Social network analysis

Lorenzo Mannocci

a.a. 2026/2027



UNIVERSITÀ DI PISA

Introduction

Complex system

Behind each complex system
there is a network,
that defines the interactions
between the components.



Example of complex system

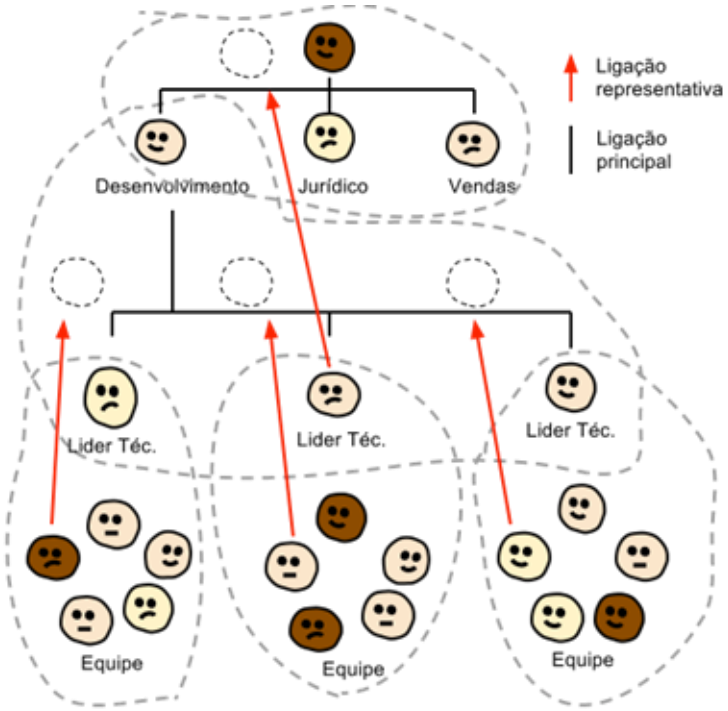
The Facebook “Social Graph”

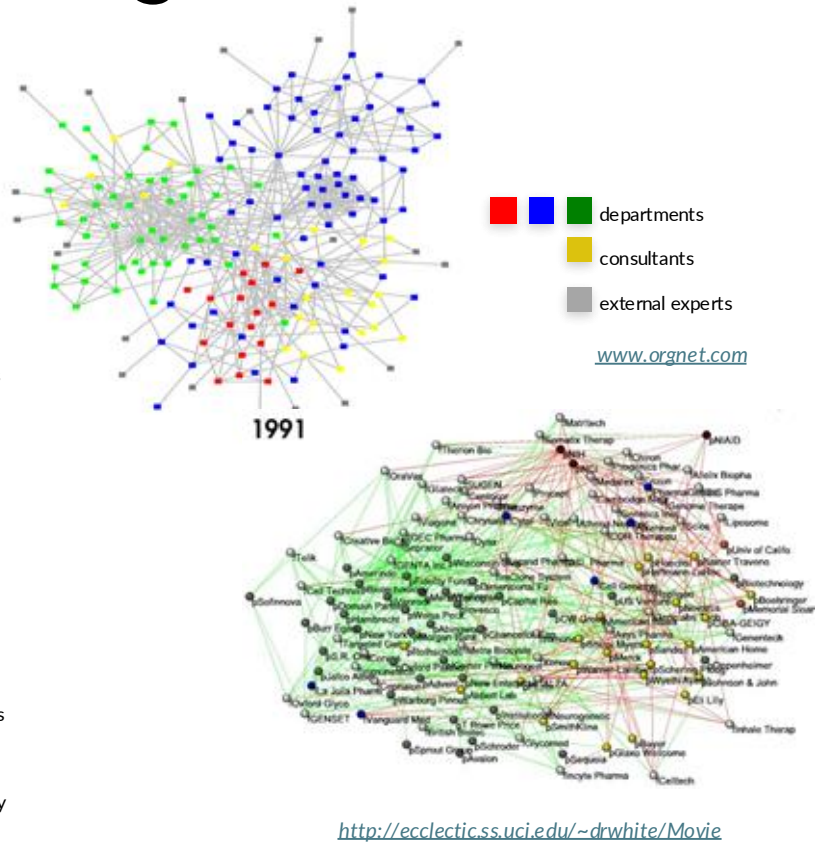




Example of complex system

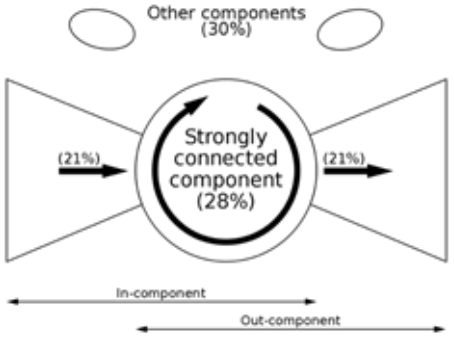
The structure of an organization



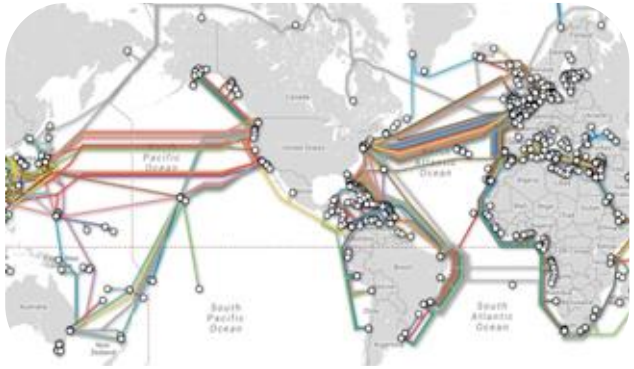


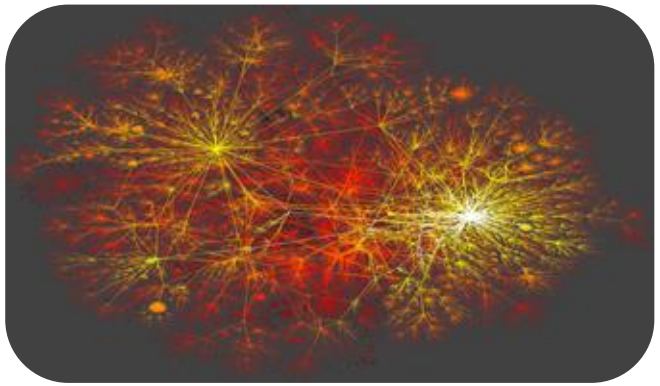
Example of complex system

The Internet backbone,
The World Wide Web…

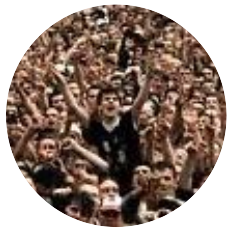








Real world network

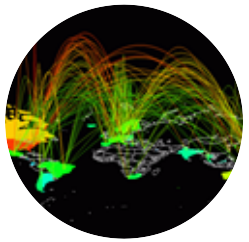


Type: Social
Nodes: Individuals
Links: Social relationship

Type: Scientific Collaborations
Nodes: Researchers
Links: Co-Authorships

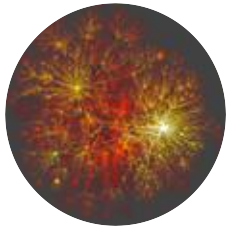


Type: Actor connectivity
Nodes: Actors
Links: Cast jointly



Type: Communication
Nodes: Phones, Airports..
Links: Phone calls, Flights..

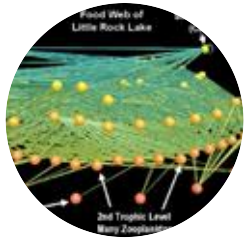
Real world network



Type: Technological
Nodes: PC, Routers
Links: Physical lines



Type: Scientific Citation
Nodes: Papers
Links: Citations



Type: Biological
Nodes: Species
Links: Trophic interactions



Type: Mobility
Nodes: Individuals, Cars...
Links: Co-Location...

Why do we need network science?

Data Availability

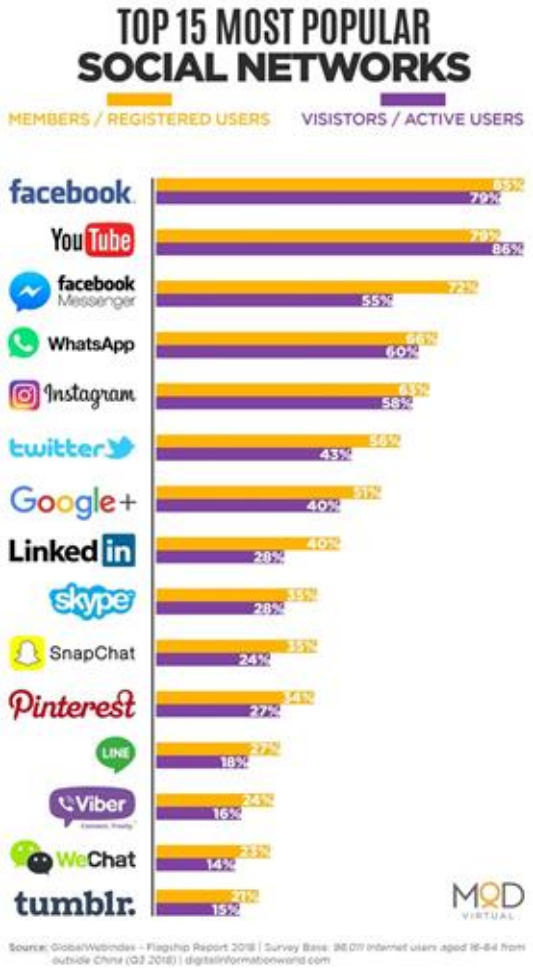
1990 C. elegans neural wiring diagram
1998 - Movie Actor Network
1998 - Citation Networks
1999 -World Wide Web
2000 - Metabolic Networks
2001 - PPI network
2008 - OSNs

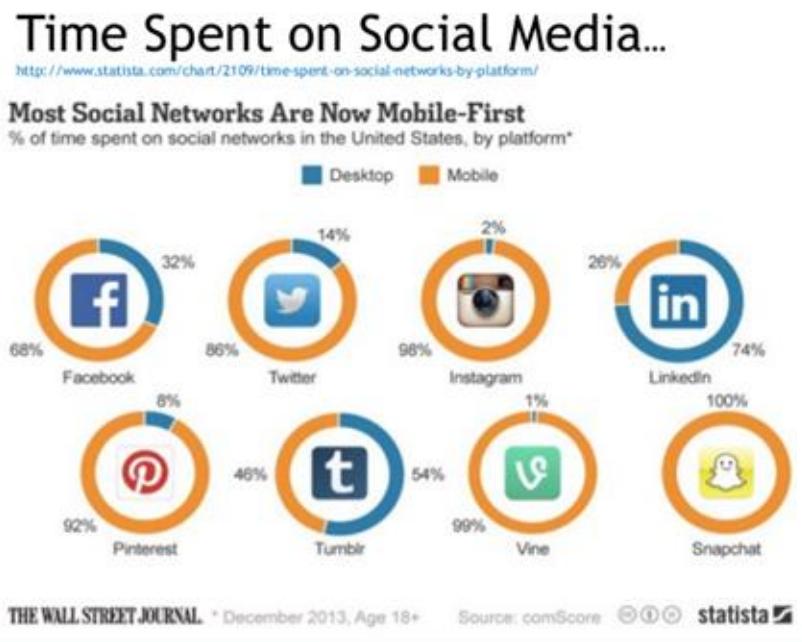


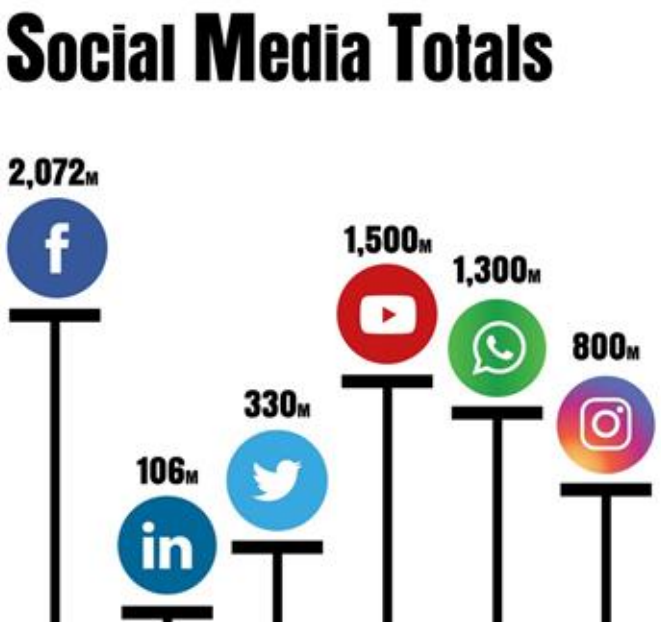
Universality

The architecture of networks emerging in various domains of science, nature, and technology are more similar to each other than one would have expected.

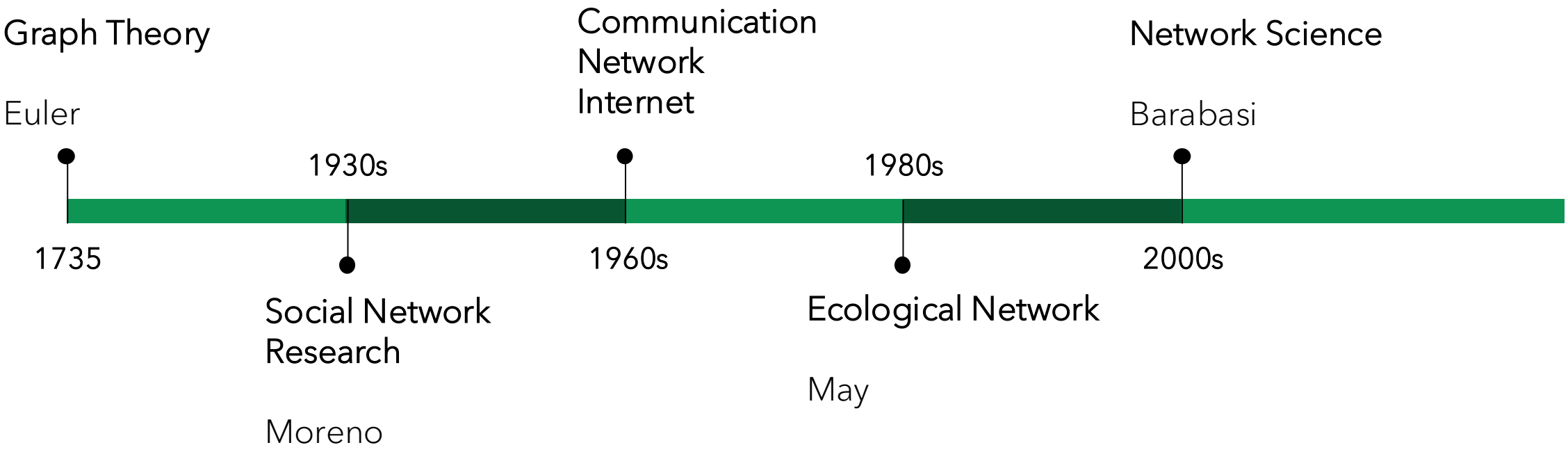
Online social network





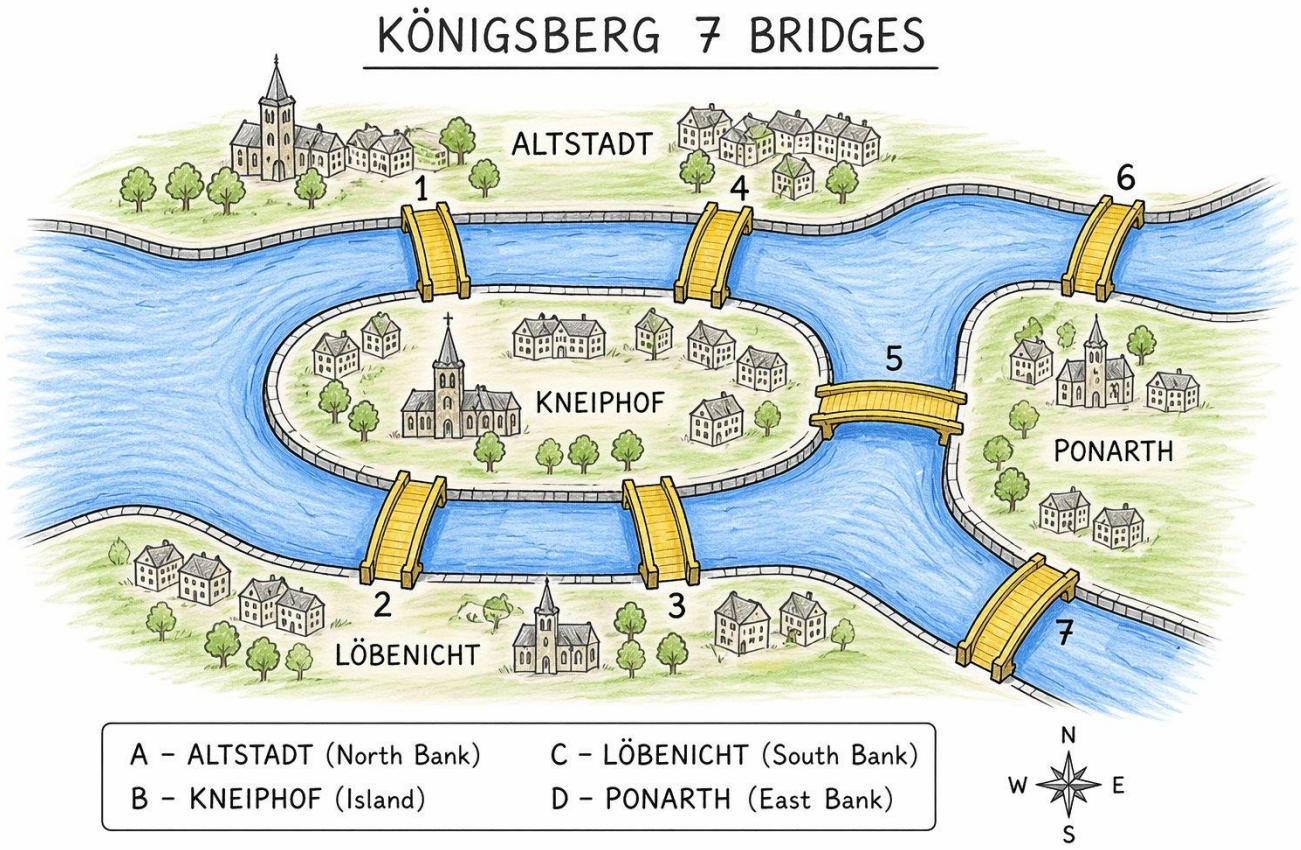


History of network science



Networks and graphs

The Bridges of Konigsberg



Königsberg (now Kaliningrad, Russia) was a city divided by the Pregel River into four land areas, connected by seven bridges.

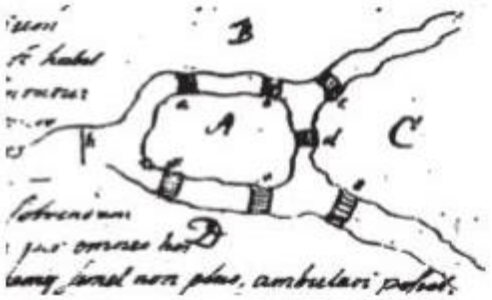
The inhabitants wondered whether it was possible to take a walk through the city satisfying two conditions:

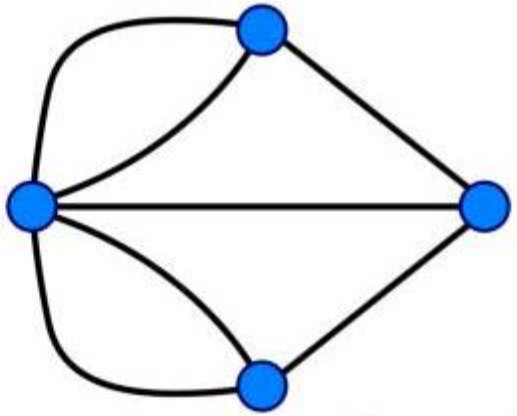
- Cross every bridge exactly once.
- Never cross the same bridge twice.

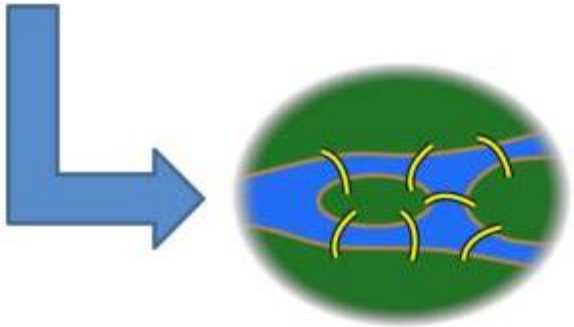
You could start and finish anywhere.
Euler demonstrated that such a walk is impossible.

The Bridges of Konigsberg

Euler's drawing of the seven bridges







The seven bridges of Königsberg

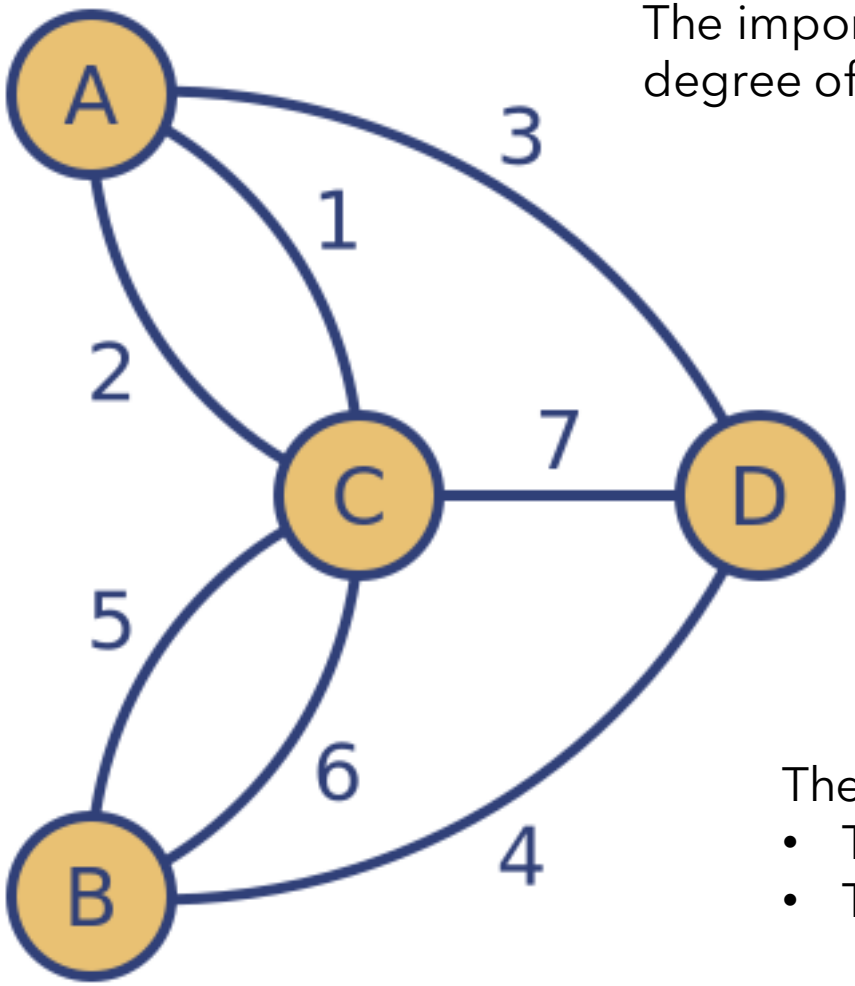
The graph associated with the seven bridges of Königsberg problem

Instead of considering the actual geography of the city, Euler simplified the problem.
Each land area becomes a node.
Each bridge becomes an edge.
The resulting graph has four nodes and seven edges.

Notice that the graph has multiple edges between the same pair of nodes. This is called a multigraph.

The important observation is the degree of each node:

The Bridges of Konigsberg



The important observation is the degree of each node:

Node	Degree
A	3
B	3
C	5
D	3

All four nodes have an odd degree.

Imagine walking through the city.
Whenever you enter a land area through one bridge, you must leave through another bridge.
Therefore, bridges are generally used in pairs:

- One edge to enter.
- One edge to exit.

This means that every intermediate node must have an even degree.

There are only two possible exceptions:

- The starting node, where you can leave without first entering.
- The ending node, where you can enter without leaving.

Consequently, at most two nodes can have an odd degree.
But in Königsberg, there are four odd-degree nodes.
Therefore, the walk is impossible.

The Bridges of Konigsberg

Euler formalized this intuition in what is now known as the Eulerian path and circuit theorem.

First, we distinguish two concepts.

- Eulerian path: A walk that traverses every edge of the graph exactly once. The starting and ending nodes may be different.
- Eulerian circuit: A closed Eulerian trail that traverses every edge exactly once and returns to its starting node.

For an undirected, connected graph:

Number of odd-degree nodes	**Result**
0	An Eulerian circuit exists
2	An Eulerian path exists, but not an Eulerian circuit
More than 2	Neither exists

The Bridges of Konigsberg

The Königsberg problem introduced a fundamental change in how we study complex systems.
Euler demonstrated that we can ignore many details of a real-world system and focus on its connectivity structure.
In particular:

- The geographical distances between land areas do not matter.
- The exact shapes of the bridges do not matter.
- The positions of the nodes do not matter.
- Only which nodes are connected and how many edges connect them matters.

This is the central idea behind graph-based representations of social networks, transportation systems, communication networks, and many other systems.

Note: An Eulerian path visits every edge exactly once, whereas a Hamiltonian path visits every node exactly once. These are different problems, and Euler's degree-based theorem applies to the former, not the latter.

Components of a complex system

Networks or Graphs?

- Network <*nodes, links*> refers to real systems (www, social network, metabolic network)
- Graph <*vertices, edges*>: mathematical representation of a network (web graph, social graph)

A graph is a mathematical structure used to represent a set of entities and the relationships between them.

Formally, a graph is defined as:

$$G = (V, E)$$

where:

- V is a set of vertices (or nodes) representing entities.
- E is a set of edges (or links) representing relationships between pairs of vertices.

For an undirected graph, each edge is an unordered pair of vertices:

$$E \subseteq \{\{u, v\} \mid u, v \in V, u \neq v\}$$

For a directed graph, each edge is an ordered pair: $E \subseteq V \times V$

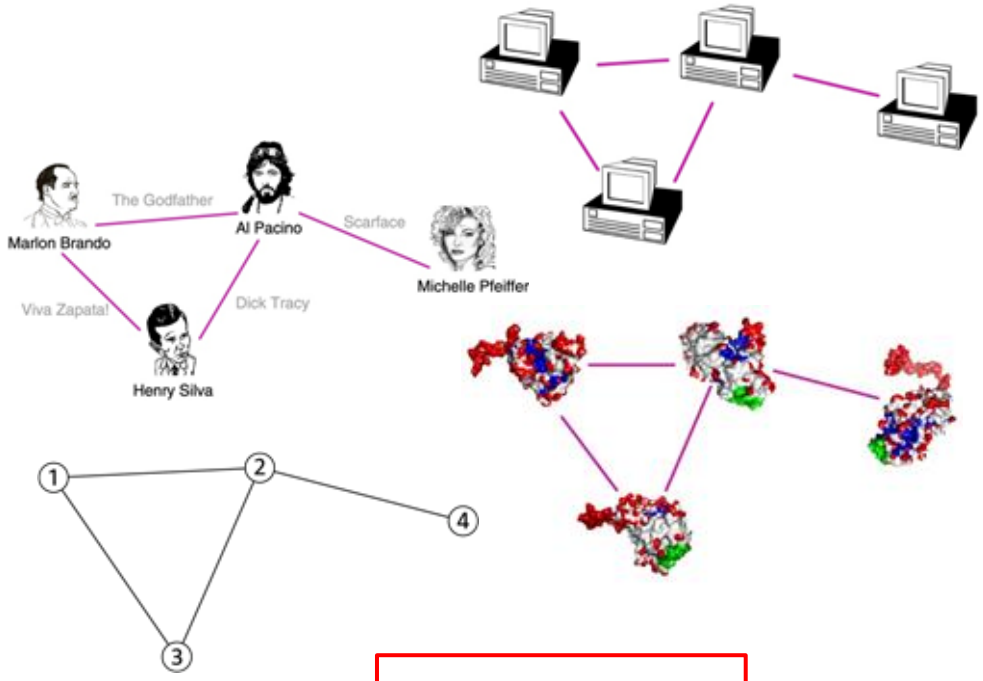
	Component	**Symbol**
Components	nodes, vertices	N (# nodes)
Interactions	edges, links	L (# edges)
System	network, graph	(N,L)

A common language

The choice of the proper network representation determines our ability to use network theory successfully.

In some cases there is a unique, unambiguous representation. In other cases, the representation is by no means unique.

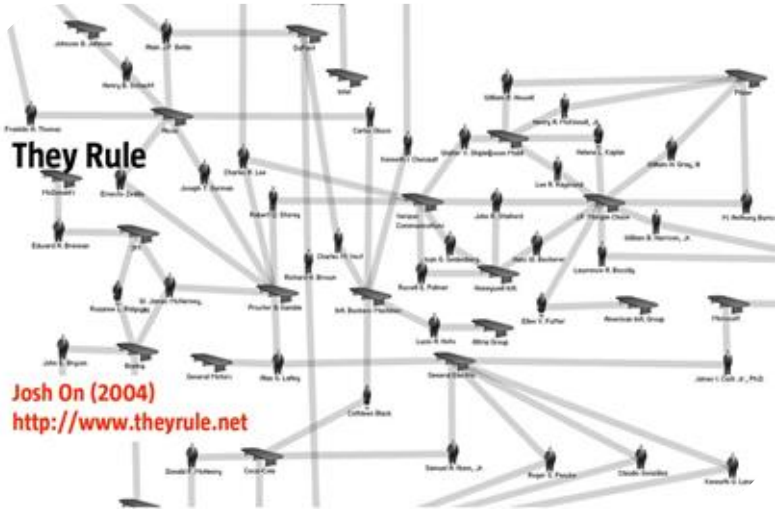
The way we assign the links between a group of individuals will determine the nature of the question we can study.



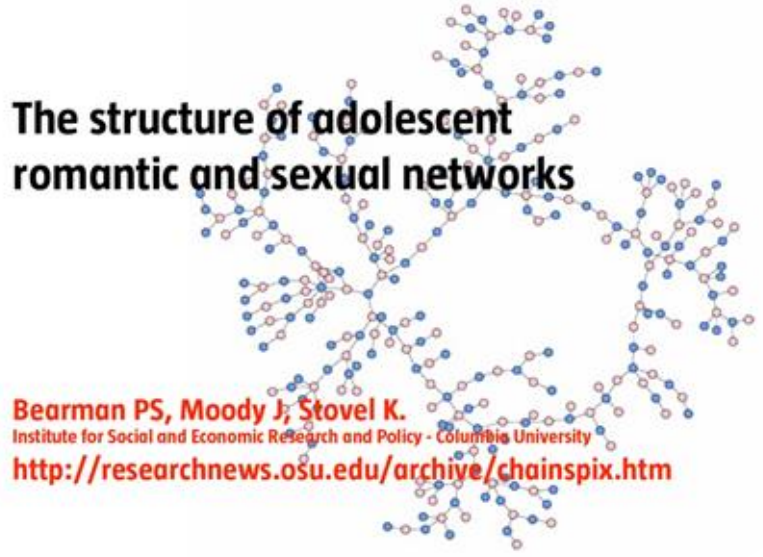
N=4
L=4

Proper representation

If you connect individuals that work with each other, you will explore the *professional network*



If you connect those that have a romantic and sexual relationship, you will be exploring the *sexual networks*



Bad representation

If you connect individuals based on their first name (e.g., *all Peters connected to each other*), you will be exploring what?

It is a network, *nevertheless*.

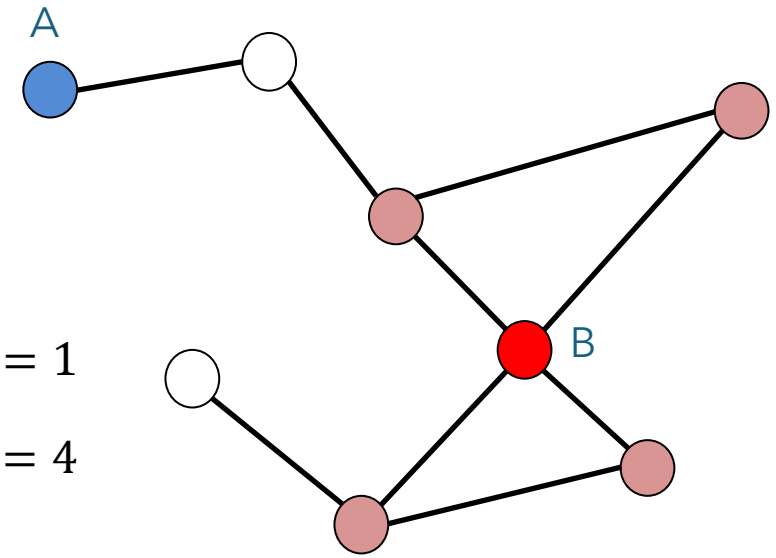
Network characteristics

Degree,
Average degree,
Degree distribution

Node degree

Undirected graphs

The number of links connected to the node

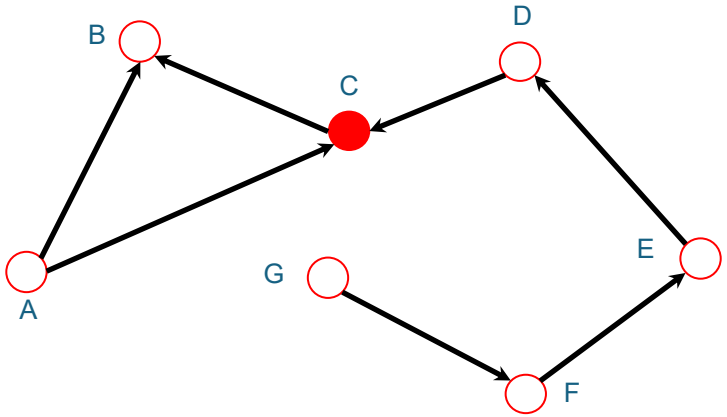


$k_a = 1$

$k_b = 4$

Directed graphs

- in-degree,
- out-degree,
- total degree: sum of in- and out-degree



$k_C^{in} = 2$

$k_C^{out} = 1$

$k_C = 3$

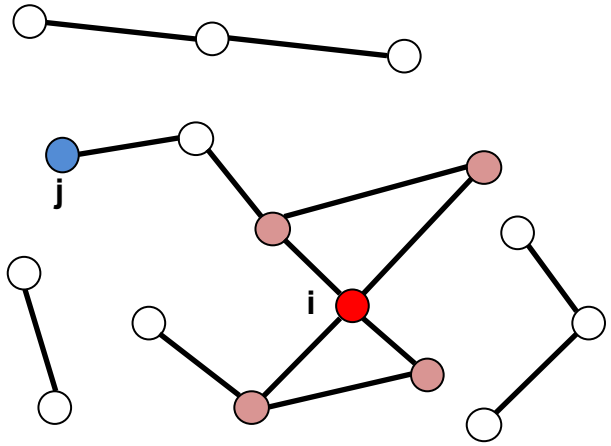
Source: a node with $k^{\text{in}} = 0$;

Sink: a node with $k^{\text{out}} = 0$.

Average degree

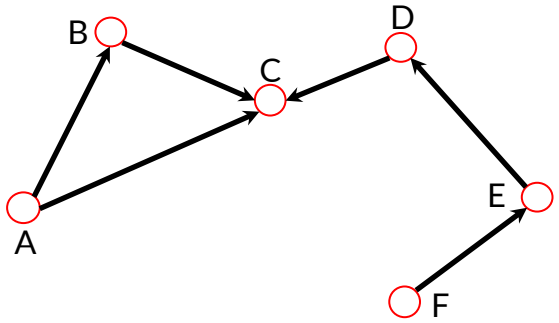
Undirected graphs

N - the number of nodes in the graph



$$\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i\,, \quad \langle k \rangle = \frac{2L}{N}$$

Directed graphs



$$\langle k^{in} \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i^{in}\,, \quad \langle k^{out} \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i^{out}$$

$$\langle k^{in} \rangle = \langle k^{out} \rangle \quad \langle k \rangle = \frac{L}{N}$$

Average degree examples

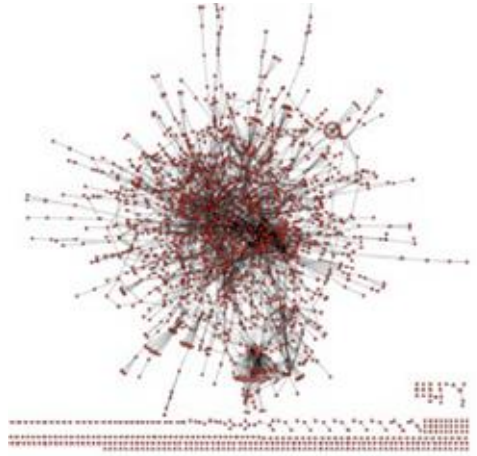
Network	Nodes	Links	Directed / Undirected	N	L		k
Internet	Routers	Internet connections	Undirected	192,244	609,066		6.33
WWW	Webpages	Links	Directed	325,729	1,497,134		4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594		2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826		2.51
Email	Email addresses	Emails	Directed	57,194	103,731		1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439		8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908		83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479		10.43
E. coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802		5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930		2.90

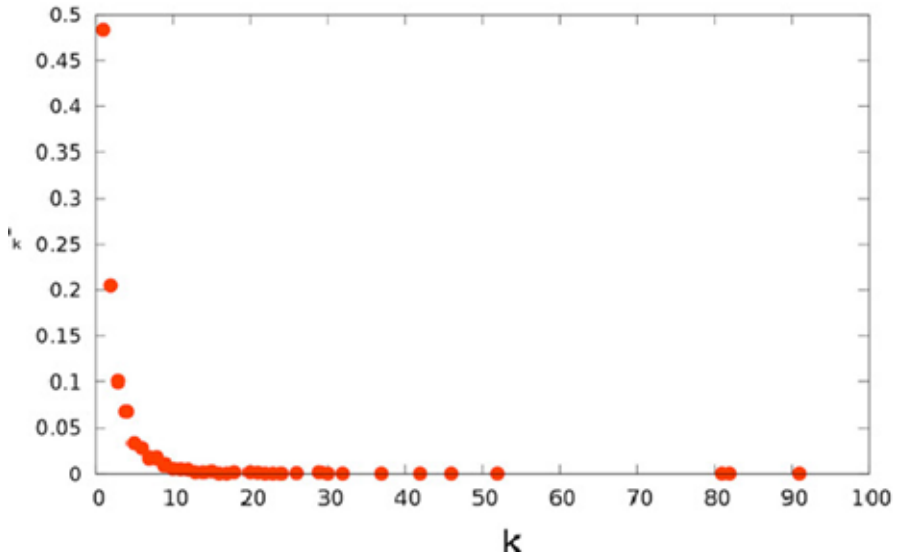
Degree distribution

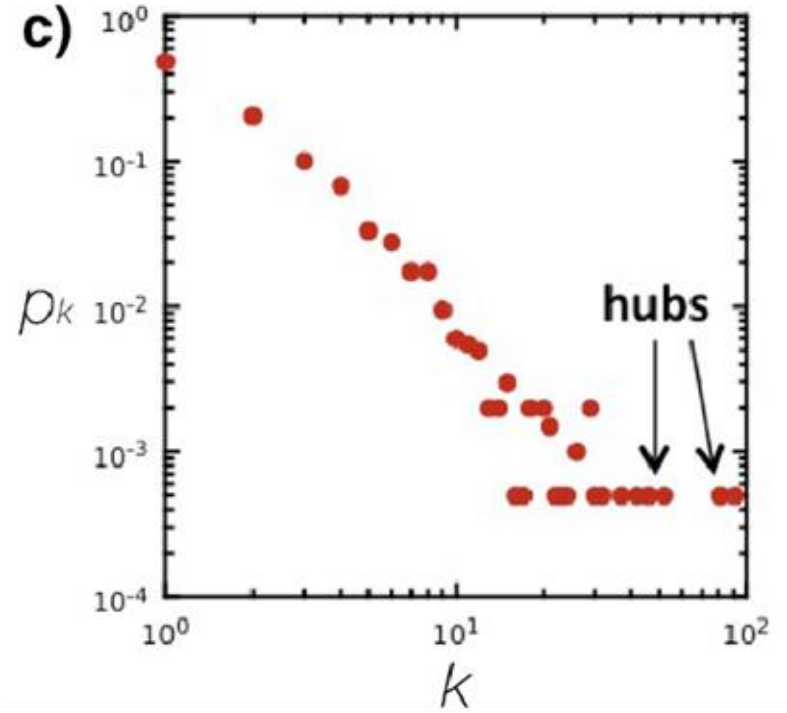
$P(k)$: probability that a randomly chosen node has degree k:

$$P(k) = \frac{N_k}{N}$$

- N_k: number of nodes with degree k
- N: total number of nodes



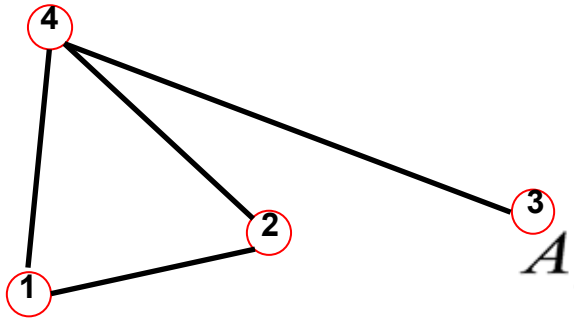




Directedness
Weightedness
Loops & Multigraphs

Adjacency matrix

Undirected graphs



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

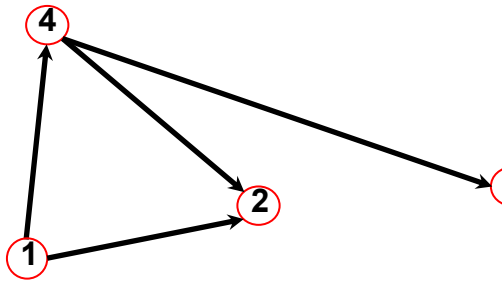
$$k_i = \sum_{j=1}^{N} A_{ij} \qquad k_j = \sum_{i=1}^{N} A_{ij}$$

$$A_{ij} = A_{ji}$$

$$A_{ii} = 0$$

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i = \frac{1}{2}\sum_{ij} A_{ij}$$

Directed graphs



$$A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$k_j^{in} = \sum_{i=1}^{N} A_{ij} \qquad k_i^{out} = \sum_{j=1}^{N} A_{ij}$$

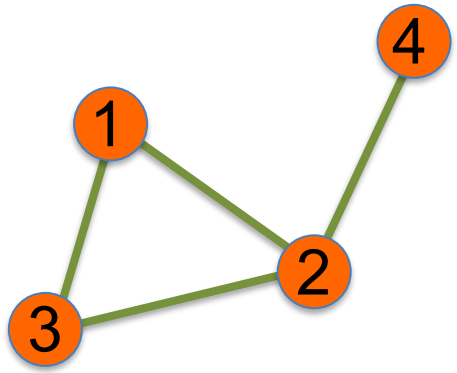
$$A_{ij} \neq A_{ji}$$

$$A_{ii} = 0$$

$$L = \sum_{j=1}^{N} k_j^{in} = \sum_{i=1}^{N} k_i^{out} = \sum_{ij} A_{ij}$$

Direct/Udnirect

Undirected graphs



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

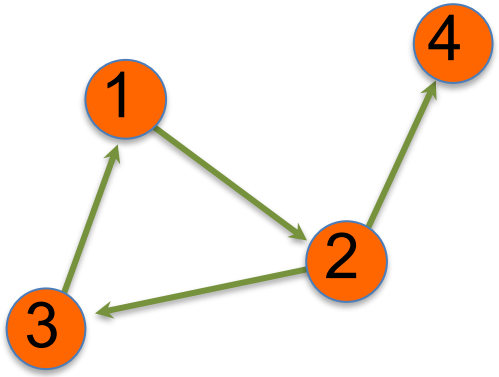
$$\mathrm{L} = \frac{1}{2}\sum_{\mathrm{i,j=1}}^{\mathrm{N}} \mathrm{A_{ij}} \qquad < \mathrm{k} > \frac{2L}{N}$$

$$A_{ij} = A_{ji}$$

$$A_{\mathrm{ii}} = 0$$

Actor network,
protein-protein interactions

Directed graphs



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathrm{L} = \sum_{\mathrm{i,j=1}}^{\mathrm{N}} \mathrm{A_{ij}} \qquad < \mathrm{k} > \frac{L}{N}$$

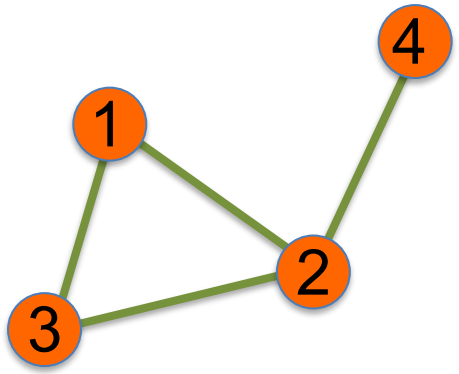
$$A_{ij} \neq A_{ji}$$

$$A_{\mathrm{ii}} = 0$$

WWW,
citation networks

Weights

Unweighted graphs



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

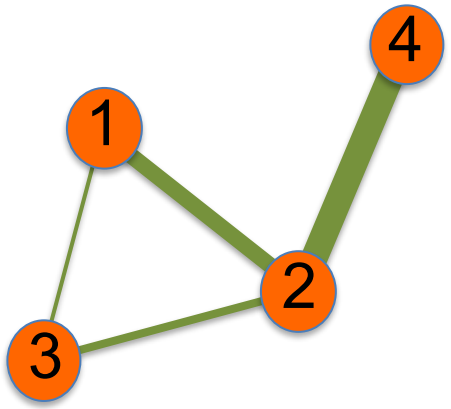
$$\mathrm{L} = \frac{1}{2}\sum_{\mathrm{i,j=1}}^{\mathrm{N}} \mathrm{A_{ij}} \qquad < \mathrm{k} > \frac{2L}{N}$$

$$A_{ij} = A_{ji}$$

$$A_{\mathrm{ii}} = 0$$

protein-protein interactions, WWW

Weighted graphs



$$A_y = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$\mathrm{L} = \frac{1}{2}\sum_{\mathrm{i,j=1}}^{\mathrm{N}} \mathrm{nonzero(A_{ij})} \qquad < \mathrm{k} > \frac{2L}{N}$$

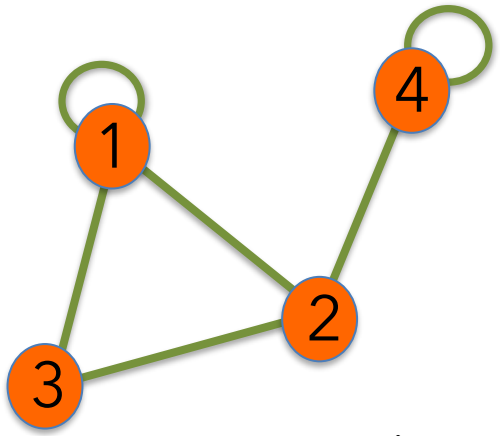
$$A_{ij} = A_{ji}$$

$$A_{\mathrm{ii}} = 0$$

Call Graph, metabolic networks

Loops and Multigraphs

Self interactions



$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

protein-protein interactions, WWW

$$L = \frac{1}{2}\sum_{i,j=1,i\neq j}^{N} A_{ij} + \sum_{i=1}^{N} A_{ij}$$

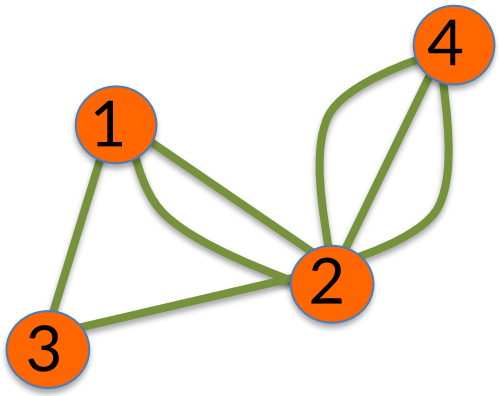
$A_{ij} = A_{ji}$

$A_{ii} \neq 0$?

Each ordinary edge is counted twice in the adjacency matrix

Each self-loop is counted only once on the diagonal

Multigraph (undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad < k > \frac{2L}{N}$$

$A_{ij} = A_{ji}$

$A_{ii} = 0$

Social Network, Collaboration Network

Real network examples

Network	Directed	Weighted	Multigraph	Self-loops
WWW	yes	no	yes	yes
Protein interactions	no	no	no	yes
Collaboration network	no	yes	yes	no
Mobile phone calls	yes	yes	no	no
Facebook Friendship	no	no	no	no

Real Networks can have multiple characteristics

Bipartite networks

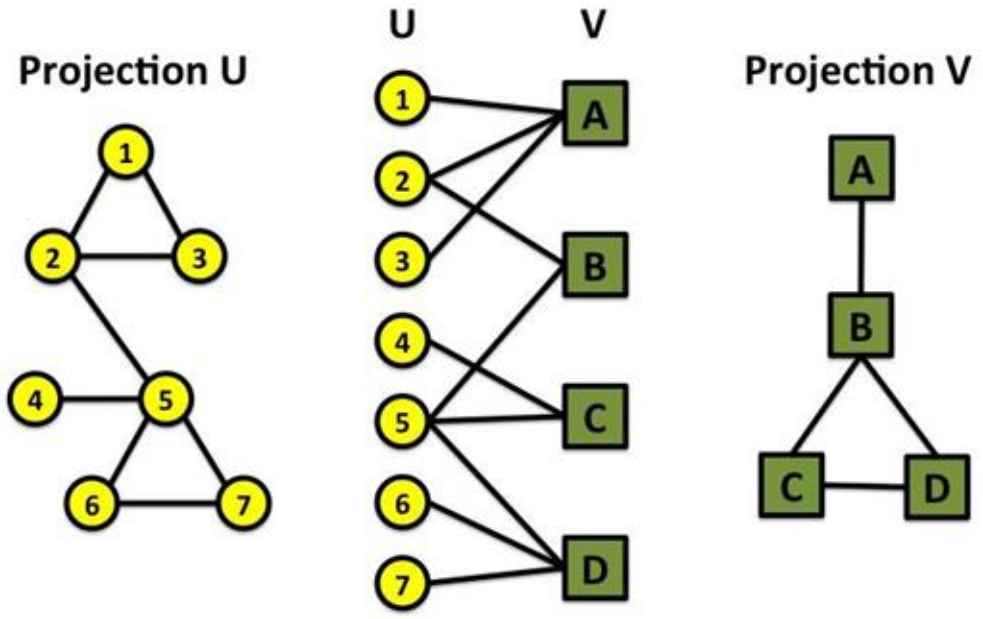
Bipartite networks

A graph whose nodes can be divided into two disjoint sets U and V such that every link connects a node in U to one in V.

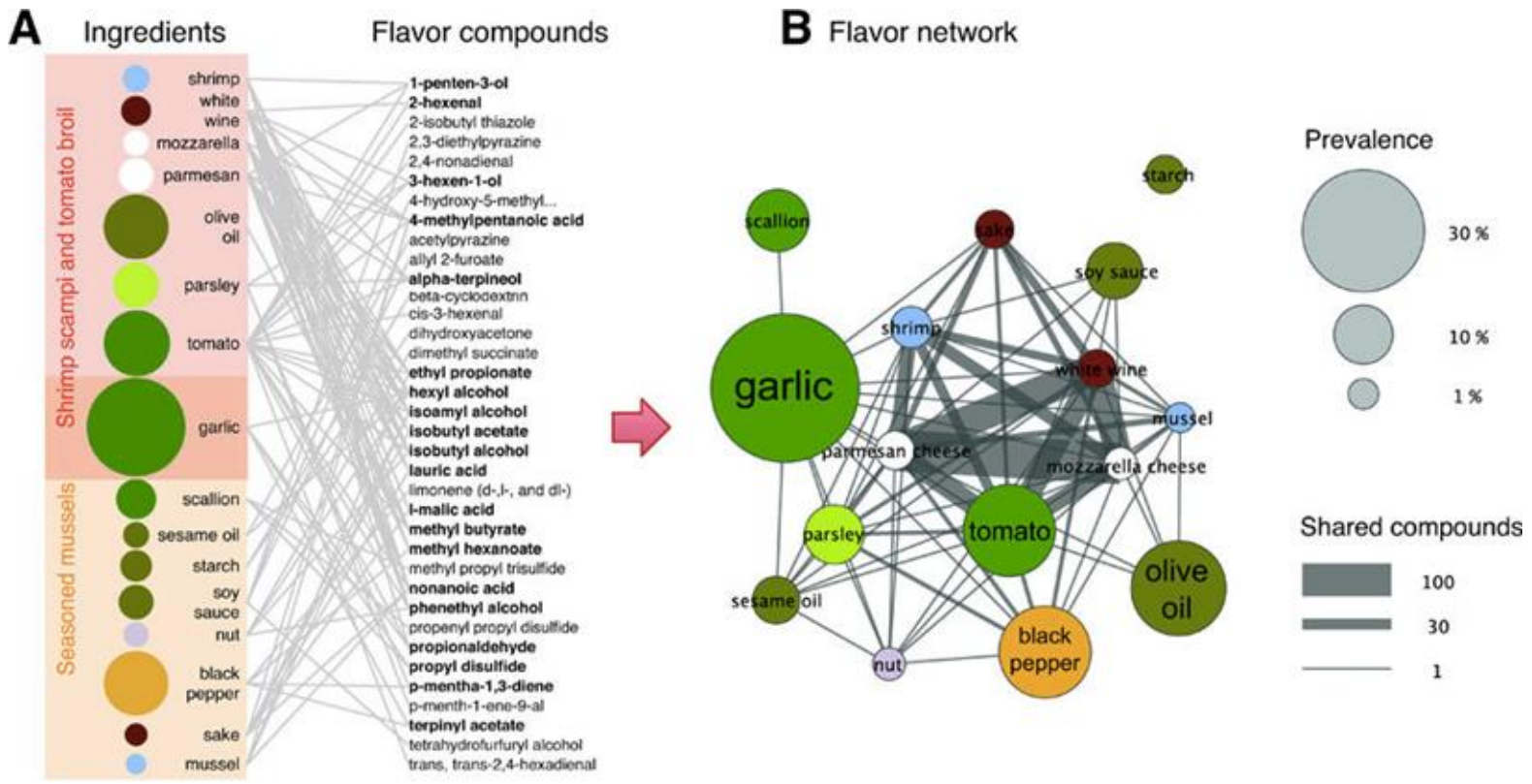
Projection: two nodes of the same class are connected by a (weighted) edge if they share at least a common neighbor.

Examples:

- Hollywood actor network
- Collaboration networks
- Disease network



Bipartite networks



Ingredient-Flavor Bipartite Network

Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, A.-L. Barabási, **Flavor network and the principles of food pairing**. *Scientific Reports* 196, (2011).

Network measures

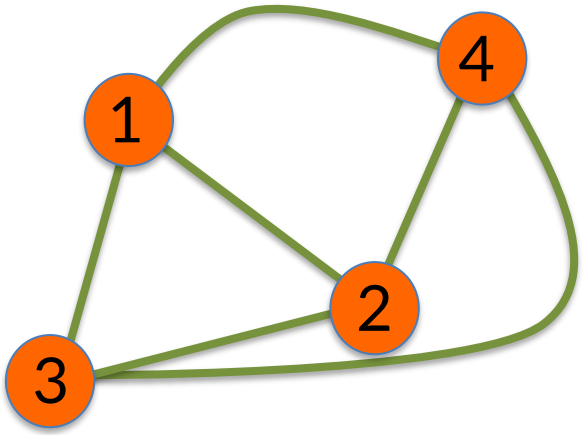
Network density

Complete graph

A graph with degree $L = L_{max}$ is called a complete graph, and its average degree is $< k >= N - 1$.

The maximum number of links an undirected network of N nodes can have:

$$L_{max} = \binom{N}{2} = \frac{N(N-1)}{2}$$



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Network density

Network density is defined as the ratio of existing edges over possible ones:

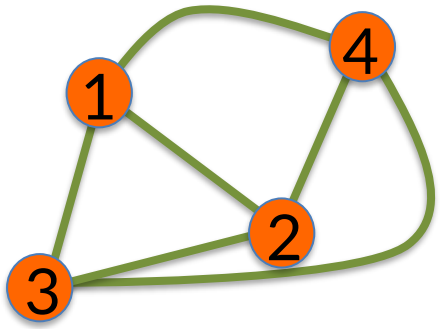
$$d(G) = \frac{L}{L_{max}}$$



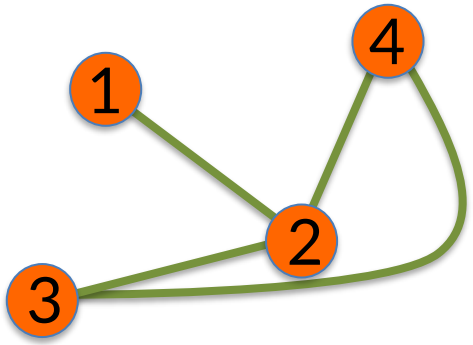
Low Density



High Density



Complete graph
$d(G) = 1$

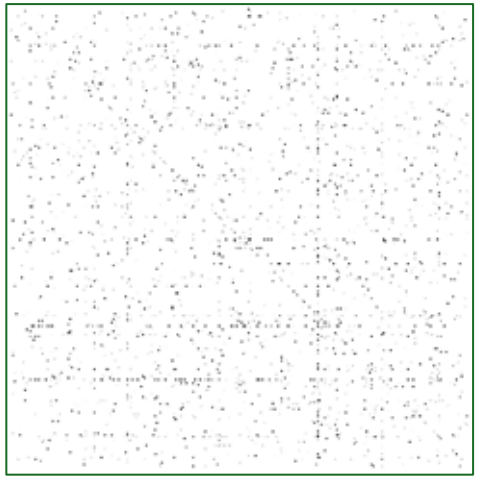


Sparse graph
$d(G) = \frac{4}{6} = \frac{2}{3}$

Real newtorks examples

Most networks observed in real systems are sparse:

- $L << L_{max}$
- $< k >=< N - 1$
- $d(G) << 1$



Sparse
Adjacency matrix

Network	**N**	**L**	L_{max}	**k**
WWW (ND Sample)	325,729	1.4 x 10^6	10^{12}	4.51
Protein (S. cerevisiae)	1,870	4,470	10^7	2.39
Coauthorship (Math)	70,975	2 x 10^5	3 x 10^{10}	3.9
Movie Actors	212,250	6 x 10^6	1.8 x 10^{13}	28.78

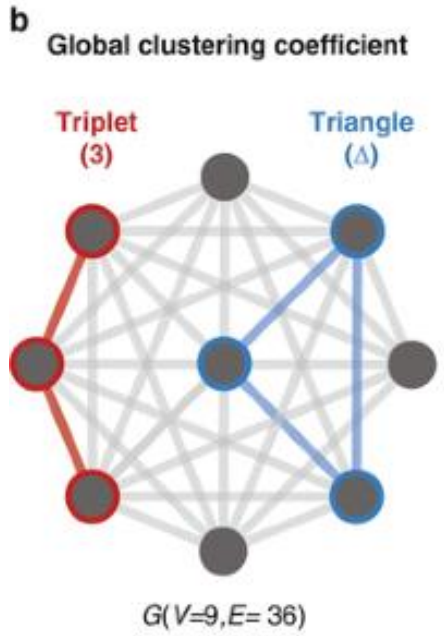
Clustering coefficient

Global clustering coefficient

Global Clustering coefficient:

$$C = \frac{3 \times number\ of\ triangles}{number\ of\ triplets}$$

- Triangles: a set of three distinct vertices that are all pairwise connected by three edges, forming a 3-cycle or a complete subgraph of size 3;
- Triplets: a set of three connected nodes
- C in$[0,1]$



Watts, D. J., & Strogatz, S. H. (1998). **Collective dynamics of 'small-world'networks**. *Nature*, 393(6684), 440-442.

Local clustering coefficient

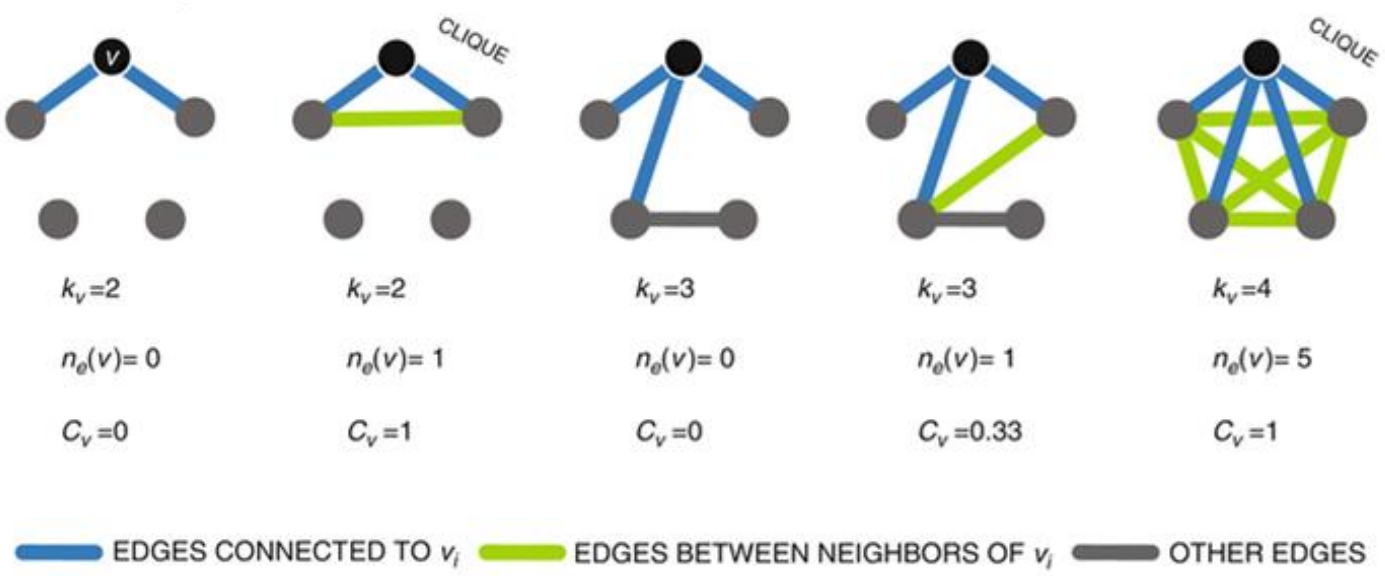
What fraction of your neighbors are connected?

Local Clustering coefficient:

$$C_i = \frac{2\mathrm{L}_i}{k_i(k_i - 1)}$$

- L_i: number of edges actually existing between the neighbors of node i
- k_i: degree of node i
- $\frac{k_i(k_i-1)}{2}$: maximum possible number of edges between its neighbors
- C_i in $[0,1]$

$$C_v = \frac{2\, n_e(v)}{k_v(k_v - 1)}$$

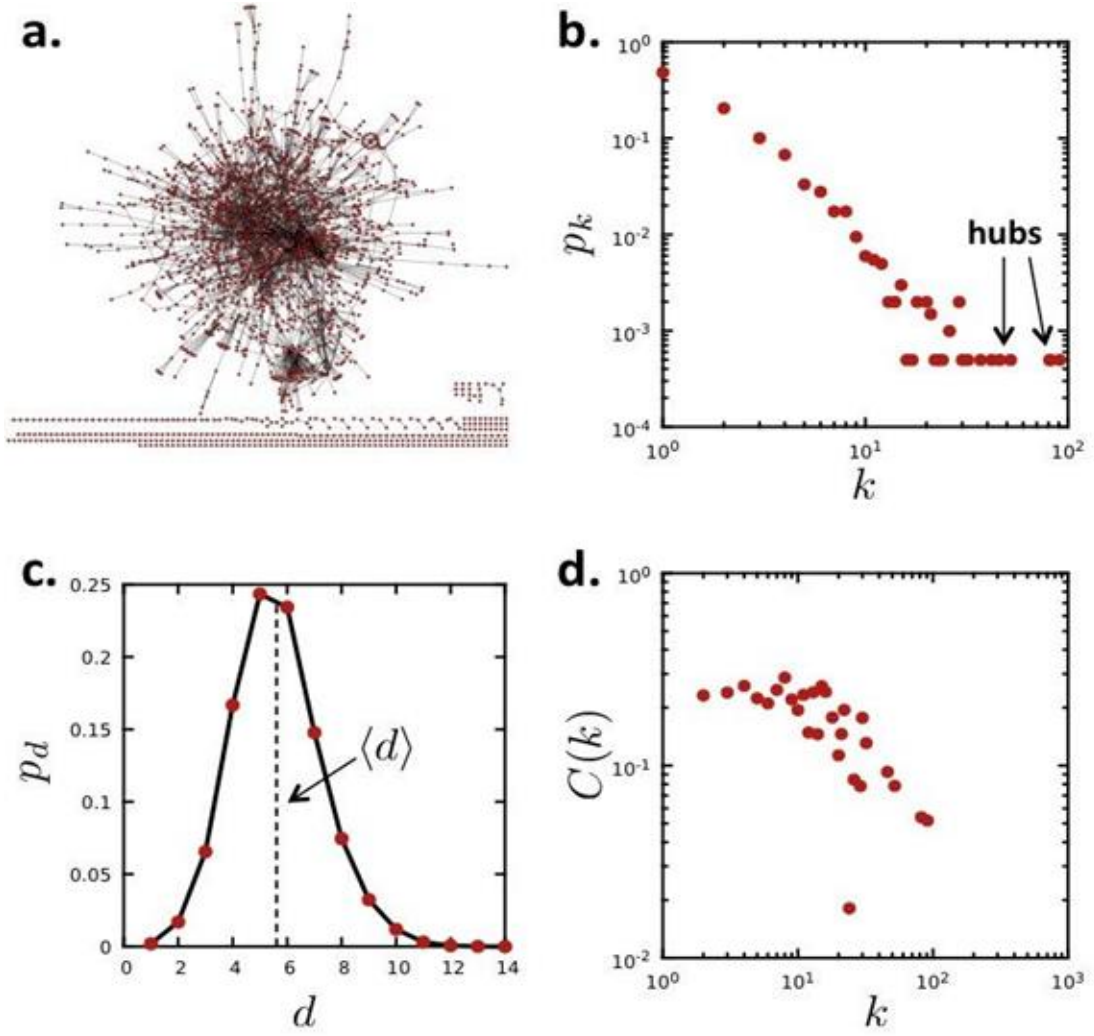


Case Study:
Protein-Protein Interaction Network

Protein-Protein Interaction Network

Undirected network
N=2,018 proteins as nodes
L=2,930 binding interactions as links.
Average degree <k>=2.90.

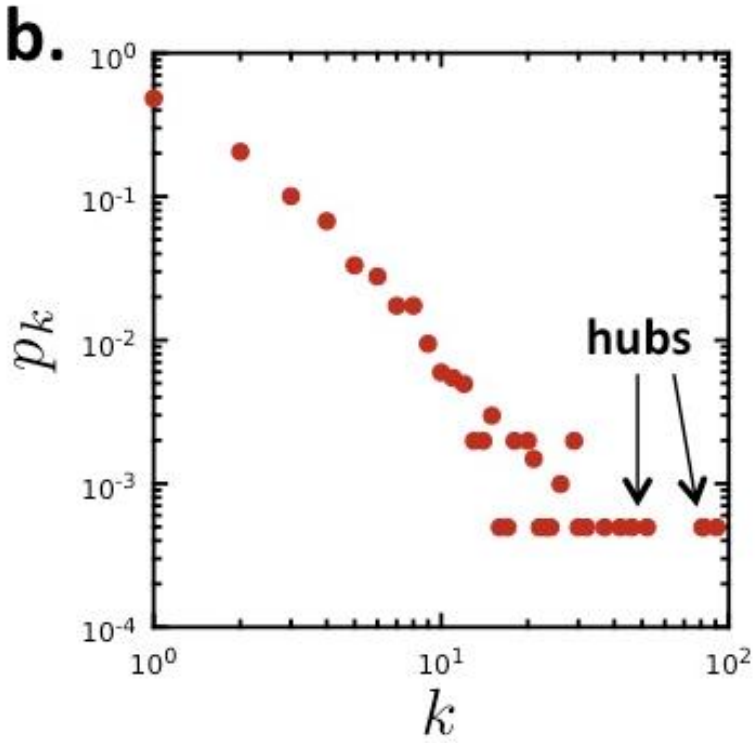
Not connected: 185 components
Largest (giant component) 1,647 nodes



Degree distribution

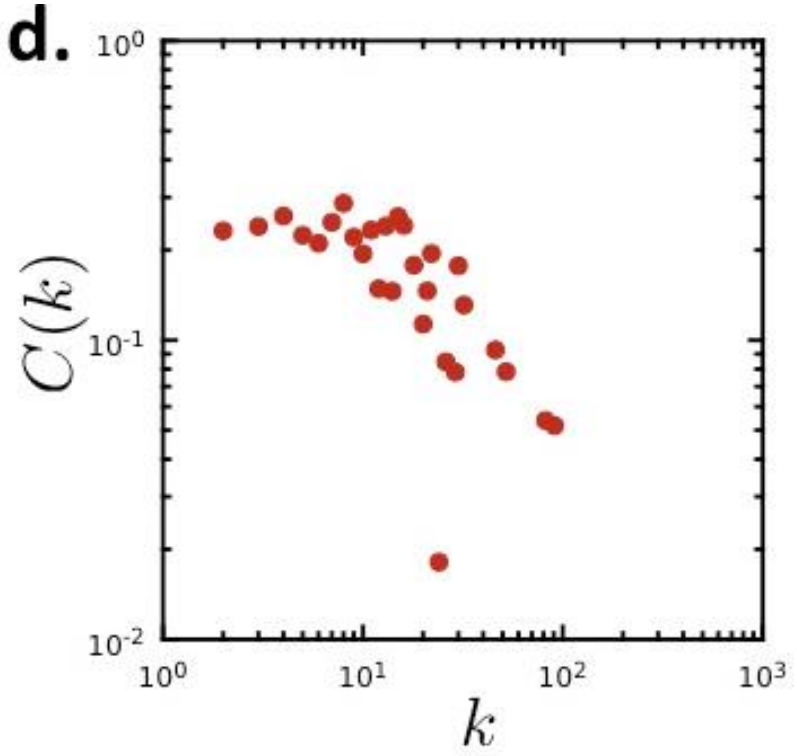
$$P_k = \frac{N_k}{N}$$

- p_k is the probability that a node has degree k
- N_k = number of nodes with degree k



Clustering Coefficiente VS. Node Degree

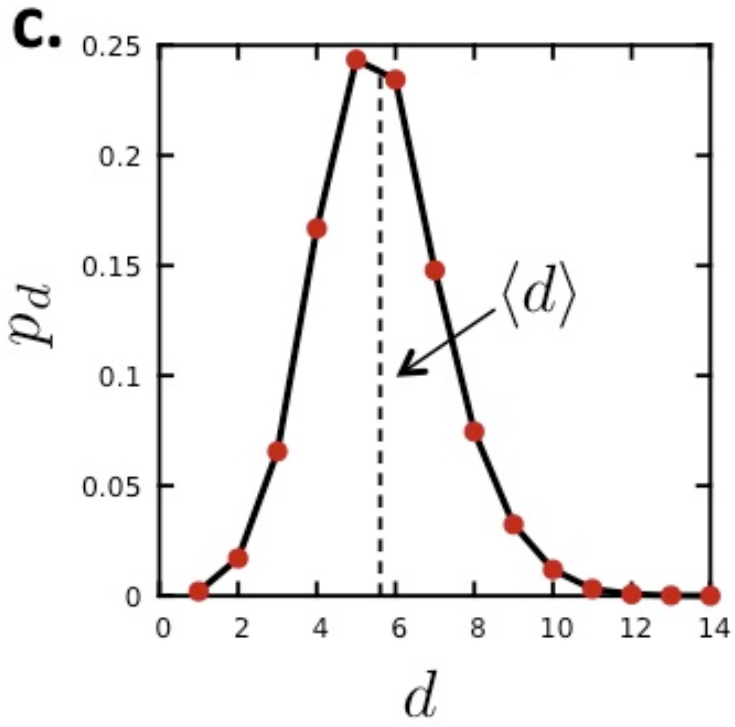
Average Clustering Coeff: <C>=0.12



Diameter

Path length distribution:

- $d_{max} = 14$
- $< d >= 5.61$



Centrality measures

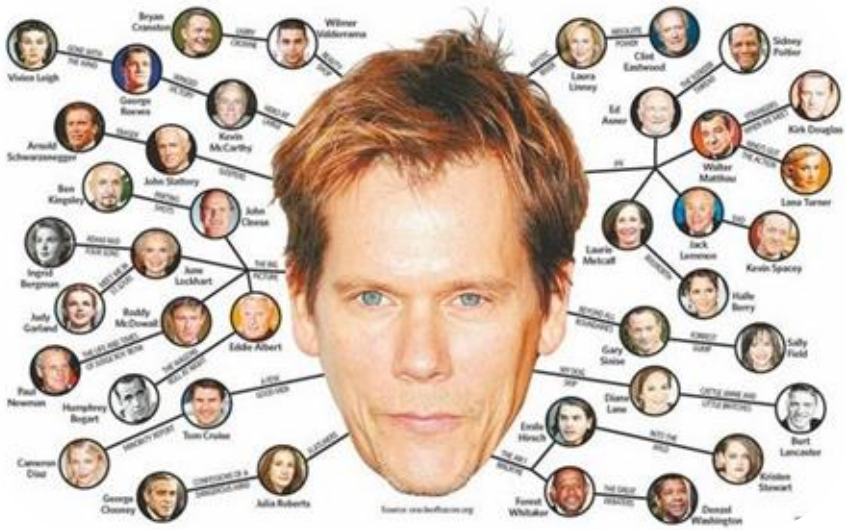
Centrality measure

We can measure nodes importance using so-called centrality.

Bad term:
nothing to do with being central in general

Usage:

- Some centralities have straightforward interpretation
- Centralities can be used as node features for machine learning on graph

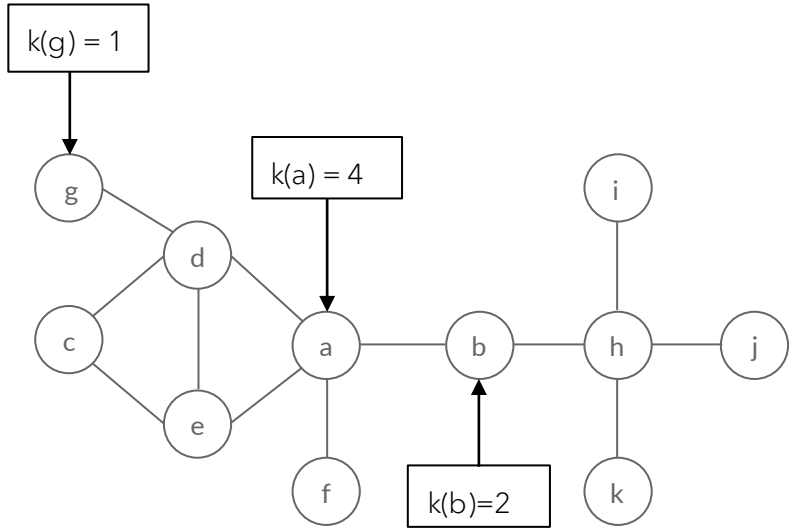


Degree centrality

How many neighbors does a node have?

Often enough to find important nodes (e.g., main characters of a tv series talk with more people)

- Twitter users with the most contacts are spam
- Webpages/wikipedia pages with most links are often lists of references



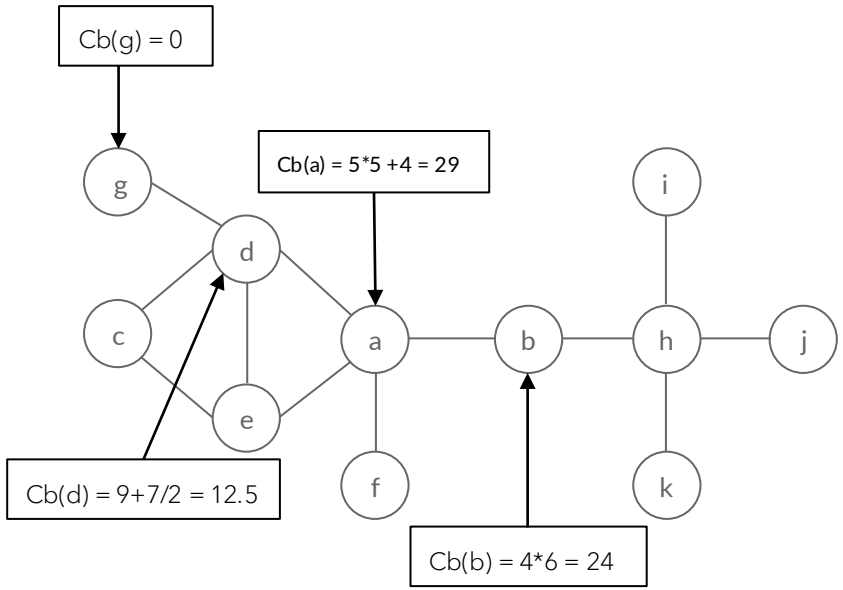
$$A_{i,j} = \begin{cases} 1, & \text{if i and j are connected} \\ 0 & \text{otherwise} \end{cases}$$

$$k_i = \sum_{j=1}^{N} A_{ij}$$

Betweeness centrality

Number of shortest paths that go through a node.

- Assumption: important vertices are bridges over which information flows
- Practically: if information spreads via shortest paths, important nodes are found on many shortest paths.



Centrality measures



- Each centrality measures is a proxy of an underlying network process.
- If such a process is irrelevant for the network than the centrality measure makes no sense
- E.g. If information does not spread through shortest paths, betweenness centrality is irrelevant
- Centrality measures should be used with caution (a) for exploratory purposes and (b) for characterization

Assortativity

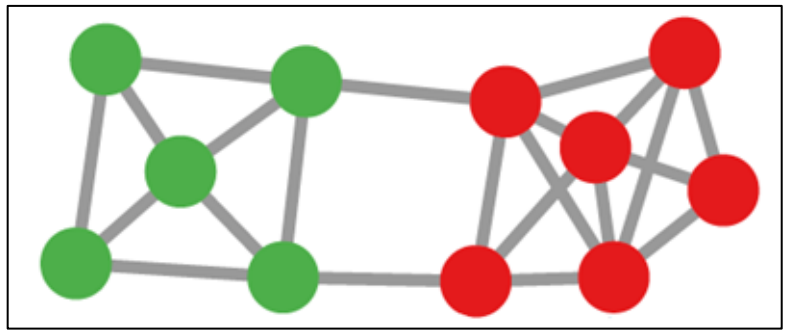
Homophily is the tendency of individuals (or nodes in a network) to establish connections with others who share similar characteristics, attributes, or behaviors.

Assortativity quantifies homophily:

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$$

where:

- e_{ij} fraction of links connecting nodes of type i and j
- a_i fraction of out-links from nodes of type a
- b_i fraction of in-links for type b nodes



Interpretation:

- $r = 0$: no assortative mixing
- $r = 1$: perfectly assortative
- $-1 < r < 0$: disassortative mixing

Newman, Mark EJ. **Mixing patterns in networks**. *Physical Review E* 67.2 (2003): 026126.

Assortativity

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$$

where:

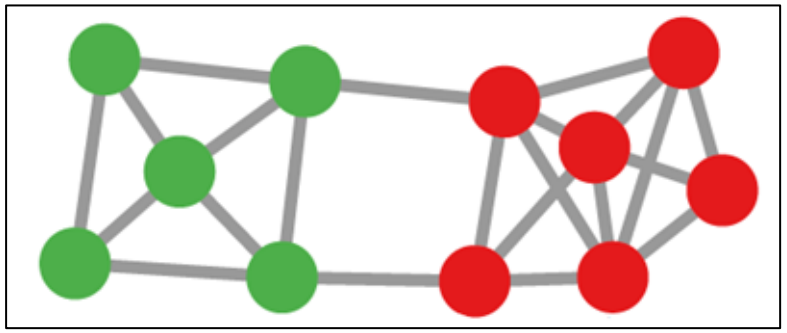
- e_{ij} fraction of links connecting nodes of type i and j
- a_i fraction of out-links from nodes of type a
- b_i fraction of in-links for type b nodes

a_g is the fraction of outgoing links when each undirected edge is considered in both directions.
For every green-green edge, both directed versions originate from a green node. That is why the 8 green-green edges contribute 16 rather than 8.
For every green-red edge, only one of its two directed versions originates from a green node.

$$a_G = \frac{2L_{GG} + L_{GR}}{2L}$$

$$a_R = \frac{2L_{RR} + L_{GR}}{2L}$$

a_i is the probability that, if we select a random edge and randomly choose one of its two endpoints, that endpoint belongs to category i.



$N = 11$
$L = 22$
$L_{gg} = 8$
$L_{rr} = 12$
$L_{gr} = 2$

In undirected network $a_i = b_i$



$$r = \frac{\left(\left(\frac{8}{22}\right) + \left(\frac{12}{22}\right) - \left(\left(\frac{9}{22}\right)^2 + \left(\frac{13}{22}\right)^2\right)\right)}{1 - \left(\left(\frac{9}{22}\right)^2 + \left(\frac{13}{22}\right)^2\right)} = \sim 0.812$$

For green nodes, we sum the first row:

$$a_G = e_{GG} + e_{GR} = \frac{16}{44} + \frac{2}{44} = \frac{18}{44} = \frac{9}{22}$$

For red nodes, we sum the second row:

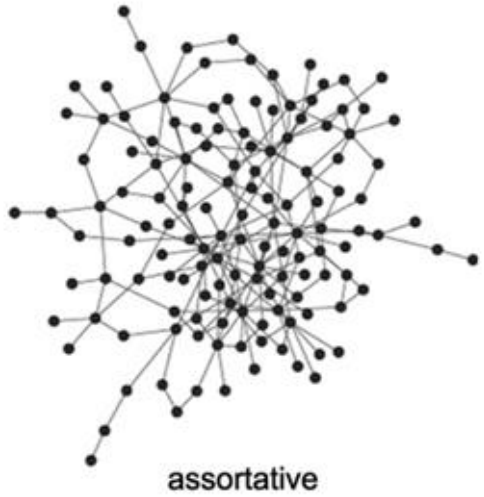
$$a_R = e_{RG} + e_{RR} = \frac{2}{44} + \frac{24}{44} = \frac{26}{44} = \frac{13}{22}$$

Newman, Mark EJ. **Mixing patterns in networks**. *Physical Review E* 67.2 (2003): 026126.

Assortativity

Assortative

Nodes tends to connect homogeneously w.r.t. their degree (e.g., hubs with hubs)



assortative

Disassortative

Nodes tends to connect in a star-like topology



disassortative

Newman, Mark EJ. **Mixing patterns in networks**. *Physical Review E* 67.2 (2003): 026126.

Community discovery

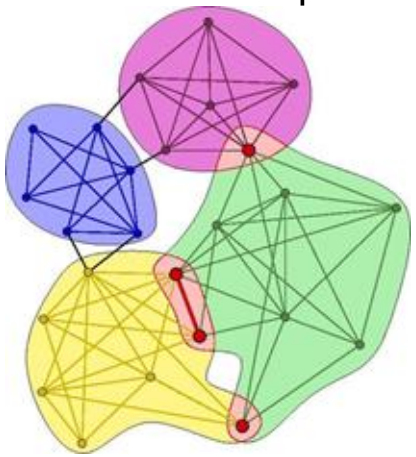
Ill-posed problem

Community discovery

The aim of Community Discovery algorithms is to identify meso-scale topologies hidden within complex network structures, i.e., cluster homogeneous nodes relying on topological information.

Major Problems:

- Community Discovery is an ill posed problem
- Each algorithm models different properties of communities
- Different approaches comparison
- Context Dependency



S. Fortunato. **Community detection in graphs**. *Physics Reports* 486 (2010).

What is a community?

Unfortunately, does not exist a universally shared definition of what a community is…

A general idea is that a community should represent:

- "A set of entities where each entity is closer, in the network sense, to the other entities within the community than to the entities outside it."

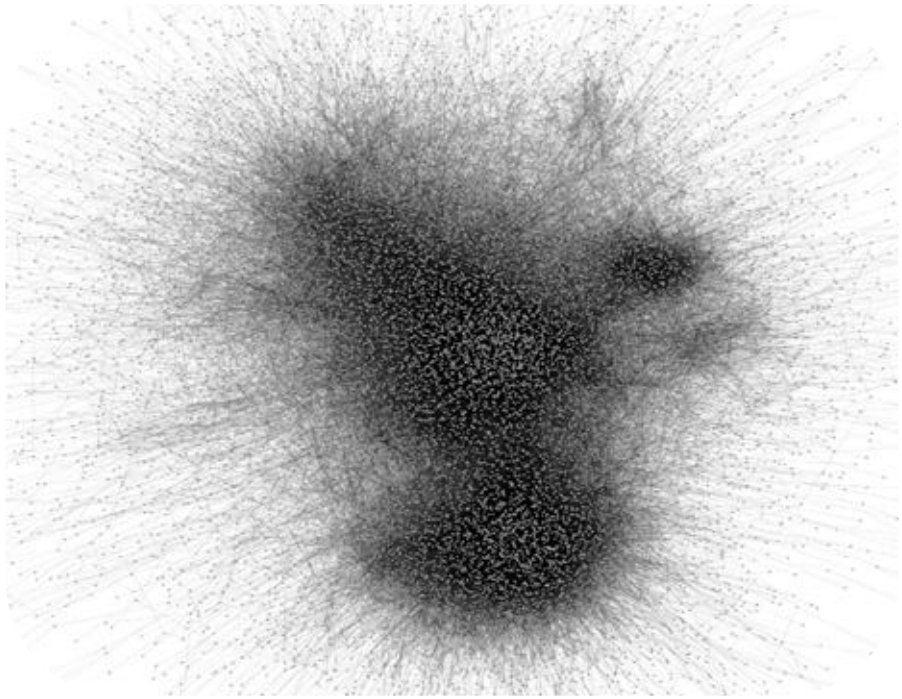
or, equivalently

- "A set of nodes more tightly connected within each other than with nodes belonging to other sets."

Community discovery in real networks

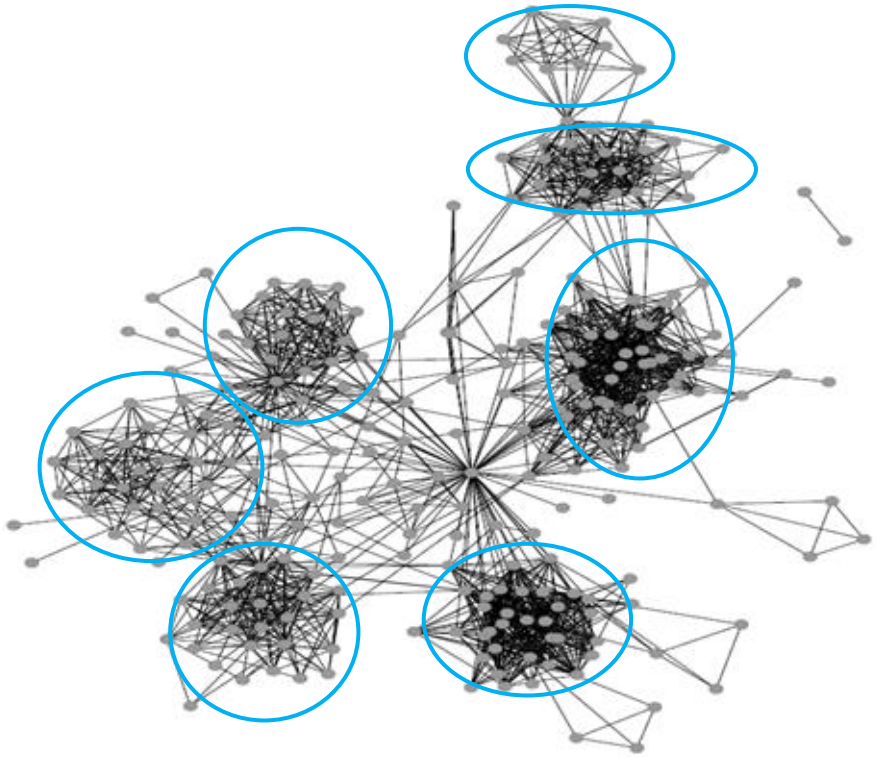
Real world networks? Too complex for visual analysis, so we can't easily identify (e.g., visually) different communities.

We need automated procedures!

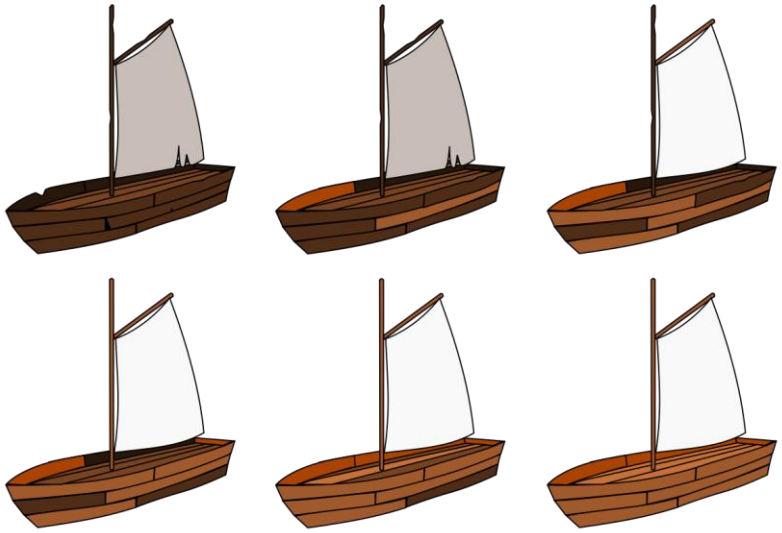


Community discovery in small networks

In simple, small, networks it is easy identify them by looking at the structure,i.e., using a Force directed layout.



Ship of Theseus Paradox



According to the ancient Greek thought experiment,
Theseus owned a ship that was preserved for many years.

Over time, its wooden planks began to deteriorate, so they were gradually replaced with new ones.

Eventually, every original plank had been replaced.

Is the restored ship still the Ship of Theseus?
There are two competing interpretations:

- Continuity: Yes, because the ship was gradually repaired and maintained its identity throughout the process.
- Composition: No, because none of the original material remains.

The paradox becomes even more interesting if someone collects the original planks and rebuilds another ship. Which of the two is the real Ship of Theseus?
There is no universally accepted answer because identity depends on the criteria we choose.

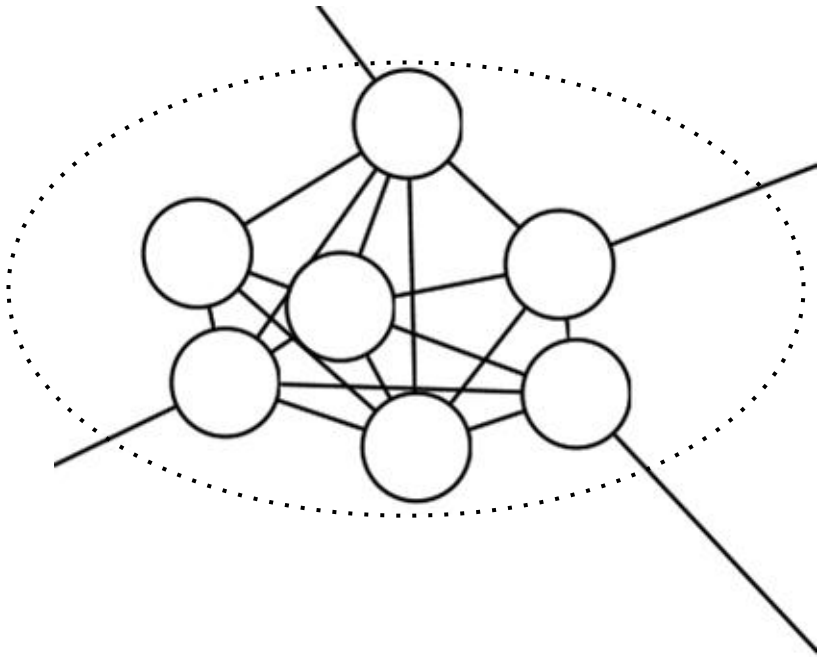
S. Fortunato. **Community detection in graphs**. *Physics Reports* 486 (2010).

Algorithm

Internal density

"Communities as a sets of densely connected entities".
Each community must have a number of edges significantly higher than what expected in a random graph.

Algorithms in this family: Greedy Modularity, Louvain, ...



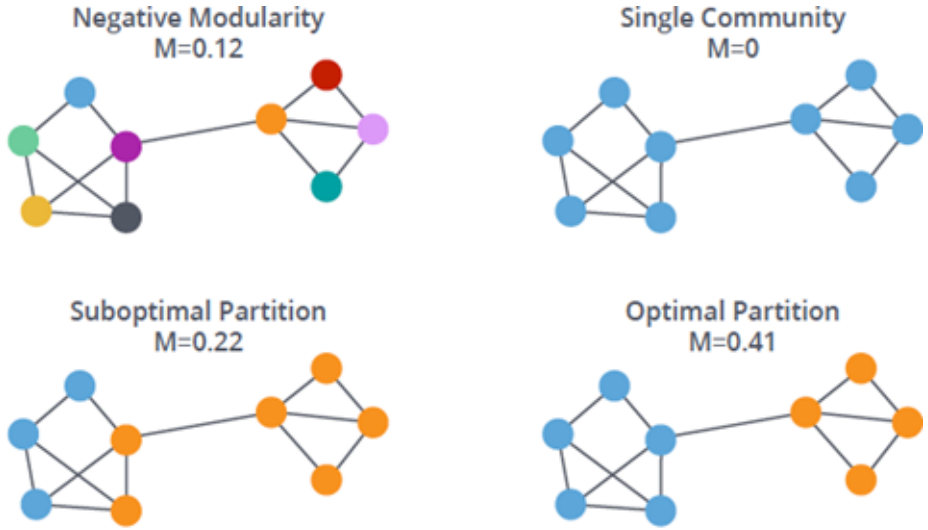
Modularity

How to assure high density?
Define a quality function that measures the density of a community and then try to maximize it, modularity

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j)$$

- Normalization factor
- Null Model expected density
- 1 if i, j in same community, 0 otherwise

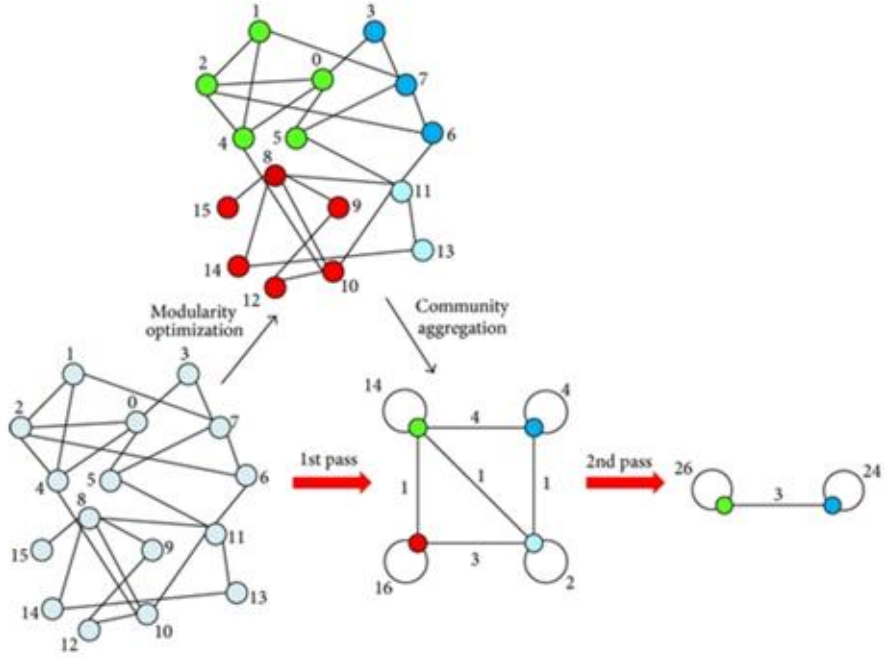
- A_{ij}: adjacency matrix. It is 1 if nodes i and j are connected, 0 otherwise
- k_i: degree of node i
- k_j: degree of node j
- m: total number of edges in the network, $2m = \sum_i k_i$
- c_i: community containing node i
- $\delta(c_i, c_j)$: equals 1 if i and j belong to the same community, and 0 otherwise
- Q in$[-1,1]$



Louvain

In order to maximize this value efficiently, the Louvain Method has two phases that are repeated iteratively.

- Initialization: Each node in the network is assigned to its own community.
- Phase 1: Each node is then moved into the adjacent community that guarantee the greatest modularity increase.
- Phase 2: A new graph is created: its nodes are the updated communities and weighted links connect them accounting for bridges in the original graph.
- Phases 1 and 2 are repeated until modularity is maximized.



Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). **Fast unfolding of communities in large networks**. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.

References

1. Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, A.-L. Barabási, **Flavor network and the principles of food pairing**. *Scientific Reports* 196, (2011).
2. Watts, D. J., & Strogatz, S. H. (1998). **Collective dynamics of 'small-world'networks**. *Nature*, 393(6684), 440-442.
3. Newman, Mark EJ. **Mixing patterns in networks**. *Physical Review E* 67.2 (2003): 026126.
4. S. Fortunato. **Community detection in graphs**. *Physics Reports* 486 (2010).
5. M. Coscia, F. Giannotti, and D. Pedreschi. **A classification for community discovery methods in complex networks**. *Statistical Analysis and Data Mining* 4, 5 (2011), 512–546.
6. Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). **Fast unfolding of communities in large networks**. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.