



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

# Social Network Analysis

A.Y. 23/24

Communication Strategies

# Graphs

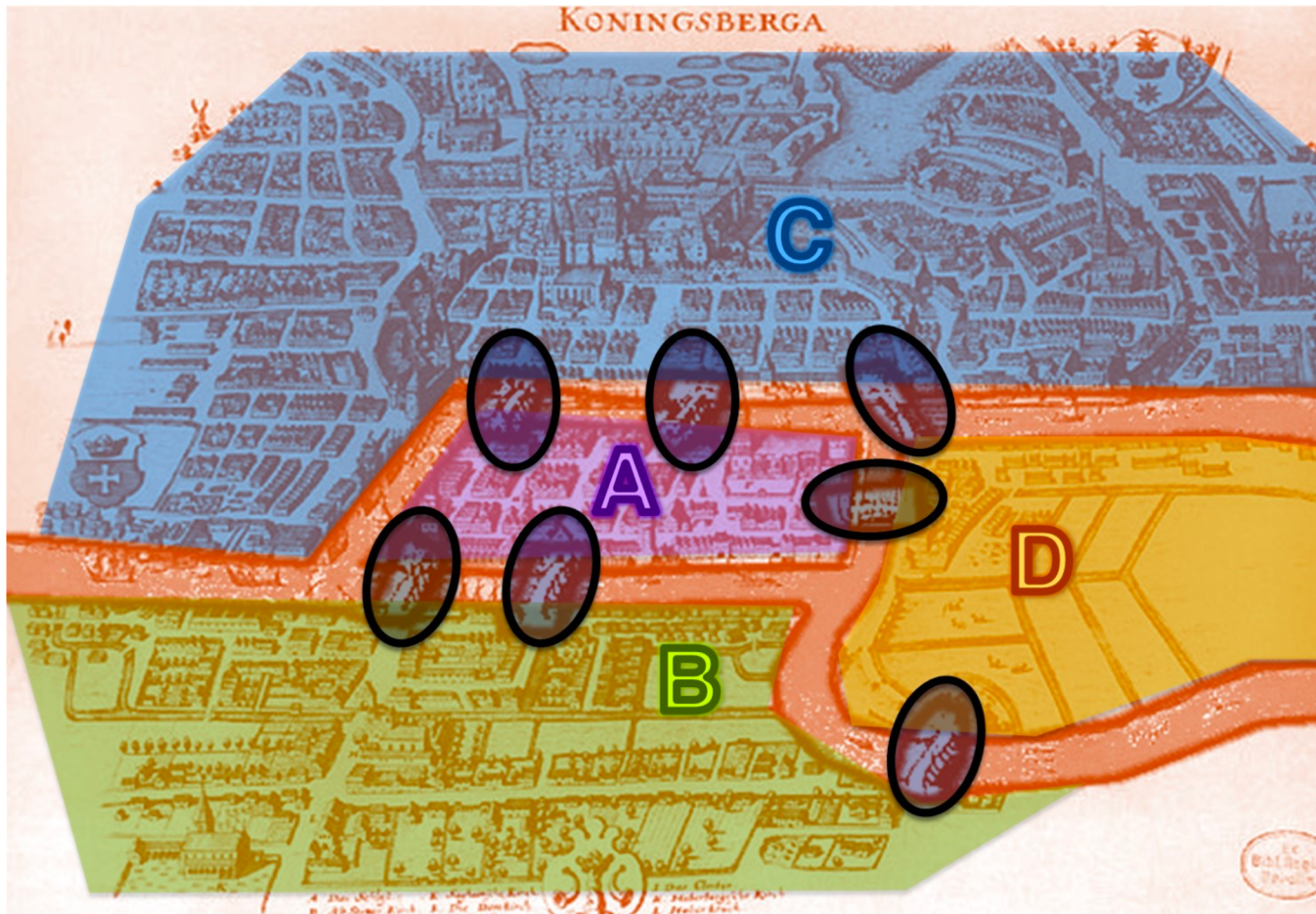
an introduction





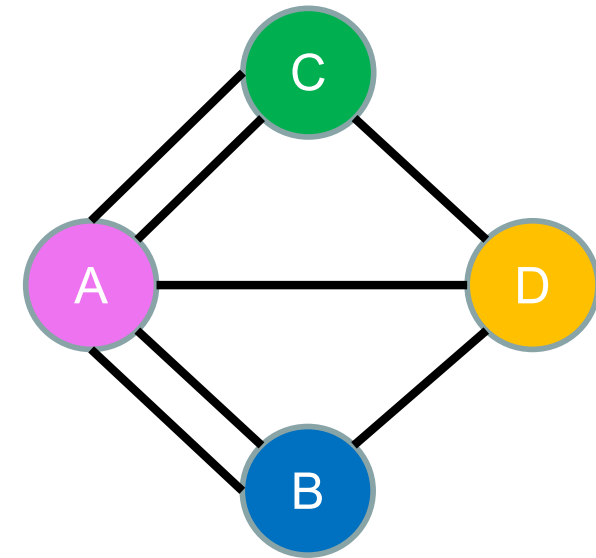
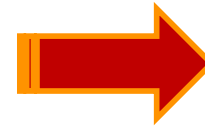
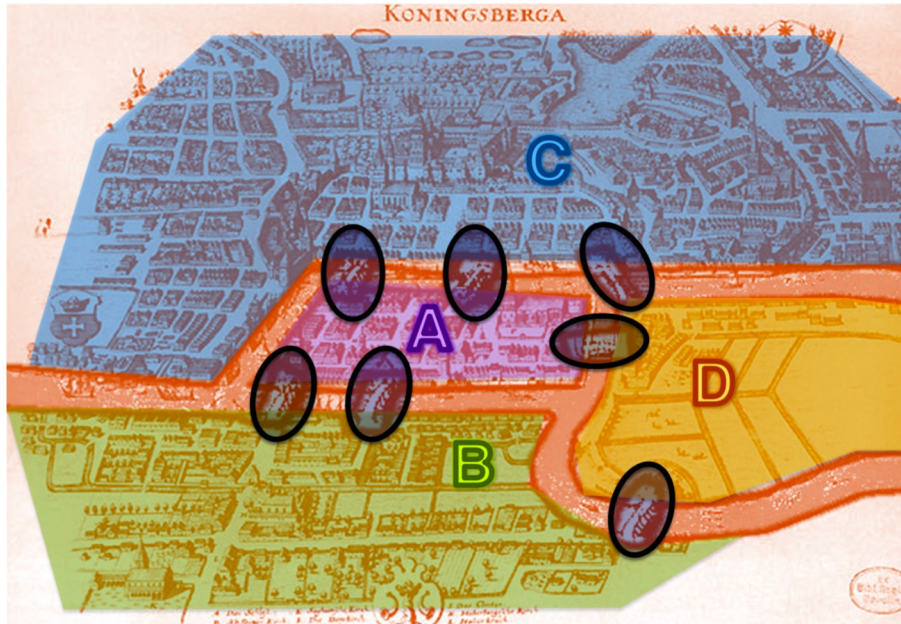
# Euler and the 7 bridges of Königsberg

(Prussia, 1736) today Kaliningrad



How to walk through the city by crossing each bridge only once?

# Networks as graphs



Graph  $\mathcal{G} (\mathcal{V}, \mathcal{E})$ : network

□ Vertices (set  $\mathcal{V}$ ): nodes, people, concepts

□ Edges (set  $\mathcal{E}$ ): links, relations, associations

↑  
*mathematics*

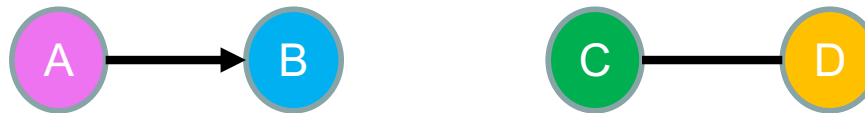
↑  
*technology*

↑  
*social  
psychology*

↑  
*social  
cognition*



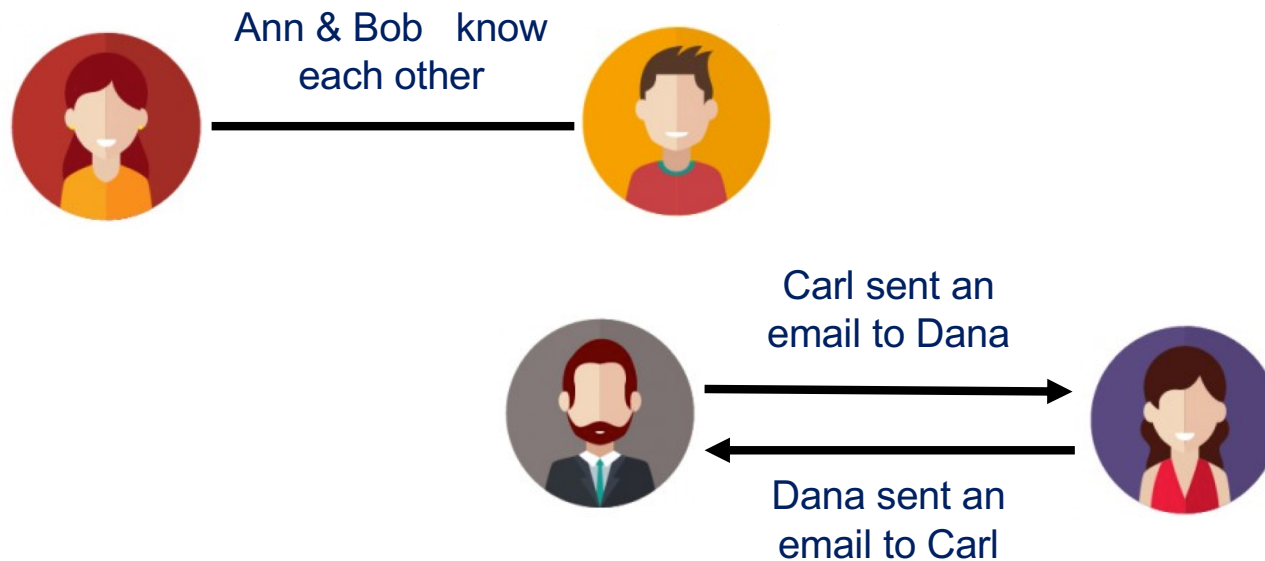
- ❑ A connection relationship can have a privileged direction or can be mutual
  - ❑ Either a **directed** or an **undirected** link



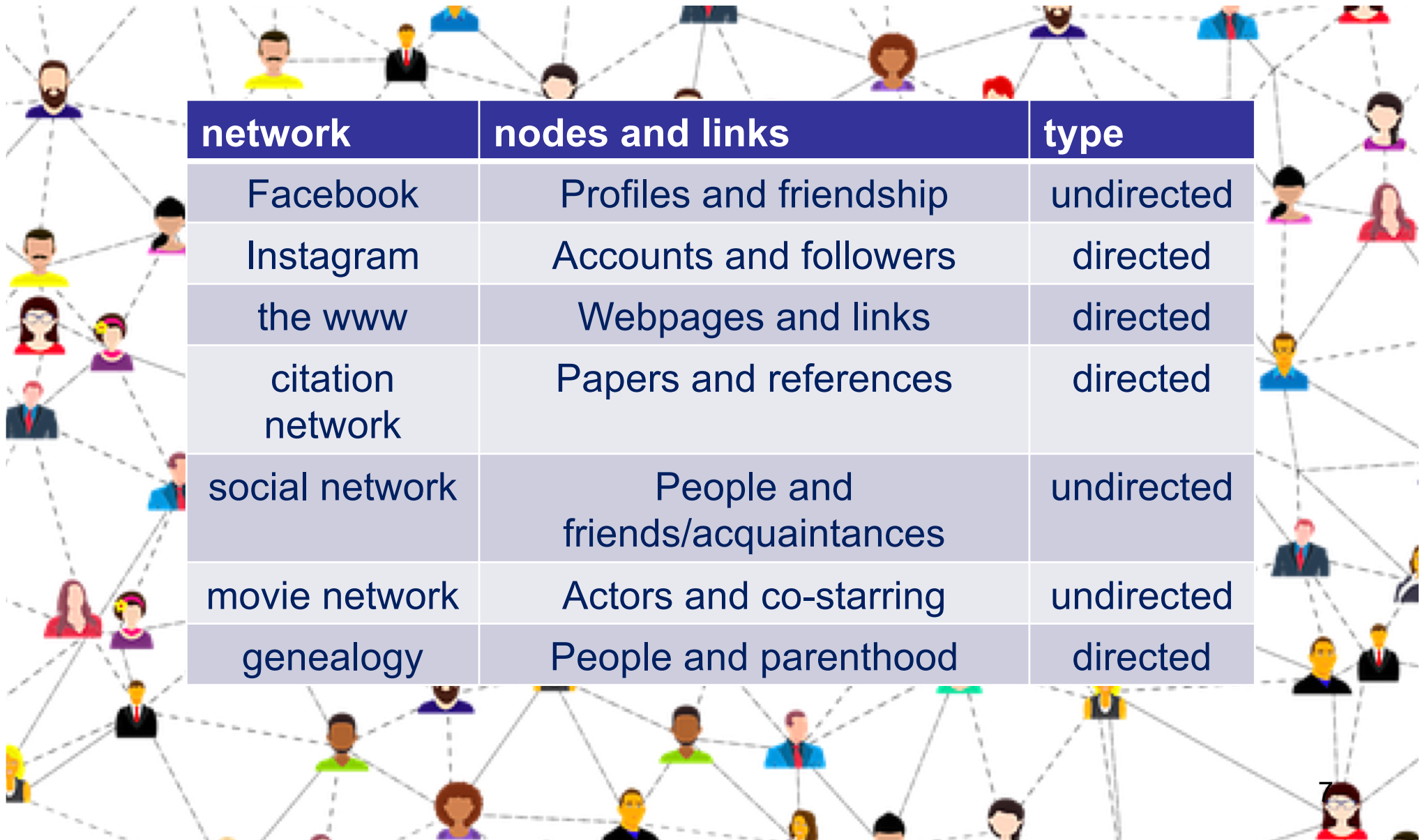
- ❑ If the network has only (un)directed links, it is also called itself (un)directed network
  - ❑ Certain networks can have both types



- At first glance **undirected** → **directed** by duplicating links, but not necessarily quite the same though

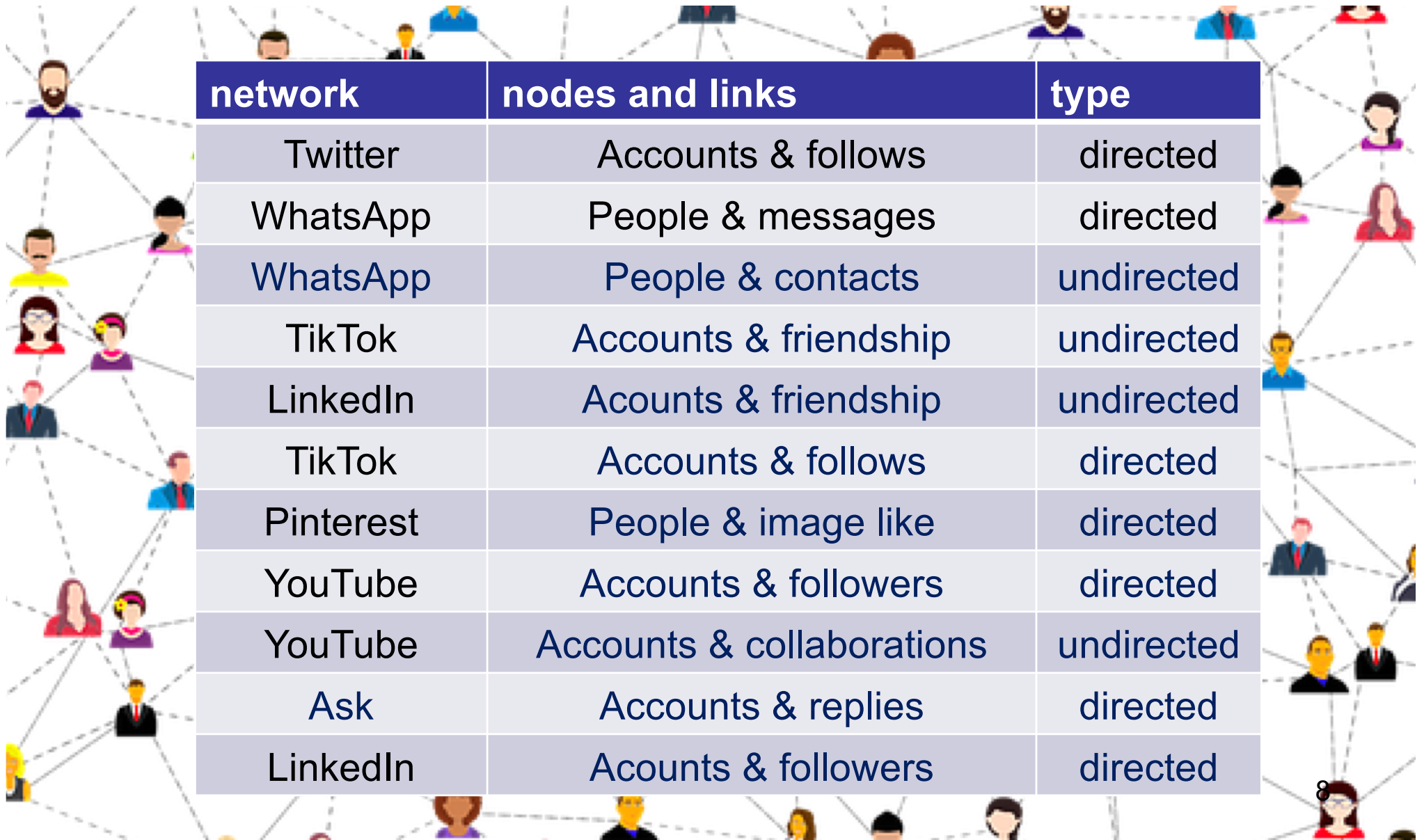








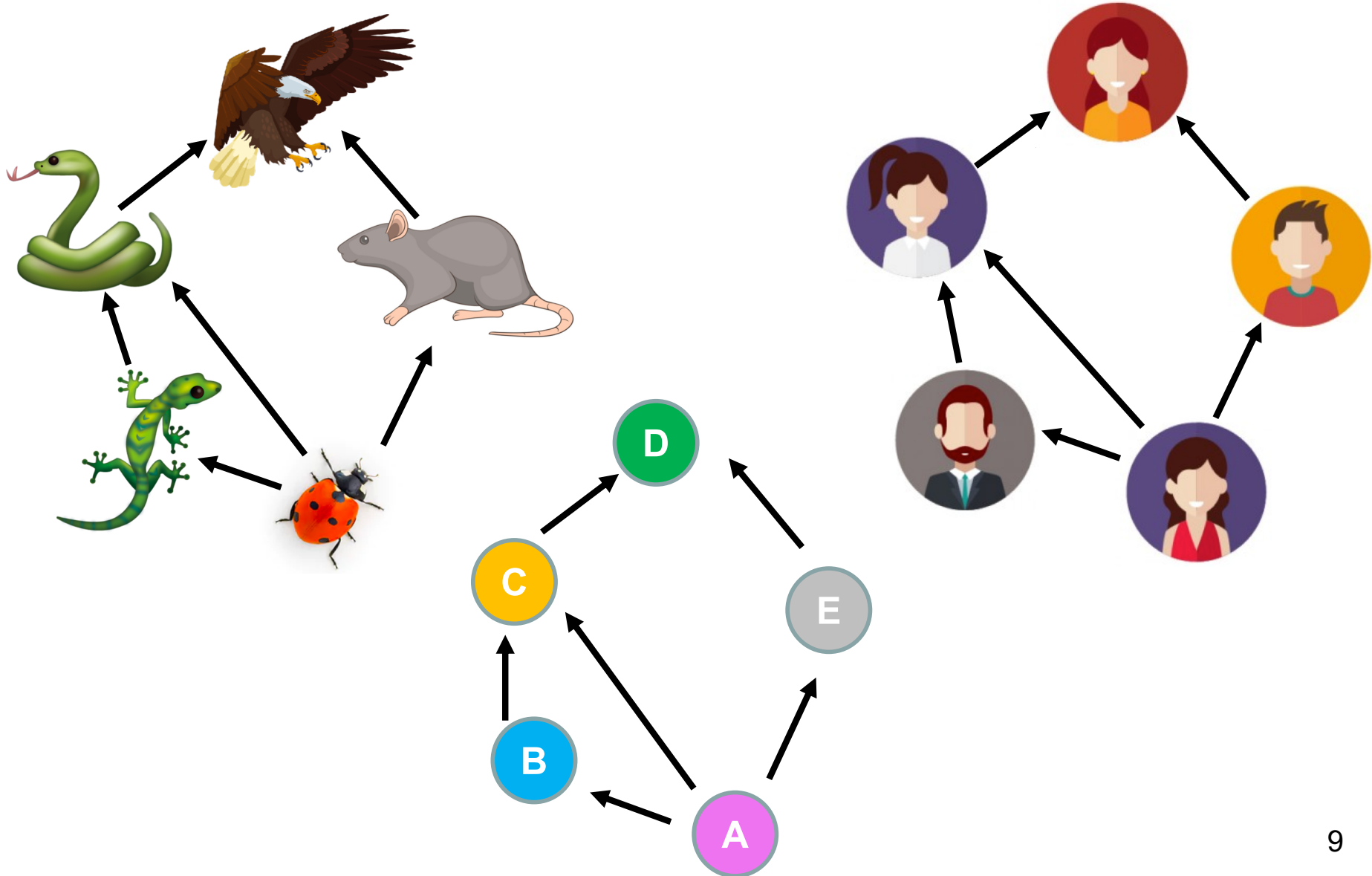
# Can U think of other social networks?



network	nodes and links	type
Twitter	Accounts & follows	directed
WhatsApp	People & messages	directed
WhatsApp	People & contacts	undirected
TikTok	Accounts & friendship	undirected
LinkedIn	Accounts & friendship	undirected
TikTok	Accounts & follows	directed
Pinterest	People & image like	directed
YouTube	Accounts & followers	directed
YouTube	Accounts & collaborations	undirected
Ask	Accounts & replies	directed
LinkedIn	Accounts & followers	directed



# Generality of representation

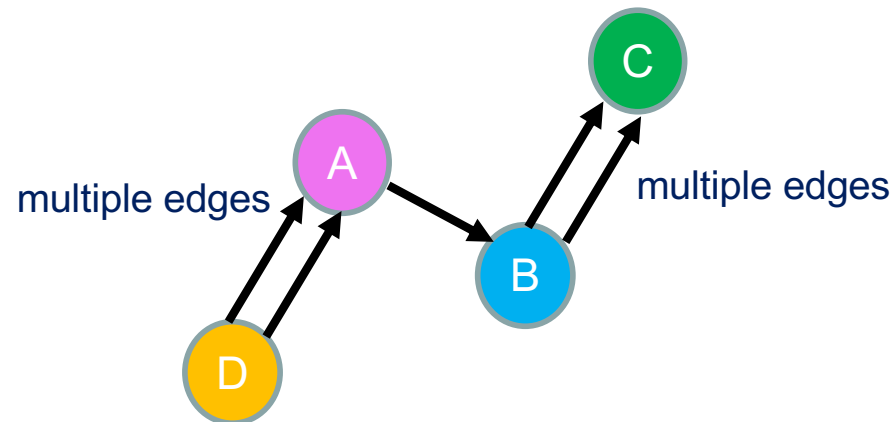


# Graph representations

visual plot, adjacency matrix, edge list



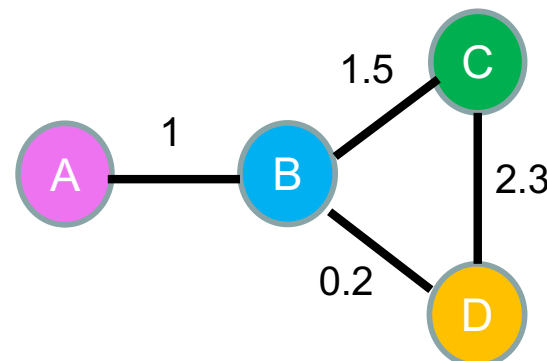
- ❑ Multi-graphs (or pseudo-graphs)  
Some network representations require **multiple** links (e.g., number of citations from one author to another)



## □ Weighted graph

Sometimes a **weight** is associated to a link, e.g., to underline that the links are not identical (strong/weak relationships)

Can be seen as a generalization of multi-graphs (weight = # of links)



e.g., **strength of a tie**

0.2 = weak (acquaintances)

1 = strong (friends)

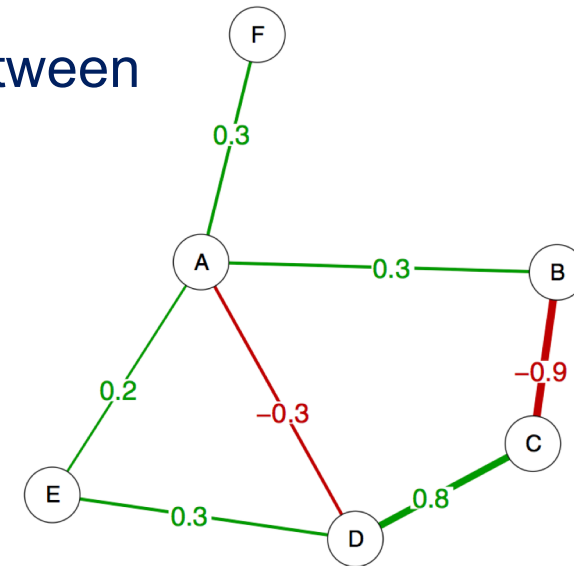
1.5 = stronger (close friends)

2.3 = very strong (best friends)

## □ Edges can have signed values

**positive** if there is an **agreement** between nodes

**negative** if there's a **disagreement**



## □ This is typical of **correlation** networks

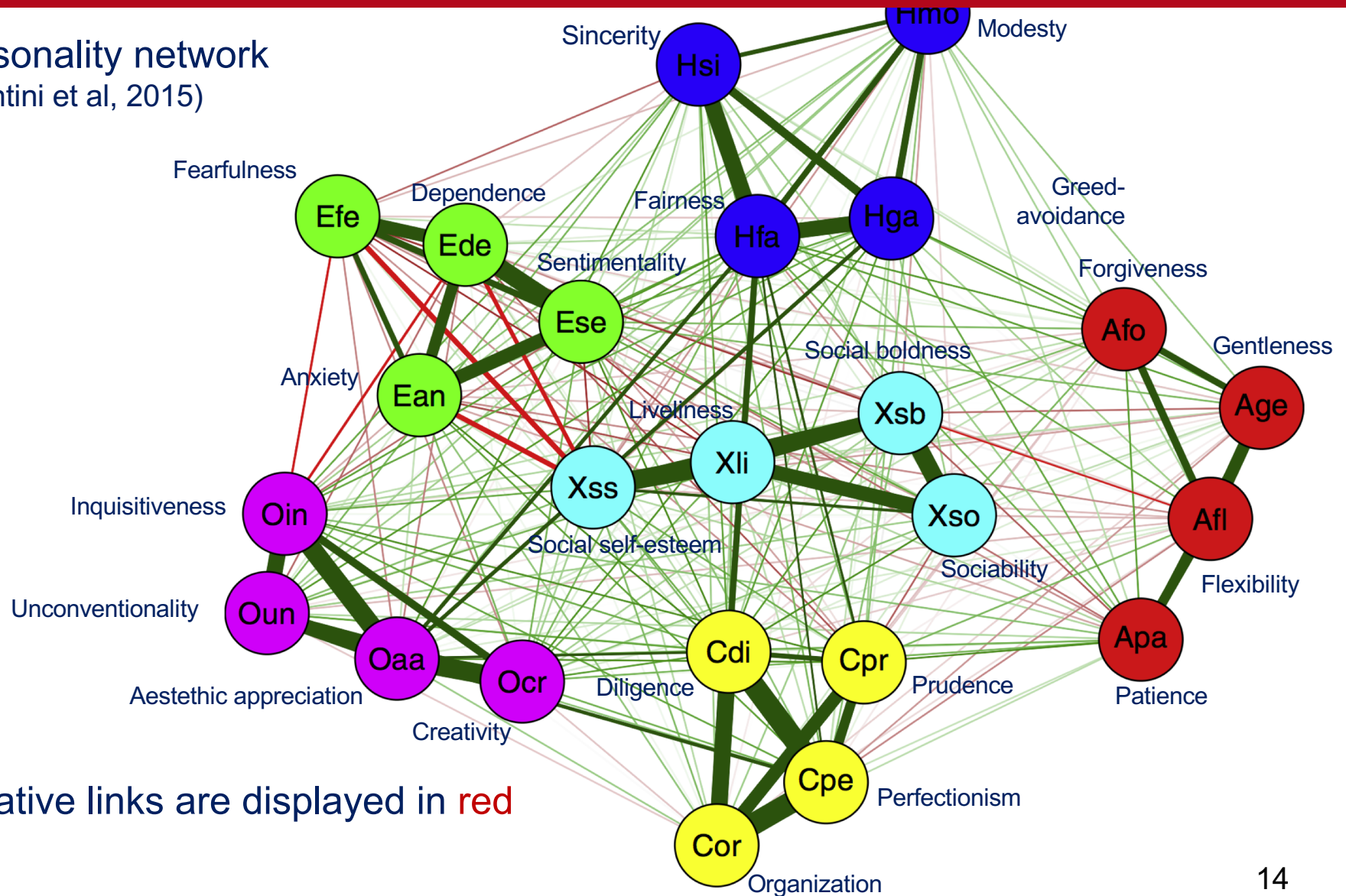
correlation = a measure of similarity

## □ More difficult to handle



# Signed graph example

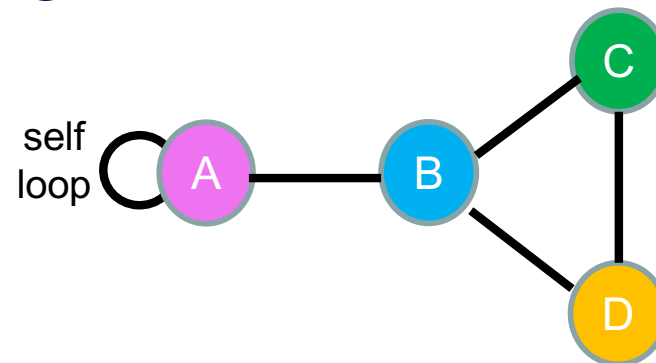
A personality network  
(Costantini et al, 2015)



Negative links are displayed in red



- ❑ In many networks nodes do not interact with themselves
- ❑ To account for self-interactions, we add **loops** to represent them

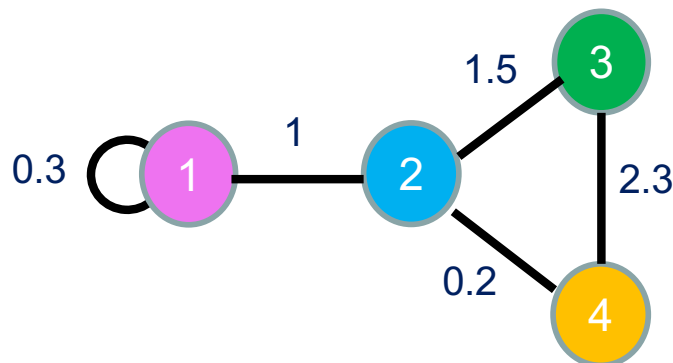




□ An adjacency matrix  $A = [a_{ij}]$  associated to graph  $\mathcal{G}$  has

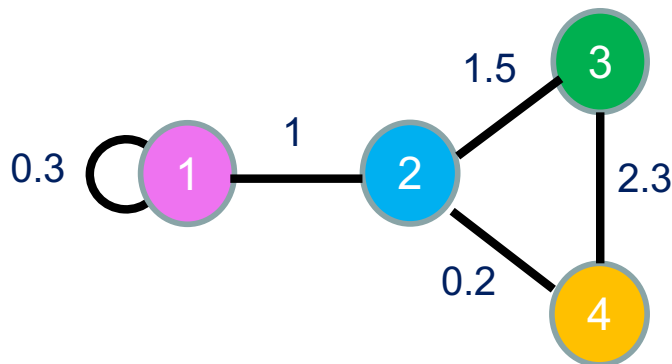
$i$  is the row index  $j$  is the column index

entries  $a_{ij} = 0$  if nodes  $i$  and  $j$  are **not connected**  
if nodes  $i$  and  $j$  are **connected** then  $a_{ij} \neq 0$



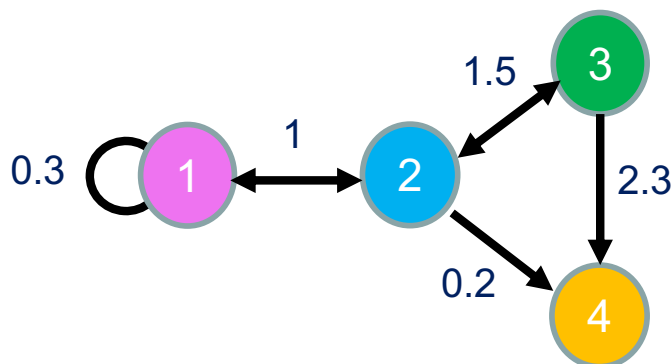
$$A = \begin{matrix} & \begin{matrix} \text{this is } a_{12} \\ \swarrow \\ 1 \end{matrix} & & & \\ \begin{matrix} 0.3 & 1 & 0 & 0 \end{matrix} & \leftarrow \text{row 1} & & & \\ \begin{matrix} 1 & 0 & 1.5 & 0.2 \\ 0 & 1.5 & 0 & 2.3 \\ 0 & 0.2 & 2.3 & 0 \end{matrix} & & & & \\ & \begin{matrix} \uparrow \\ \text{column 2} \end{matrix} & & & \end{matrix}$$

- Undirected graph = **symmetric** matrix



$$A = \begin{bmatrix} 0.3 & 1 & 0 & 0 \\ 1 & 0 & 1.5 & 0.2 \\ 0 & 1.5 & 0 & 2.3 \\ 0 & 0.2 & 2.3 & 0 \end{bmatrix}$$

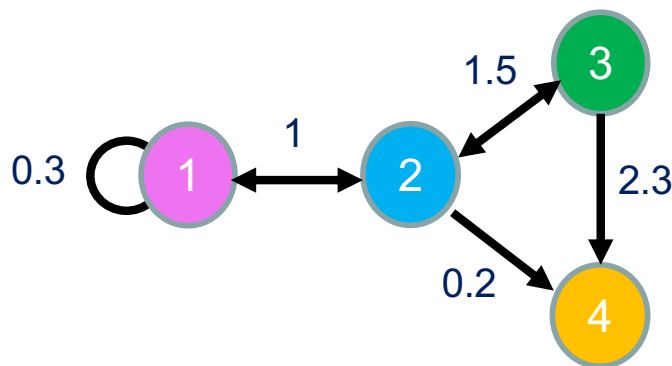
- Directed graph = **asymmetric** matrix



$$A = \begin{bmatrix} 0.3 & 1 & 0 & 0 \\ 1 & 0 & 1.5 & 0 \\ 0 & 1.5 & 0 & 0 \\ 0 & 0.2 & 2.3 & 0 \end{bmatrix}$$



- The weight  $a_{ij}$  is associated to
  - $i$  th row
  - $j$  th column
  - directed edge  $j \rightarrow i$  starting from node  $j$  and leading to node  $i$



$$A = \begin{bmatrix} 0.3 & 1 & 0 & 0 \\ 1 & 0 & 1.5 & 0 \\ 0 & 1.5 & 0 & 0 \\ 0 & 0.2 & 2.3 & 0 \end{bmatrix}$$

Labels in the matrix:  $a_{24}$  (circled 0),  $a_{34}$  (circled 0),  $a_{42}$  (circled 0.2),  $a_{43}$  (circled 2.3). A dashed red line is drawn from the top-right to the bottom-left.

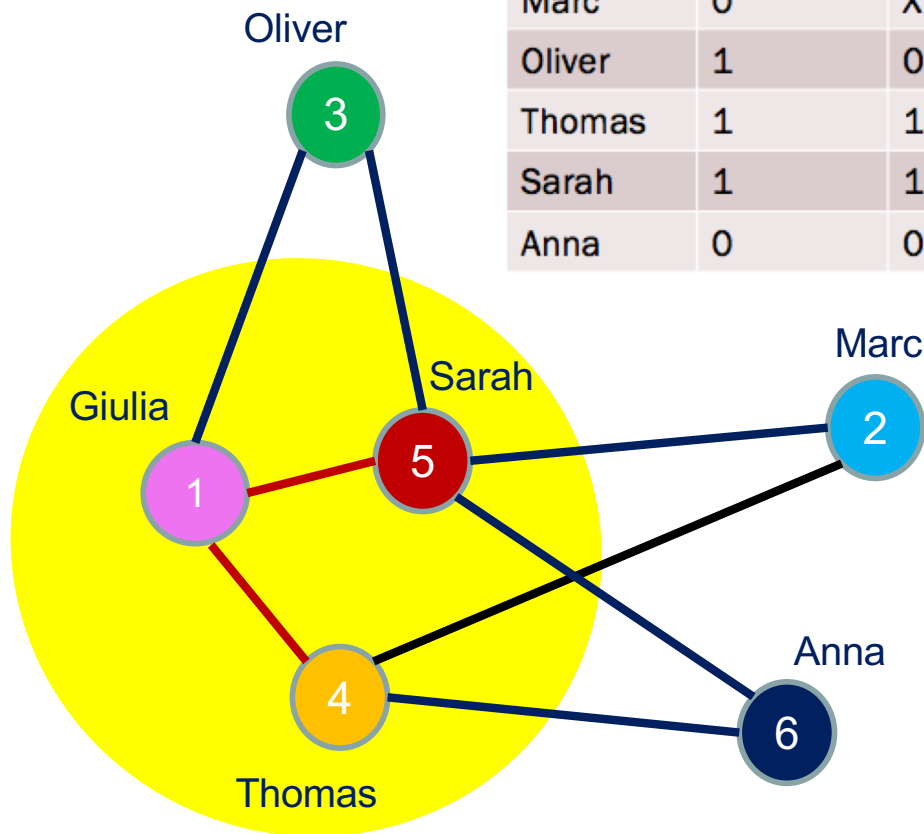




# An example

which of these representations do you like best?

	Giulia	Marc	Oliver	Thomas	Sarah	Anna
Giulia	X					
Marc	0	X				
Oliver	1	0	X			
Thomas	1	1	0	X		
Sarah	1	1	1	0	X	
Anna	0	0	0	1	1	x



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

which of these representations do you like best?



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

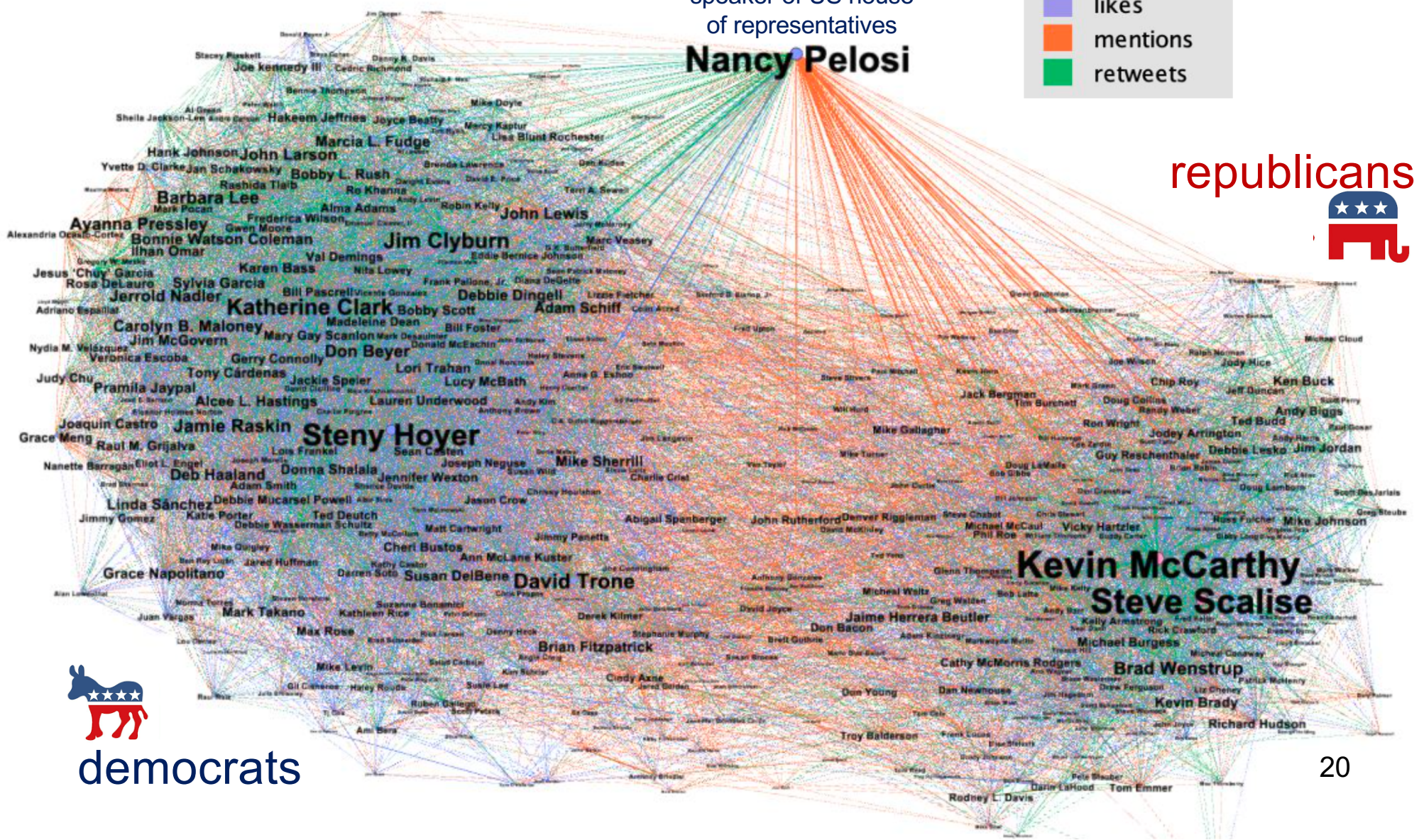
# Graph plots may carry relevant info...

US republicans and democrats interactions on Twitter (2020)

speaker of US house  
of representatives  
**Nancy Pelosi**

- likes
- mentions
- retweets

republicans



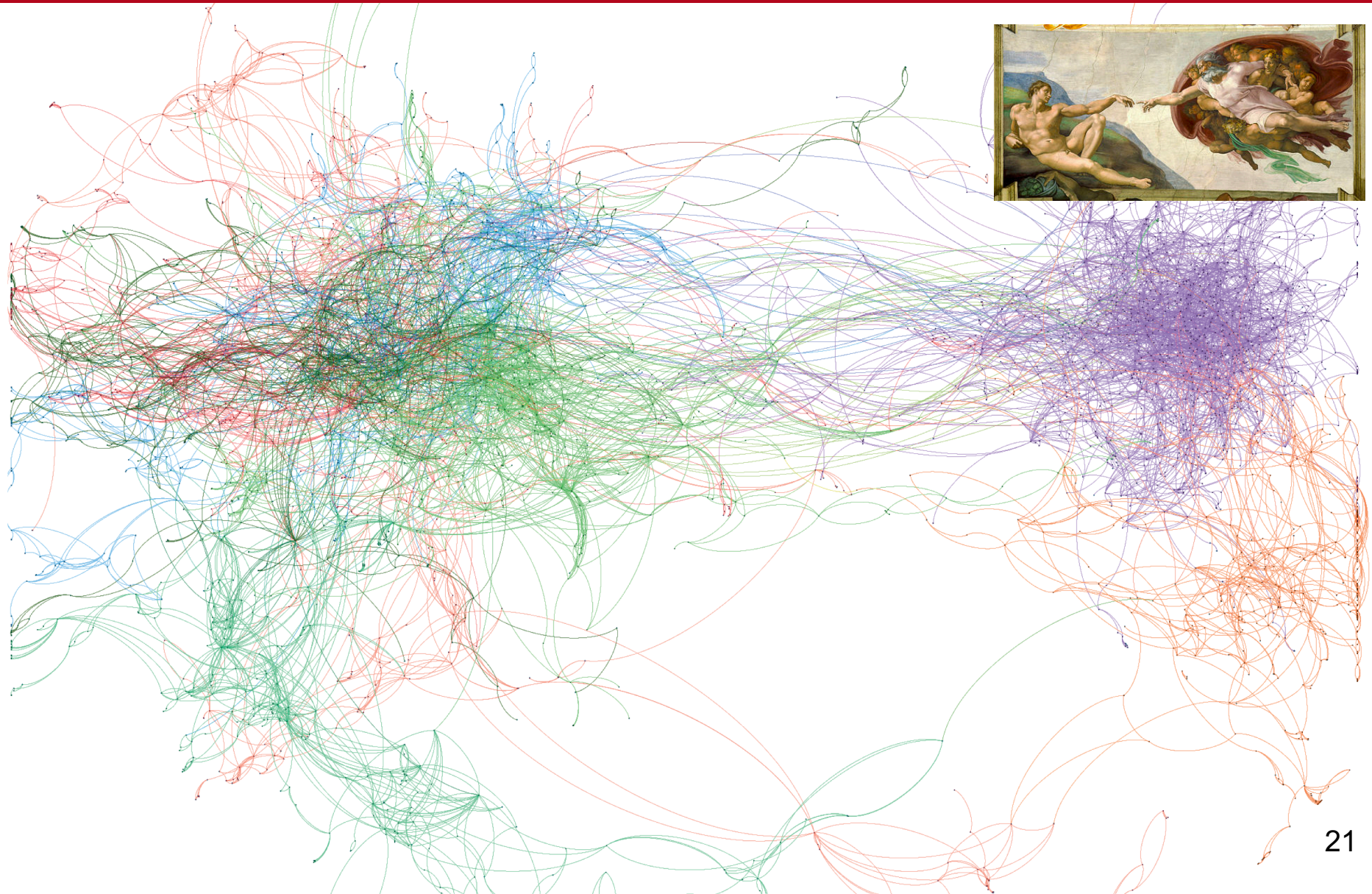
democrats





UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

... or may not!

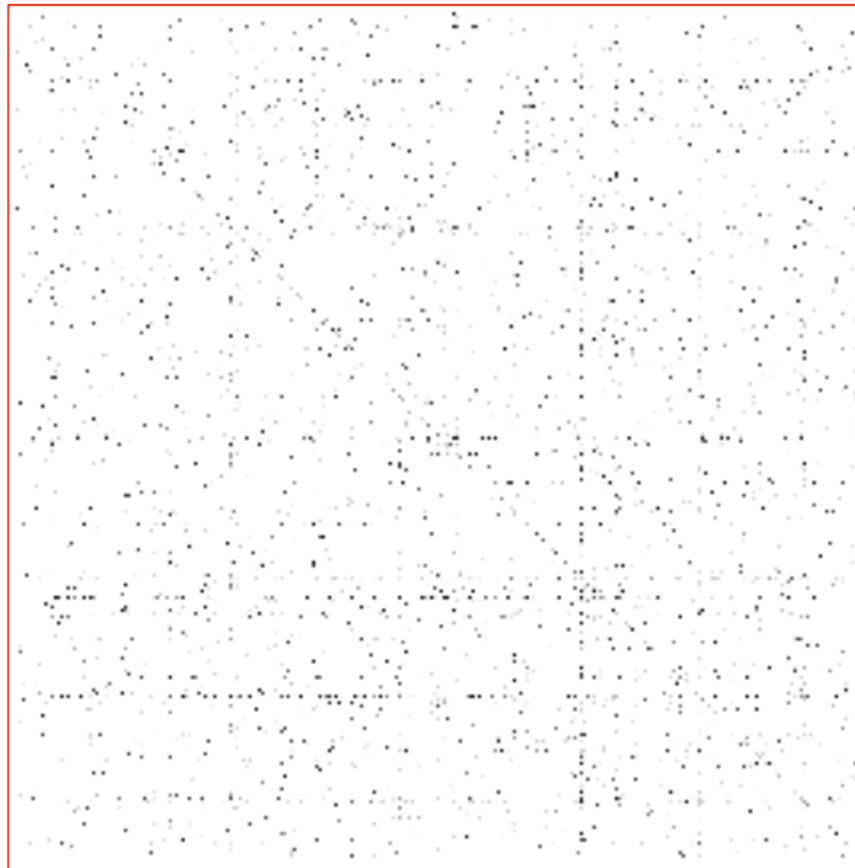




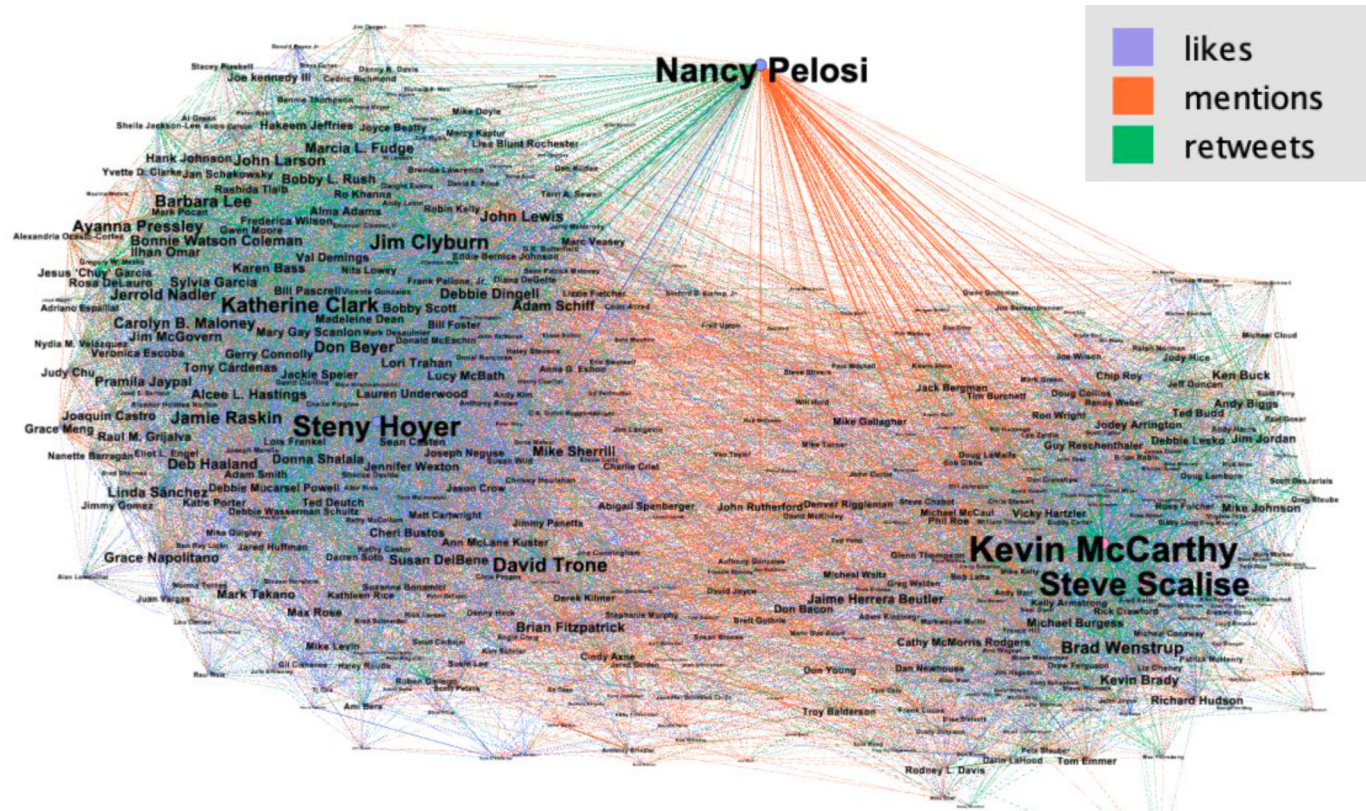
□ The adjacency matrix is typically sparse

good for tractability !

$A =$







described by a **set** of adjacency matrices  $A_\ell$

e.g., one for likes, one for mentions, and one for retweets



□ So, what's the take-away so far?



# Storing network data adjacency matrix versus edge list

adjacency matrix

$A = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$   $N^2$  entries

edge list

$L$  entries

- 1 → 3
- 1 → 4
- 1 → 5
- 2 → 4
- 2 → 5
- 3 → 5
- 4 → 6
- 5 → 6

Which one do U think is better?

# Distances in graphs

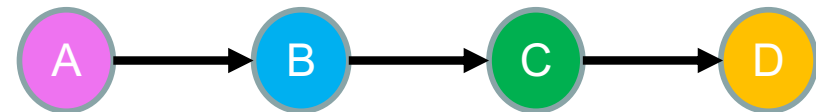
and related concepts





## □ Path

a sequence of interconnected nodes (meaning each pair of nodes adjacent in the sequence are connected by a link)

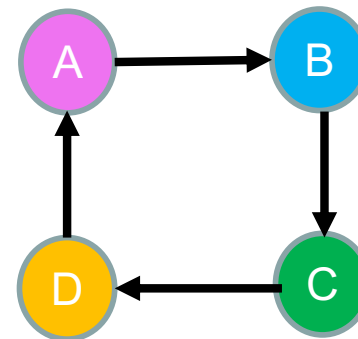


## □ Path length

# of links involved in the path (if the path involves  $n$  nodes then the path length is  $n-1$ )

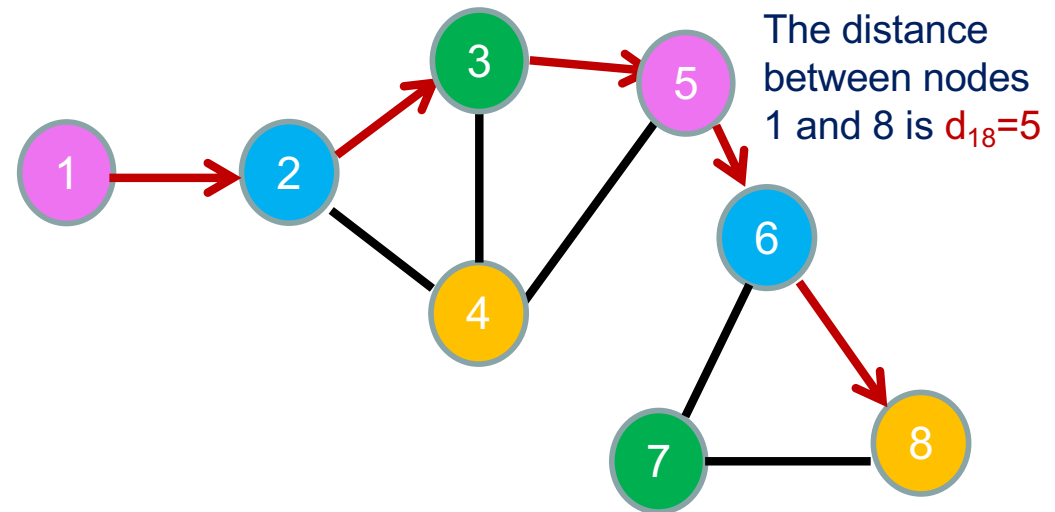
## □ Cycle

path where starting and ending nodes coincide



## □ Shortest path (between any two nodes)

the path with the minimum length, which is called the **distance**



it is **not** unique!

## □ Diameter (of the network)

the highest distance in the network

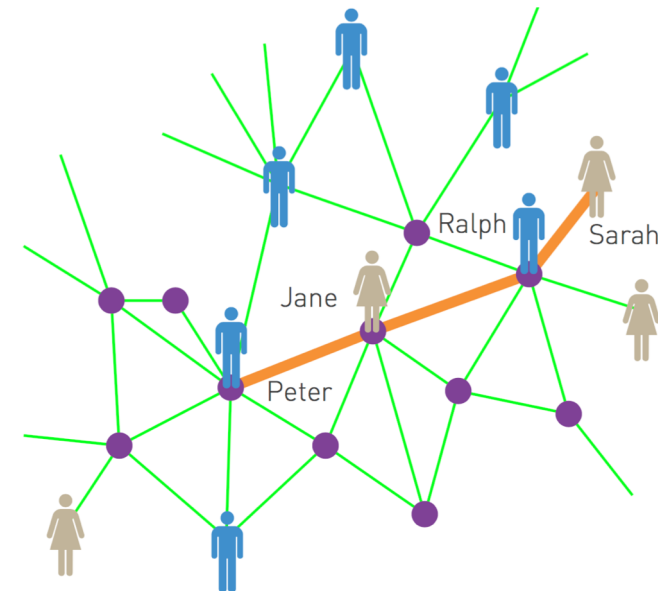
The diameter is  $d=5$

## □ Average path length

average distance between all nodes pairs (apply an algorithm to all node couples, and take the average)

- In real networks distance between two randomly chosen nodes is generally **short**
- Milgram [1967]: *6 degrees of separation*

- What does this mean?  
We are more connected than we think





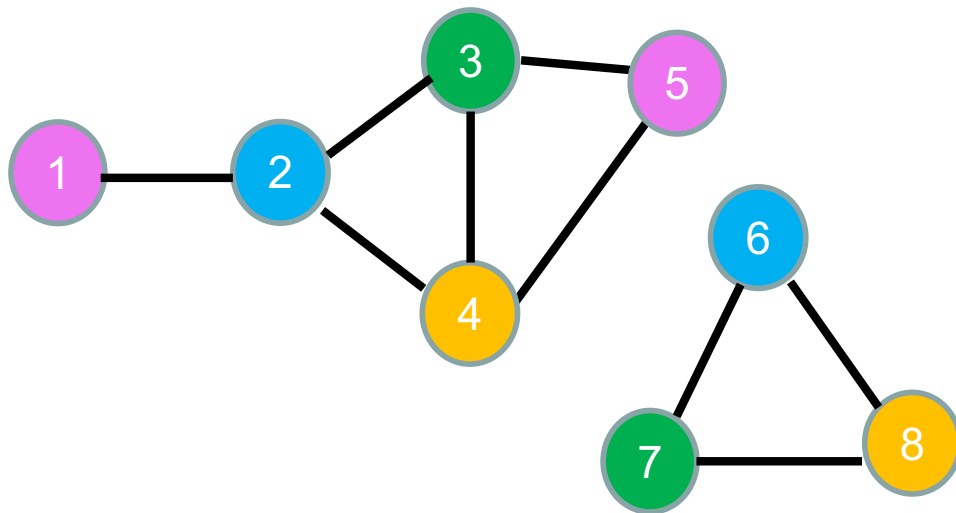
## □ Connected graph (undirected)

for all couples  $(i,j)$  there exists a path connecting them

if **disconnected**, we count the # of connected components  
(e.g., use BFS and iterate)

## □ Giant component (the biggest one)

## □ Isolates (the other ones)



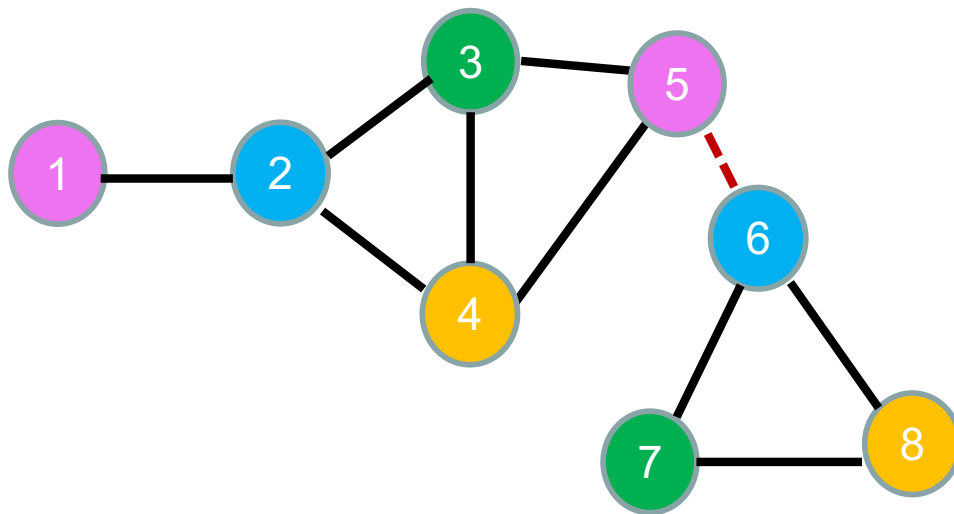
$A =$

0	1	0	0	0	0	0	0
1	0	1	1	0	0	0	0
0	1	0	1	1	0	0	0
0	1	1	0	1	0	0	0
0	0	1	1	0	0	0	0
0	0	0	0	0	0	1	1
0	0	0	0	0	1	0	1
0	0	0	0	0	1	1	0

block-diagonal matrix

□ A **bridge** is a link between two connected components

its removal would make the network disconnected

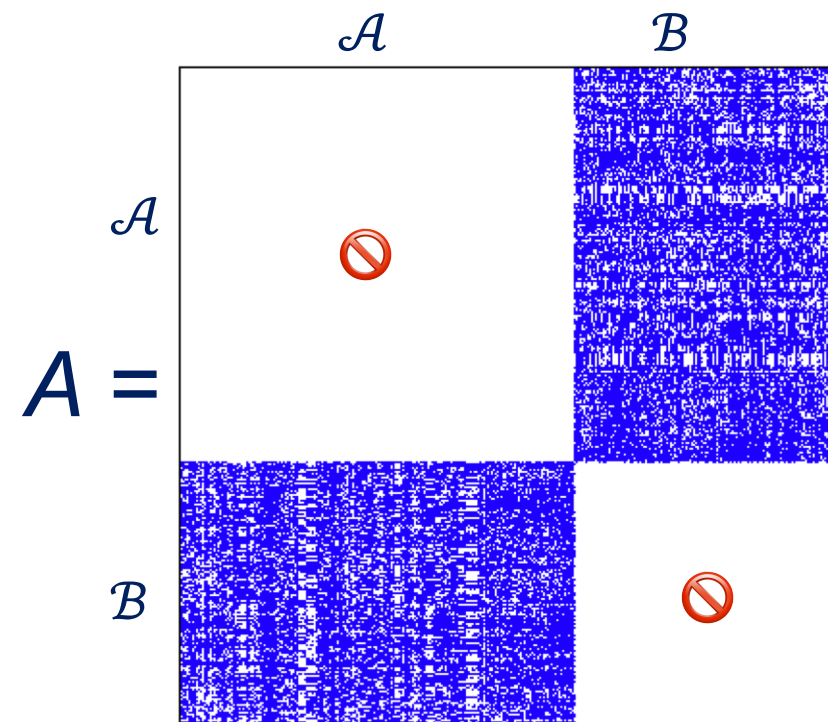
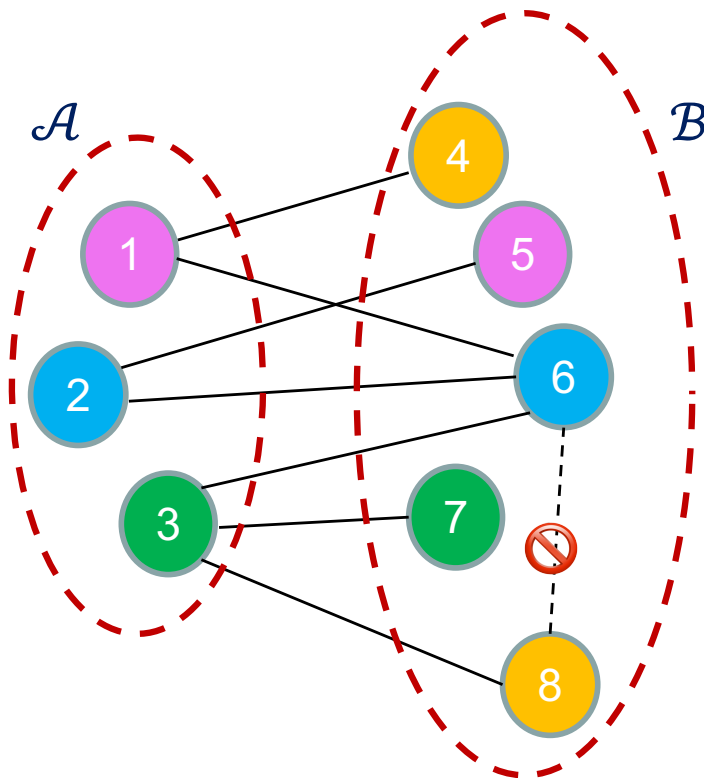


$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

# Bipartite graphs

and semantic networks

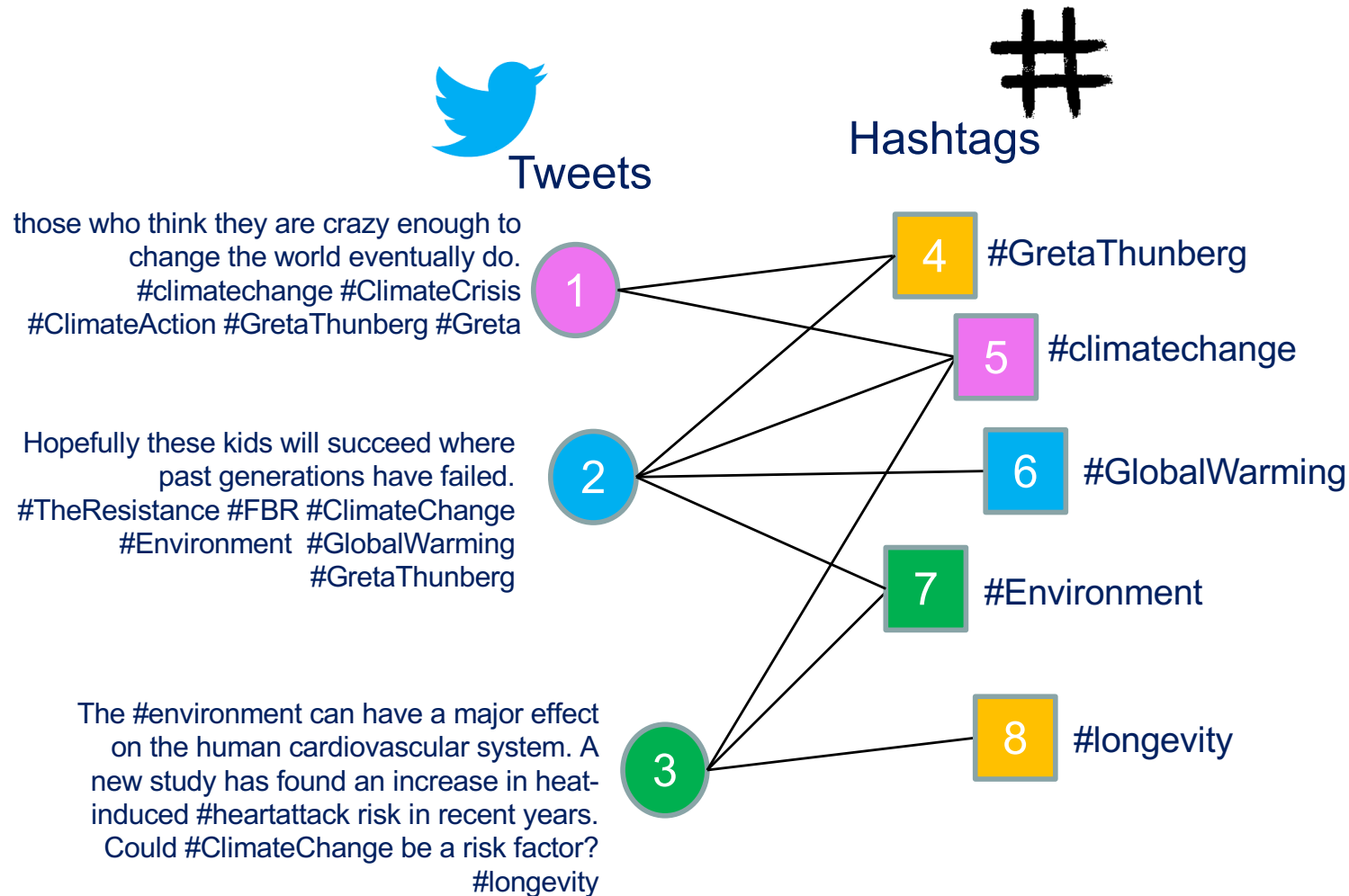
- Connections are available only between the groups  $A$  and  $B$







# Bipartite graph example





- ❑ Bipartite graphs represent **memberships**/relationships, e.g., groups ( $\mathcal{A}$ ) to which people ( $\mathcal{B}$ ) belong

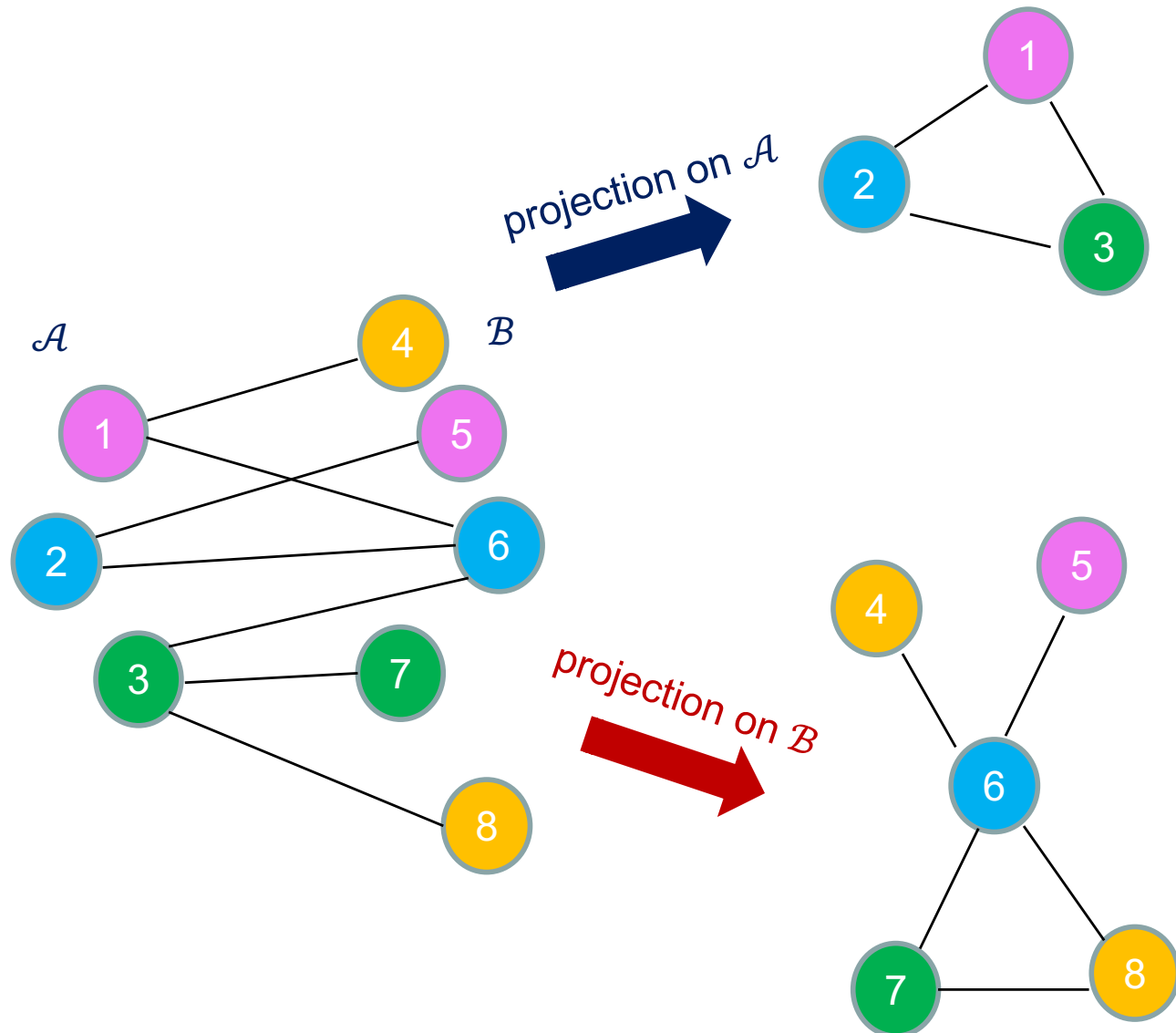
examples: movies/actors, classes/students, conferences/authors

- ❑ We can build separate networks (**projections**) for  $\mathcal{A}$  and  $\mathcal{B}$  (sometimes this is useful)

in the **movies/actors** example being linked can be interpreted in two ways: “**actors in the same movie**” (projection on  $\mathcal{B}$ ), or “**movies sharing the same actor**” (projection on  $\mathcal{A}$ )



# Abstract example



Nodes are linked if they have a **common neighbour** in  $\mathcal{B}$

PS: we say that nodes  $i$  and  $j$  have a common neighbour  $k$  if both  $i$  and  $j$  are connected to  $k$

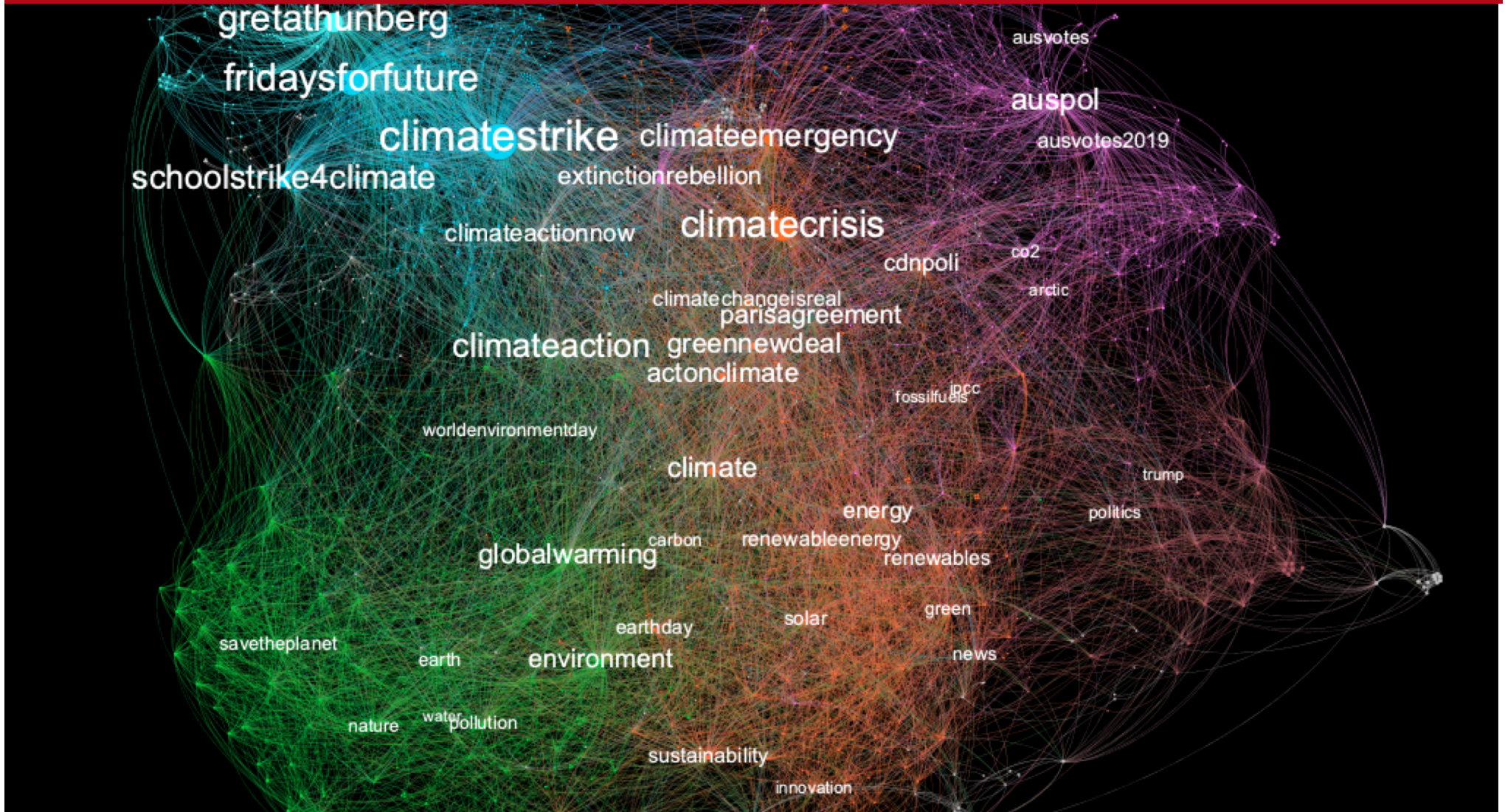
Nodes are linked if they have a **common neighbour** in  $\mathcal{A}$





# Projection on a semantic network

#hashtags that appear in the same tweet are linked



#climateaction tweets after Greta Thunberg







- ❑ (un)Directed graphs
- ❑ Weighted and signed graphs
- ❑ Adjacency matrix & edge list
- ❑ Distances
- ❑ Giant component, isolates, bridges
- ❑ Bipartite graphs & projections

# Degree centrality

a first approach to node importance



## Centrality

---

From Wikipedia, the free encyclopedia

*For the statistical concept, see [Central tendency](#).*

In [graph theory](#) and [network analysis](#), indicators of **centrality** identify the most important [vertices](#) within a graph.

Applications include identifying the most influential person(s) in a [social network](#), key infrastructure nodes in the [Internet](#) or [urban networks](#), and [super-spreaders](#) of disease. Centrality concepts were first developed in [social network analysis](#), and many of the terms used to measure centrality reflect their [sociological](#) origin.<sup>[1]</sup> They should not be confused with [node influence metrics](#), which seek to quantify the influence of every node in the network.



[Degree centrality](#) [\[ edit \]](#)

*Main article: [Degree \(graph theory\)](#)*

[PageRank centrality](#) [\[ edit \]](#)

*Main article: [PageRank](#)*

[Betweenness centrality](#) [\[ edit \]](#)

*Main article: [Betweenness centrality](#)*

[Eigenvector centrality](#) [\[ edit \]](#)

*Main article: [Eigenvector centrality](#)*

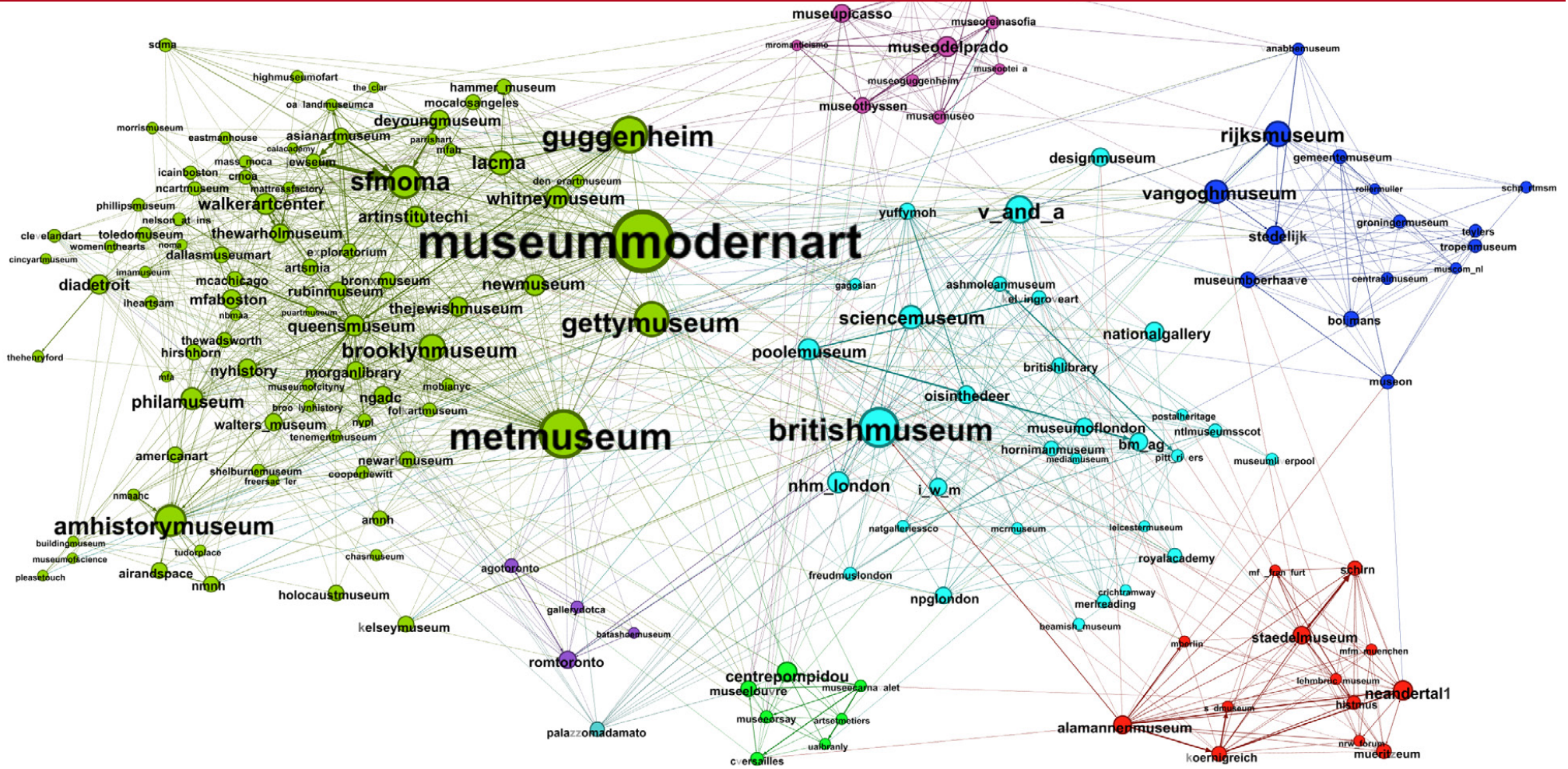
[Closeness centrality](#) [\[ edit \]](#)

*Main article: [Closeness centrality](#)*



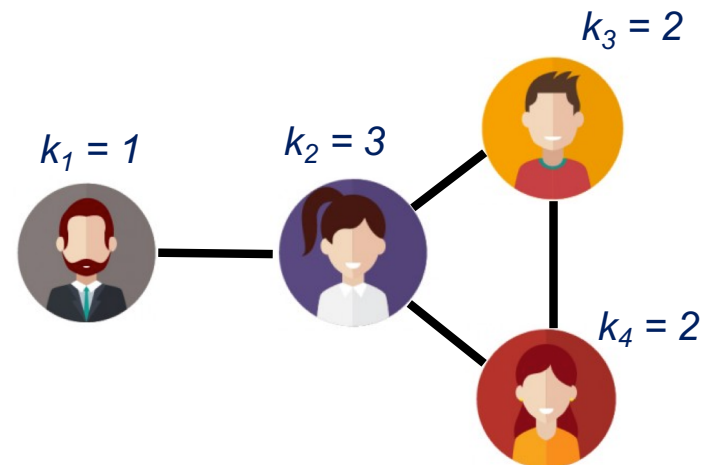


# An example of node centrality museums network



□ The **degree**  $k_i$  of node  $i$  in an **undirected** networks is

the # of links  $i$  has to other nodes, or  
the # of nodes  $i$  is linked to

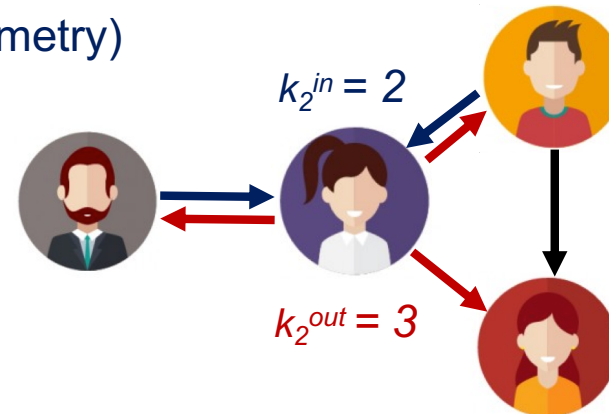


The **average** degree is

$$\langle k \rangle = \sum_i k_i / N = (1+3+2+2)/4 = 2$$

- For **directed** networks we distinguish between
  - in-degree**  $k_i^{in}$  = # of entering links
  - out-degree**  $k_i^{out}$  = # of exiting links

(undirected:  $k_i^{in} = k_i^{out}$  due to the symmetry)



The **average** degree is

$$\begin{aligned}
 \langle k \rangle &= \sum k_i^{out} / N = (1+3+2+0)/4 \\
 &= \sum k_i^{in} / N = (1+2+1+2)/4 \\
 &= 3/2
 \end{aligned}$$

- ❑ A social-capital measure of **cohesion**
- ❑ In-degree = importance as an **Authority**
- ❑ Out-degree = importance as a **Hub**

In www:

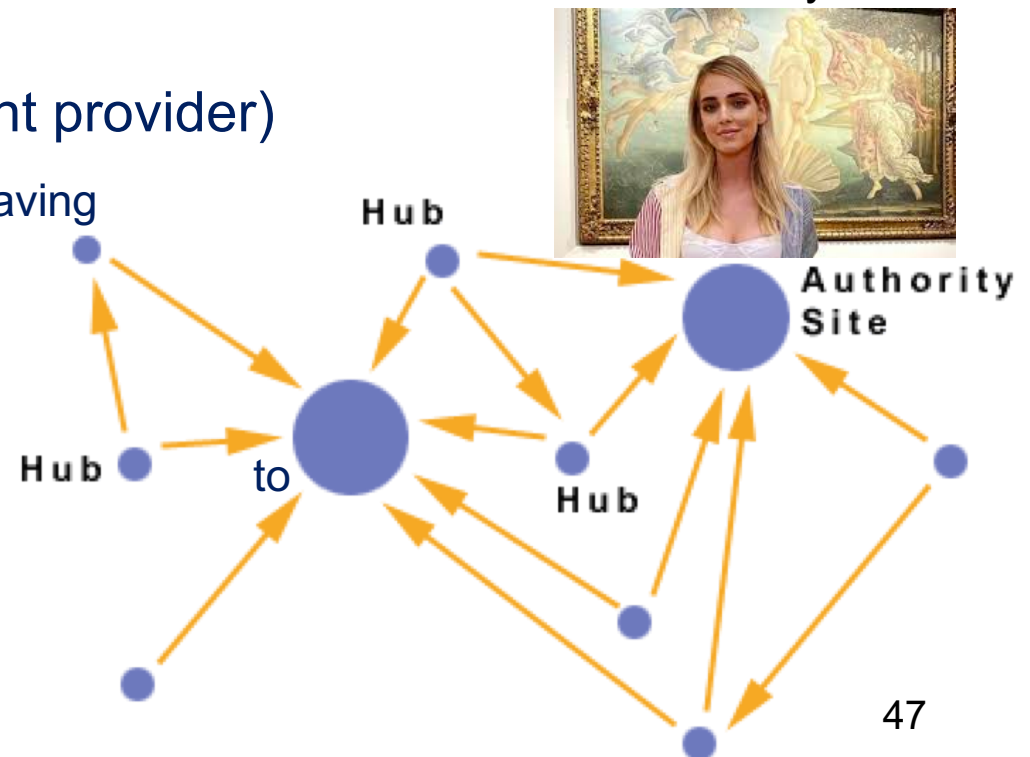
- ❑ **Authorities** (quality as a content provider)

nodes that contain useful information, or having a high number of edges pointing to them (e.g., course homepages)

- ❑ **Hubs** (quality as an expert)

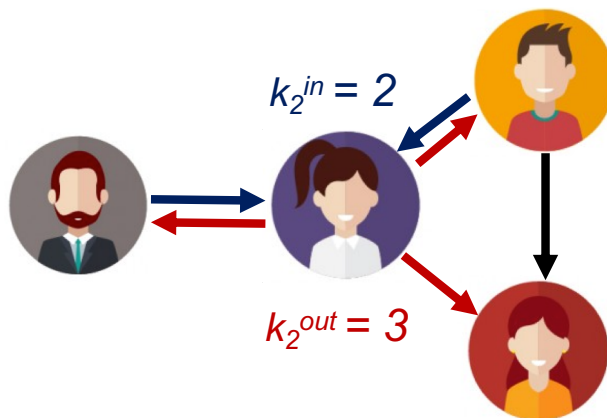
trustworthy nodes, or nodes that link many authorities (e.g., course bulletin)

an influencer:  
authority or hub?





- The in (out) degree can be obtained by **summing** the adjacency matrix over rows (columns)



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

Annotations:  $k_2^{in} = 2$  (sum of row 2),  $k_2^{out} = 3$  (sum of column 2).



# Real networks are sparse

- The maximum degree is  $N-1$
- In real networks  $\langle k \rangle \ll N-1$

NETWORK	$N$	$L$	$\langle k \rangle$
Internet	192,244	609,066	6.34
WWW	325,729	1,497,134	4.60
Mobile Phone Calls	36,595	91,826	2.51
Email	57,194	103,731	1.81
Science Collaboration	23,133	93,439	8.08
Actor Network	702,388	29,397,908	83.71
Citation Network	449,673	4,689,479	10.43

# Visualizing degree centrality

how to get useful insights on centrality

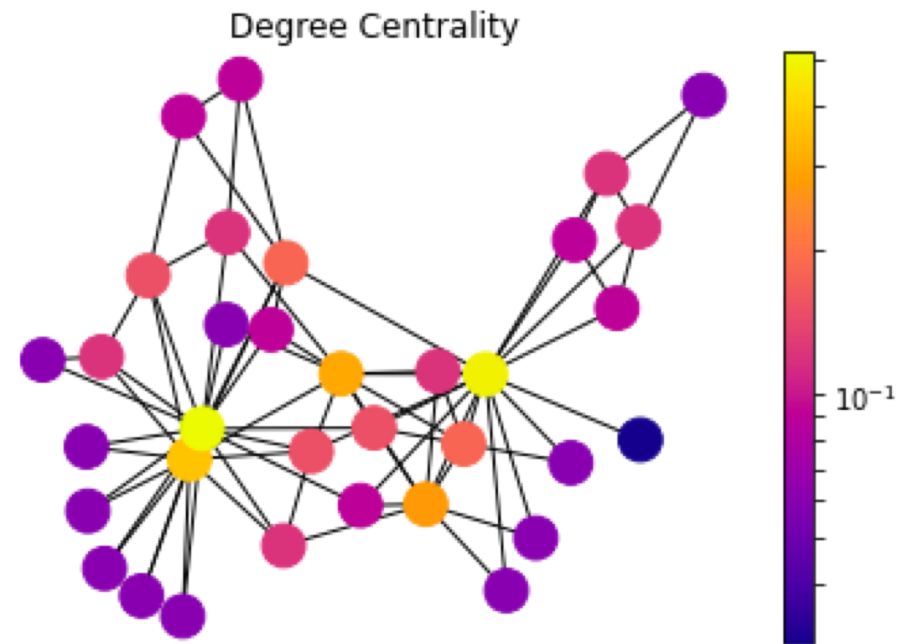


# Graphical representations of degree centrality

by size

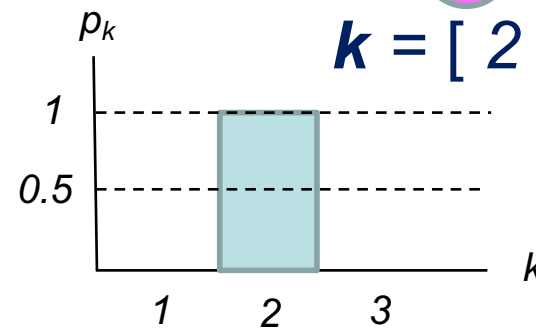
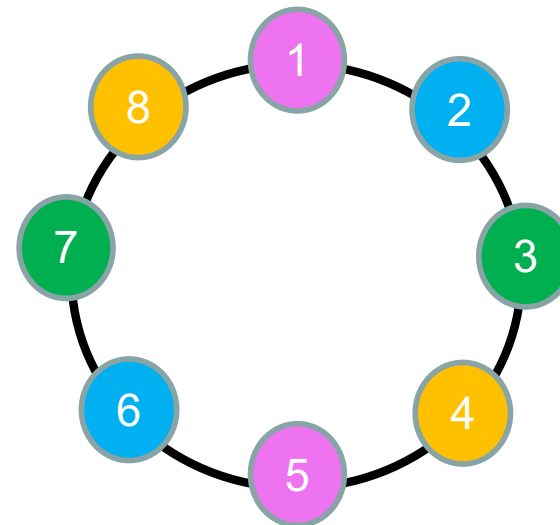
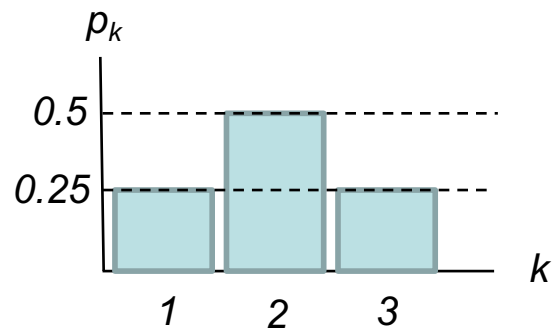
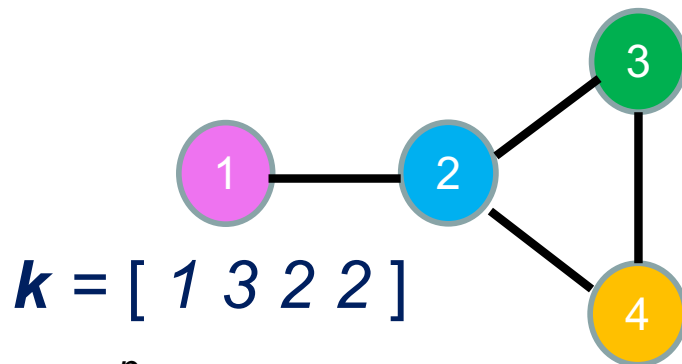


by colour



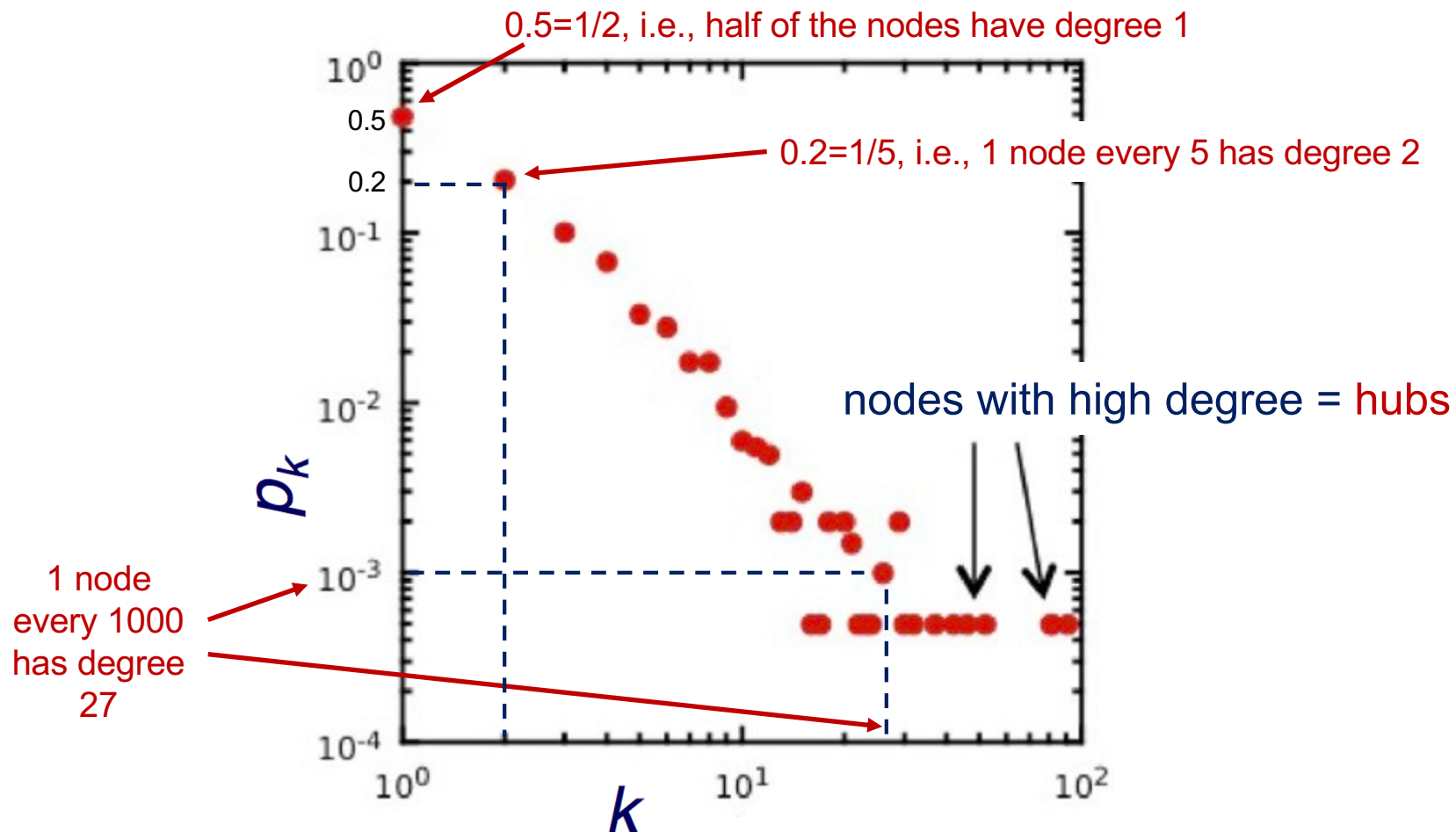


- ✓ a probability distribution  $p_k$
- ✓  $p_k$  = the **fraction** of nodes that have degree equal to  $k$
- ✓  $p_k$  = # of nodes with degree  $k$ , divided by  $N$





- In real (large) networks, degrees have a large range  $\rightarrow$  log representation

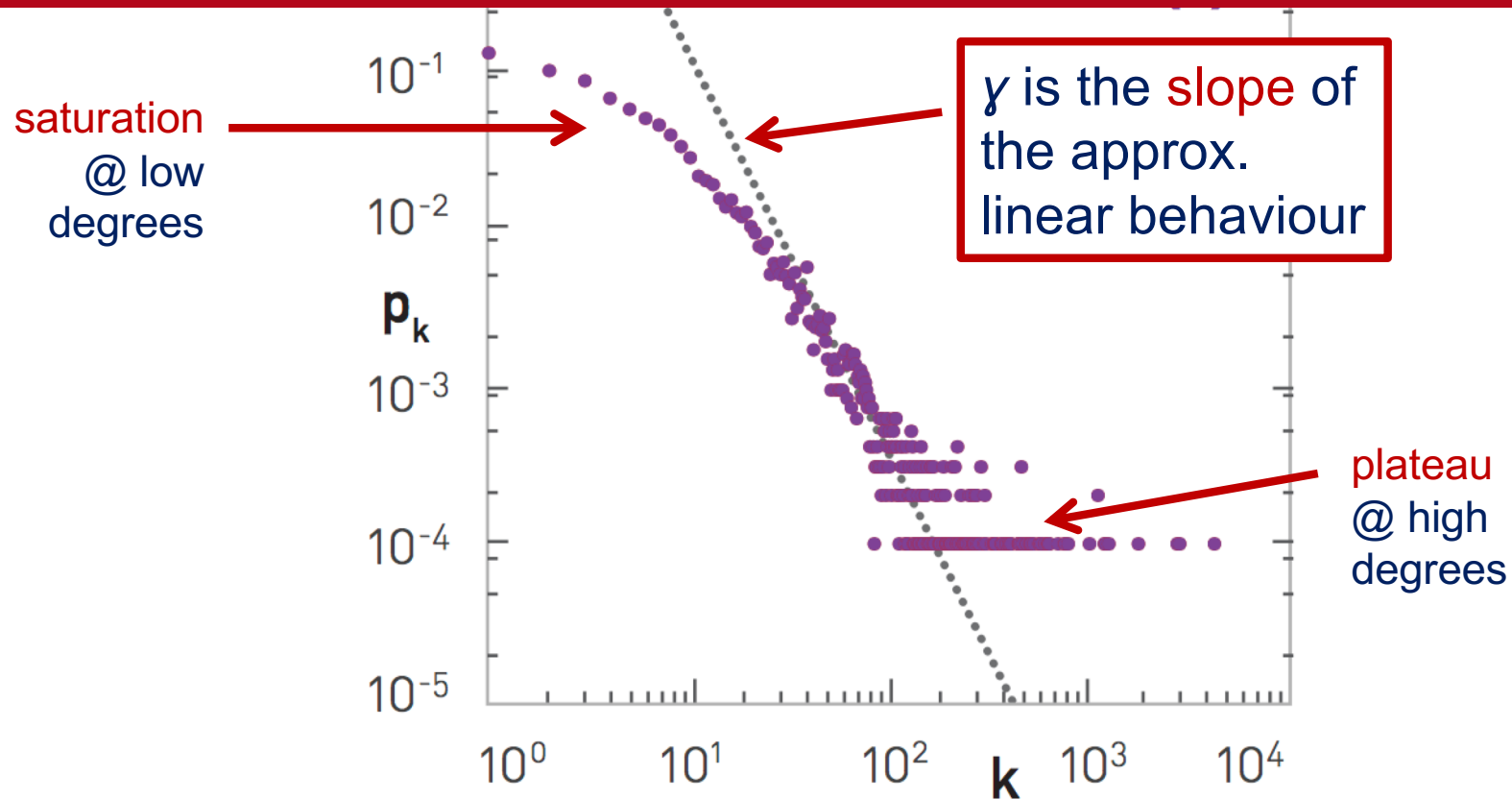


# Scale-free networks

those that follow a power-law

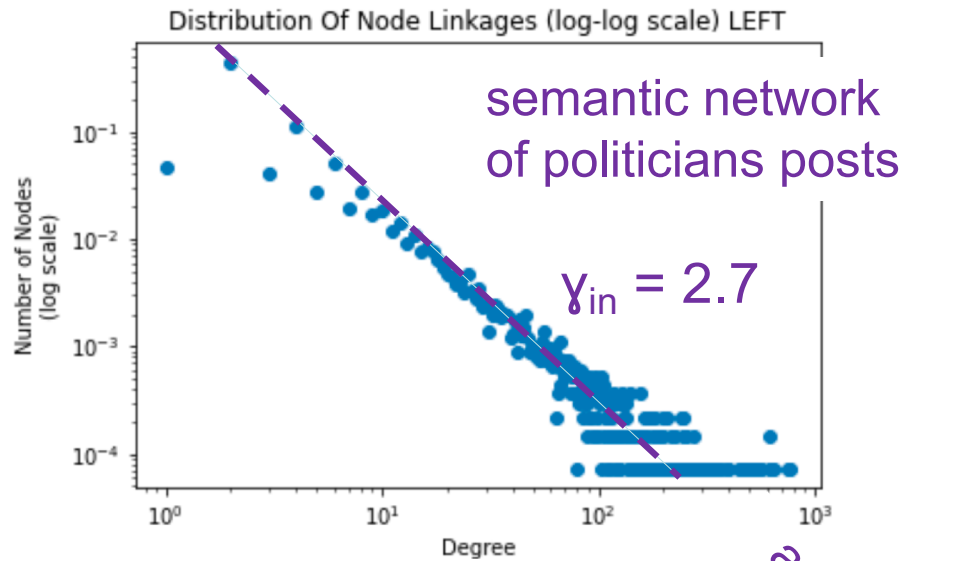


# The power law typical of social networks

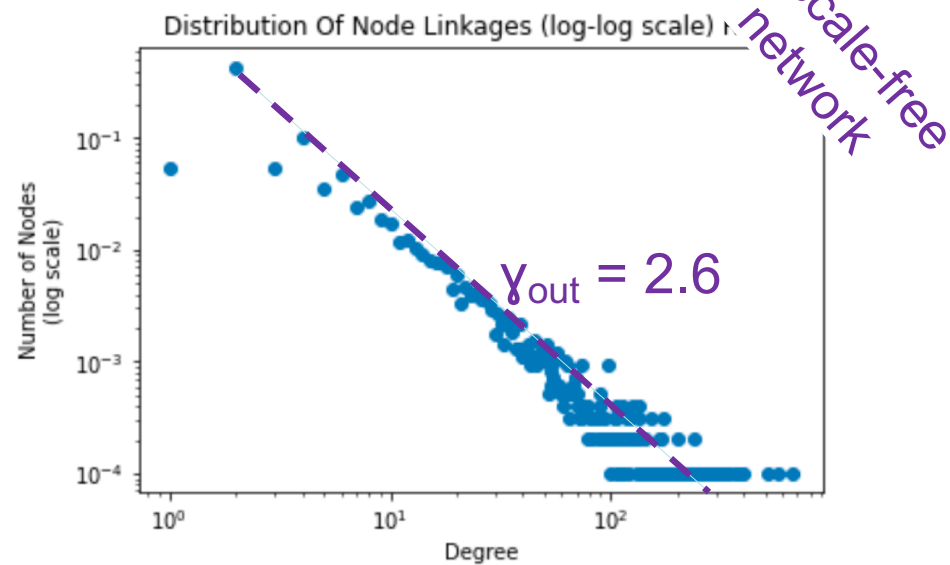
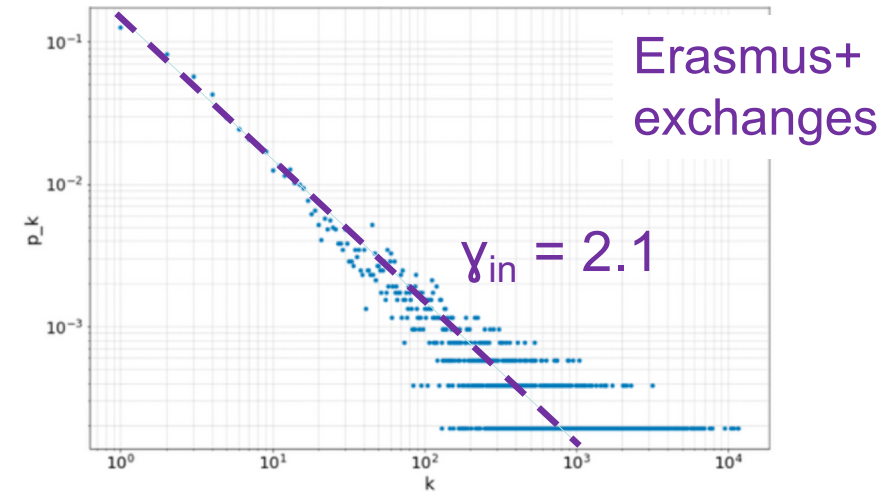


Why the name **power-law**? Because the (approx.) linear behaviour in the log domain ensures

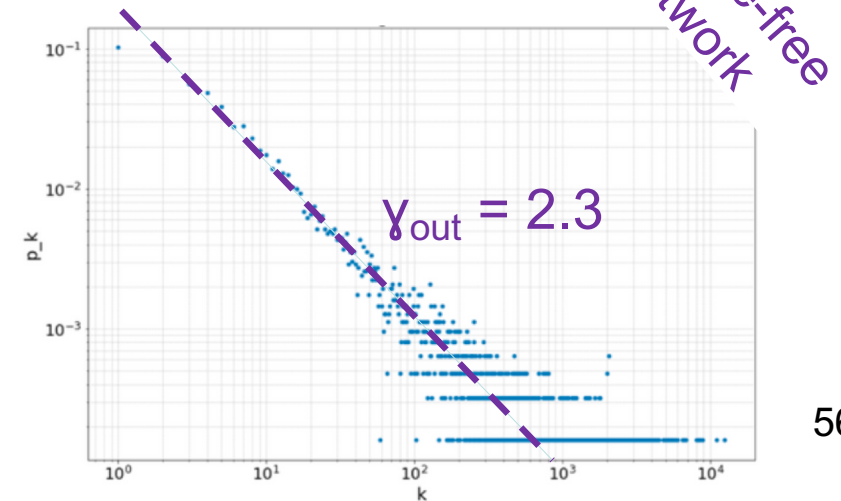
$$\ln(p_k) = c - \gamma \cdot \ln(k) \quad \rightarrow \quad p_k = C k^{-\gamma}$$



### In Degrees Distribution

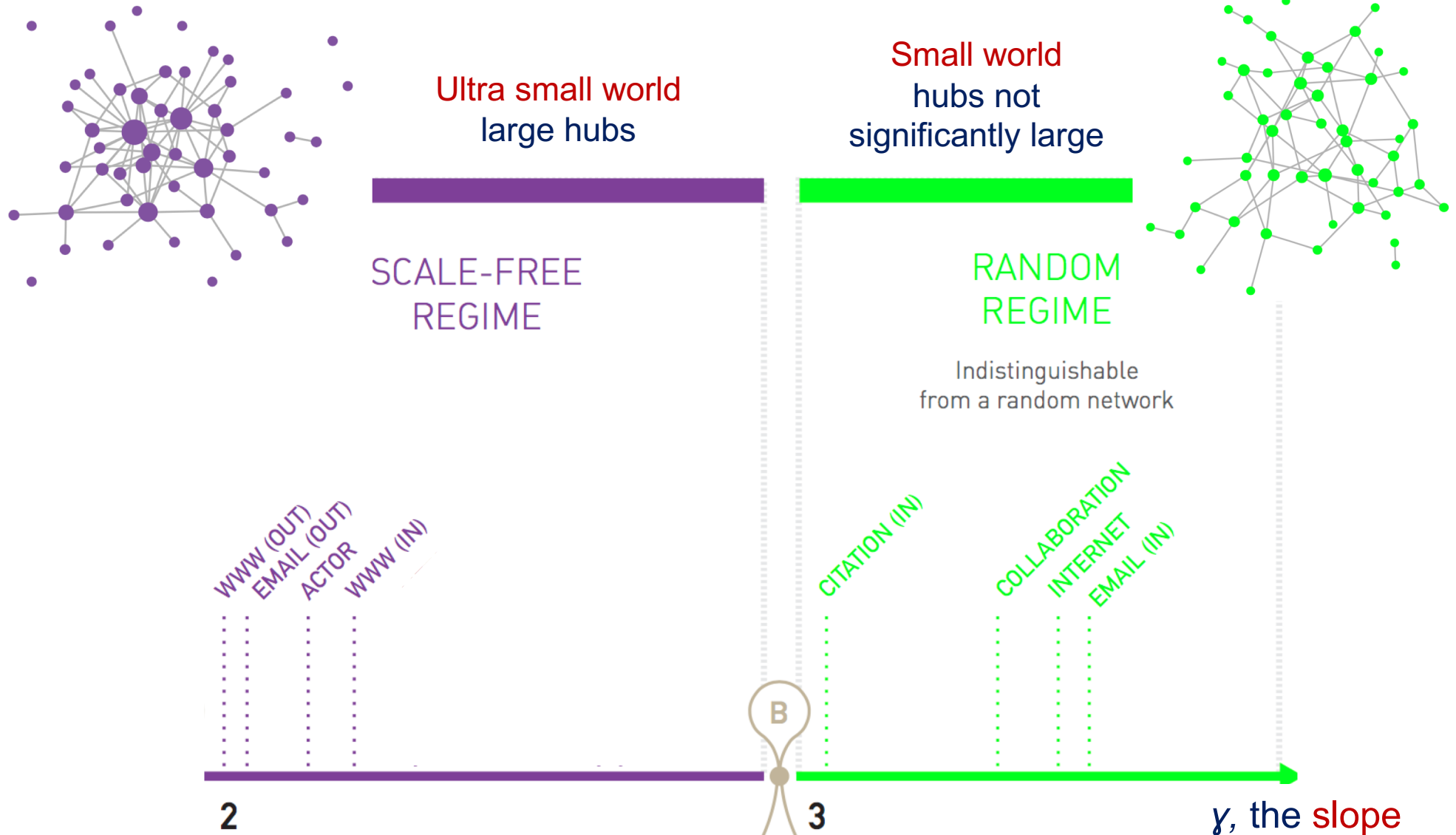


### Out Degrees Distribution



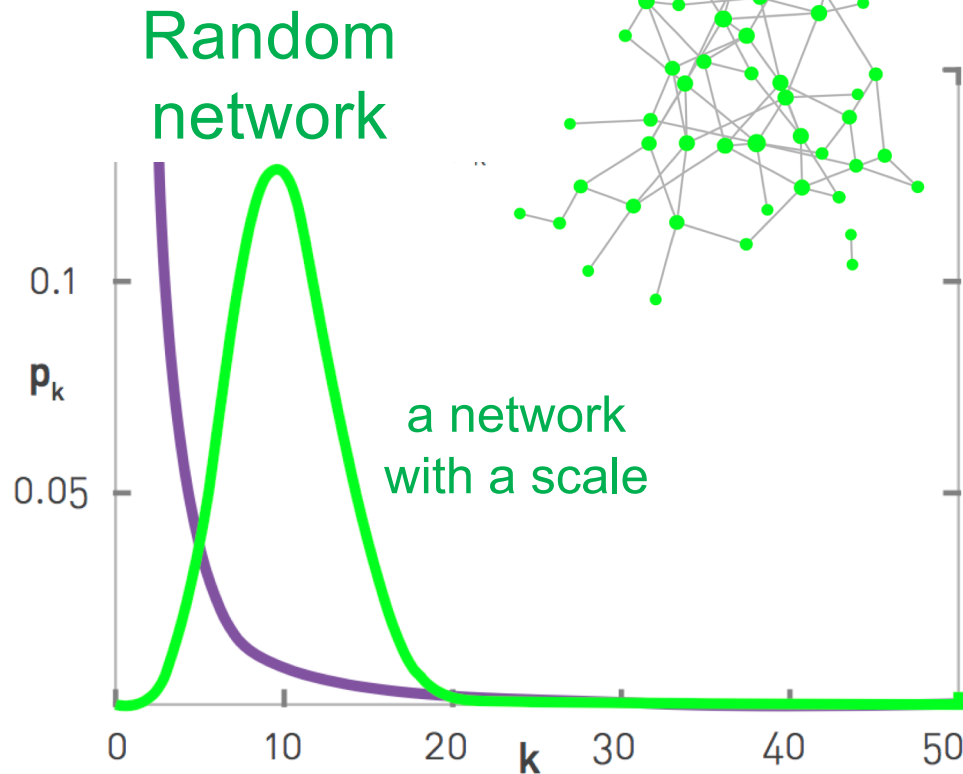


# The ultra-small-world of scale-free networks

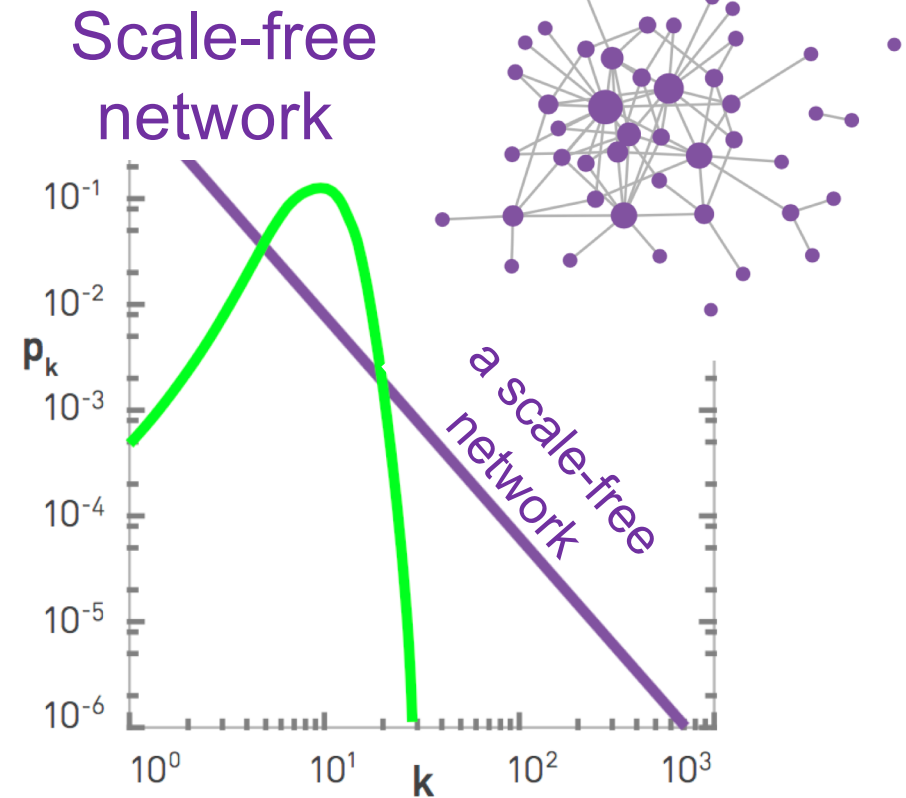




# Scale-free networks versus random networks



- Randomly wired network
- Has smaller hubs
- Needs a linear plot



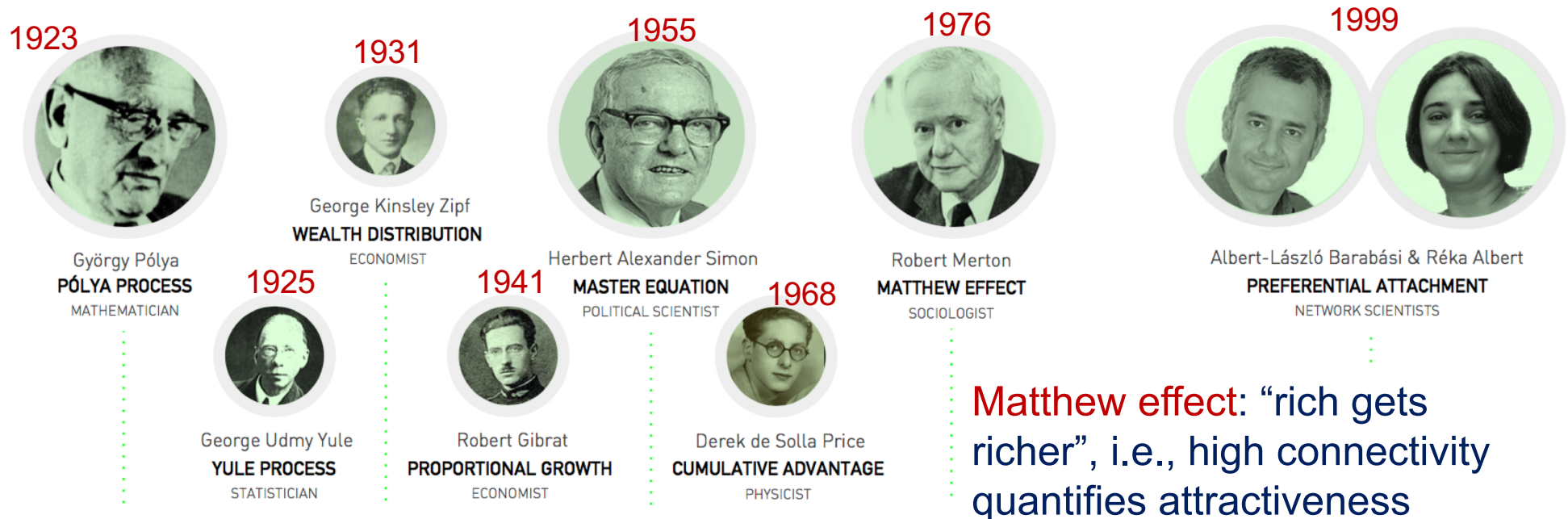
- Power-law network
- Has big hubs
- Needs a log-log plot



## Nodes link to the **more connected** nodes

e.g., think of www

## This idea has a long history







## □ Citation network

researchers decide what papers to read and cite by “**copying**” **references** from papers they have read → papers with more citations are more likely to be cited

## □ Social network

the more acquaintances an individual has, the higher the chance of getting new friends, i.e., we “**copy**” the **friends** of friends → difficult to get friends if you have none

## □ Semantic network

does the model apply here?



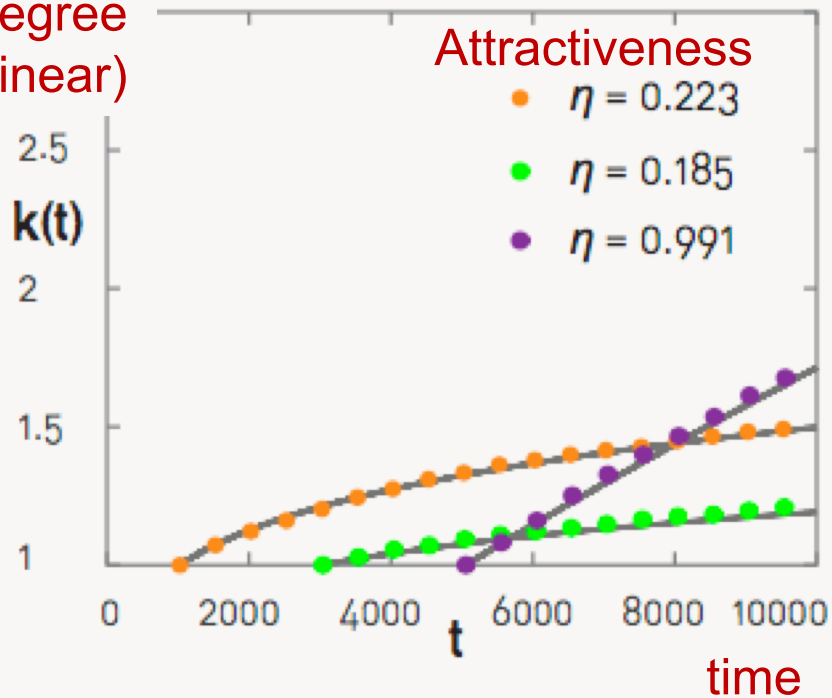
- ❑ There is an innate ability of a node to **attract** links just a quality assessment of the individual
- ❑ Otherwise oldest nodes would have an inherent advantage and cannot be defeated (*first mover's advantage*), which is in contrast with intuition and evidence

e.g., Altavista [1990] → Google [2000] → Facebook [2011] → Instagram [202?]

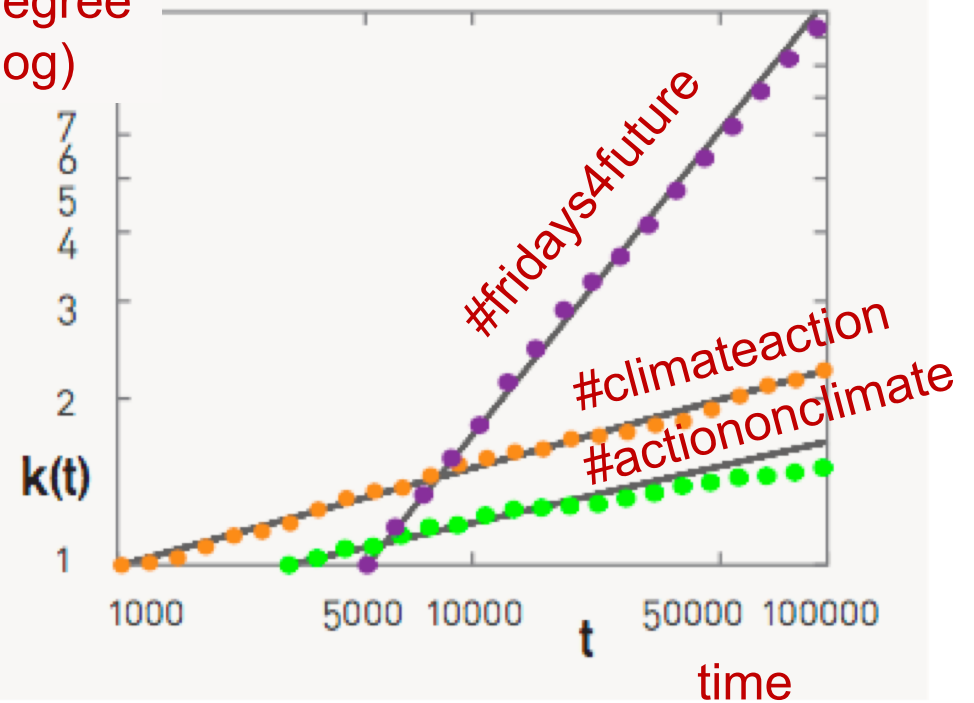
e.g., #parisagreement [2018] → #fridays4future [2019]



node  
degree  
(linear)



node  
degree  
(log)



$\eta_i$  can be measured by data scientists !



- ❑ Degree, degree distribution, loglog plot
- ❑ Authorities and hubs
- ❑ Power law, scale-free networks
- ❑ Slope, Ultra-small-world regime
- ❑ Preferential attachment
- ❑ Attractiveness