DESCRIPTIVE NETWORK ANALYSIS

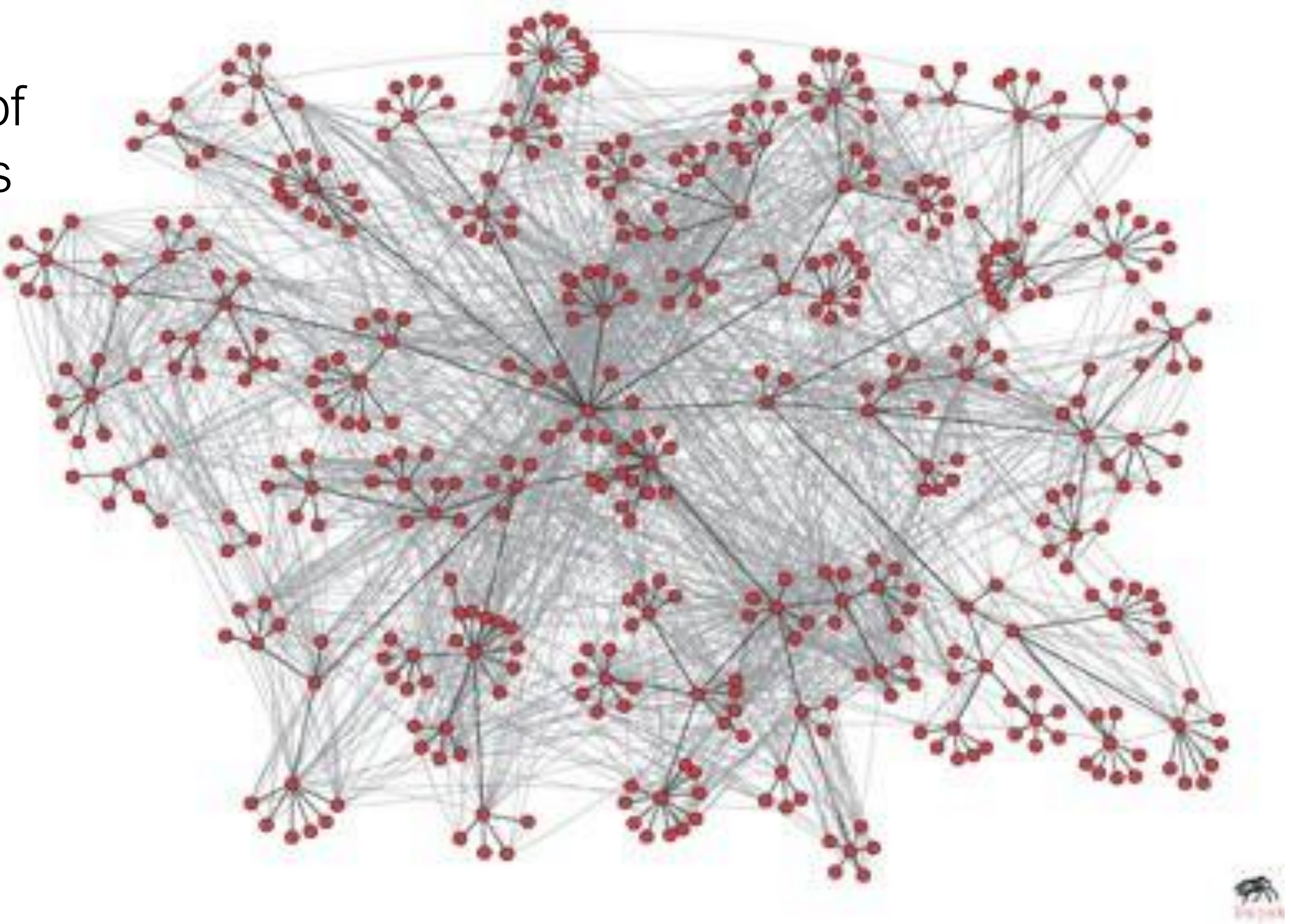
Ilia Karpov (karpovilia@gmail.com)

Moscow, 2023

BASIC CONCEPTS

STRUCTURE OF NETWORKS

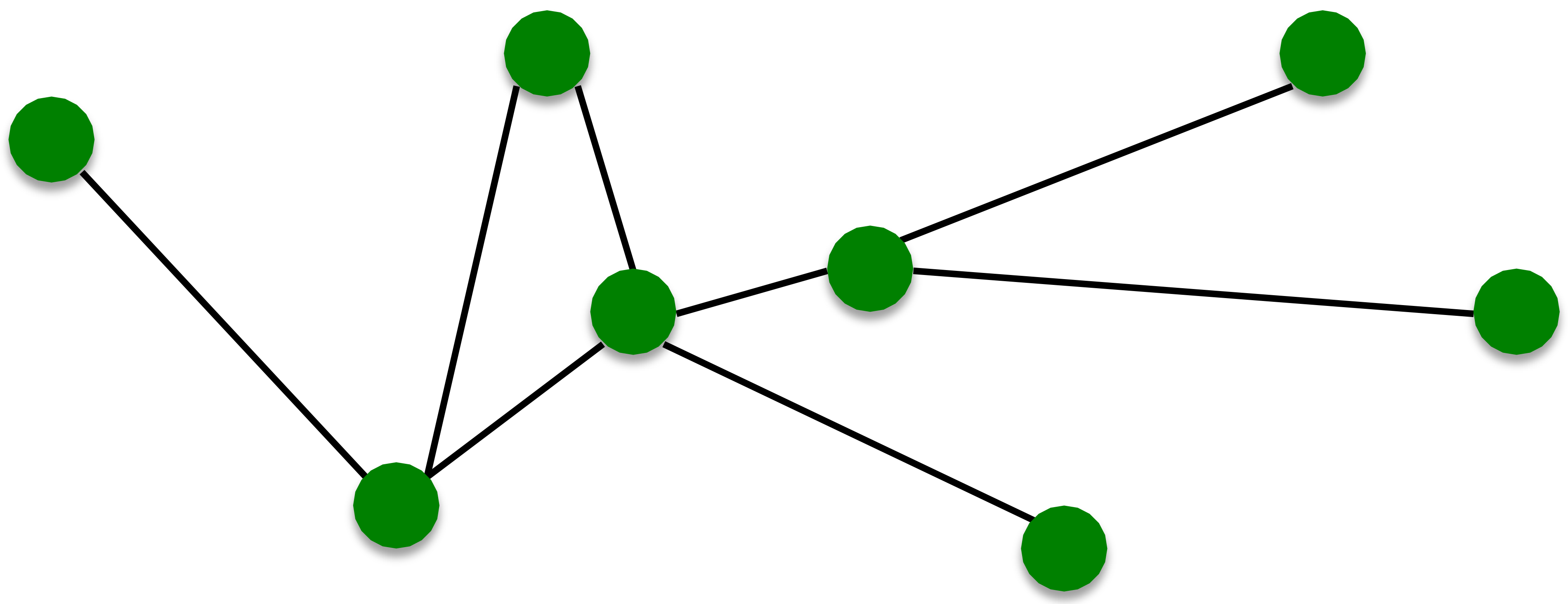
- A network is a collection of objects where some pairs of objects are connected by links



BASIC CONCEPTS

COMPONENTS OF A NETWORK

- Objects: nodes, vertices $\boldsymbol{N}$
- Interactions: links, edges $\boldsymbol{E}$
- System: network, graph $\boldsymbol{G(N,E)}$



A ***graph*** $G = (V, E)$ is an ordered pair of sets: a set of vertices V and a set edges E, where $n = |V|$, $m = |E|$

An ***edge*** $e_{ij} = (v_i, v_j)$ is pair of vertices (ordered pair for directed graph)

BASIC CONCEPTS

COMPONENTS OF A NETWORK

- Network often refers to real systems

  Web, Social network, Metabolic network

  Language: Network, node, link

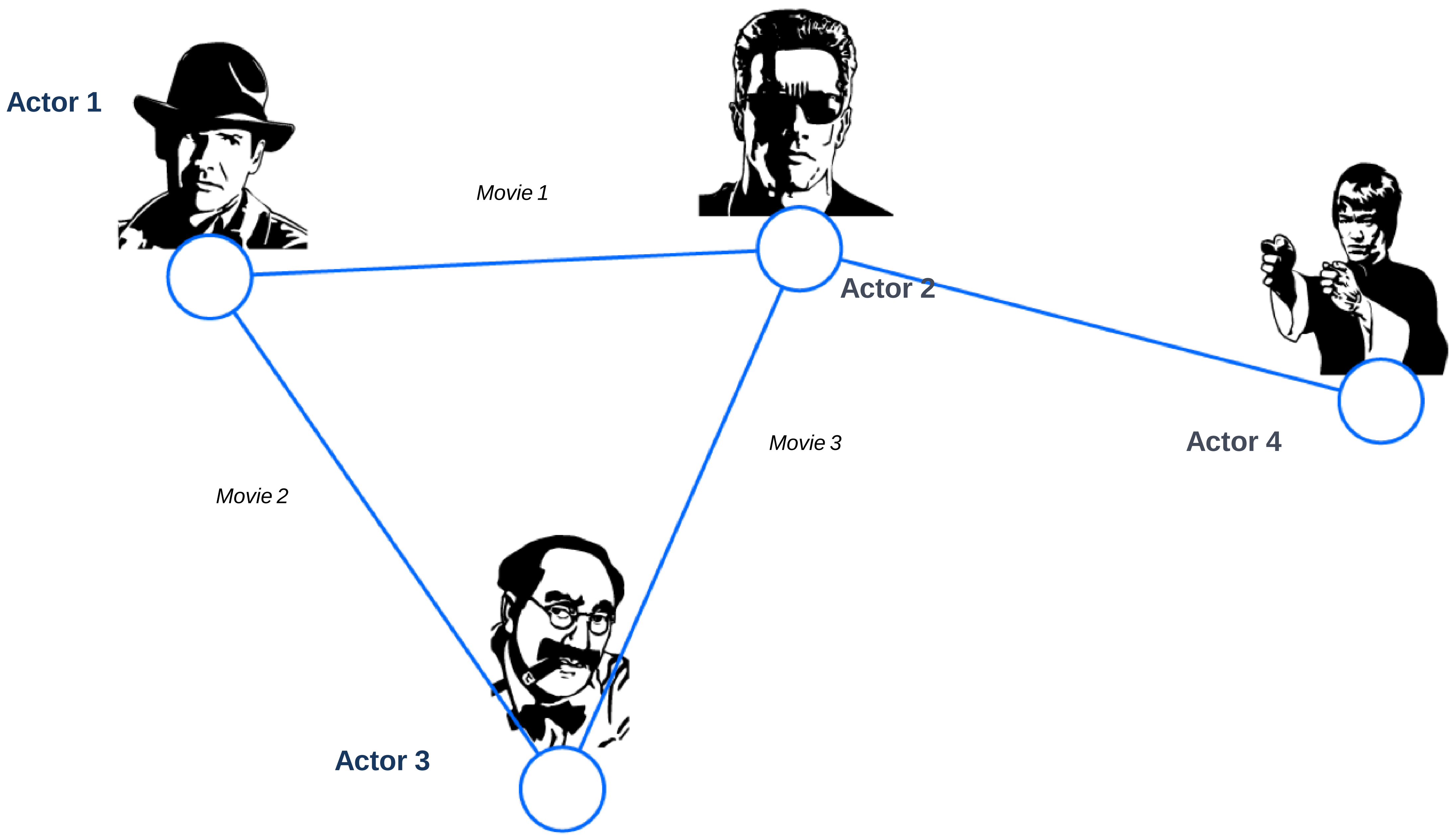
- Graph is a mathematical representation of a network

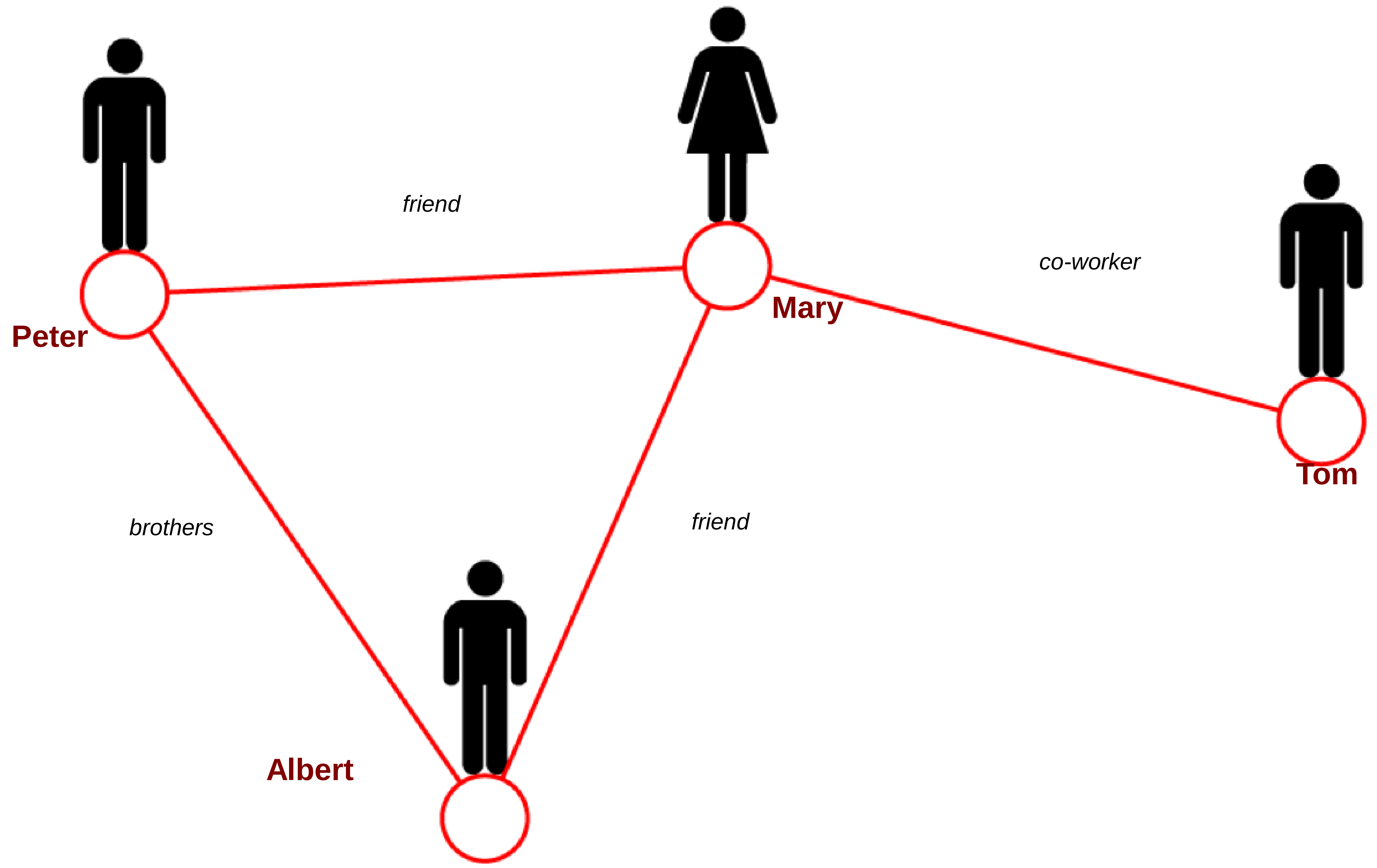
  Web graph, Social graph (a Facebook term)

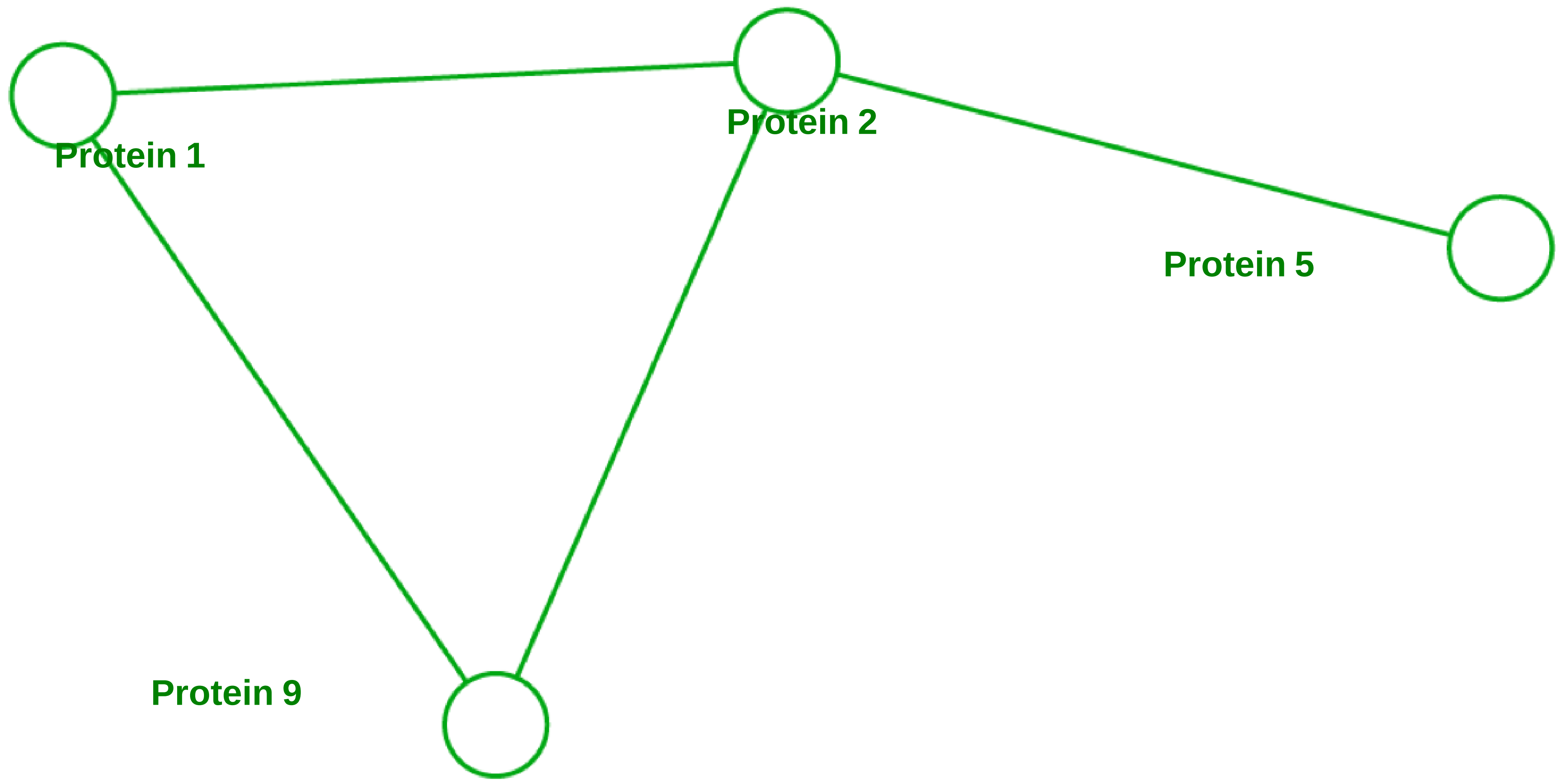
  Language: Graph, vertex, edge

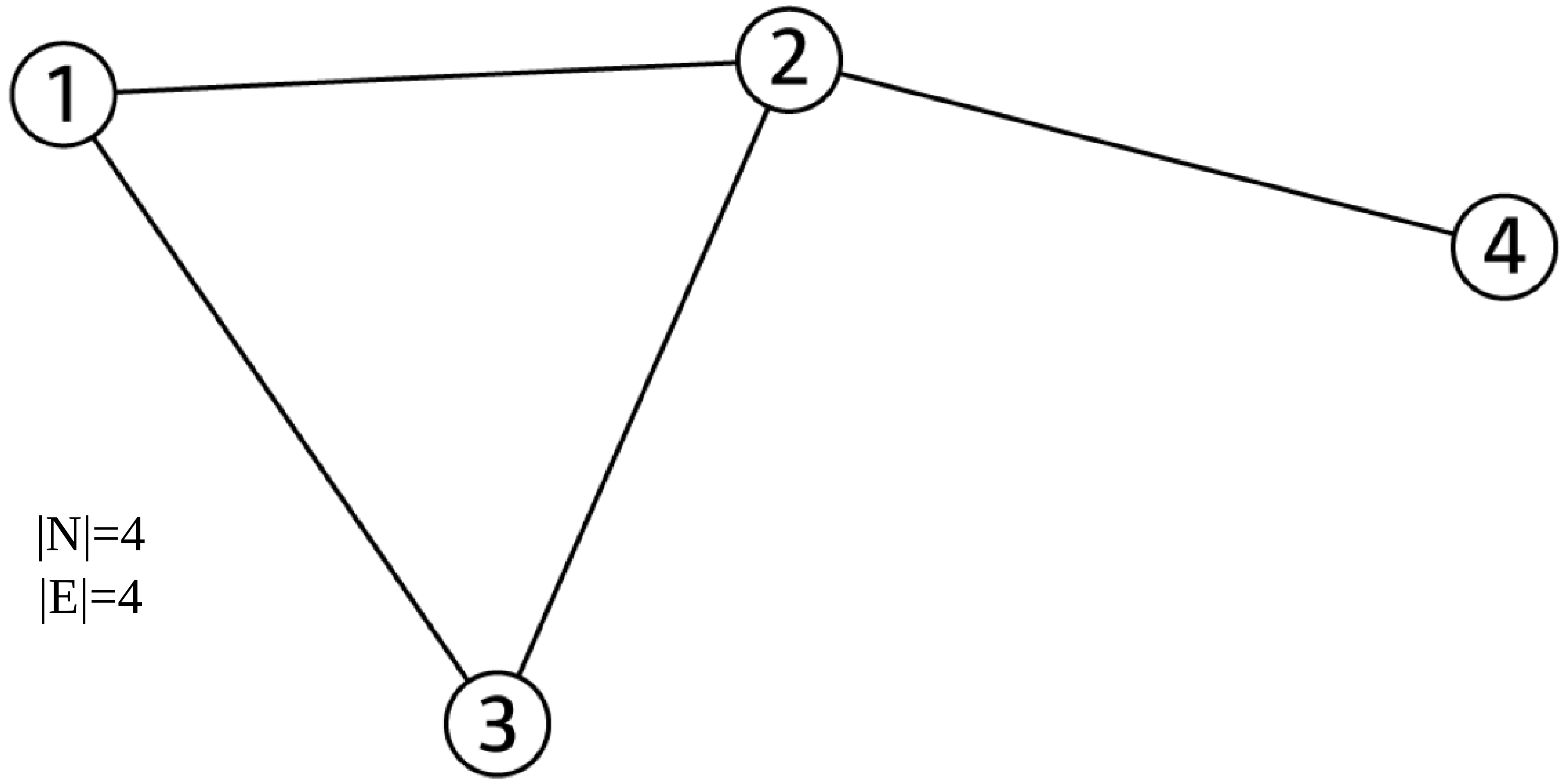
BASIC CONCEPTS

COMPONENTS OF A NETWORK









$|N|=4$
$|E|=4$

BASIC CONCEPTS

HOW TO DEFINE A NETWORK

- How to build a graph:
  What are nodes?
  What are edges?

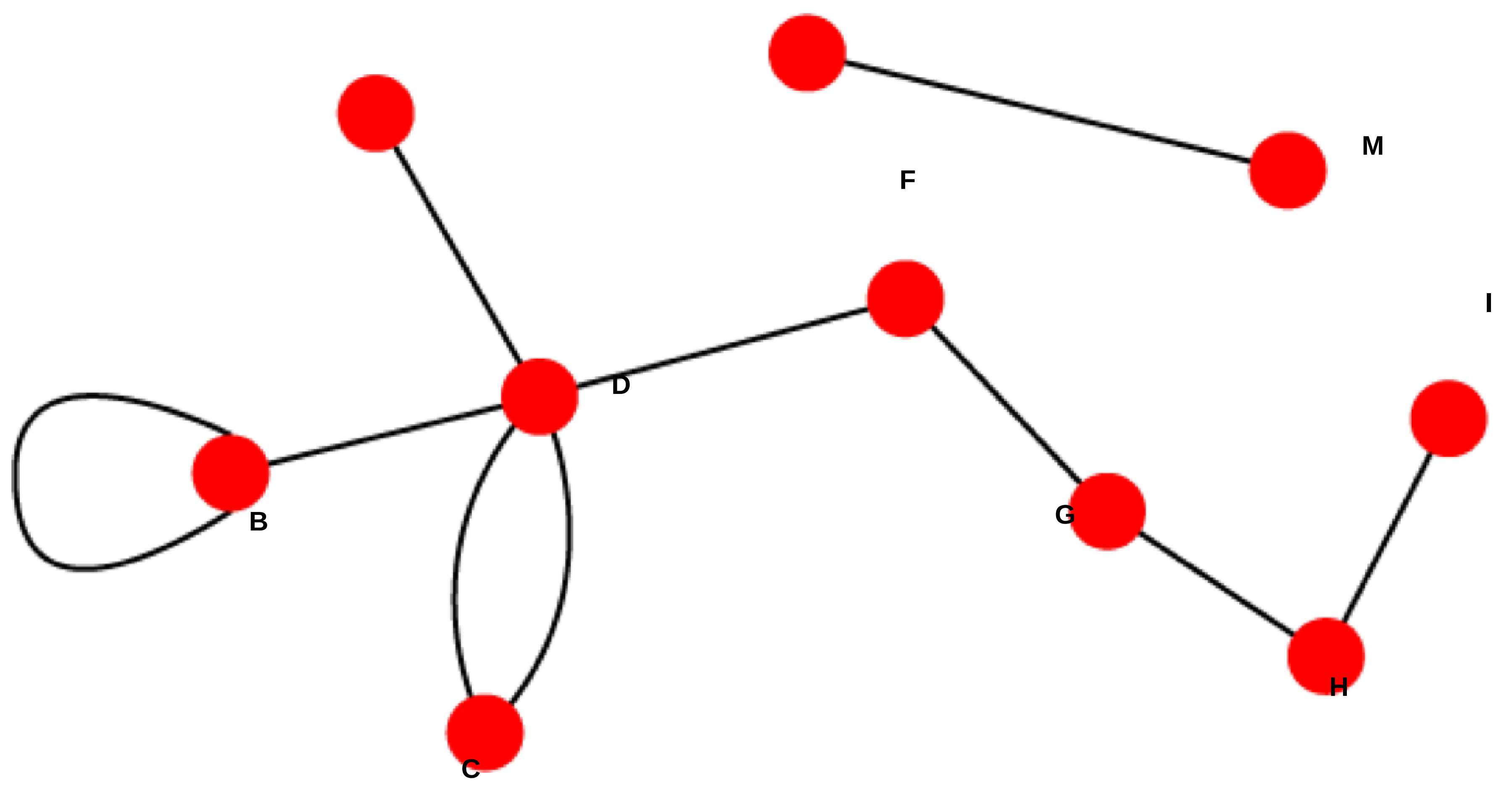
- Choice of the proper network representation of a given domain/problem determines our ability to use networks successfully:
  In some cases there is a unique, unambiguous representation
  In other cases, the representation is by no means unique
  The way you assign links will determine the nature of the question you can study

BASIC CONCEPTS

DIRECTED VS. UNDIRECTED GRAPHS

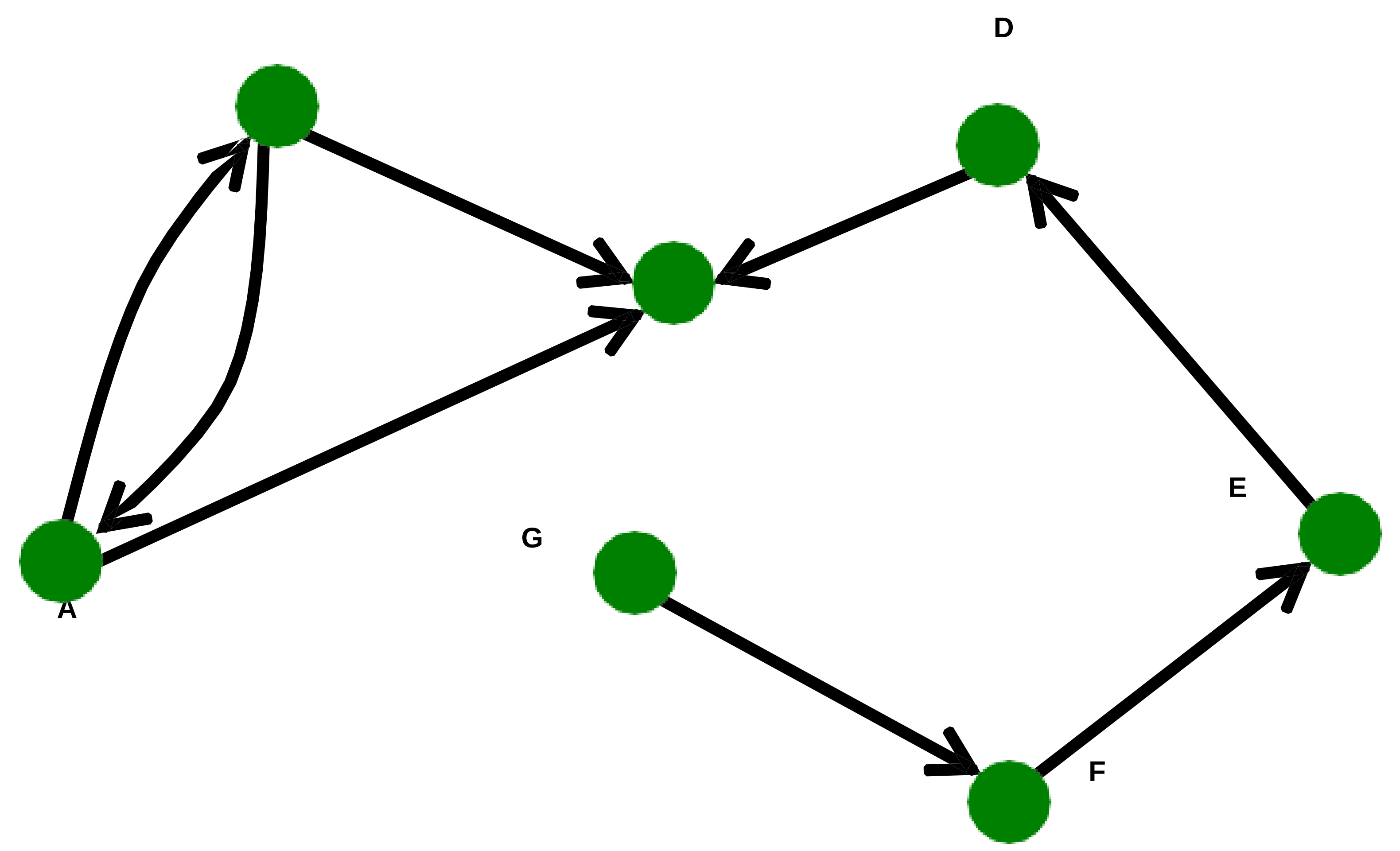
- Undirected
  Links: undirected (symmetrical, reciprocal)

- Examples:
  Collaborations
  Friendship on Facebook



- Directed
  Links: directed (arcs)

- Examples:
  Phone calls
  Following on Twitter



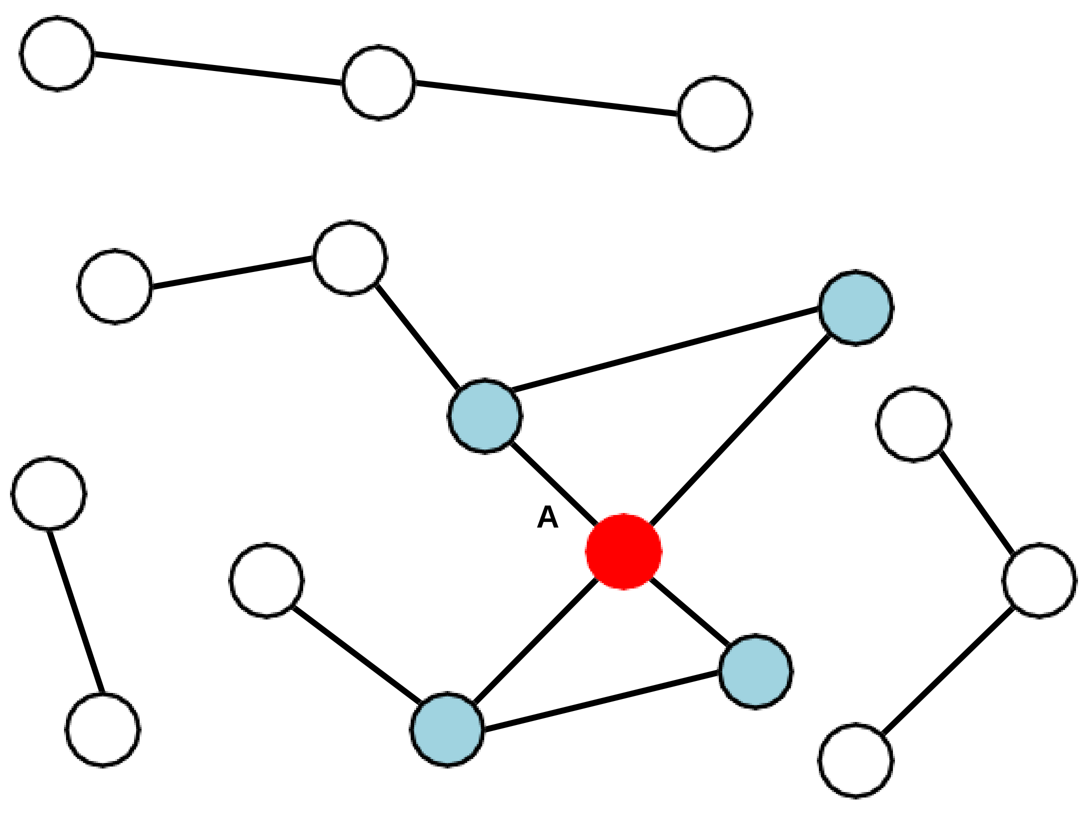
Two nodes/vertices are *adjacent* if they share a common edge
An edge and a node on that edge are called *incident*.

BASIC CONCEPTS
NODE DEGREES

- Undirected
- Node degree, k_i: the number of edges adjacent to node i

$k_A = 4$

Avg. degree: $\bar{k} = \langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2E}{N}$



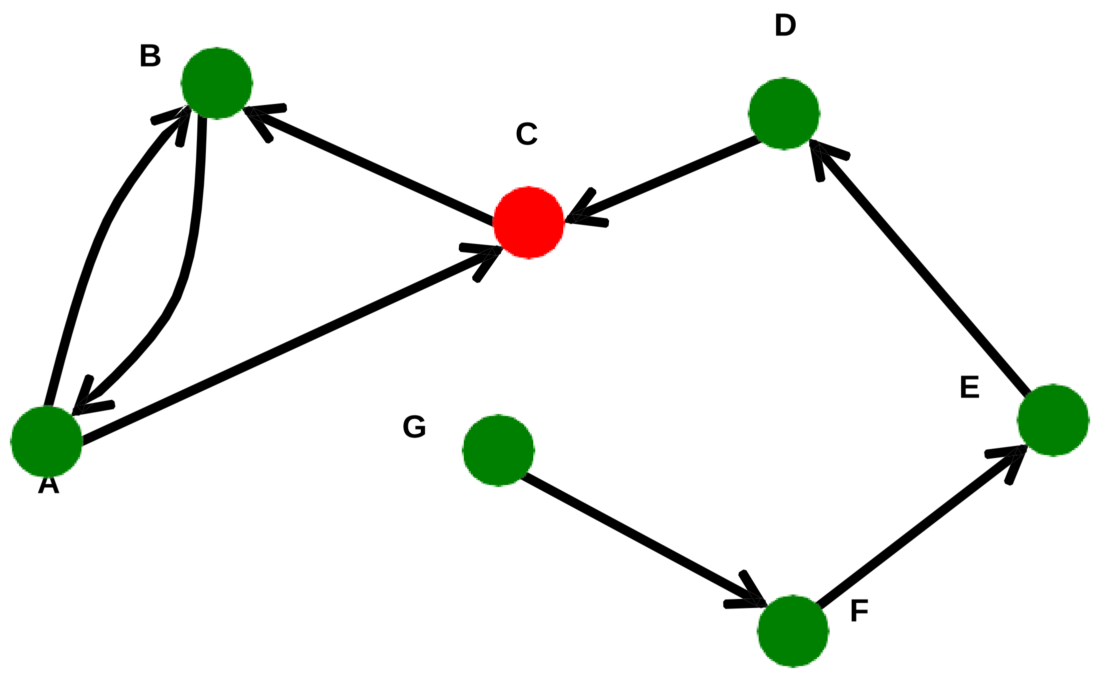
Source: Node with $k^{in} = 0$
Sink: Node with $k^{out} = 0$

- Directed

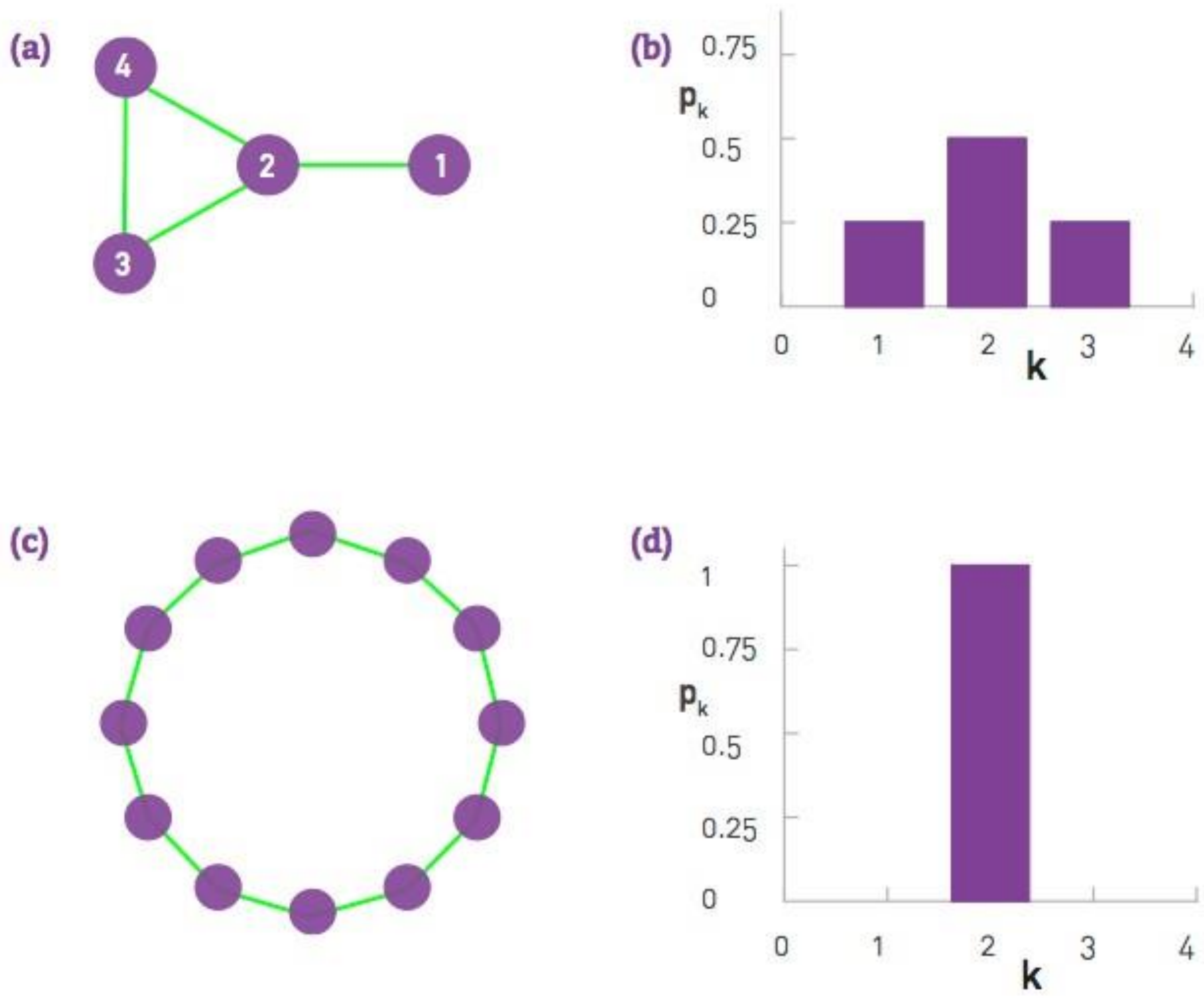
In directed networks we define an in-degree and out-degree.
The (total) degree of a node is the sum of in- and out-degrees.

$k^{in} = 2$

$k_C^{out} = 1 \qquad k_C = 3$



BASIC CONCEPTS
DEGREE DISTRIBUTION

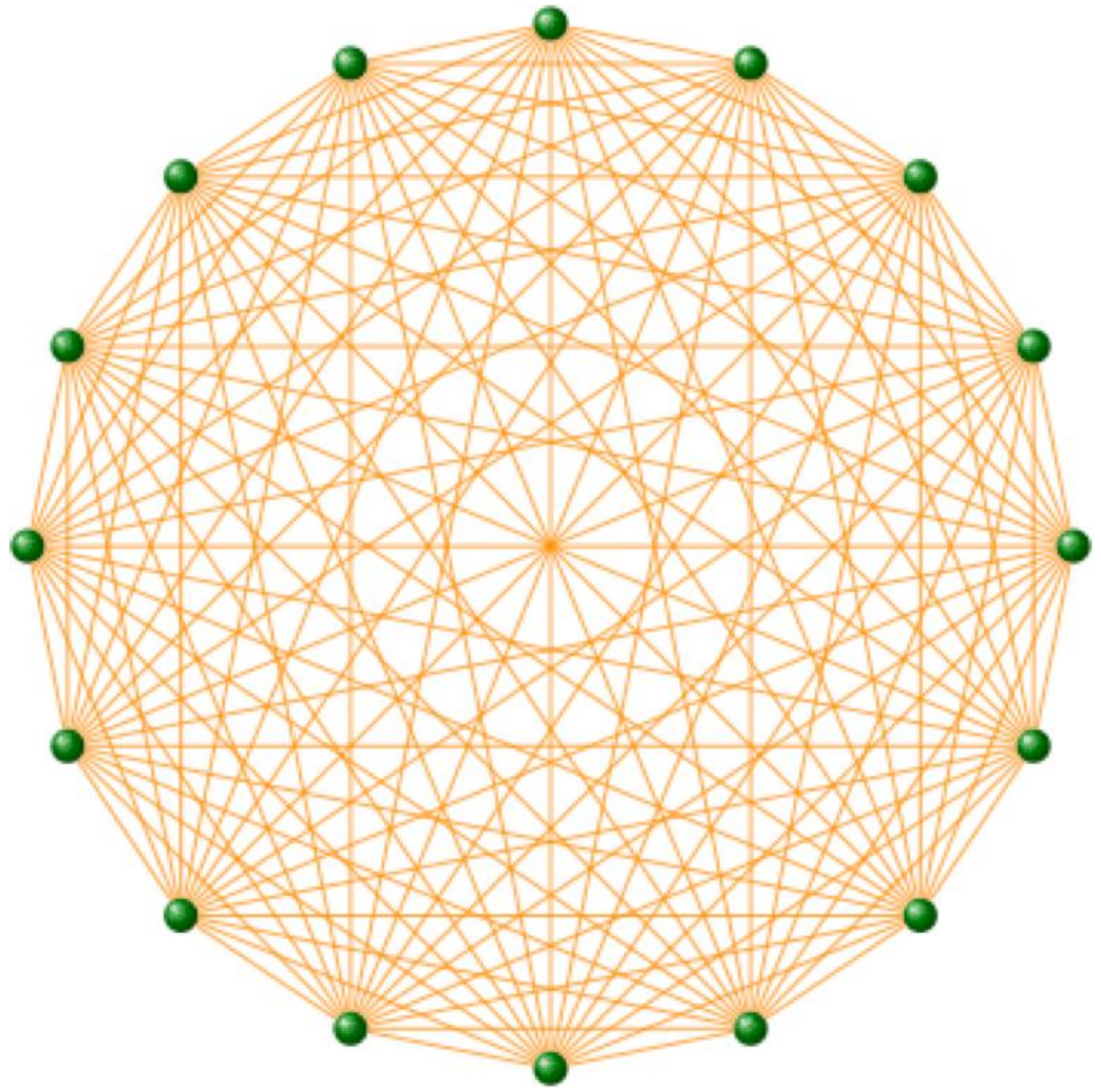


BASIC CONCEPTS

COMPLETE GRAPH

- The maximum number of edges in an undirected graph on N nodes is

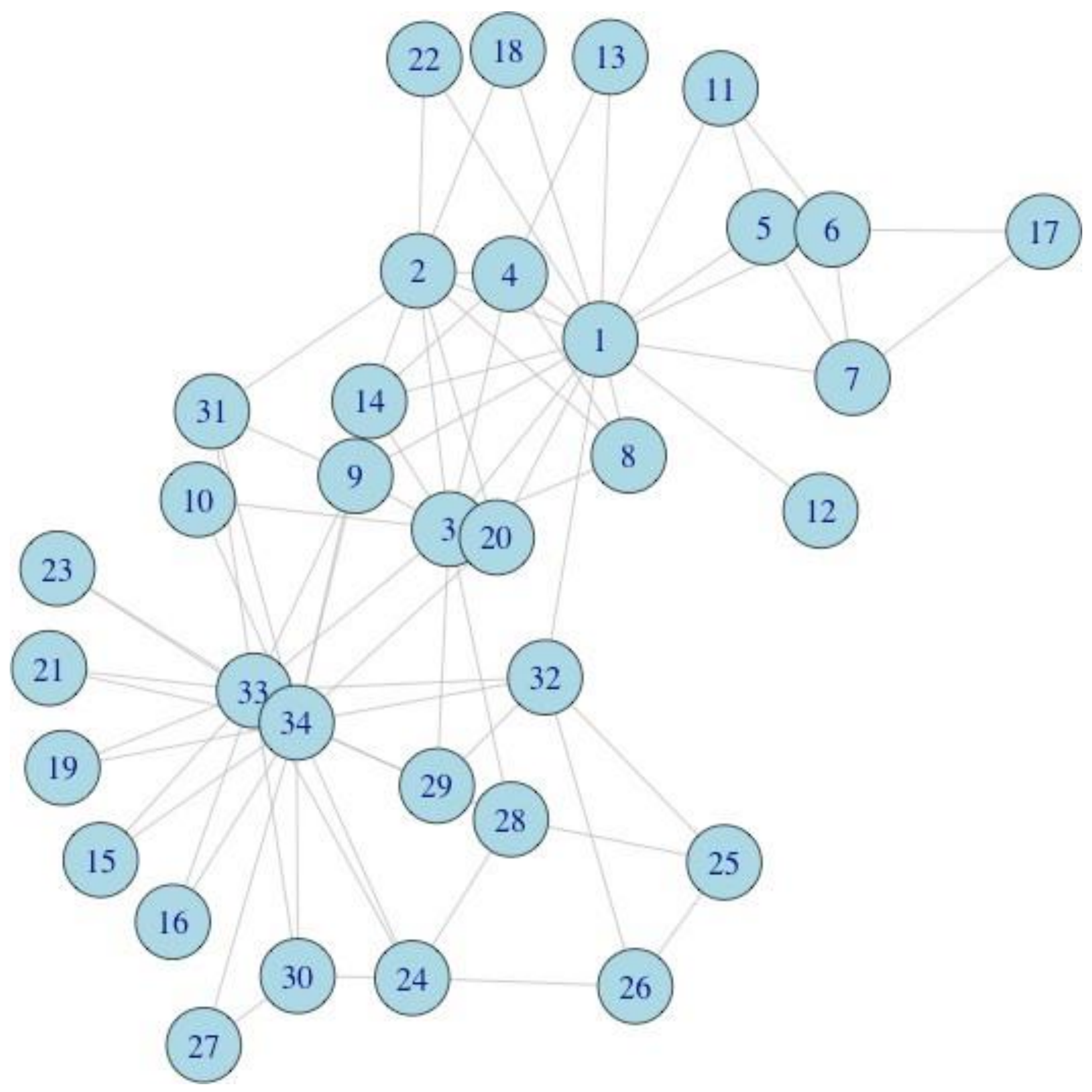
$$E_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$$



An undirected graph with the number of edges $E = E_{max}$ is called a **complete graph**, and its average degree is $N\text{-}1$

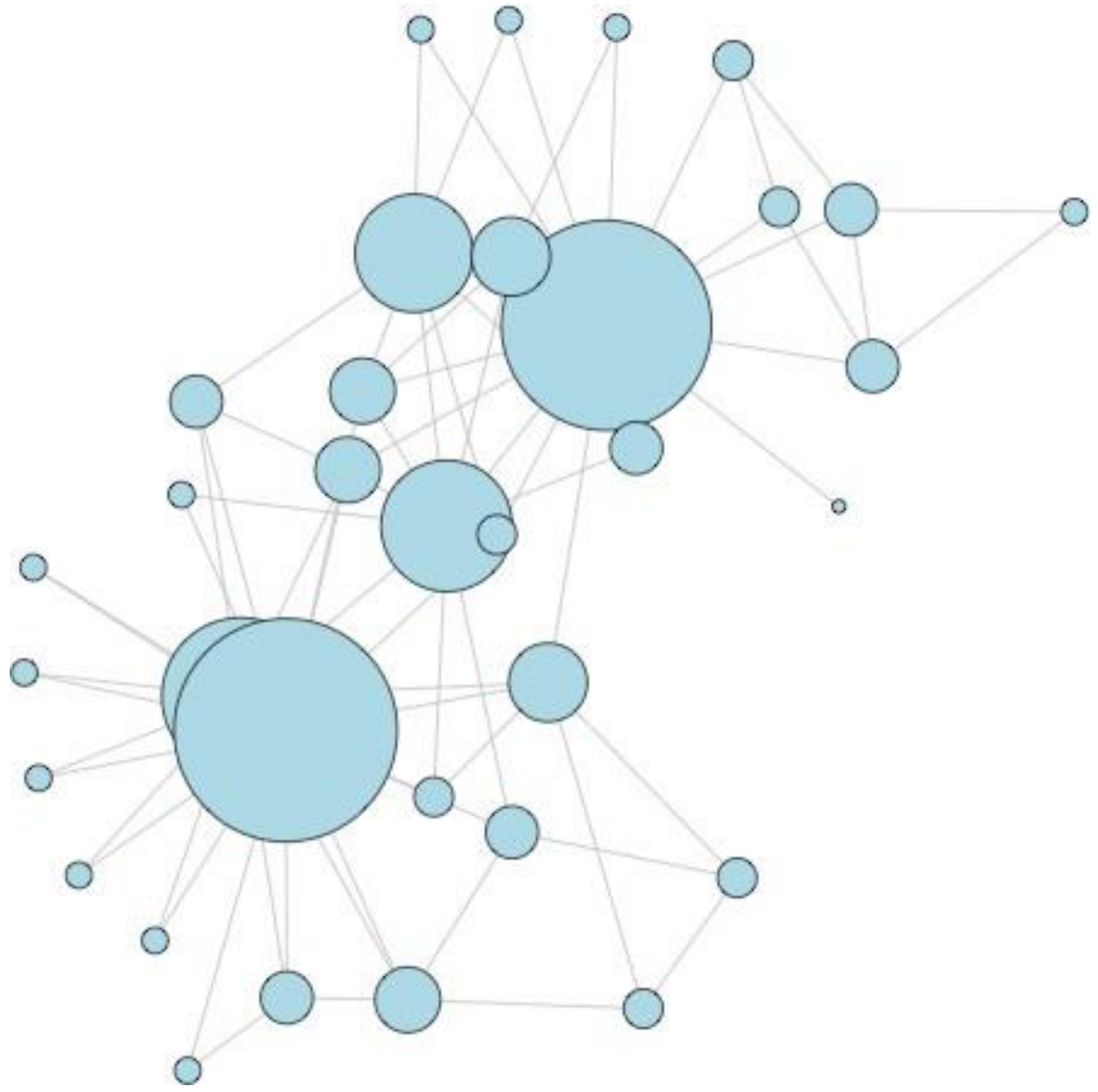
BASIC CONCEPTS

DEGREE DISTRIBUTION



BASIC CONCEPTS

DEGREE DISTRIBUTION



BASIC CONCEPTS

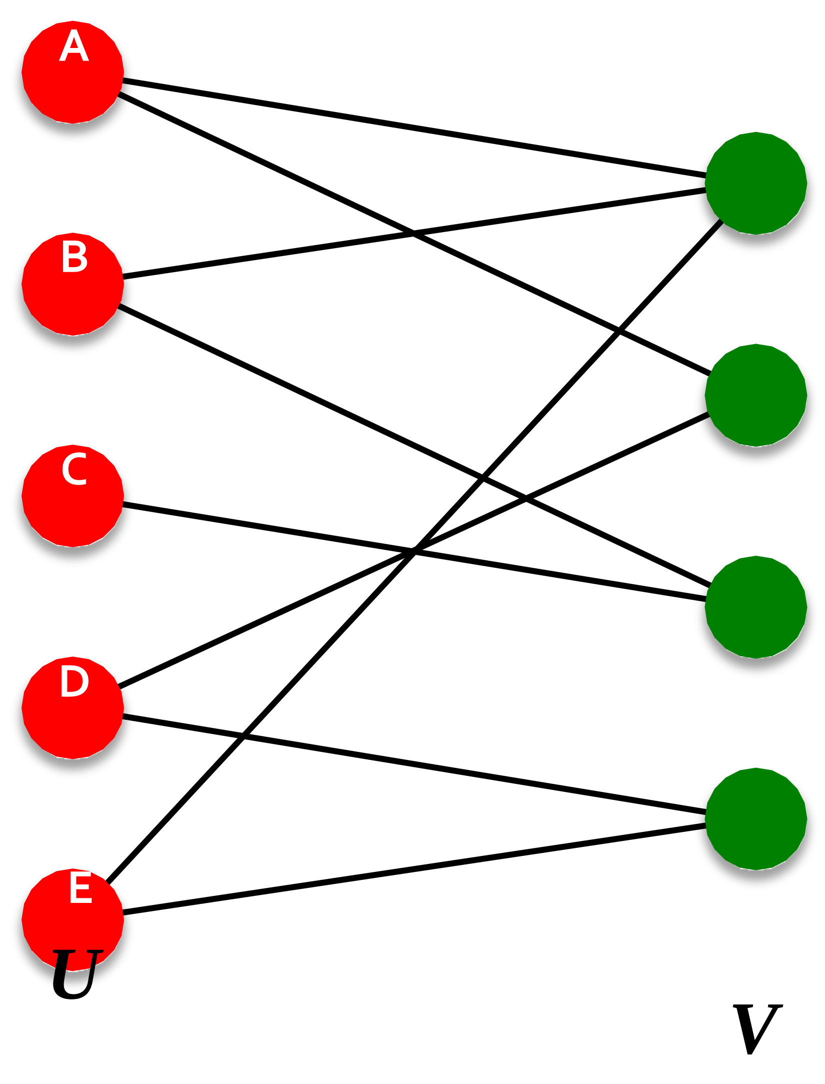
AVERAGE DEGREES

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	$\langle k \rangle$
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

BASIC CONCEPTS
BIPARTITE GRAPH

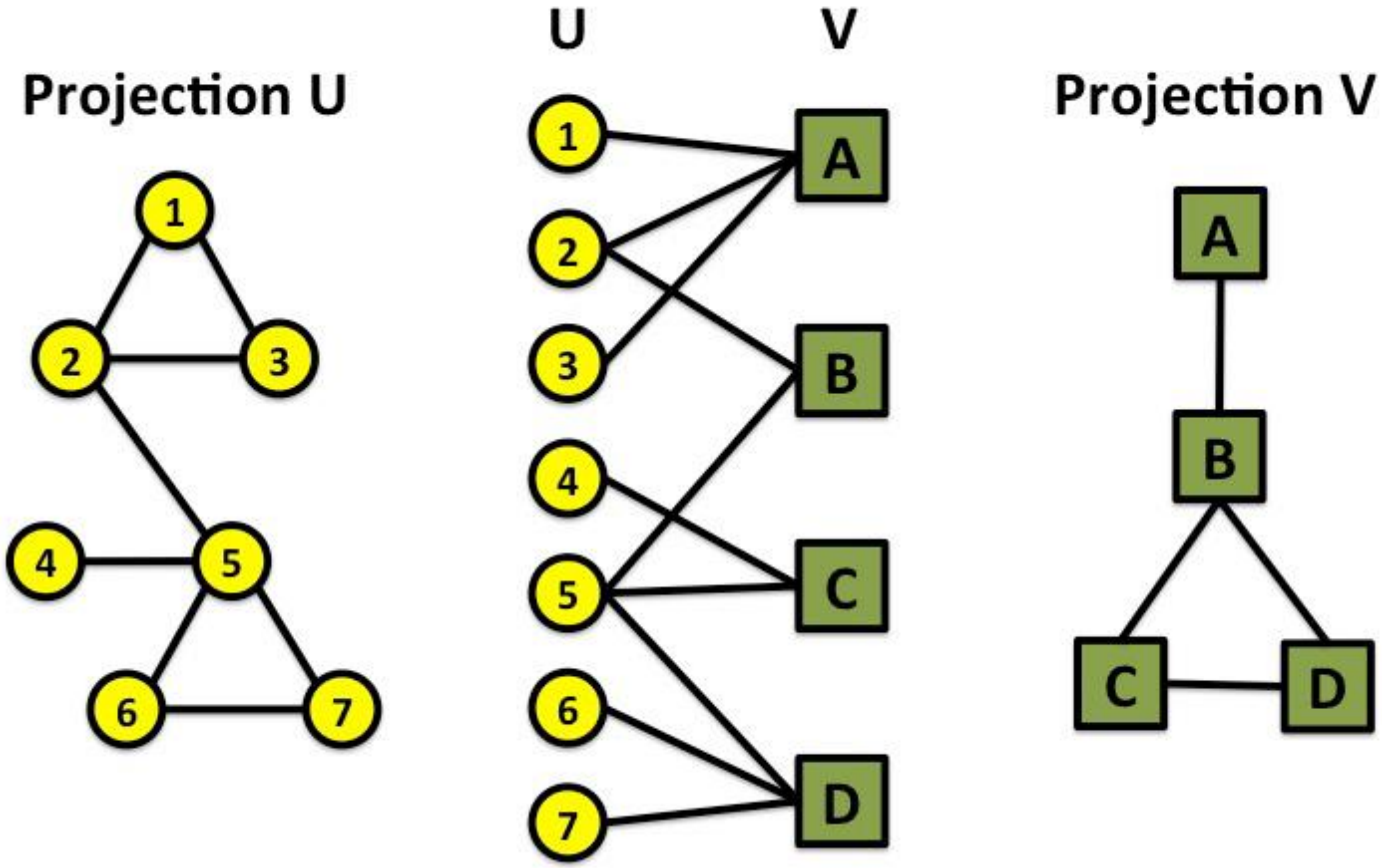
- Bipartite graph is a graph whose nodes can be divided into two disjoint sets U and V such that every link connects a node in U to one in V; that is, U and V are independent sets

- Examples:
  - Authors-to-Papers (they authored)
  - Actors-to-Movies (they appeared in)
  - Users-to-Movies (they rated)
  - Recipes-to-Ingredients (they contain)
- "Folded" networks:
  - Author collaboration networks
  - Movie co-rating networks



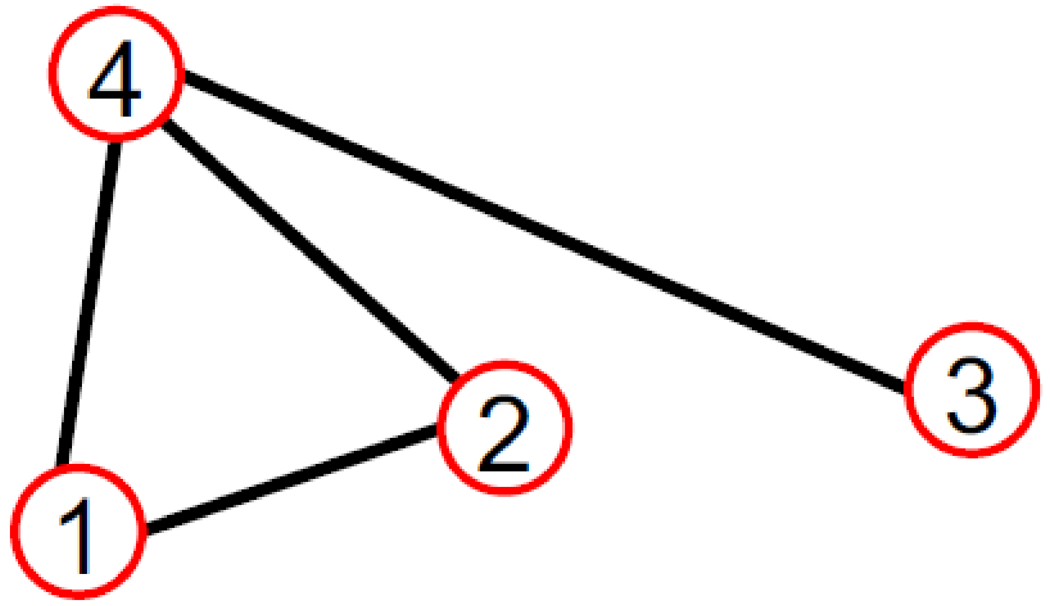
BASIC CONCEPTS

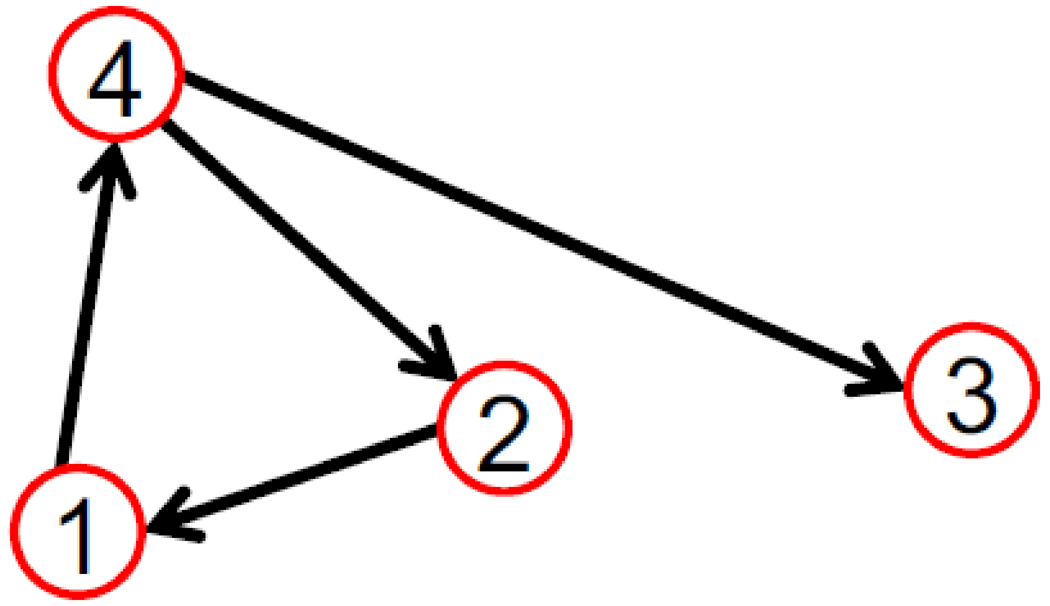
BIPARTITE GRAPH



BASIC CONCEPTS

GRAPH REPRESENTATION: ADJACENCY MATRIX





$A_{ij} = 1$ if there is a link from node ***i*** to node ***j***

$A_{ij} = 0$ otherwise

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

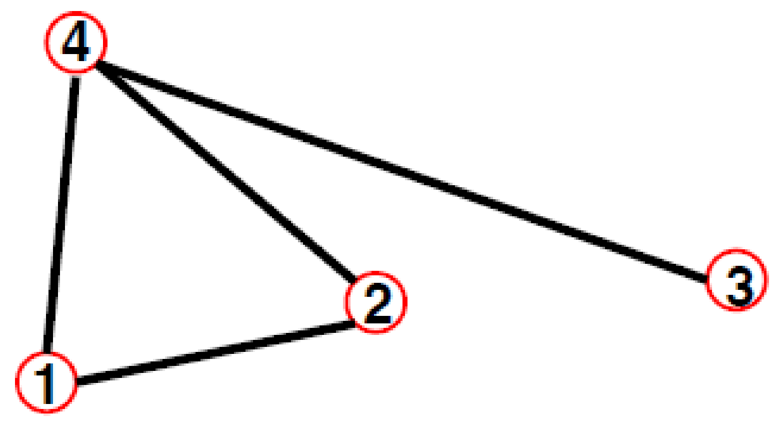
$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Adjacency matrix $A^{n \times n}$ is a matrix with nonzero element a_{ij} when there is an edge e_{ij}

BASIC CONCEPTS

GRAPH REPRESENTATION: ADJACENCY MATRIX

Undirected



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

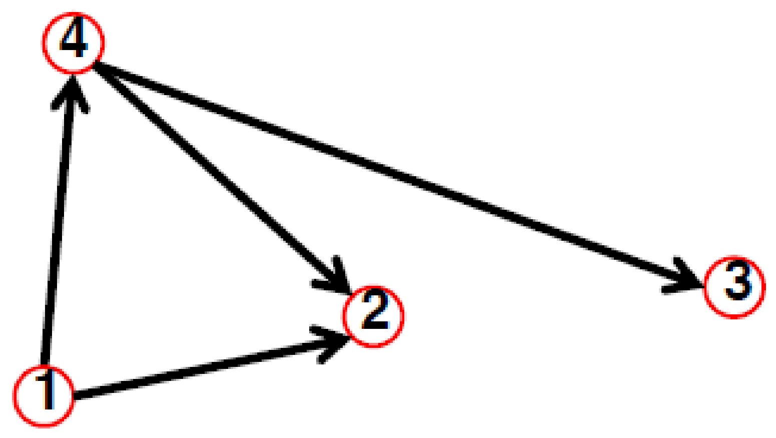
$$A_{ij} = A_{ji}$$
$$A_{ii} = 0$$

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

$$\boxed{k_j = \sum_{i=1}^{N} A_{ij}}$$

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

Directed



$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ij} \neq A_{ji}$$
$$A_{ii} = 0$$

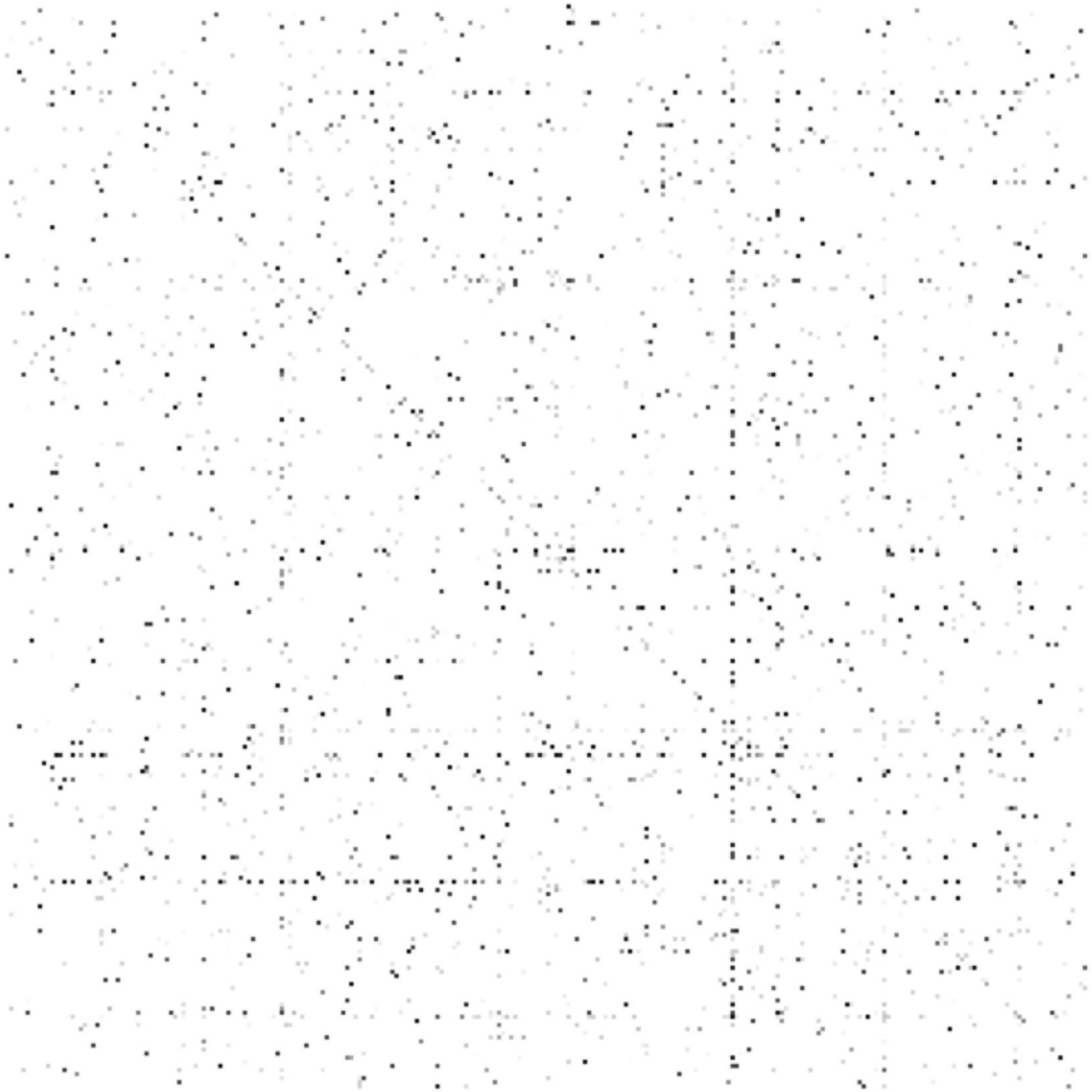
$$\boxed{k_i^{out} = \sum_{j=1}^{N} A_{ij}}$$

$$\boxed{k_j^{in} = \sum_{i=1}^{N} A_{ij}}$$

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

BASIC CONCEPTS

GRAPH REPRESENTATION: ADJACENCY MATRIX

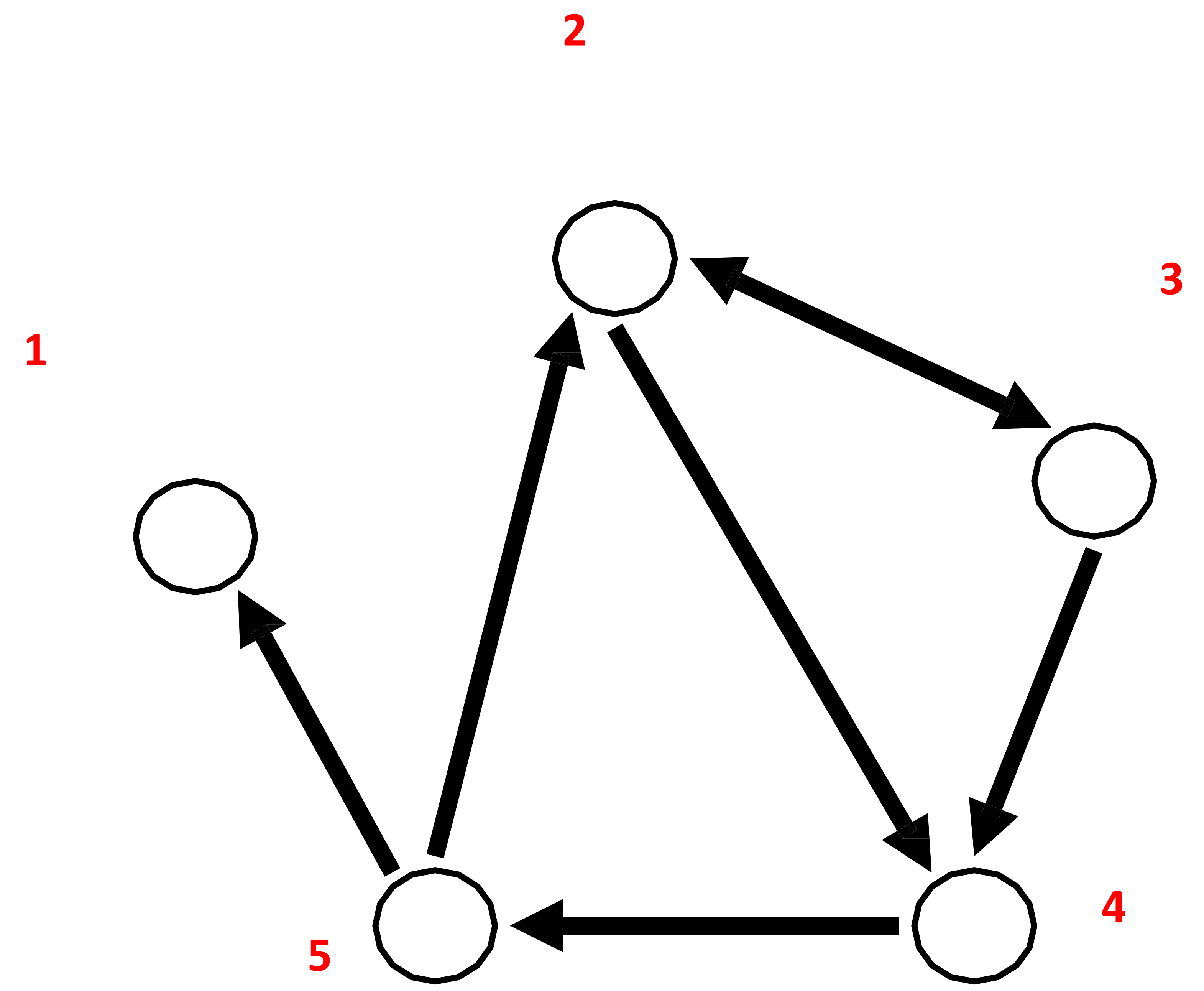


BASIC CONCEPTS

GRAPH REPRESENTATION: EDGE LIST

Represent graph as a set of edges:

- (2, 3)
- (2, 4)
- (3, 2)
- (3, 4)
- (4, 5)
- (5, 2)
- (5, 1)

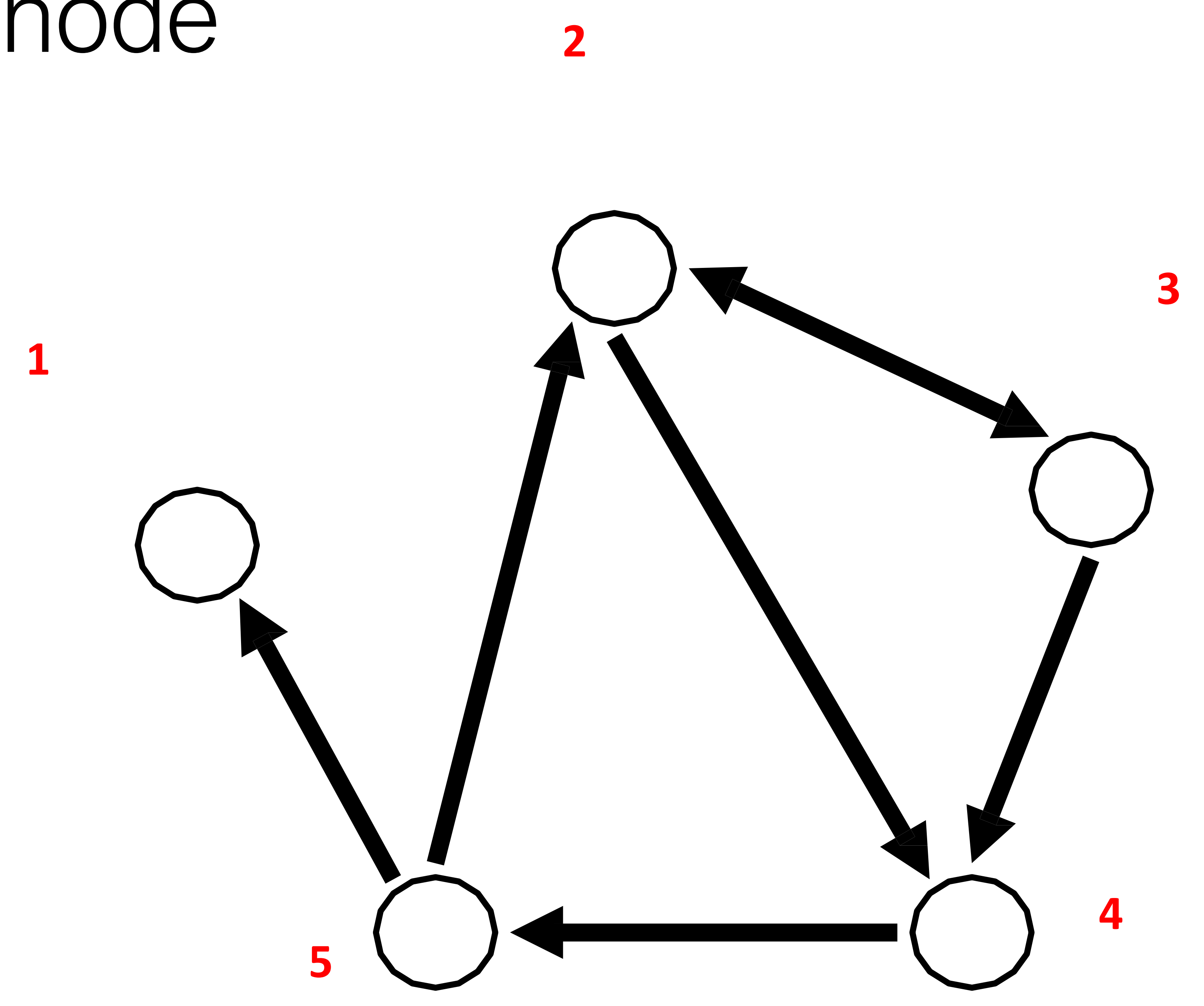


BASIC CONCEPTS

GRAPH REPRESENTATION: EDGE LIST

Adjacency list:

- Easier to work with if network is
  - Large
  - Sparse
- Allows us to quickly retrieve all neighbors of a given node
  - 1:
  - 2: 3, 4
  - 3: 2, 4
  - 4: 5
  - 5: 1, 2



BASIC CONCEPTS

EDGE ATTRIBUTES

Possible options:

- Weight (e.g. frequency of communication)
- Ranking (best friend, second best friend…)
- Type (friend, relative, co-worker)
- Sign: Friend vs. Foe, Trust vs. Distrust
- Properties depending on the structure of the rest of the graph: number of common friends

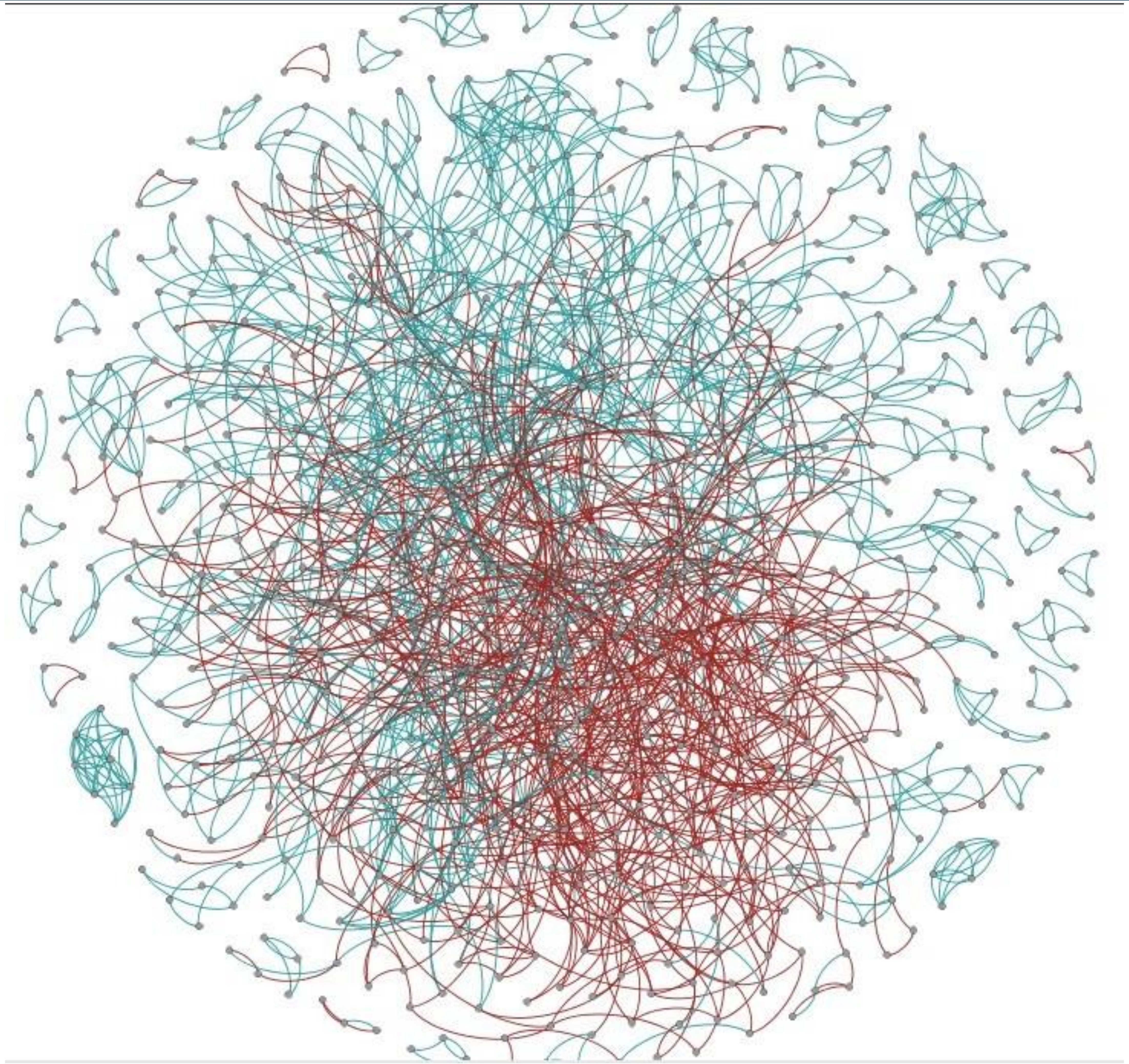
BASIC CONCEPTS

EDGE ATTRIBUTES

Possible options:
- Weight (e.g. frequency of communication)
- Ranking (best friend, second best friend…)
- Type (friend, relative, co-worker)
- Sign: Friend vs. Foe, Trust vs. Distrust
- Properties depending on the structure of the rest of the graph: number of common friends

BASIC CONCEPTS

POSITIVE AND NEGATIVE WEIGHTS



BASIC CONCEPTS

POSITIVE AND NEGATIVE WEIGHTS

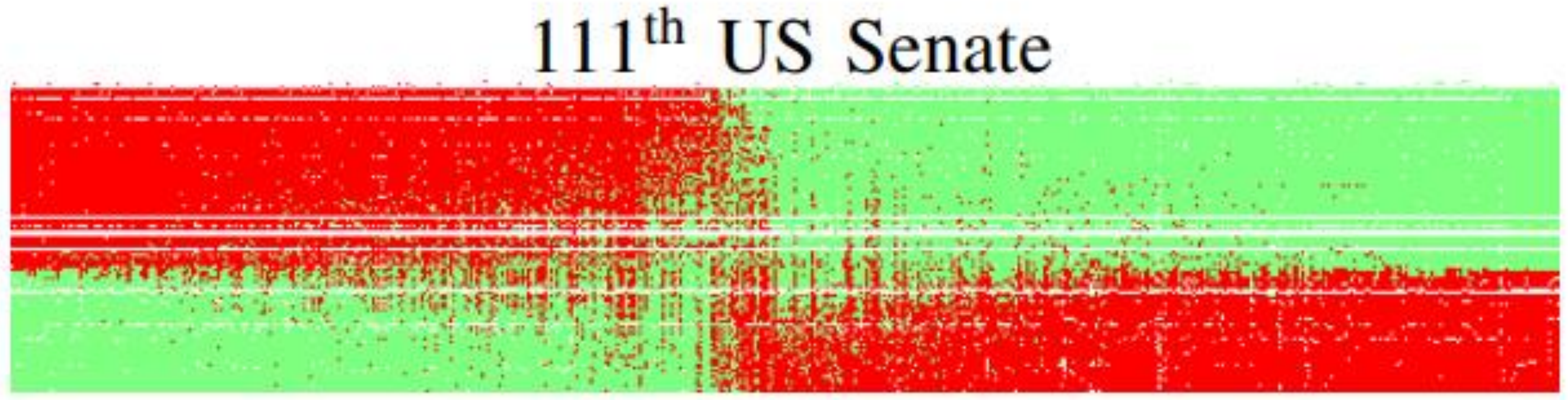


Fig. 4. Vote matrix of the 111^{th} US Senate after scaling with ANCO-HITS

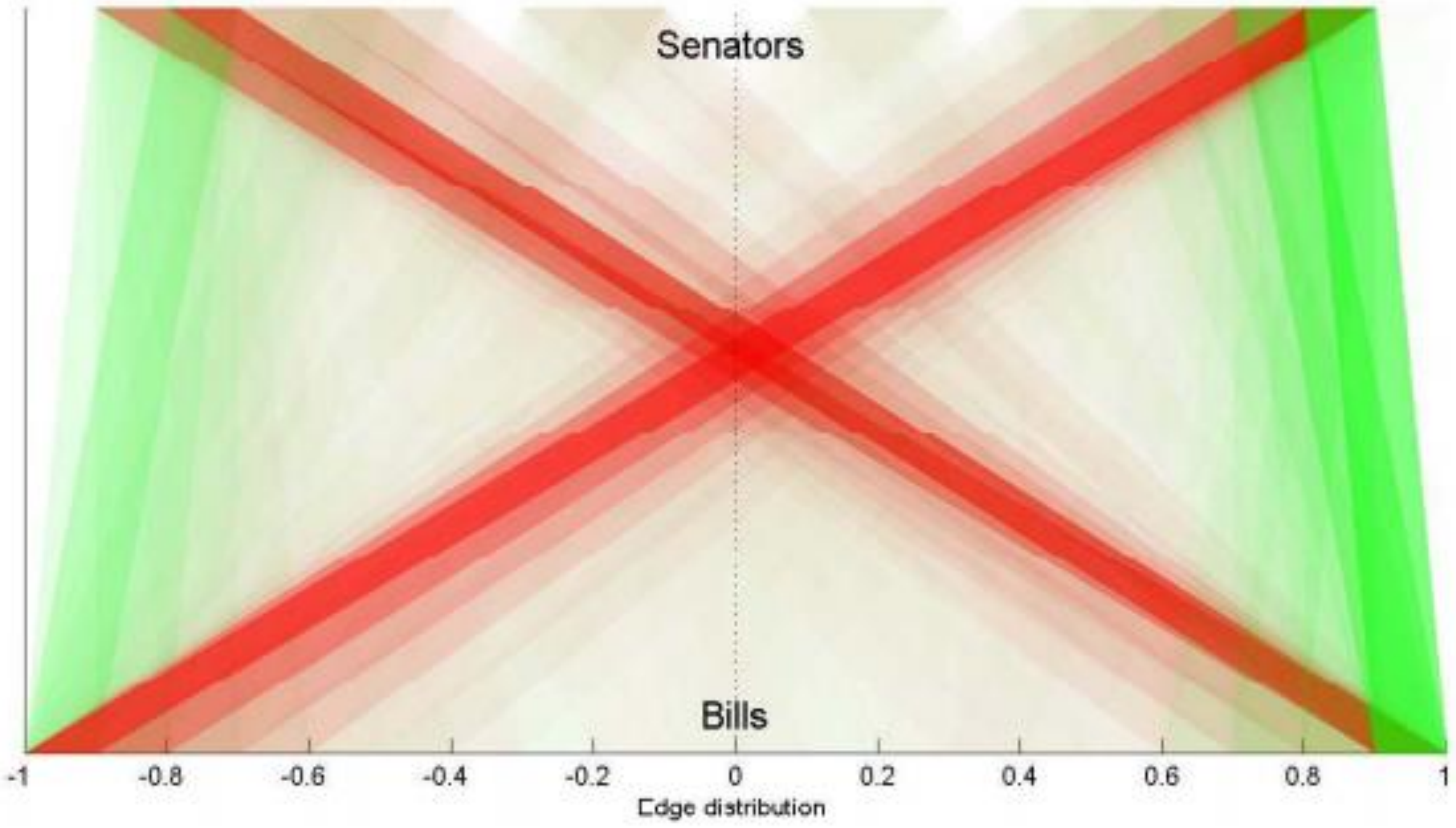


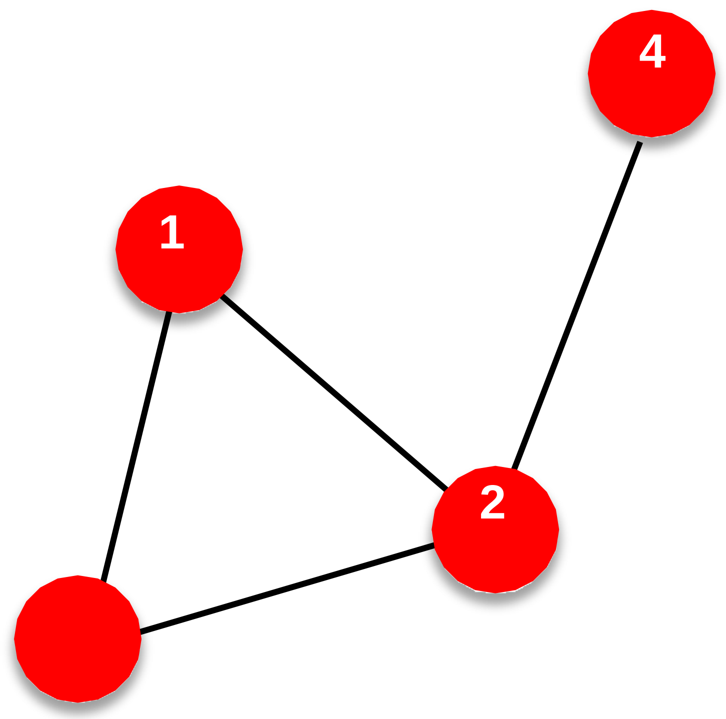
Fig. 5. Bipartite graph of the 111^{th} US Senate after scaling with ANCO-HITS

BASIC CONCEPTS

POSITIVE AND NEGATIVE WEIGHTS

Unweighted

(undirected)



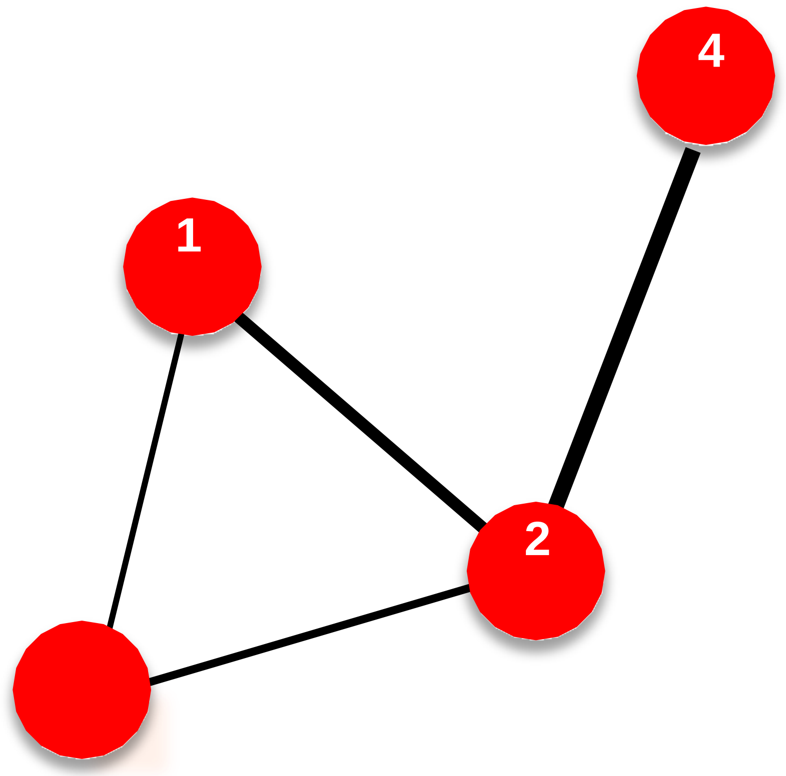
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$E = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad \bar{k} = \frac{2E}{N}$$

Weighted

(undirected)



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

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

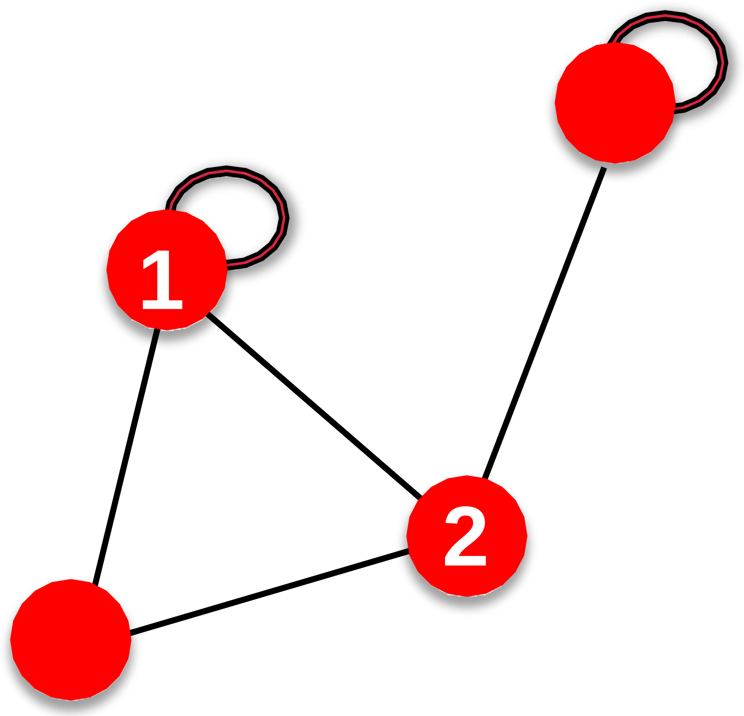
$$E = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad \bar{k} = \frac{2E}{N}$$

BASIC CONCEPTS

POSITIVE AND NEGATIVE WEIGHTS

Self-edges (self-loops)

(undirected)



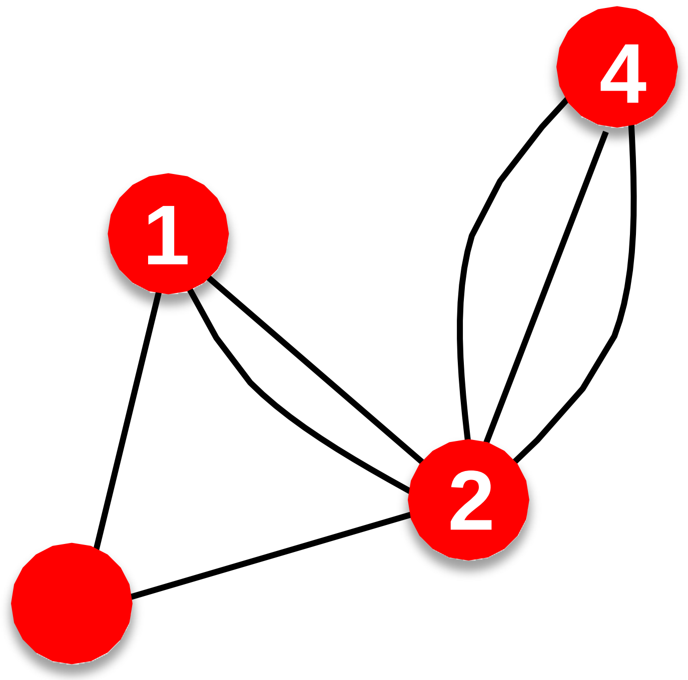
$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$A_{ii} \neq 0 \qquad A_{ij} = A_{ji}$$

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

Multigraph

(undirected)



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

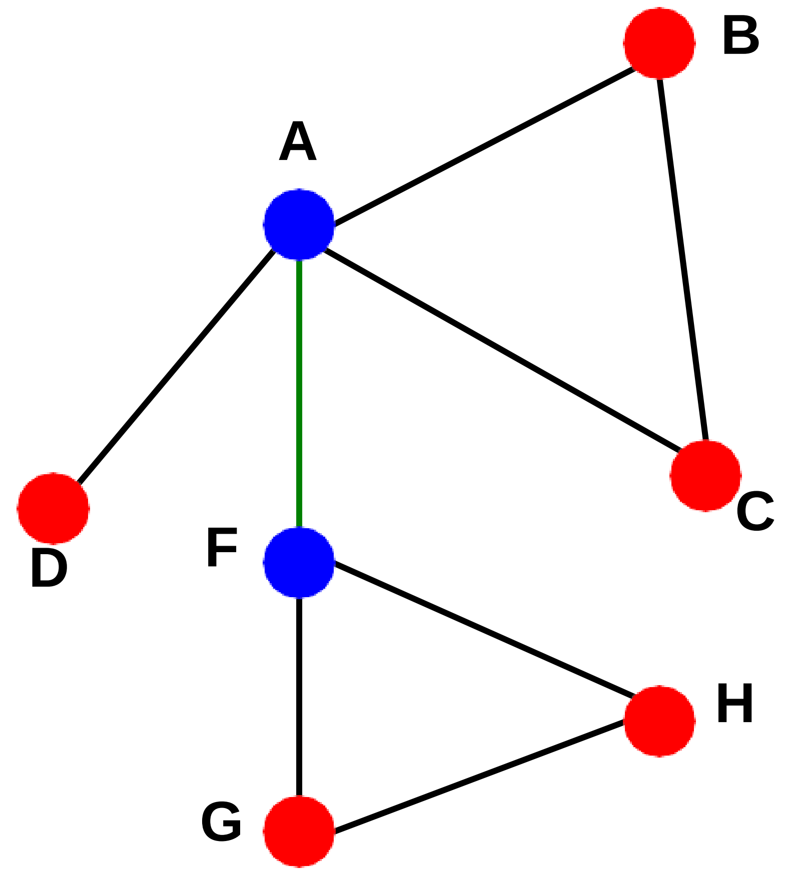
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

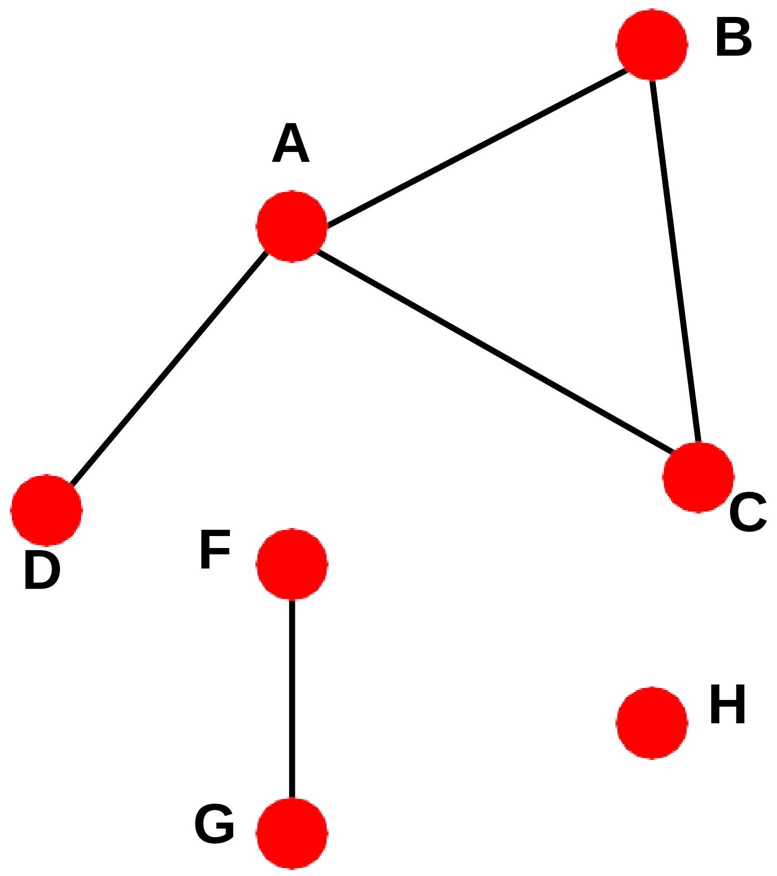
$$E = \frac{1}{2} \sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad \bar{k} = \frac{2E}{N}$$

BASIC CONCEPTS

CONNECTIVITY OF UNDIRECTED GRAPHS

- *A path* from v_i to v_j is a sequence of edges that joins two vertices. (It also ordered list of vertices such that that there is an edge to the next vertex on the list)
- A graph is *connected* if there a paths between any two vertices.
- *Connected component* is a maximal connected subgraph of *G*





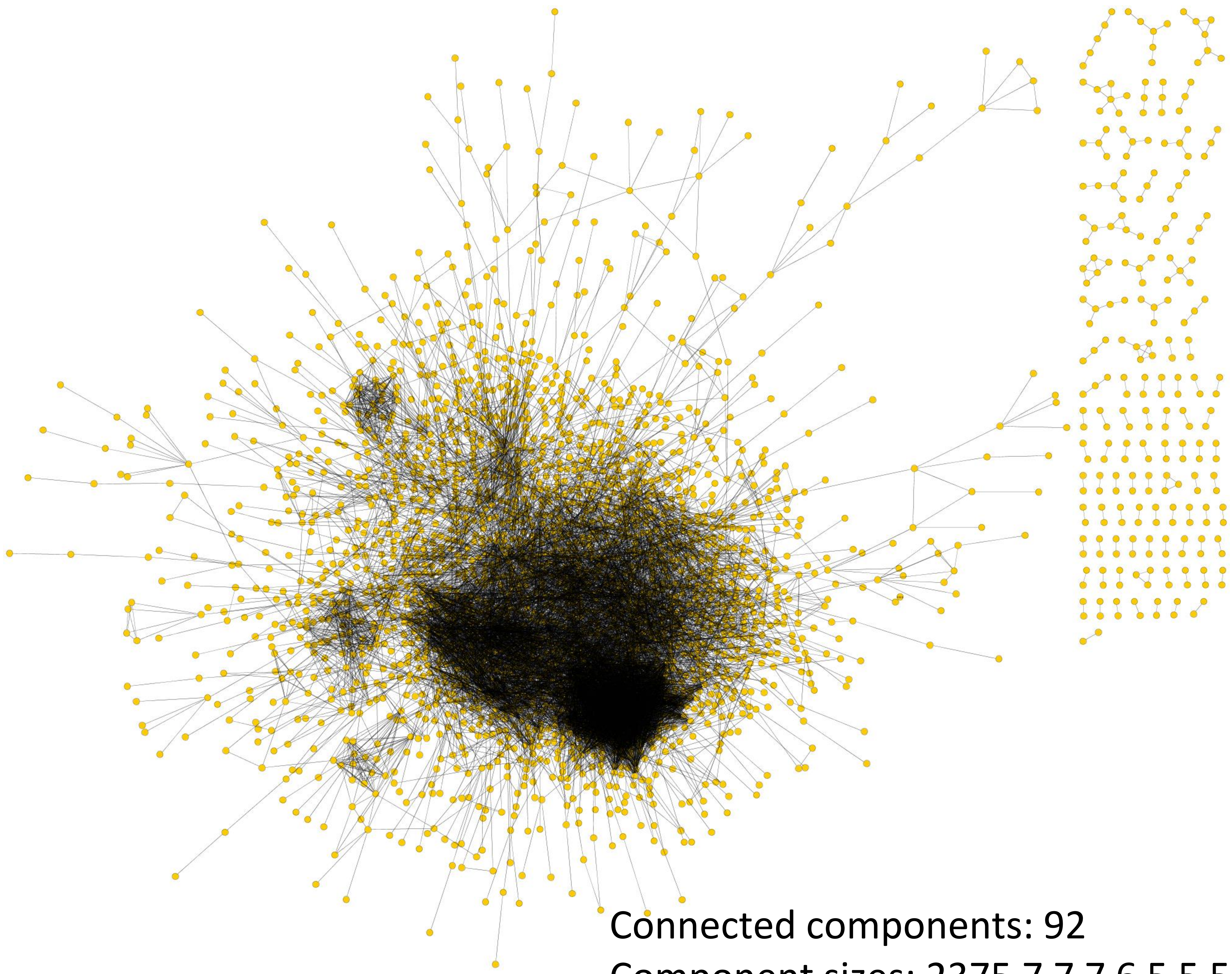
Largest Component:
Giant Component

Isolated node (node H)

Bridge edge: If we erase the **edge**, the graph becomes disconnected
Articulation node: If we erase the **node**, the graph becomes disconnected

BASIC CONCEPTS

CONNECTIVITY OF UNDIRECTED GRAPHS



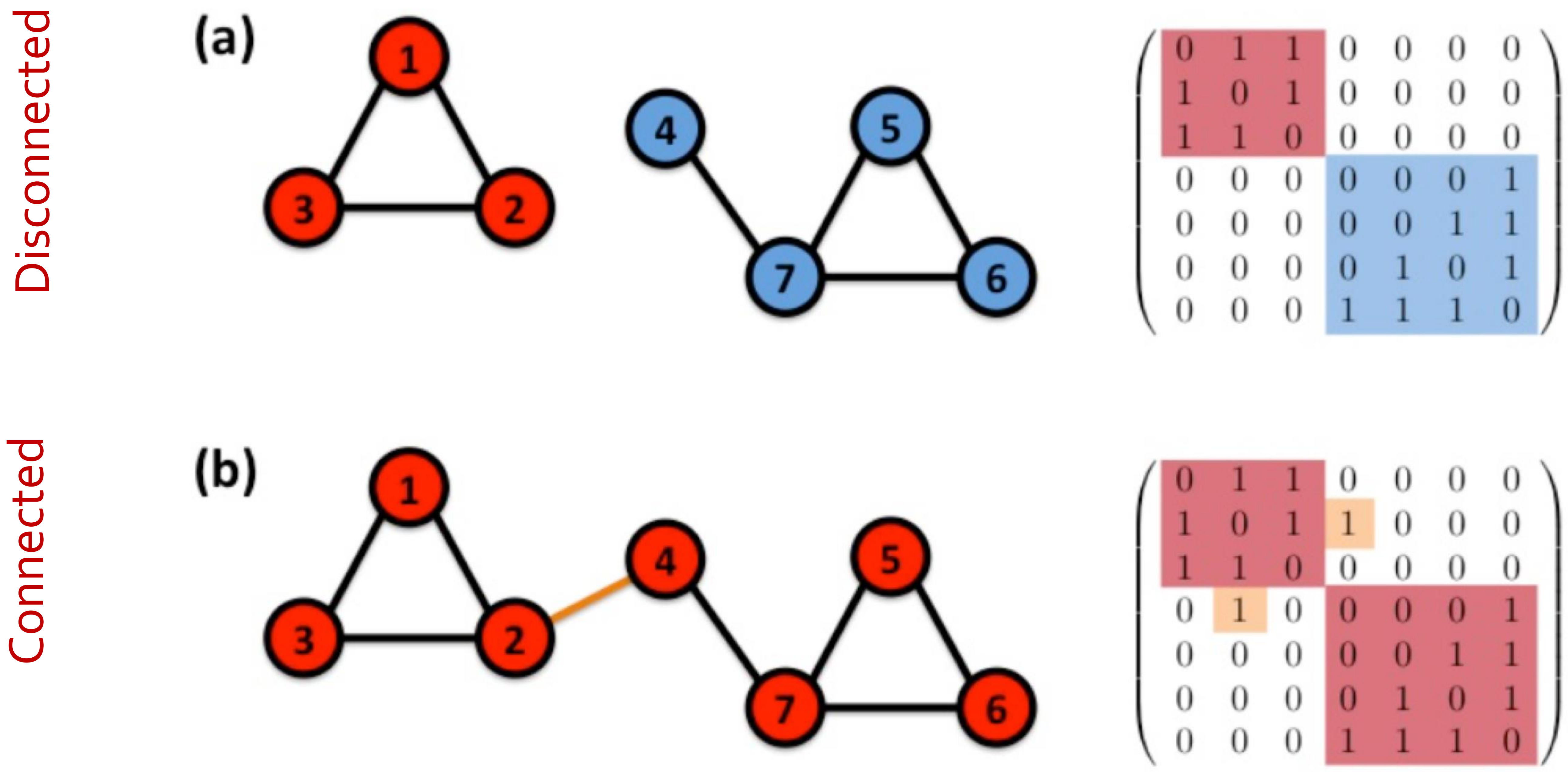
Connected components: 92

Component sizes: 2375 7 7 7 6 5 5 5 5 5 5 4 4 4 4

BASIC CONCEPTS

CONNECTIVITY EXAMPLE

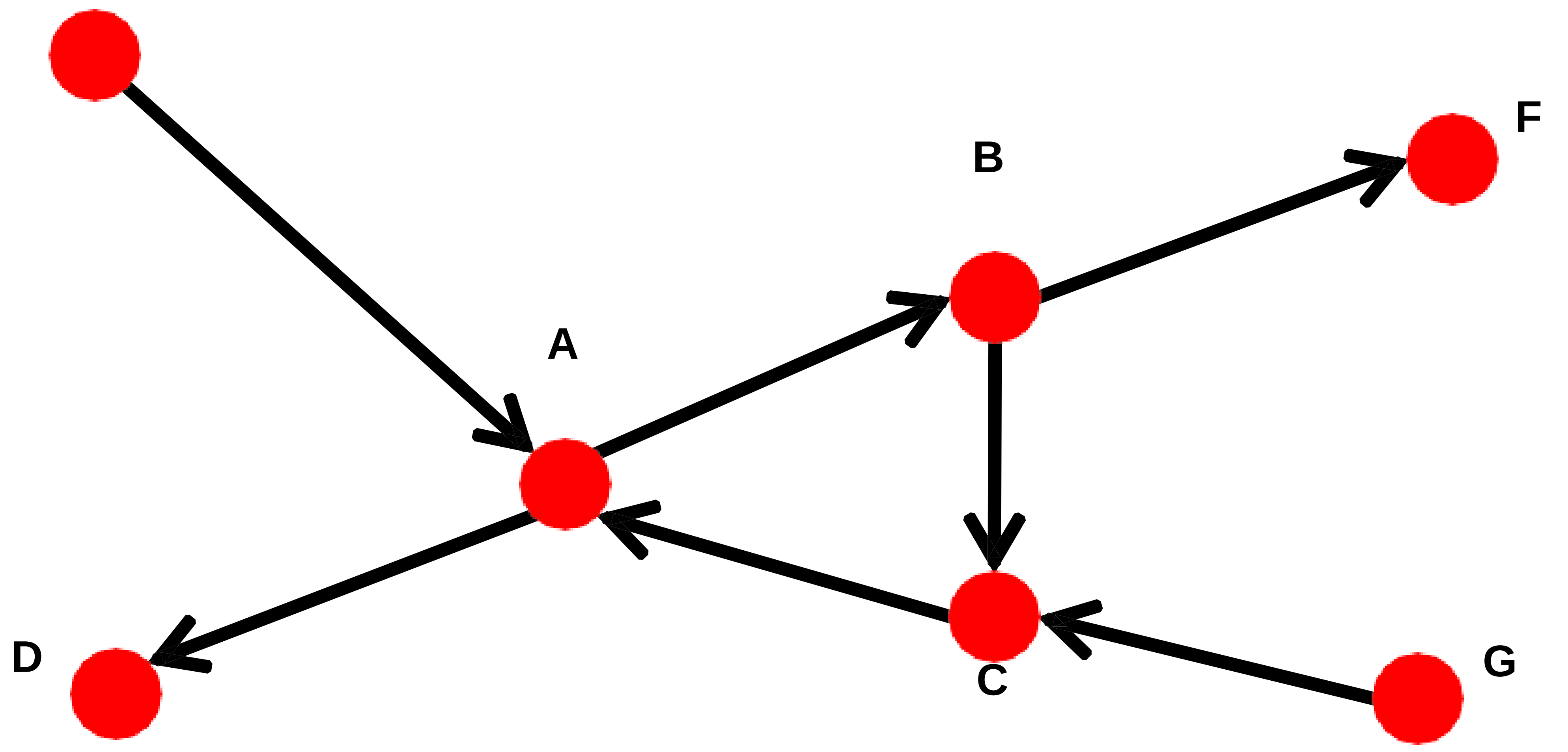
The adjacency matrix of a network with several components can be written in a block- diagonal form, so that nonzero elements are confined to squares, with all other elements being zero:



BASIC CONCEPTS

CONNECTIVITY OF DIRECTED GRAPHS

- Strongly connected directed graph has a path from each node to every other node and vice versa e.g., A-B path and B-A path)
- Weakly connected directed graph is connected if we disregard the edge directions



Graph on the left is connected but not strongly connected (e.g., there is no way to get from F to G by following the edge directions).

NETWORK PROPERTIES

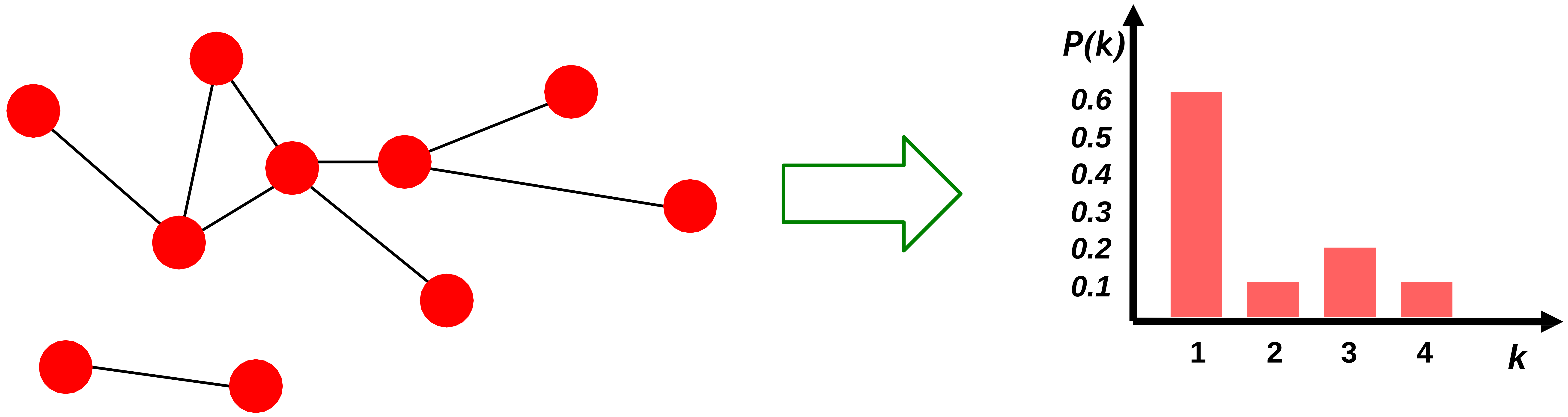
DEGREE DISTRIBUTION

Degree distribution *P(k)*: Probability that a randomly chosen node has degree k

N_k = # nodes with degree k

Normalized histogram:

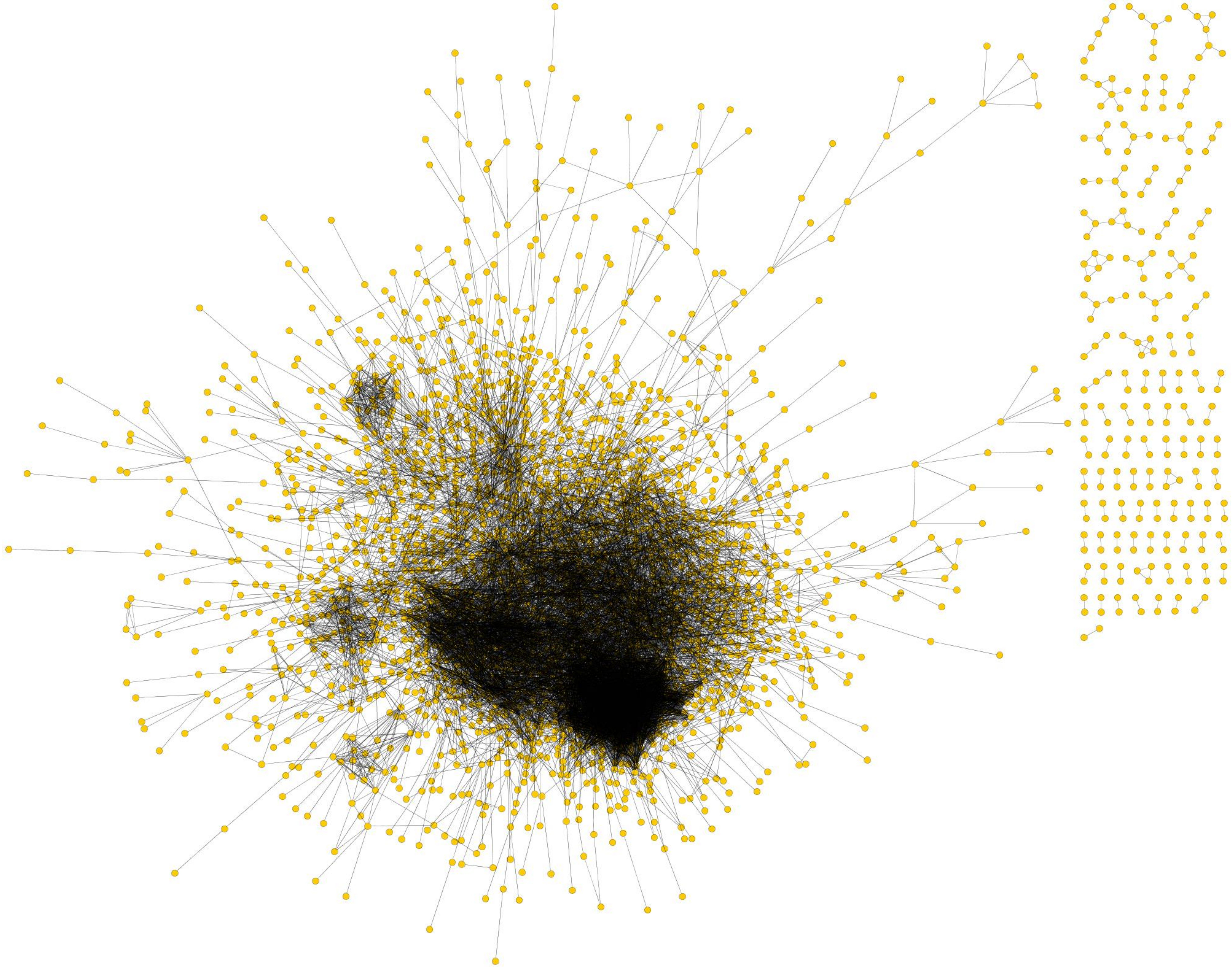
$$P(k) = N_k / N \quad \rightarrow \quad \textbf{plot}$$



For directed graphs we have separate in- and out-degree distributions.

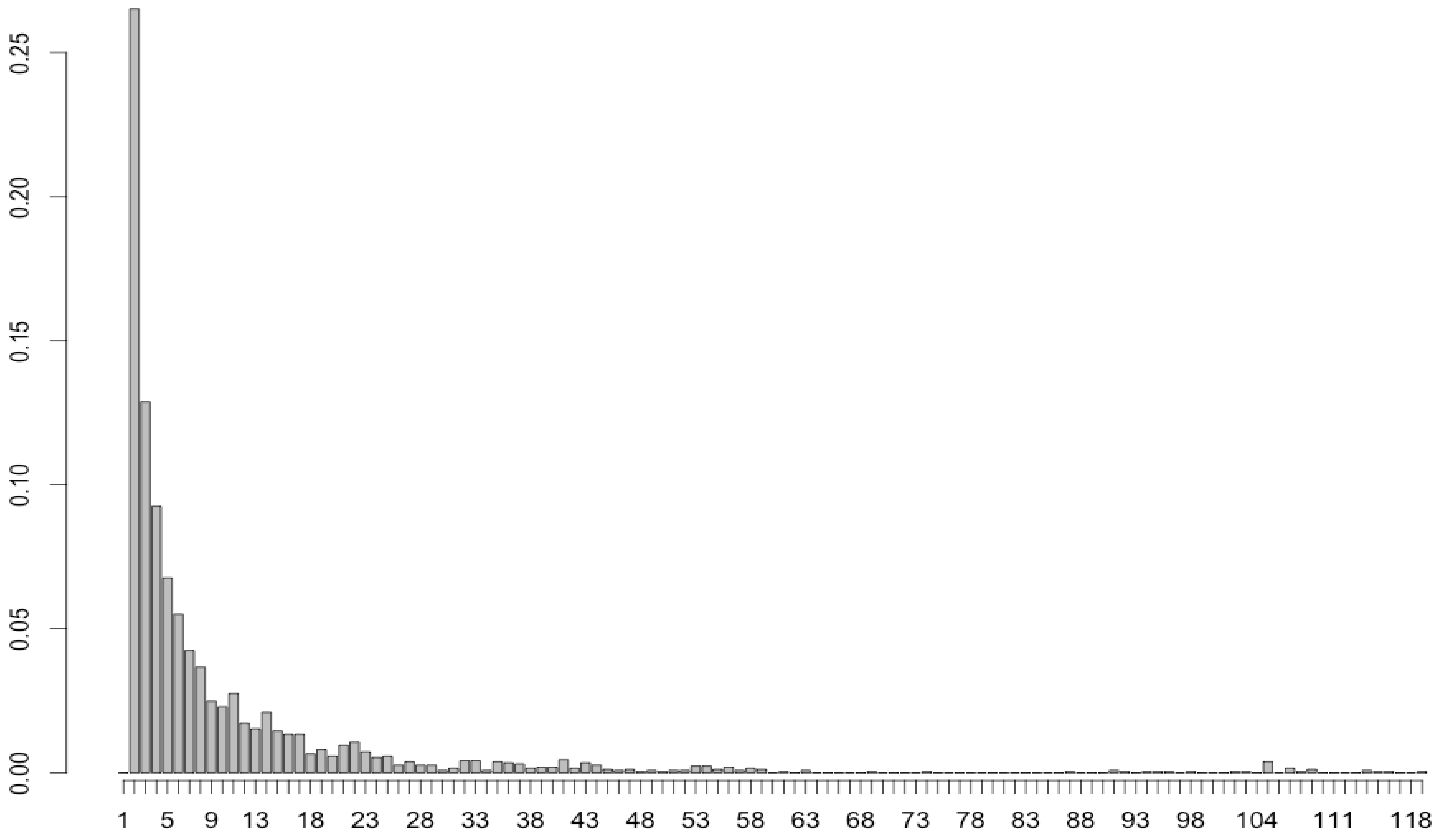
NETWORK PROPERTIES

DEGREE DISTRIBUTION



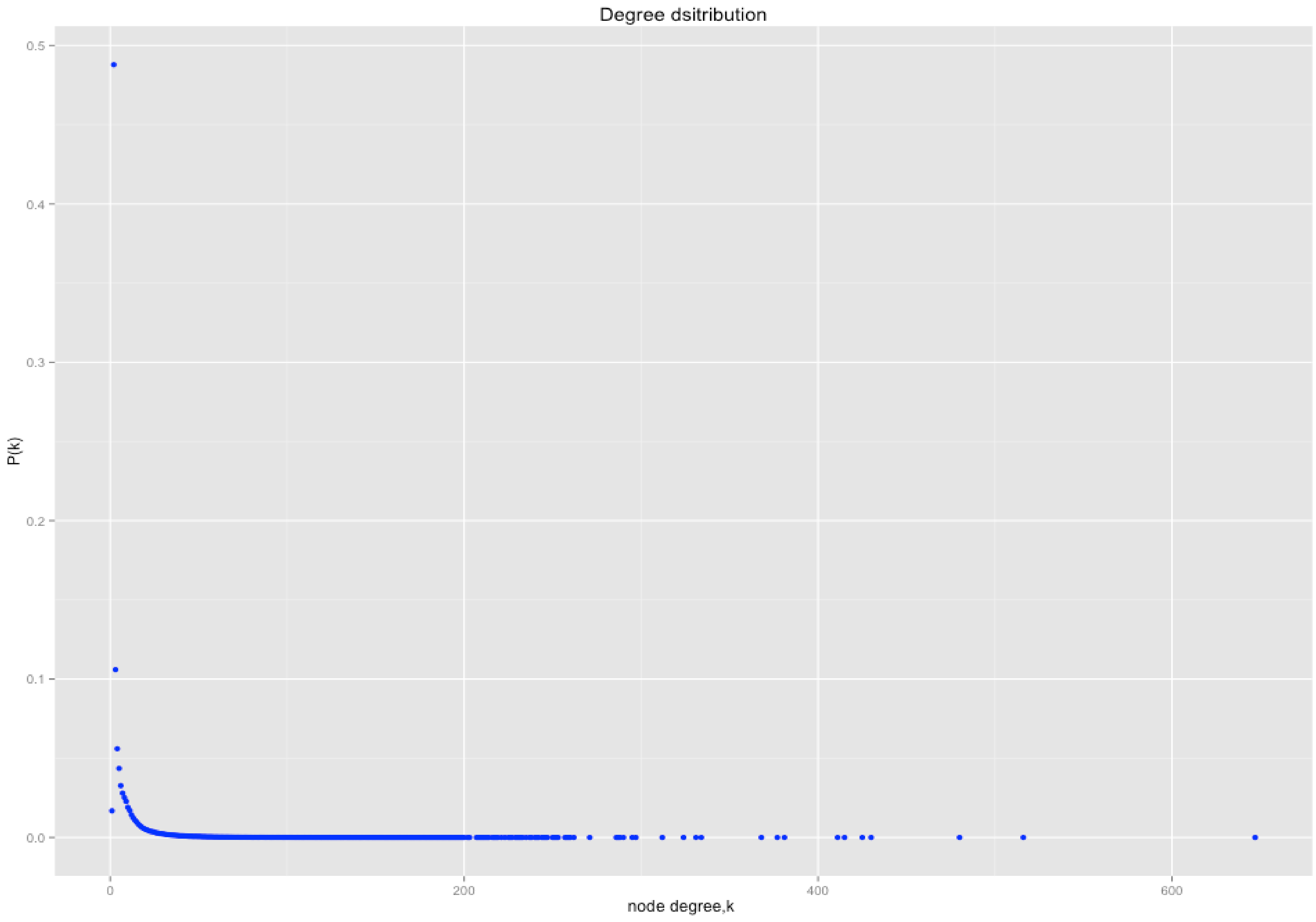
NETWORK PROPERTIES

DEGREE DISTRIBUTION

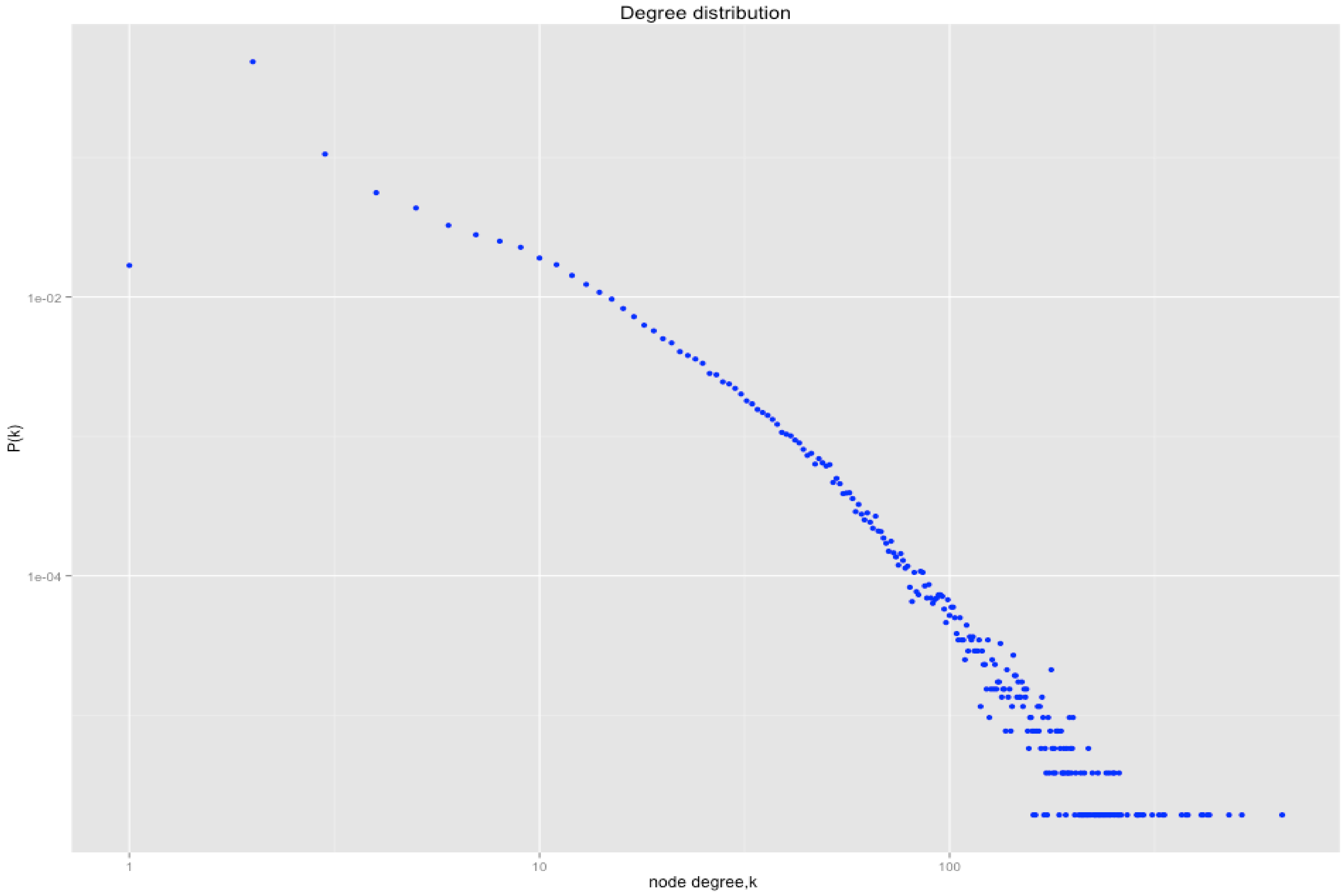


NETWORK PROPERTIES

DEGREE DISTRIBUTION

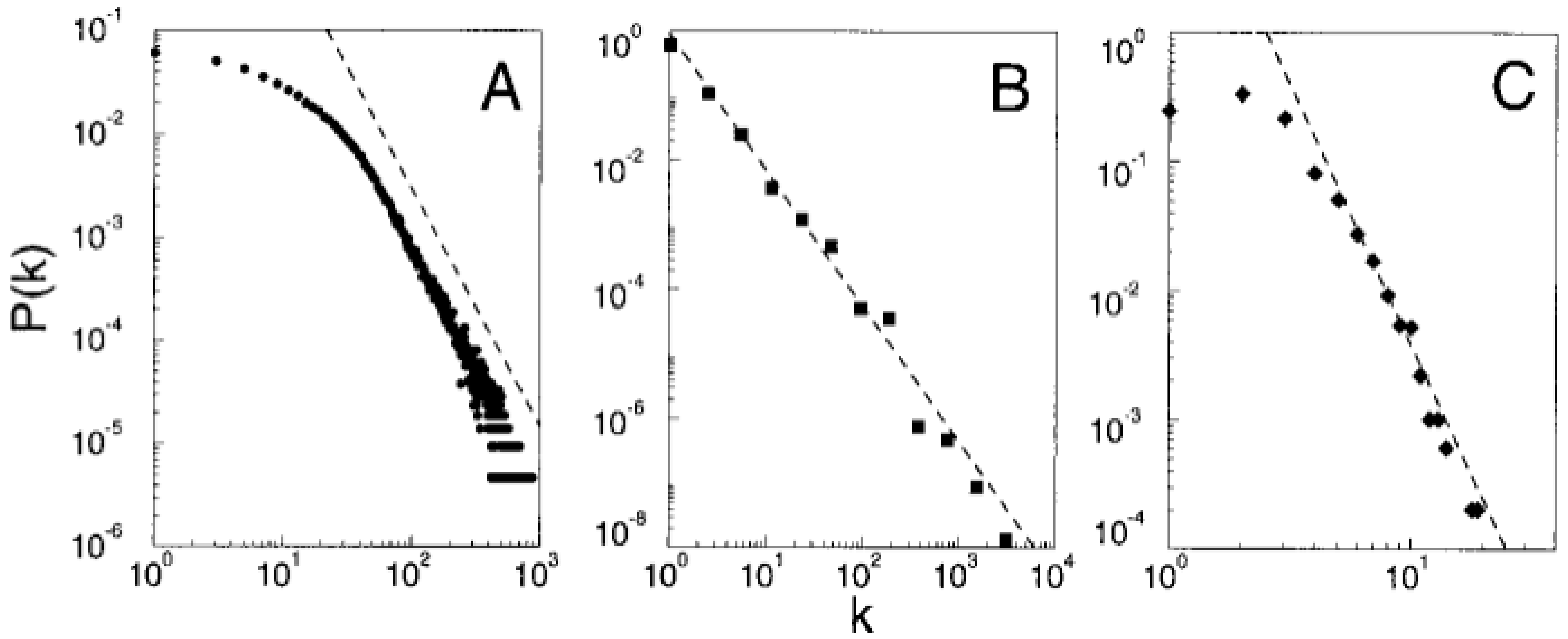


NETWORK PROPERTIES
DEGREE DISTRIBUTION



NETWORK PROPERTIES

POWER LAW NETWORKS



Actor collaboration graph, N=212,250 nodes, $\langle k \rangle = 28{:}8$, = 2:3

WWW, N = 325,729 nodes, $\langle k \rangle = 5{:}6$, = 2:1

Power grid data, N = 4941 nodes, $\langle k \rangle = 5{:}5$, = 4

NETWORK PROPERTIES

DEGREE DISTRIBUTION

- Power law distribution

$$P(k) = Ck^{-\gamma} = \frac{1}{k^{\gamma}}C$$

- Log-log coordinates

$$\log P(k) = -\gamma \log k + \log C$$

$$y = -\gamma x + b$$

NETWORK PROPERTIES

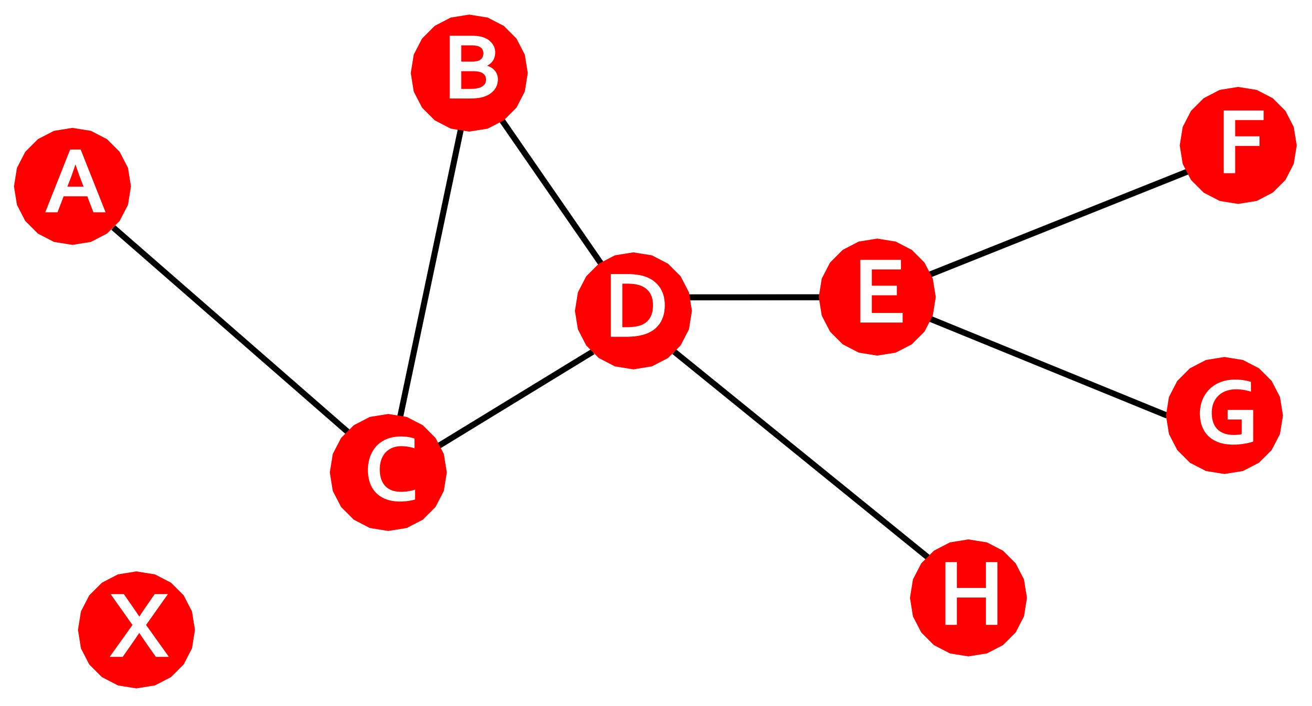
PATH IN GRAPHS

A ***path*** is a sequence of nodes in which each node is linked to the next one

$$P_n = \{i_0, i_1, i_2, \ldots, i_n\} \qquad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$$

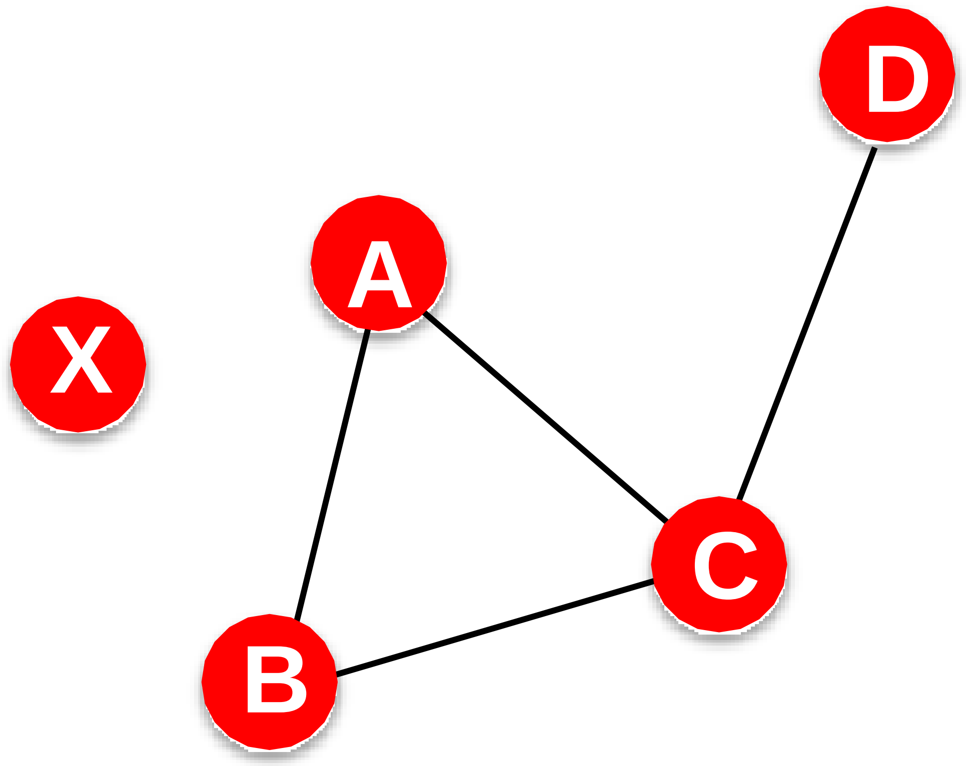
A path can intersect itself and pass through the same edge multiple times

- E.g.: ACBDCDEG



NETWORK PROPERTIES

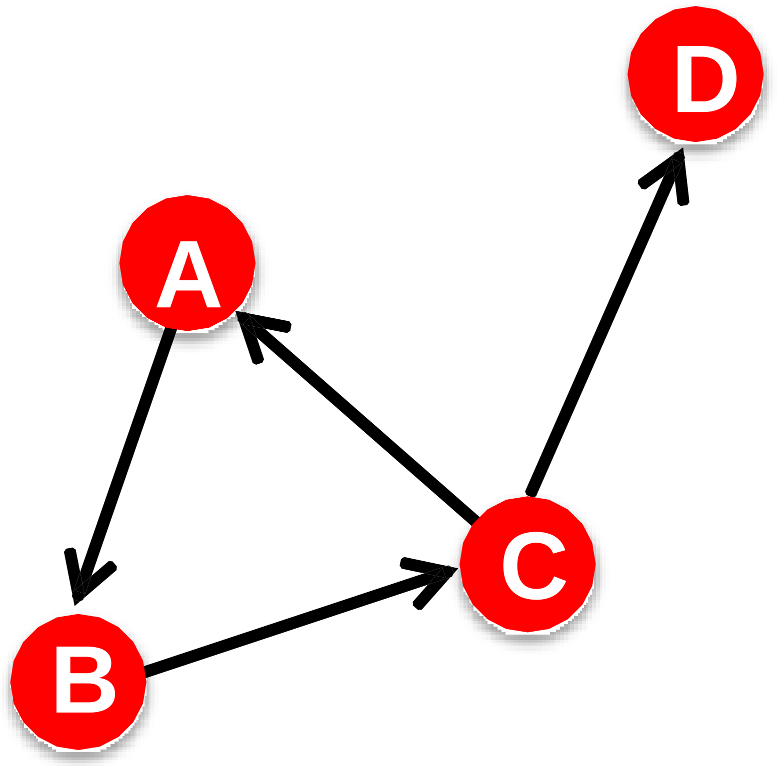
DISTANCE IN GRAPHS



$h_{B,D} = 2$
$h_{A,X} = \infty$

Distance (shortest path, geodesic) between a pair of nodes is defined as the number of edges along the shortest path connecting the nodes

*If the two nodes are not connected, the distance is usually defined as infinite (or zero)



$h_{B,C} = 1$, $h_{C,B} = 2$

In **directed graphs,** paths need to follow the direction of the arrows

Consequence: Distance is **not symmetric**:

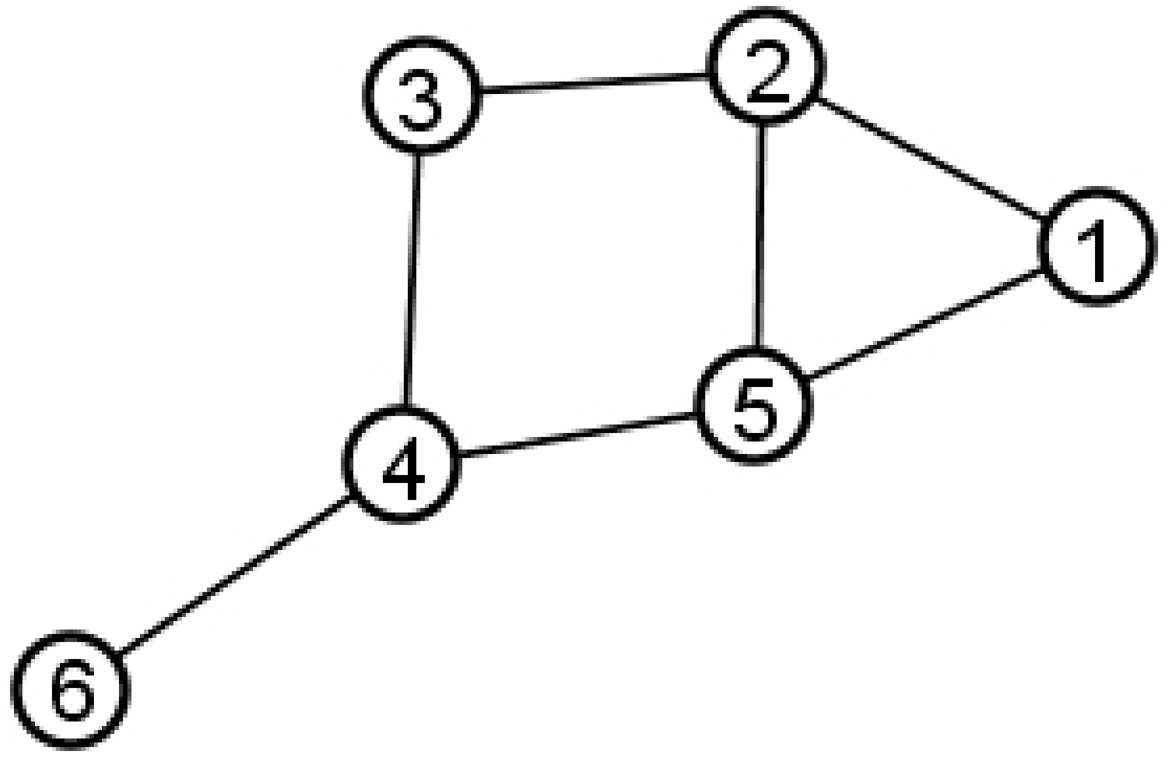
$$h_{B,C} \neq h_{C,B}$$

NETWORK PROPERTIES

NETWORK DIAMETER

- The **distance** $d_{G(v_i;\, v_j)}$ between two vertices is the number of edges in the shortest path from v_i to v_j
- Graph **diameter** is the largest shortest path: $D = max_{i;j}\, d_{G(v_i;\, v_j)}$
- **Average path length**:

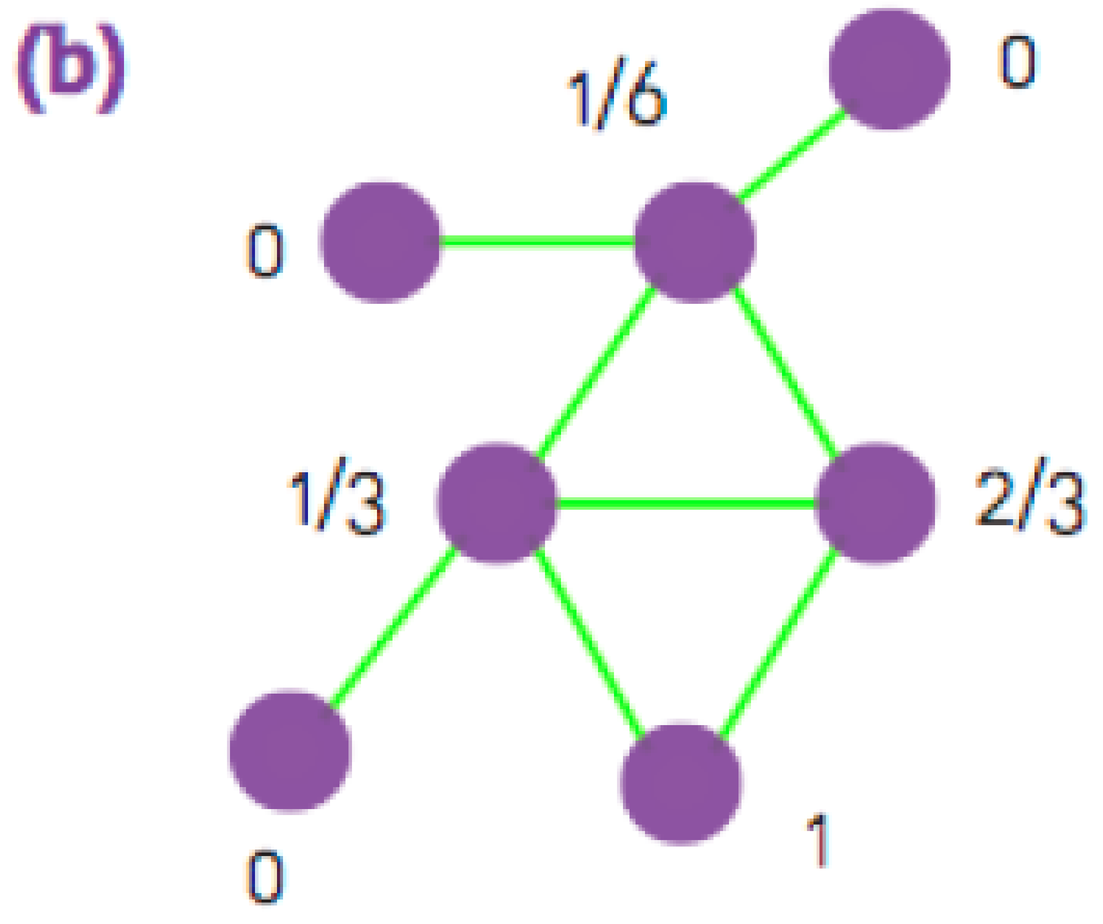
$$\langle L \rangle = \frac{1}{n(n-1)} \sum_{i \neq j} d_G(v_i, v_j)$$



NETWORK PROPERTIES
CLUSTERING COEFFICIENT

- Global clustering coefficient:

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



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

$$C_{\Delta} = \frac{3}{8} = 0.375$$

NETWORK PROPERTIES
CLUSTERING COEFFICIENT

- Local clustering coefficient (per vertex):

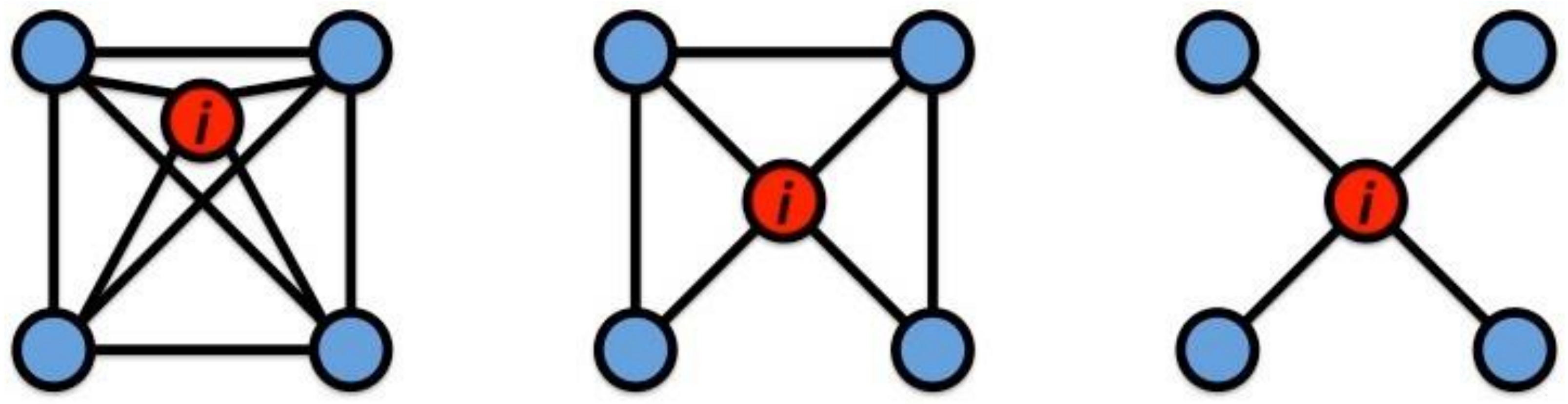
How connected are ***i***'s neighbors to each other?

$C_i = [0,1]$ Node ***i*** with degree $\boldsymbol{k_i}$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

where e_i is the number of edges between the neighbors of node i

Note $k_i(k_i - 1)$ is max number of edges between the k_i neighbors



$C_i = 1$ $C_i = 1/2$ $C_i = 0$

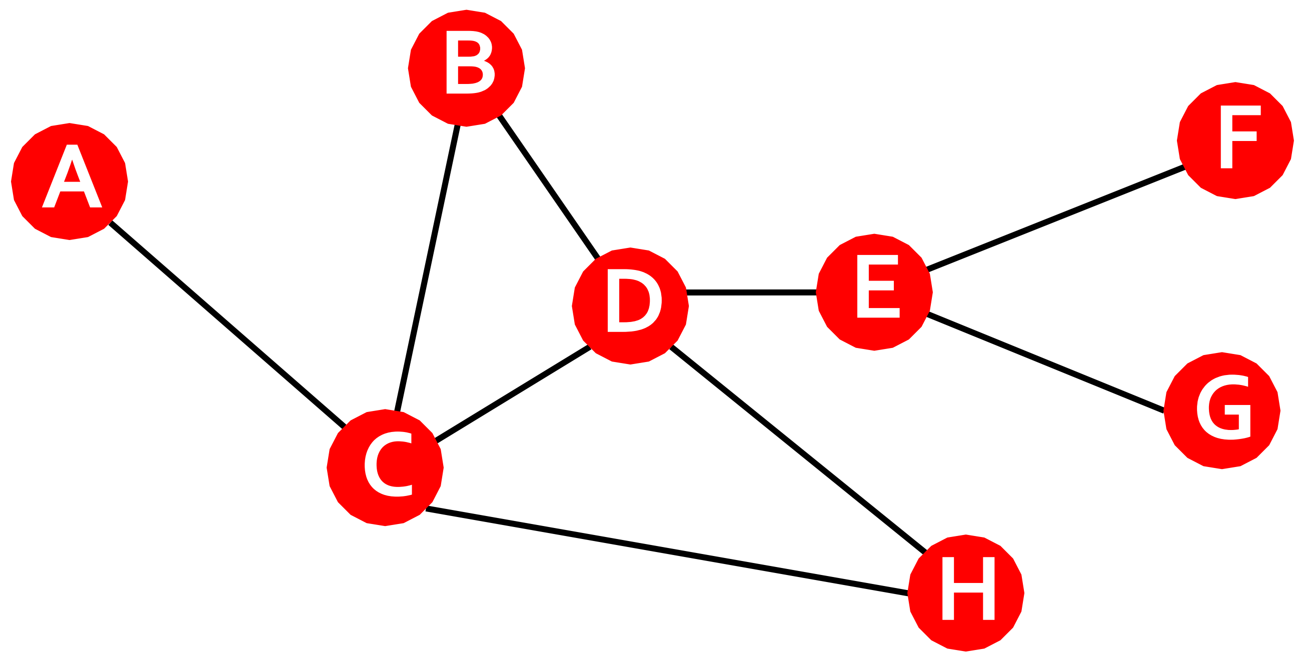
Clustering coefficient is undefined (or defined to be 0) for nodes with degree 0 or 1

NETWORK PROPERTIES

CLUSTERING COEFFICIENT

- Average clustering coefficient

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$



$k_B=2,\ e_B=1,\ C_B=2/2 = 1$

$k_D=4,\ e_D=2,\ C_D=4/12 = 1/3$

Avg. clustering: $C=0.33$

NETWORK PROPERTIES

NETWORK FRAMING