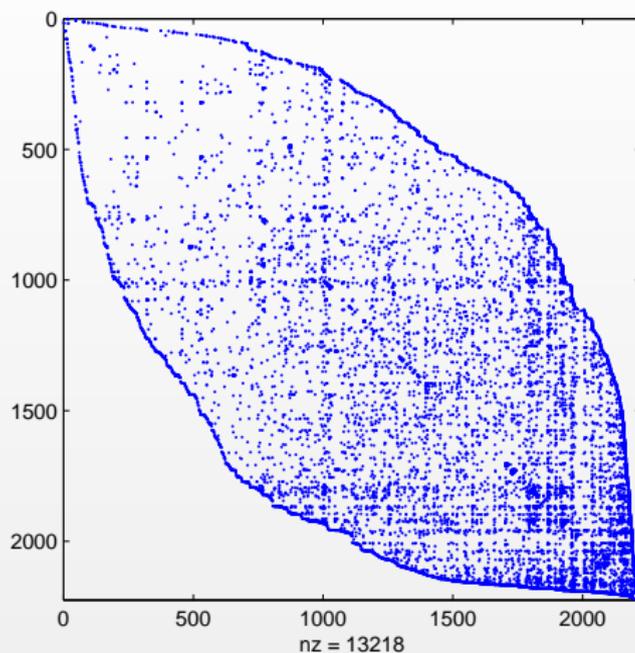
F. Chung and L. Lu, *Complex Graphs and Networks*, AMS, 2006.

## Example: PPI network of *Saccharomyces cerevisiae*



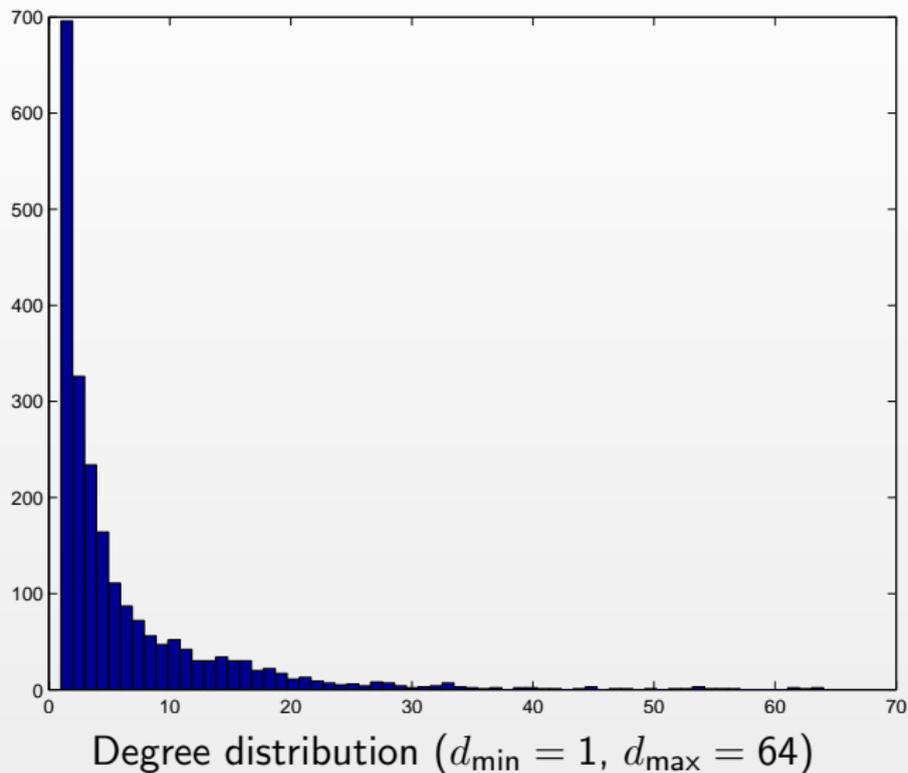
Adjacency matrix,  $|V| = 2224$ ,  $|E| = 6609$ .

## Example: PPI network of *Saccharomyces cerevisiae*

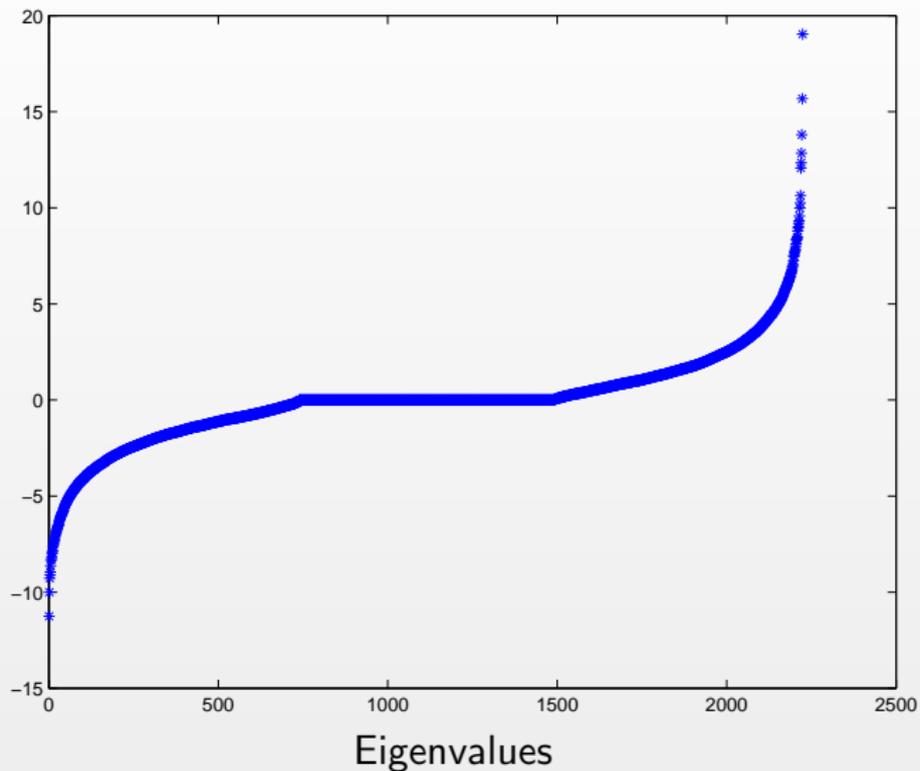


Same, reordered with Reverse Cuthill–McKee

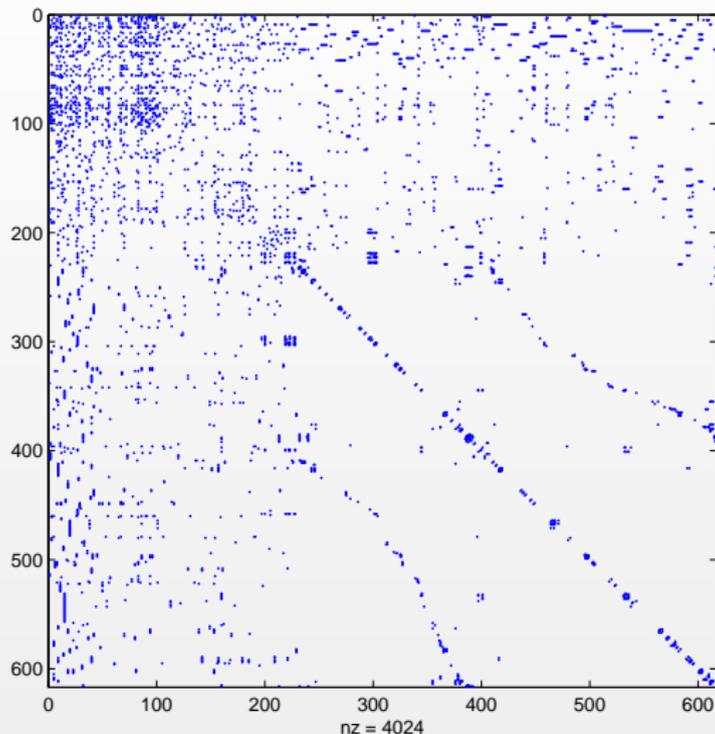
## Example: PPI network of *Saccharomyces cerevisiae*



# Example: PPI network of *Saccharomyces cerevisiae*

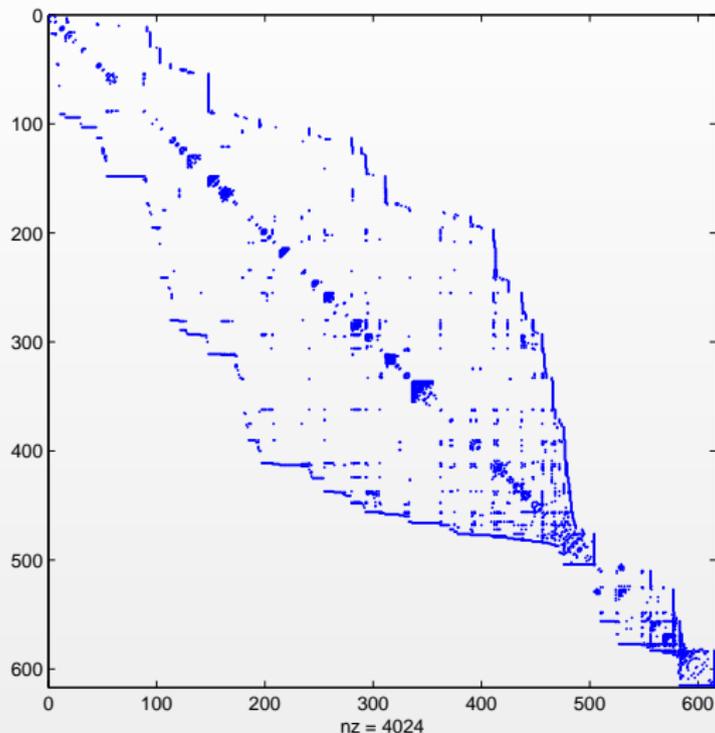


# Example: Intravenous drug users network



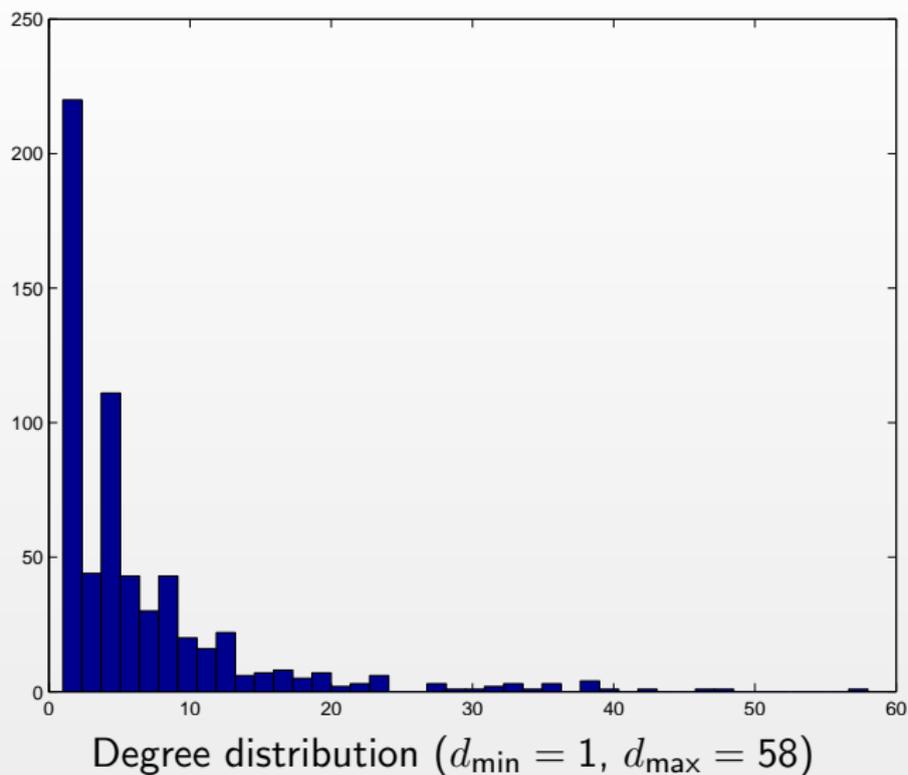
Adjacency matrix,  $|V| = 616$ ,  $|E| = 2012$ .

# Example: Intravenous drug users network

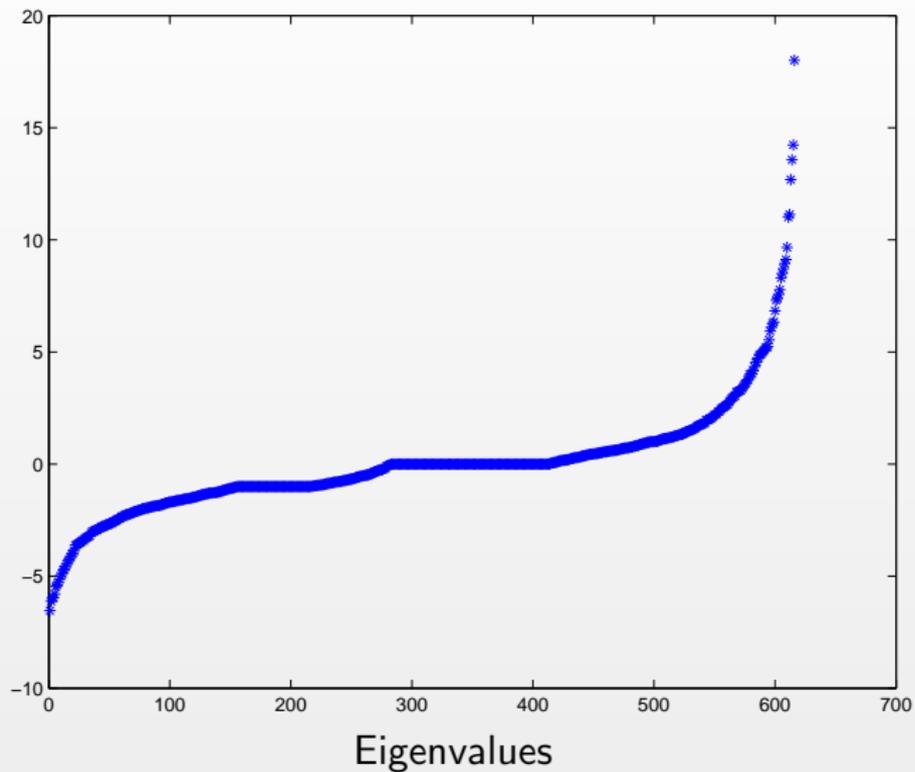


Same, reordered with Reverse Cuthill–McKee

## Example: Intravenous drug users network



# Example: Intravenous drug users network



# Graph Laplacian spectra

For an undirected network, the spectrum of the graph Laplacian  $L = D - A$  also plays an important role. For instance, the connectivity properties of  $G$  can be characterized in terms of spectral properties of  $L$ ; moreover,  $L$  plays an important role in the study of [diffusion processes](#) on  $G$ .

The eigenvalues of  $L$  are all real and nonnegative, and since  $L\mathbf{1} = \mathbf{0}$ , where  $\mathbf{1}$  denotes the vector of all ones,  $L$  is singular. Hence,  $0 \in \sigma(L)$ .

[The graph Laplacian  \$L\$  is a singular  \$M\$ -matrix.](#)

**Theorem.** The multiplicity of 0 as an eigenvalue of  $L$  coincides with the number of connected components of the network.

**Corollary.** For a connected network, the null space of the graph Laplacian is 1-dimensional and is spanned by  $\mathbf{1}$ . Thus,  $\text{rank}(L) = N - 1$ .

## Graph Laplacian spectra (cont.)

Let  $G$  be a simple, connected graph.

The eigenvector of  $L$  associated with the smallest nonzero eigenvalue is called the **Fiedler vector** of the network. Since this eigenvector must be orthogonal to the vector of all ones, it must contain both positive and negative entries.

There exist elegant **graph partitioning** algorithms that assign nodes to different subgraphs based on the sign of the entries of the Fiedler vector.

These methods, however, tend to work well only in the case of fairly regular graphs, such as those arising from the discretization of PDEs. In general, **partitioning complex graphs** (scale-free graphs in particular) **is very hard!**

## Graph Laplacian spectra (cont.)

If the graph  $G$  is regular of degree  $d$ , then  $L = dI_N - A$  and the eigenvalues of the Laplacian are just

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

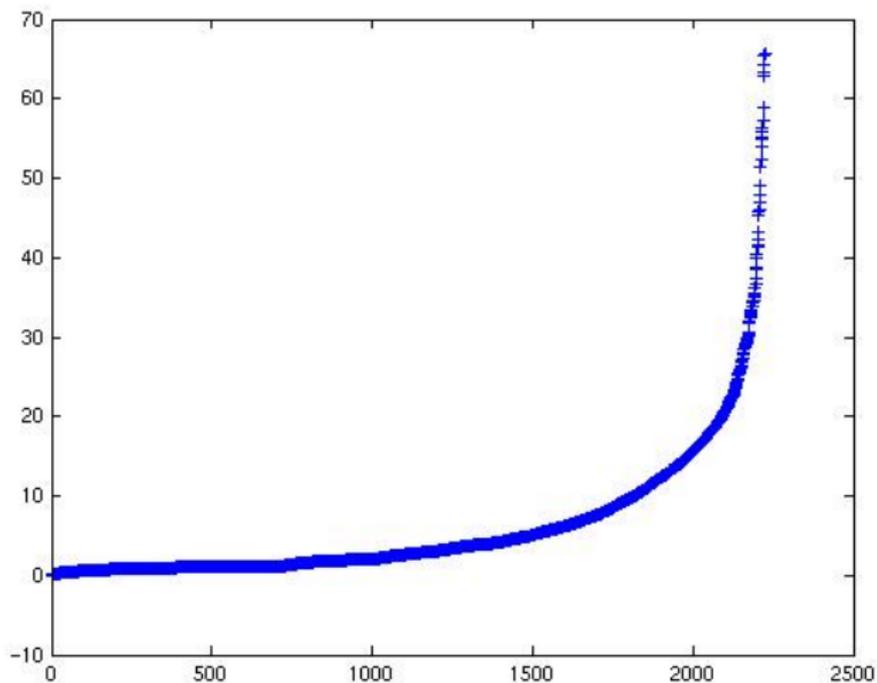
If  $G$  is not a regular graph, there is no simple relationship between the eigenvalues of  $L$  and those of  $A$ .

Also useful is the notion of **normalized Laplacian**:

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

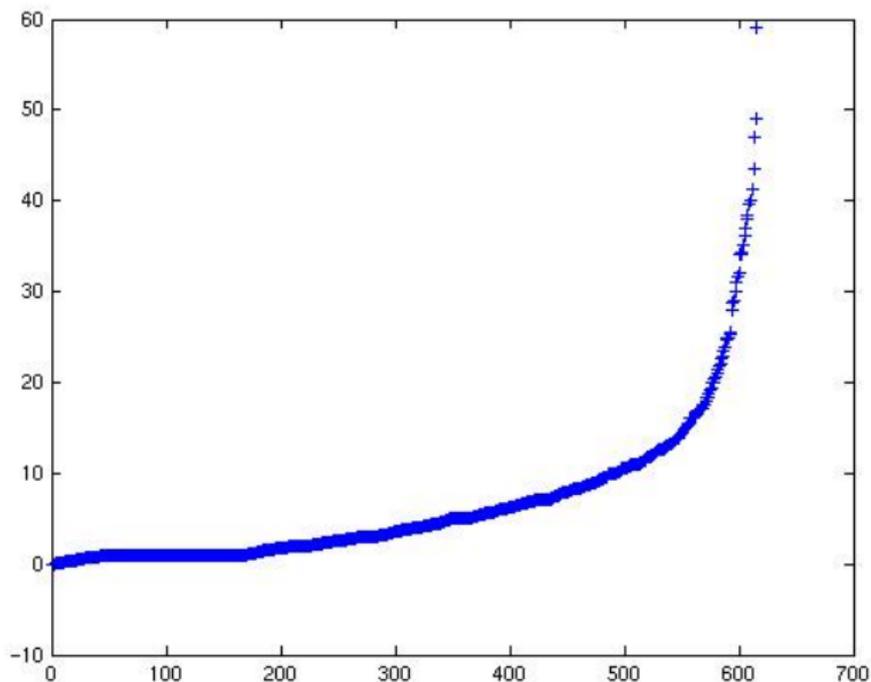
In the case of **directed graphs**, there are several distinct notions of graph Laplacian in the literature.

## Example: PPI network of *Saccharomyces cerevisiae*



Laplacian eigenvalues ( $\lambda_2(L) = 0.0600$ ,  $\lambda_N(L) = 65.6077$ )

## Example: Intravenous drug users network



- 1 Complex networks: motivation and background
- 2 Network structure: centrality and communicability measures
- 3 Bibliography

# Centrality measures

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.

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

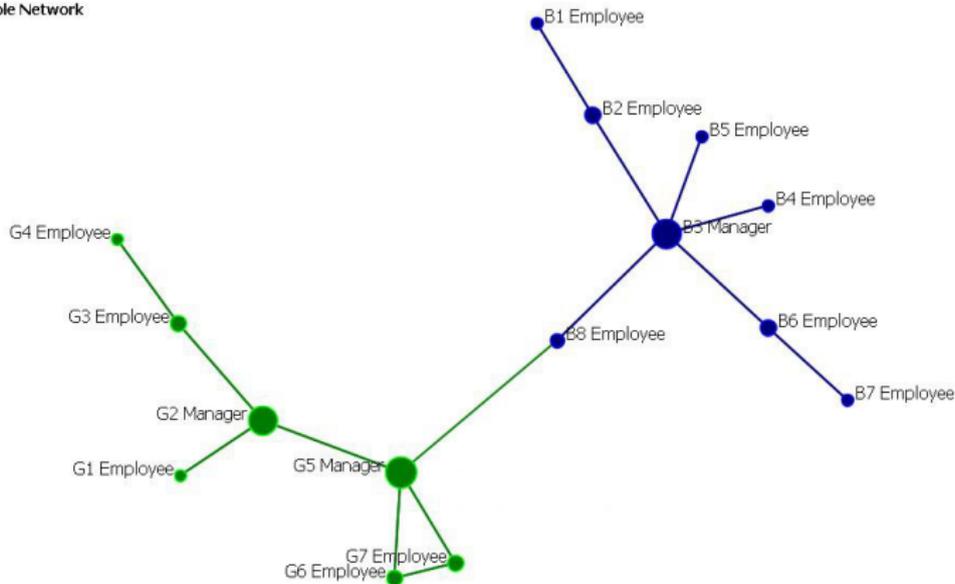
$$C_B(i) := \sum_{j \neq i} \sum_{k \neq i} \delta_{jk}(i),$$

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

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

# Centrality measures (cont.)

Example Network



powered by ORA, CASOS Center @ CMU

The degree is very cheap to compute but is unable to recognize the centrality of certain nodes: it's a **purely local** notion.

## Centrality measures (cont.)

Another centrality measure popular in social network analysis is **closeness centrality** (Freeman, 1979), defined as

$$C_C(i) = \frac{1}{\sum_{j \in V} d(i, j)}.$$

Betweenness and closeness centrality assume 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.

## Centrality measures (cont.)

We mention here that for **directed graphs** it is often necessary to distinguish between **hubs** and **authorities**. Indeed, in a directed graph a node plays two roles: **broadcaster** and **receiver** of information.

Crude broadcast and receive centrality measures are provided by the out-degree  $d_i^{out}$  and by the in-degree  $d_i^{in}$  of the node, respectively.

Other, more refined receive and broadcast centrality measures will be introduced later.

# Spectral centrality measures

Bonacich's **eigenvector centrality** (1987) uses the entries of the **dominant eigenvector**  $\mathbf{x}$  to rank the nodes in the network in order of importance: the larger  $x_i$  is, the more important node  $i$  is considered to be. By the Perron–Frobenius Theorem, the vector  $\mathbf{x}$  is positive and 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

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}, \quad k = 0, 1, \dots$$

which converges, upon normalization, to the dominant eigenvector of  $A$  as  $k \rightarrow \infty$ . The rate of convergence depends on the **spectral gap**  $\gamma = \lambda_1 - \lambda_2$ . The larger  $\gamma$ , the faster the convergence.

In the case of directed networks, the dominant left and right eigenvectors of  $A$  provide authority and hub scores, respectively.

## Spectral centrality measures (cont.)

Google's [PageRank algorithm](#) (Brin & 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 (in the form of a rank-one modification to the hyperlink matrix) are needed to have a unique PageRank eigenvector.

PageRank has a [probabilistic interpretation](#) in terms of [random walks on the web graph](#), a special type of [Markov chain](#). The PageRank eigenvector is the stationary probability distribution of this Markov chain.

An alternative approach (HITS), proposed by J. Kleinberg in 1998, 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 this topic in Part II.

# Subgraph centrality

We now turn to centrality measures that are based on [matrix functions](#).

[Subgraph centrality](#) (Estrada & Rodríguez-Velásquez, *Phys. Rev. E*, 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  $k = 1, 2, \dots$  the number of [closed walks](#) in  $G$  starting and ending at node  $i$ , with longer walks being penalized (given a smaller weight).

It is sometimes useful to introduce a tuning parameter  $\beta > 0$  (“inverse temperature”) to simulate external influences on the network, for example, increased tension in a social network, financial distress in the banking system, etc.

## Subgraph centrality (cont.)

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  $\beta^k/k!$  as weights leads to the notion of **subgraph centrality**:

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

Note that  $SC(i) \geq 1$ . Subgraph centrality has been used successfully in various settings, including **proteomics** and **neuroscience**.

**Note:** the weights are needed to “penalize” longer walks, and to make the power series converge.

## Subgraph centrality (cont.)

It is sometimes desirable to **normalize** the subgraph centrality of a node by the sum

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

of all the subgraph centralities. The quantity  $EE(G, \beta)$  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.

## Subgraph centrality (cont.)

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

Analogous to the Gibbs–Shannon entropy, one can then define the **walk entropy** of a graph  $G$  by

$$S(G, \beta) := - \sum_{i=1}^N p(i) \log_2 p(i).$$

The walk entropy provides a useful measure of the complexity of a graph (Estrada, de la Peña, & Hatano, *Lin. Algebra Appl.*, 2014.)

## Subgraph centrality (cont.)

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.

In a bipartite graph there cannot be any closed walks of odd length; hence, all the diagonal entries of  $A^{2k+1}$  must be zero, for all  $k \geq 0$ . Therefore  $\text{Tr}(\sinh(A)) = 0$ . Hence, the quantity

$$B(G) := \frac{\text{Tr}(\cosh(A))}{\text{Tr}(e^A)}$$

provides a measure of how “close” a given graph is to being bipartite: if  $B(G)$  is close to 1, then  $G$  is “nearly bipartite”.

Hyperbolic matrix functions are also used to define the notion of **returnability** in digraphs (E. Estrada & N. Hatano, *Lin. Algebra Appl.*, 2009).

# Katz centrality

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

$$(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \dots = \sum_{k=0}^{\infty} \alpha^k A^k, \quad 0 < \alpha < 1/\lambda_1.$$

Note that  $I - \alpha A$  is a nonsingular  $M$ -matrix, in particular  $(I - \alpha A)^{-1} \geq 0$ .

Originally, Katz proposed to use the row sums of  $(I - \alpha A)^{-1}$  as a centrality measure:

$$C_K(i) = \mathbf{e}_i^T (I - \alpha A)^{-1} \mathbf{1}.$$

Resolvent-based subgraph centrality uses instead the diagonal entries of  $(I - \alpha A)^{-1}$  (E. Estrada & D. Higham, SIAM Rev., 2010).

In the case of a **directed network** one can use the solution vectors of the linear systems

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

to rank **hubs** and **authorities**. These are the row and column sums of the matrix resolvent  $(I - \alpha A)^{-1}$ , respectively.

## Comparing centrality measures

Different centrality measures correspond to different notions of what it means for a node to be “influential”; some (like degree) only look at local influence, others (like eigenvector centrality and PageRank) emphasize long distance influences, and others (like subgraph and Katz centrality) try to take into account short, medium and long range influences.

Others yet, like closeness and betweenness centrality, emphasize nodes that are a short distance away from most other nodes, or that are central to the flow of information along shortest paths in  $G$ .

And these are just a few of the many centrality measures proposed in the literature!

## Comparing centrality measures (cont.)

A natural question at this point is: **which centrality measure is better?**  
**Which measure should be used?**

Unfortunately, it is not possible to give a general, satisfactory answer to this question. Some centrality measures may work very well on a network and not so well on others.

Competing methods can be tested on networks for which some kind of **ground truth** is known, for example, from experiments (e.g., PPI network of yeast), and the “best one” can then be used on similar networks, but with few guarantees of success.

Often, however, multiple centrality measures are used and one looks at the overlap between lists of top-ranked nodes (a form of consensus).

## Comparing centrality measures (cont.)

Some centrality measures have been shown to have **greater discriminating power** than others.

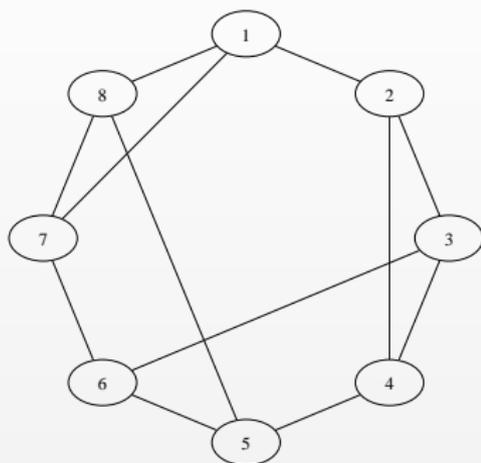
For example, if a graph  $G$  is **regular** ( $d_i = d$  for all  $i = 1 : N$ ), degree centrality, eigenvector centrality and Katz centrality (as well as all other centrality measures of the form  $\mathbf{c} = f(A)\mathbf{1}$ ) are unable to distinguish between the nodes in  $G$ : all nodes are given equal status.

In contrast, there are regular graphs for which subgraph centrality is able to identify nodes that are more "central" than others.

Hence, subgraph centrality has a **higher discriminating power** than other centrality measures.

A recent result of Estrada and de la Peña (LAA, 2014) states that in a graph  $G$  all nodes have the same subgraph centrality ( $[e^A]_{ii} = \text{const.}$ ) if and only if  $G$  is **walk regular**, i.e.,  $[A^k]_{ii} = c(k) = \text{const.}$  (independent of  $i$ ) for all  $k = 0, 1, \dots$

## Comparing centrality measures (cont.)



**Figure:** A 3-regular graph on 8 nodes which is not walk-regular.

Example of a 3-regular graph on 8 nodes for which subgraph centrality can discriminate among nodes. Node 4 has the highest subgraph centrality, node 5 the lowest.

## Comparing centrality measures (cont.)

In the case of very large-scale networks (with millions or even billions of nodes/edges), **computational cost** becomes a limiting factor.

Of course, degree centrality is very cheap, but as we know it is not very satisfactory.

At the other end of the spectrum, subgraph centrality (which scales roughly like  $\mathcal{O}(N^2)$ ) is very satisfactory but is quite expensive for very large networks.

For large networks, methods like eigenvector and Katz centrality, while still potentially expensive, are often used, together with closeness and betweenness centrality.

## Comparing centrality measures (cont.)

In practice, one often finds a good degree of correlation between the rankings provided by (apparently very different) centrality measures. How can we explain this?

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

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

and therefore

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

where  $x_{k,i}$  denotes the  $i$ th entry of  $\mathbf{x}_k$ .

Similar spectral formulas exist for other centrality measures based on matrix functions (e.g., the resolvent).

## Comparing centrality measures (cont.)

Dividing all the subgraph centralities by  $e^{\beta\lambda_1}$  has no effect on the centrality rankings:

$$e^{-\beta\lambda_1} [e^{\beta A}]_{ii} = x_{1,i}^2 + \sum_{k=2}^N e^{\beta(\lambda_k - \lambda_1)} x_{k,i}^2.$$

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.

The same is true if  $\beta$  is chosen "large".

Hence, in this case subgraph centrality can be expected to give similar results to eigenvector centrality. The same applies to Katz centrality.

In contrast, the various measures can give significantly different results **when the spectral gap is small**. In this case, going beyond the dominant eigenvector can result in **significant differences** in the corresponding rankings.

## Comparing centrality measures (cont.)

**Theorem:** Let  $A$  be the adjacency matrix for a simple, connected, undirected graph  $G$ . Then

- For  $\alpha \rightarrow 0+$ , Katz centrality reduces to degree centrality;
- For  $\alpha \rightarrow \frac{1}{\lambda_1}-$ , Katz centrality reduces to eigenvector centrality;
- For  $\beta \rightarrow 0+$ , subgraph centrality reduces to degree centrality;
- For  $\beta \rightarrow \infty$ , subgraph centrality reduces to eigenvector centrality.

**Note:** A similar result holds for directed networks, in which case we need to distinguish between in-degree, out-degree, left/right dominant eigenvectors, and row/column sums of  $(I - \alpha A)^{-1}$  and  $e^{\beta A}$ .

M. Benzi & C. Klymko, *A matrix analysis of different centrality measures*, submitted (2014).

## Comparing centrality measures (cont.)

The previous result helps explain why Katz and subgraph centrality are most informative and useful when the parameters  $\alpha$ ,  $\beta$  are chosen to be neither too small, nor too large.

In practice, we found that a good choice for Katz's parameter is  $\alpha = \delta/\lambda_1$ , with  $0.5 \leq \delta \leq 0.9$ ; for subgraph centrality it is best to use  $0.5 \leq \beta \leq 2$ .

As default parameters we recommend  $\alpha = 0.85/\lambda_1$  and  $\beta = 1$ , although smaller values should be used for networks with large spectral gap  $\gamma = \lambda_1 - \lambda_2$ .

# 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):

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

Note that as the temperature increases ( $\beta \rightarrow 0$ ) the communicability between nodes decreases ( $e^{\beta A} \rightarrow I$ ).

Communicability has been successfully used to identify bottlenecks in networks (e.g., regions of low communicability in brain networks) and for community detection.

E. Estrada & N. Hatano, *Phys. Rev. E*, 77 (2008), 036111;

E. Estrada, N. Hatano, & M. Benzi, *Phys. Rep.*, 514 (2012), 89–119.

The **total communicability** of a node is defined as

$$TC(i) := \sum_{j=1}^N C(i, j) = \sum_{j=1}^N [e^{\beta A}]_{ij}.$$

This is another node centrality measure that can be computed very efficiently, even for large graphs, since it involves computing  $e^{\beta A} \mathbf{1}$  and, as we will see in Part II, **Krylov subspace methods** are very good at this!

## Communicability (cont.)

The **normalized total communicability** of  $G$ :

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

provides a global measure of how “well-connected” a network is, and can be used to compare different network designs.

Thus, highly connected networks, such as small-world networks without bottlenecks, can be expected to have a high total communicability.

Conversely, large-diameter networks with a high degree of locality (such as regular grids, road networks, etc.) or networks containing bottlenecks are likely to display a low value of  $TC(G)$ .

Note that  $TC(G)$  can be easily bounded from below and from above:

$$\frac{1}{N}EE(G) \leq TC(G) \leq e^{\beta\lambda_1}$$

where  $EE(G) = \text{Tr}(e^{\beta A})$  is the Estrada index of the graph. These bounds are sharp, as can be seen from trivial examples.

In the next Table we present the results of some calculations (for  $\beta = 1$ ) of the normalized Estrada index, global communicability and  $e^{\lambda_1}$  for various real-world networks.

M. Benzi & C. F. Klymko, *J. Complex Networks*, 1 (2013), 124–149.

## Communicability (cont.)

**Table:** Comparison of the normalized Estrada index  $E(G) = \text{Tr}(e^A)/N$ , the normalized total network communicability  $TC(G)$ , and  $e^{\lambda_1}$  for various real-world networks.

Network	$E(G)$	$TC(G)$	$e^{\lambda_1}$
Zachary Karate Club	30.62	608.79	833.81
Drug User	1.12e05	1.15e07	6.63e07
Yeast PPI	1.37e05	3.97e07	2.90e08
Pajek/Erdos971	3.84e04	4.20e06	1.81e07
Pajek/Erdos972	408.23	1.53e05	1.88e06
Pajek/Erdos982	538.58	2.07e05	2.73e06
Pajek/Erdos992	678.87	2.50e05	3.73e06
SNAP/ca-GrQc	1.24e16	8.80e17	6.47e19
SNAP/ca-HepTh	3.05e09	1.06e11	3.01e13
SNAP/as-735	3.00e16	3.64e19	2.32e20
Gleich/Minnesota	2.86	14.13	35.34

## 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

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

Most real 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.

- 1 Complex networks: motivation and background
- 2 Network structure: centrality and communicability measures
- 3 **Bibliography**

- 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.
- Excellent accounts by Barabási (2002) e Watts (2003) for a general readership.

# Specialized journals

Two new journals launched in 2013:

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

<http://comnet.oxfordjournals.org/>

<http://journals.cambridge.org/action/displayjournal?jid=NWS>

# Singular Value Decomposition (SVD)

For any matrix  $A \in \mathbb{R}^{m \times n}$  there exist orthogonal matrices  $U \in \mathbb{R}^{m \times m}$  and  $V \in \mathbb{R}^{n \times n}$  and a “diagonal” matrix  $\Sigma \in \mathbb{R}^{m \times n}$  such that

$$U^T A V = \Sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$$

where  $p = \min\{m, n\}$ .

The  $\sigma_i$  are the singular values of  $A$  and satisfy

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

where  $r = \text{rank}(A)$ . The matrix  $\Sigma$  is uniquely determined by  $A$ , but  $U$  and  $V$  are not.

The columns  $u_i$  and  $v_i$  of  $U$  and  $V$  are left and right singular vectors of  $A$  corresponding to the singular value  $\sigma_i$ .

# Singular Value Decomposition (cont.)

Note that

$$Av_i = \sigma_i u_i \quad \text{and} \quad A^T u_i = \sigma_i v_i, \quad 1 \leq i \leq p.$$

From  $AA^T = U\Sigma\Sigma^T U^T$  and  $A^T A = V\Sigma^T \Sigma V^T$  we deduce that the singular values of  $A$  are the (positive) square roots of the SPD matrices  $AA^T$  and  $A^T A$ ; the left singular vectors of  $A$  are eigenvectors of  $AA^T$ , and the right ones are eigenvectors of  $A^T A$ .

Moreover,

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T,$$

showing that any matrix  $A$  of rank  $r$  is the sum of exactly  $r$  rank-one matrices.

## Singular Value Decomposition (cont.)

It is also important to recall that if we form the symmetric matrix

$$\mathcal{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \in \mathbb{R}^{(n+m) \times (n+m)},$$

then the eigenvalues of  $\mathcal{A}$  are of the form  $\lambda_i = \pm\sigma_i$ , and the corresponding eigenvectors are given by

$$x_i = \begin{bmatrix} u_i \\ \pm v_i \end{bmatrix}.$$

The matrix  $\mathcal{A}$  plays an important role in computations and in the analysis of bipartite and directed networks.

# Geršgorin's Theorem

Let  $A \in \mathbb{C}^{n \times n}$ . For all  $i = 1 : N$ , let

$$r_i := \sum_{j \neq i} |a_{ij}|, \quad D_i = D_i(a_{ii}, r_i) := \{z \in \mathbb{C} : |z - a_{ii}| \leq r_i\}.$$

The set  $D_i$  is called the *i*th Geršgorin disk of  $A$ .

Geršgorin's Theorem (1931) states that  $\sigma(A) \subset \cup_{i=1}^N D_i$ . Moreover, each connected component of  $\cup_{i=1}^N D_i$  consisting of  $p$  Geršgorin disks of  $A$  contains exactly  $p$  eigenvalues of  $A$ , counted with their multiplicities.

Of course, the same result holds replacing the off-diagonal row-sums with off-diagonal column-sums. The spectrum is then contained in the intersection of the two resulting regions.