
Faculty of Computer Science

Data Science

Moscow
31 January 2024

# Introduction to network science

Network Science
Vitaliy Pozdnyakov

# Networks

Sociology (Networks)

Mathematics, graph theory (Graphs)

Computer Science (Graphs)

Statistical Physics (Complex networks)

Economics (Networks)

Bioinformatics (Networks)

# Topics

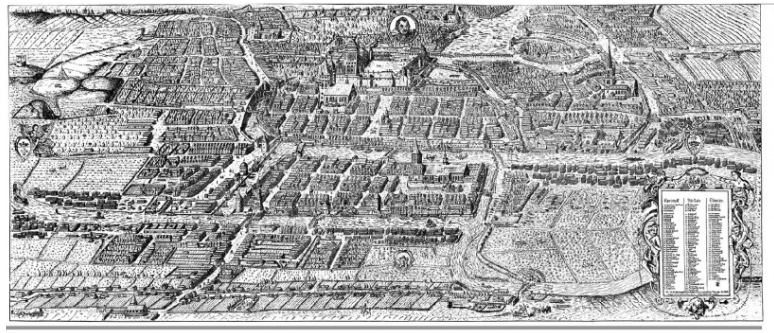
- Statistical properties and modelling of the network
- Network structure and dynamics
- Processes on networks
- Predictions on networks (Machine Learning)
- Network embeddings (Deep Learning)
- Graph neural networks (Deep Learning)

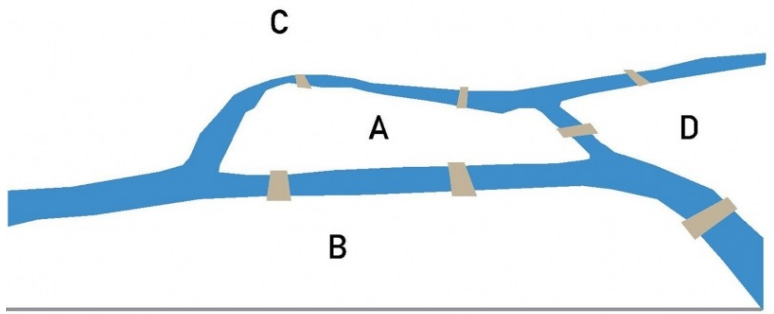
# Konigsberg Bridges

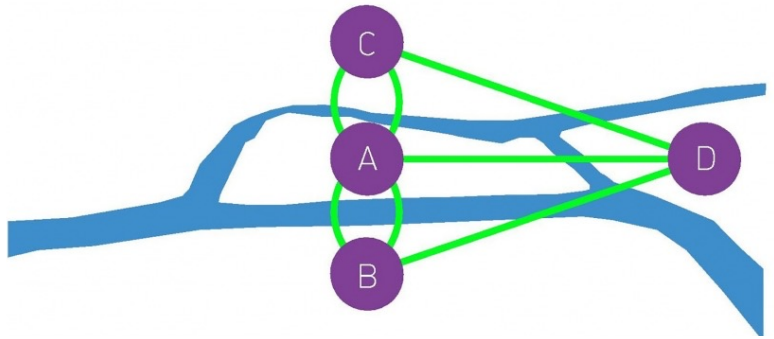
How to walk along Konigsberg in 1735 to pass each bridge exactly once?

Leonard Euler formulated the problem in the form of a graph, where parts of the city are vertices and bridges are edges

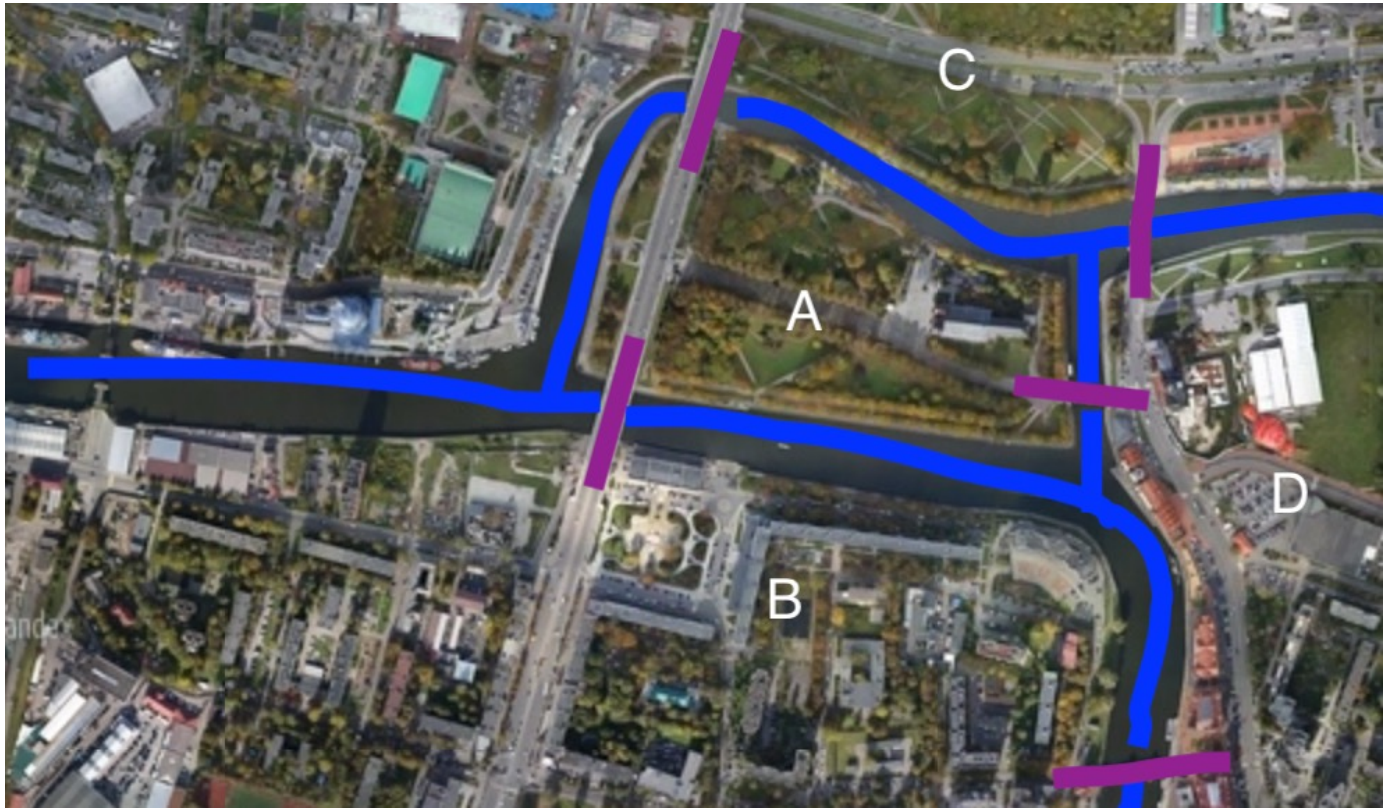
Find such a route in the graph on the right





# Today



# Definitions

Graph is a set of nodes (vertices) and a set of edges (links)

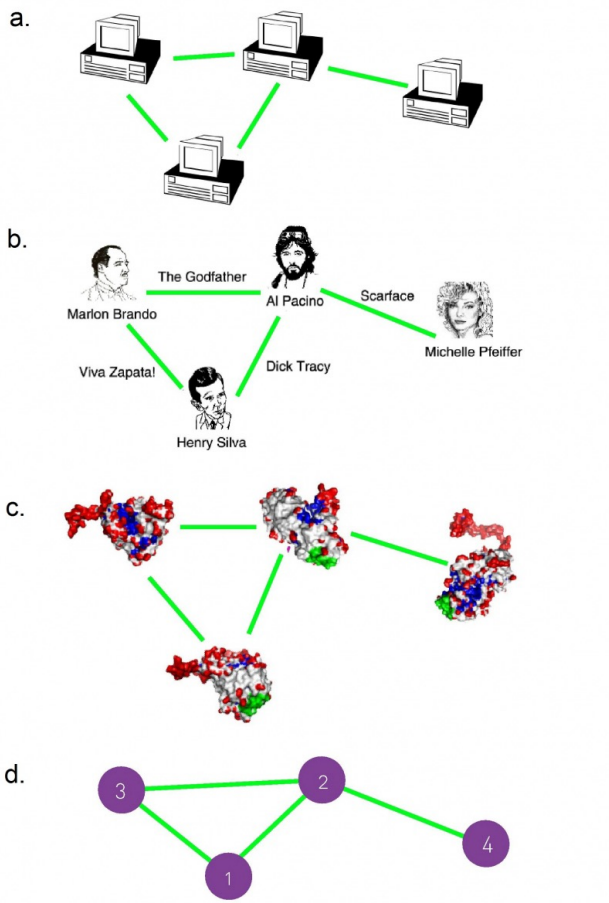
$$G = (V, E)$$
$$V = \{v_1, v_2, \dots, v_n\}$$
$$e_{ij} = (v_i, v_j) \in E$$

Degree of a node $i$ is the number of edges connected to the node $i$

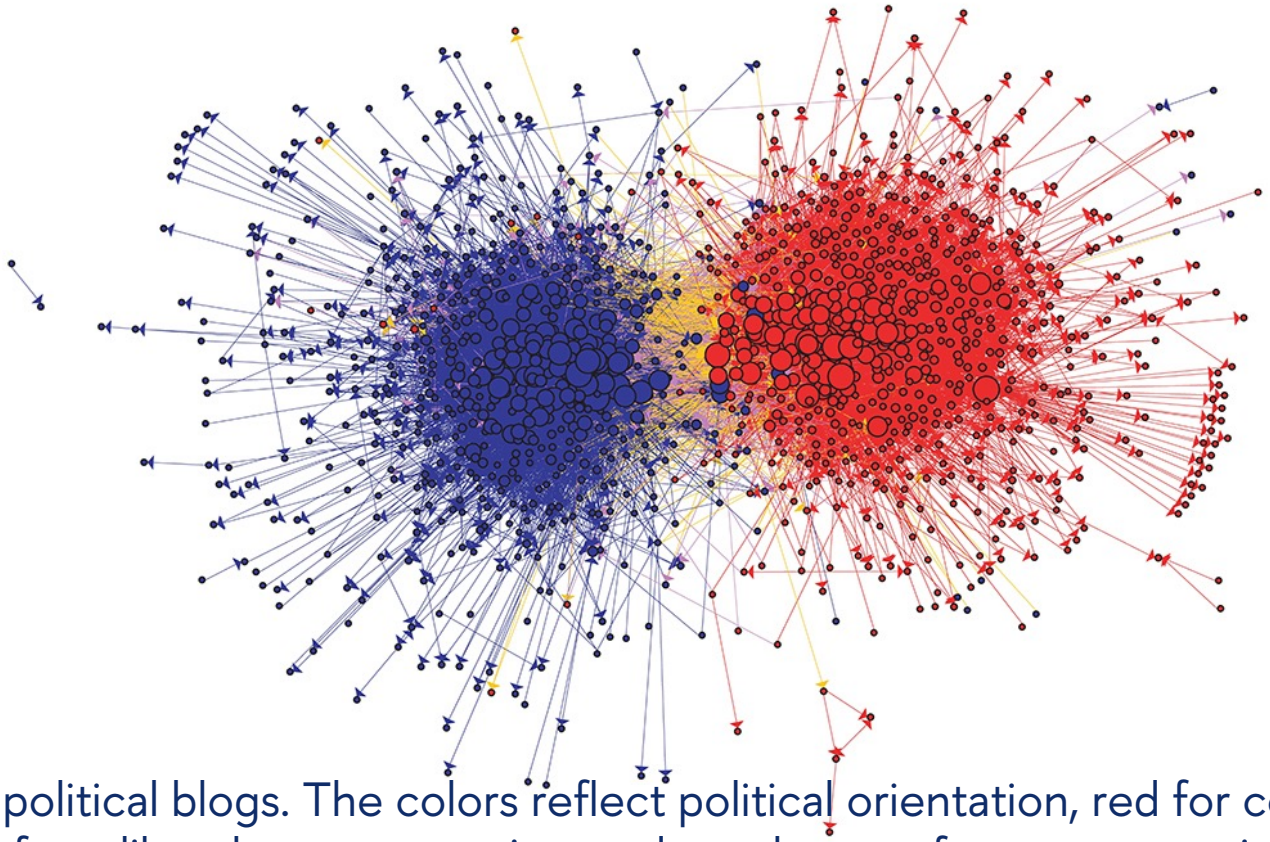
$$k_i = |\{e_{ij} | e_{ij} \in E\}|$$

Neighbours of a node $i$ are nodes that are connected by an edge to the node $i$

$$N(i) = \{v_j | e_{ij} \in E\}$$

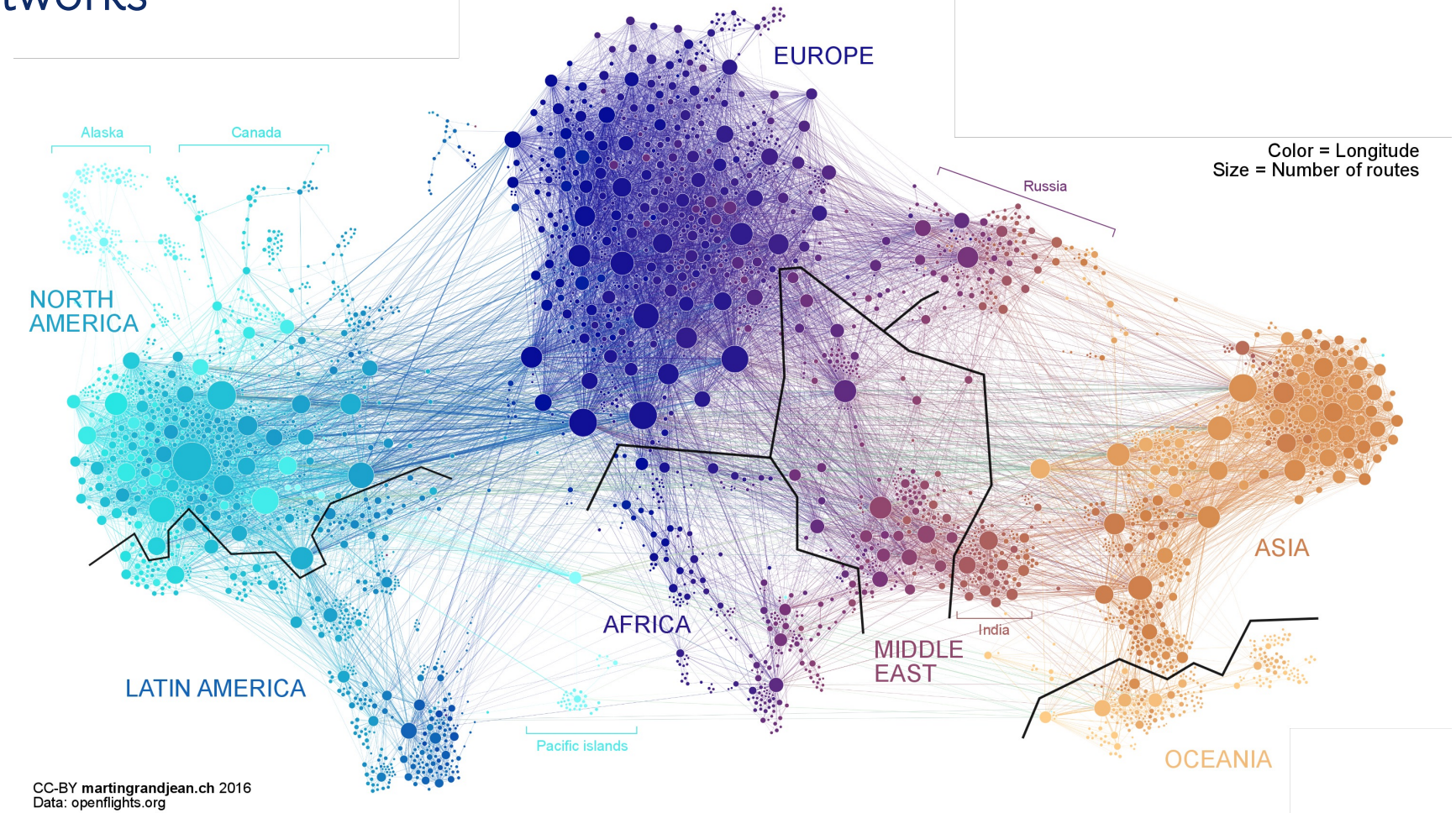


# Political networks in the 2004 US election

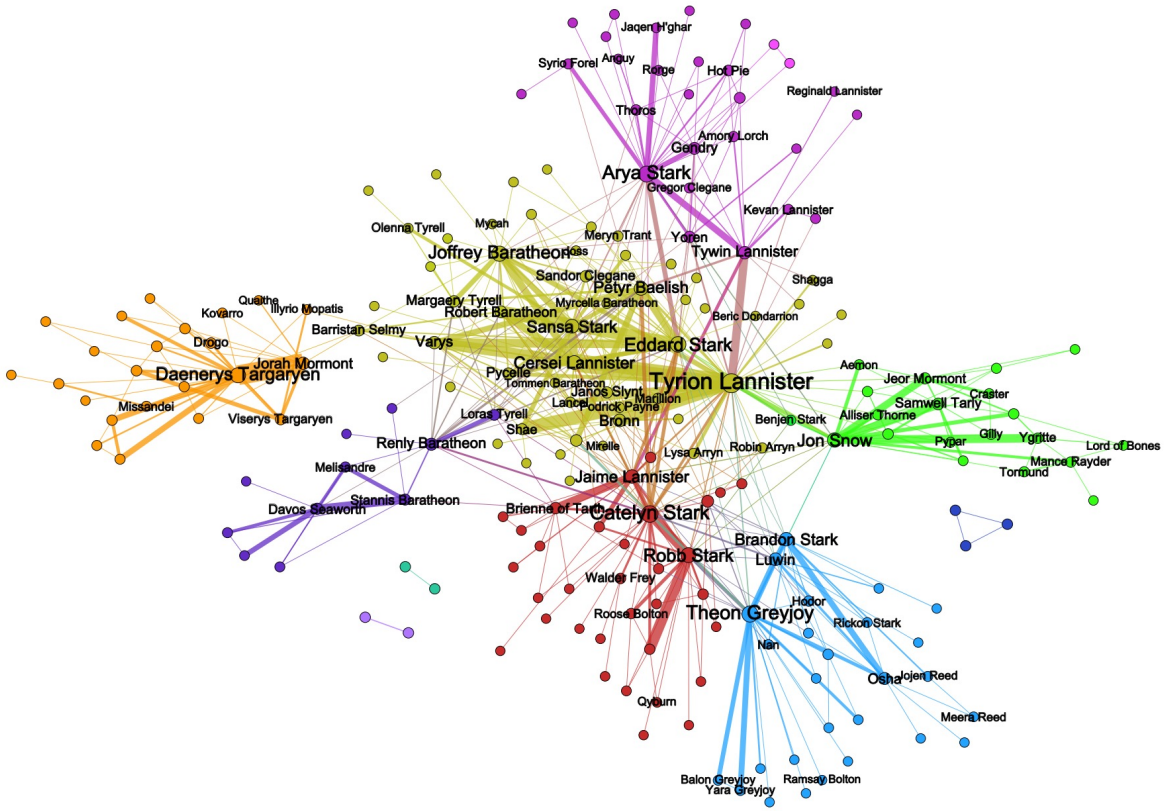


Community structure of political blogs. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

# Airlines networks



# Character networks




Communities in the TV series Game of Thrones (seasons 1 & 2)

# Fake news networks

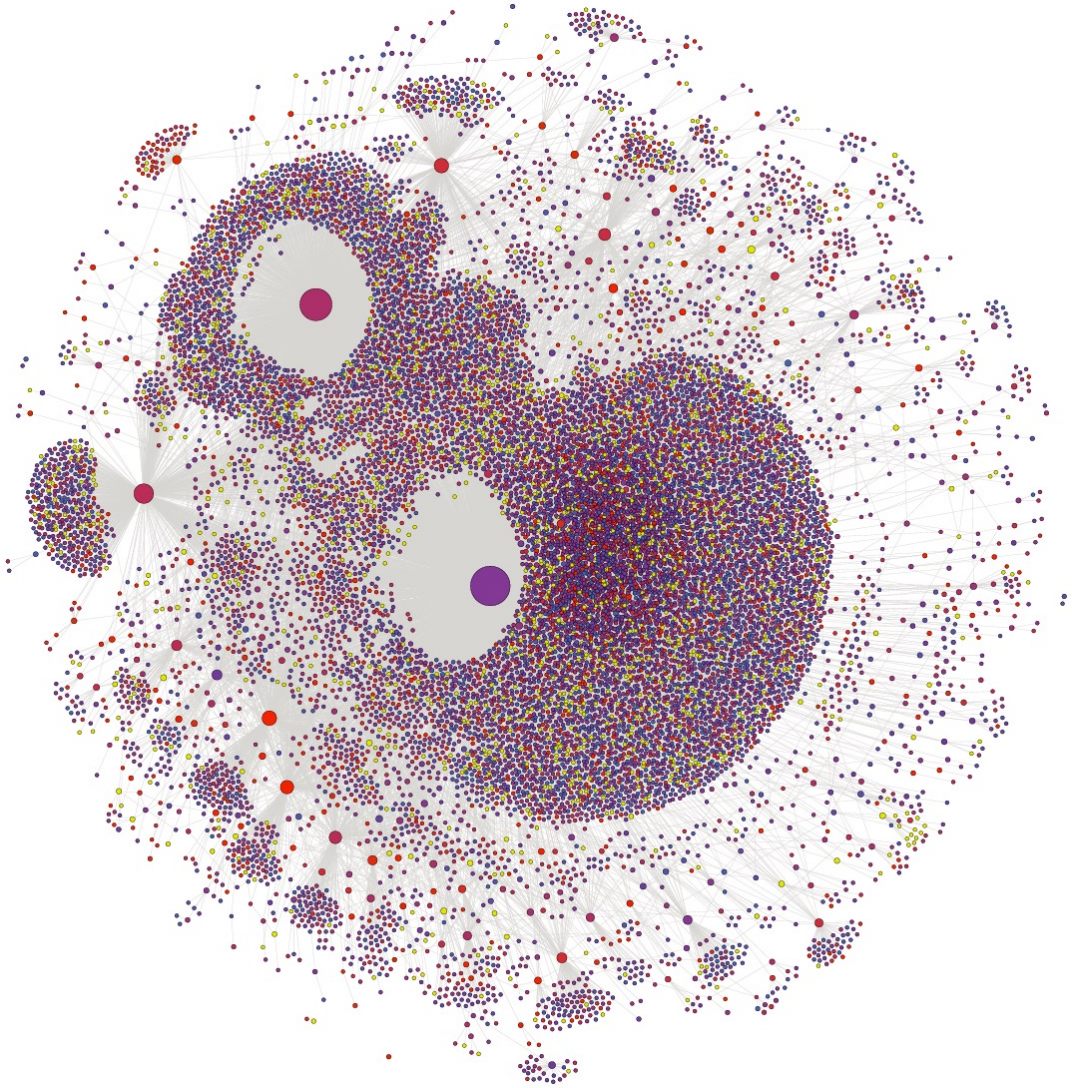
Fake article "Spirit cooking: Clinton campaign chairman practices bizarre occult ritual"

30,000 retweets 4 days before the 2016 US election

Botometer result:

- Blue peaks are more like people
- Red tops are more like bots

Bots launch a distribution chain through mentions of popular bloggers

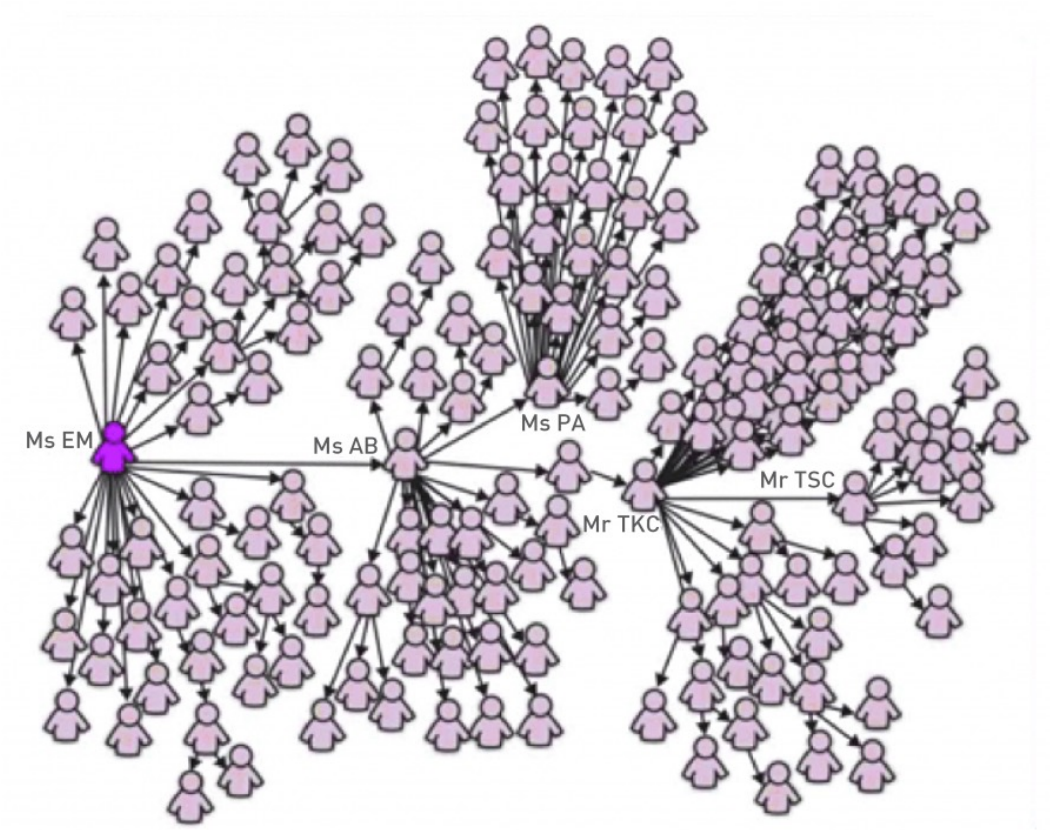


# Epidemic networks

A network of 144 cases of SARS-Cov-1 in Singapore in 2003

Signed nodes are supersreaders

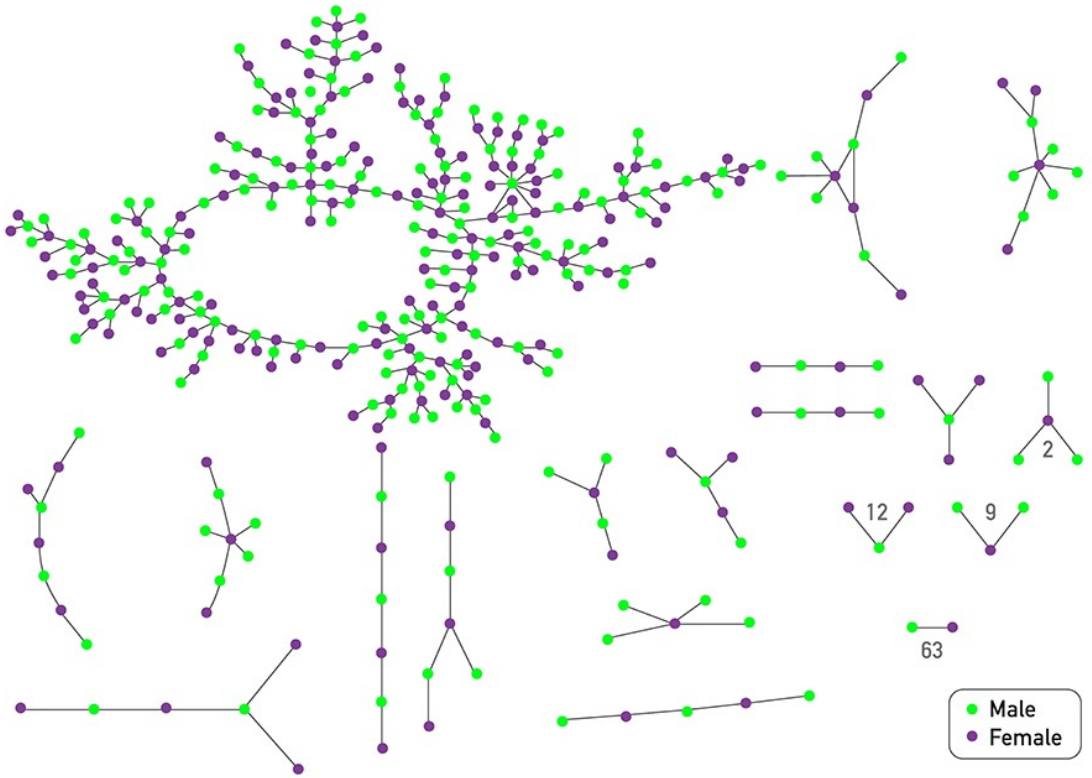
Who do you think the signed people?



# Sexual networks

573 high school students in the United States surveyed during the period 1993-1995

52% in the isolated component — the risk of infection is higher, even for nodes with a small degree



# More examples

Road networks

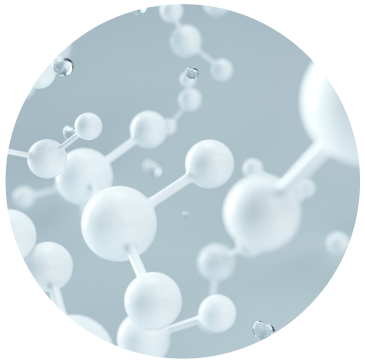
**nodes**
cities

**edges**
roads

City networks

**nodes**
cross-roads

**edges**
streets



Molecules

**nodes**
atoms

**edges**
relations


Transactions

**nodes**
agents

**edges**
transactions

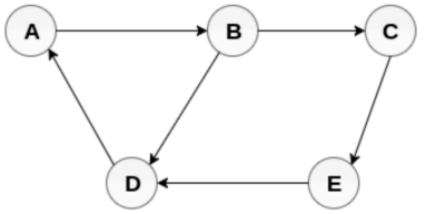

Wikipedia

**nodes**
pages

**edges**
links

# Adjacency matrix

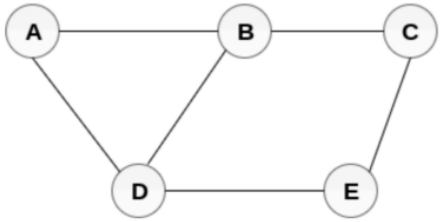
The adjacency matrix $A^{n \times n}$ — matrix with nonzero elements $a_{ij}$ for edges $e_{ij}$




Directed Graph

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 1 | 0 | 0 | 0 |
| B | 0 | 0 | 1 | 1 | 0 |
| C | 0 | 0 | 0 | 0 | 1 |
| D | 1 | 0 | 0 | 0 | 0 |
| E | 0 | 0 | 0 | 1 | 0 |

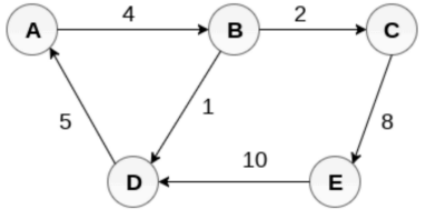
Adjacency Matrix




Undirected Graph

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 1 | 0 | 1 | 0 |
| B | 1 | 0 | 1 | 1 | 0 |
| C | 0 | 1 | 0 | 0 | 1 |
| D | 1 | 1 | 0 | 0 | 1 |
| E | 0 | 0 | 1 | 1 | 0 |

Adjacency Matrix




Weighted Directed Graph

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 4 | 0 | 0 | 0 |
| B | 0 | 0 | 2 | 1 | 0 |
| C | 0 | 0 | 0 | 0 | 8 |
| D | 5 | 0 | 0 | 0 | 0 |
| E | 0 | 0 | 0 | 10 | 0 |

Adjacency Matrix

# Basic statistics

Number of nodes: $N = |V|$

Number of edges: $L = |E|$

Maximal number of edges:

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

Density: $L/L_{\max}$

Path length: $l(v_i, v_j)$ is a minimal path between $v_i$ and $v_j$

Node eccentricity:

$$\varepsilon(v_i) = \max_{v_j \in V} l(v_i, v_j)$$

Radius: $r = \min \varepsilon$

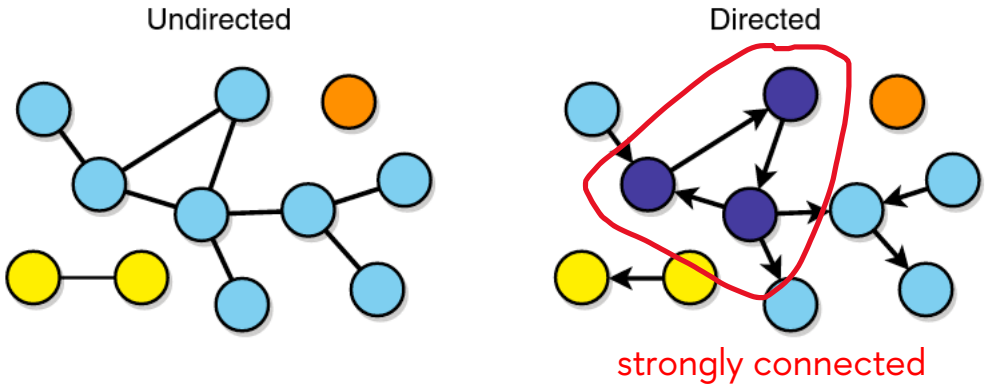
Diameter: $d = \max \varepsilon$

# Connected components

Connected components — a subset of nodes, such that there is a path between any pair of these nodes, but there is no path connected them to other components

Directed graphs

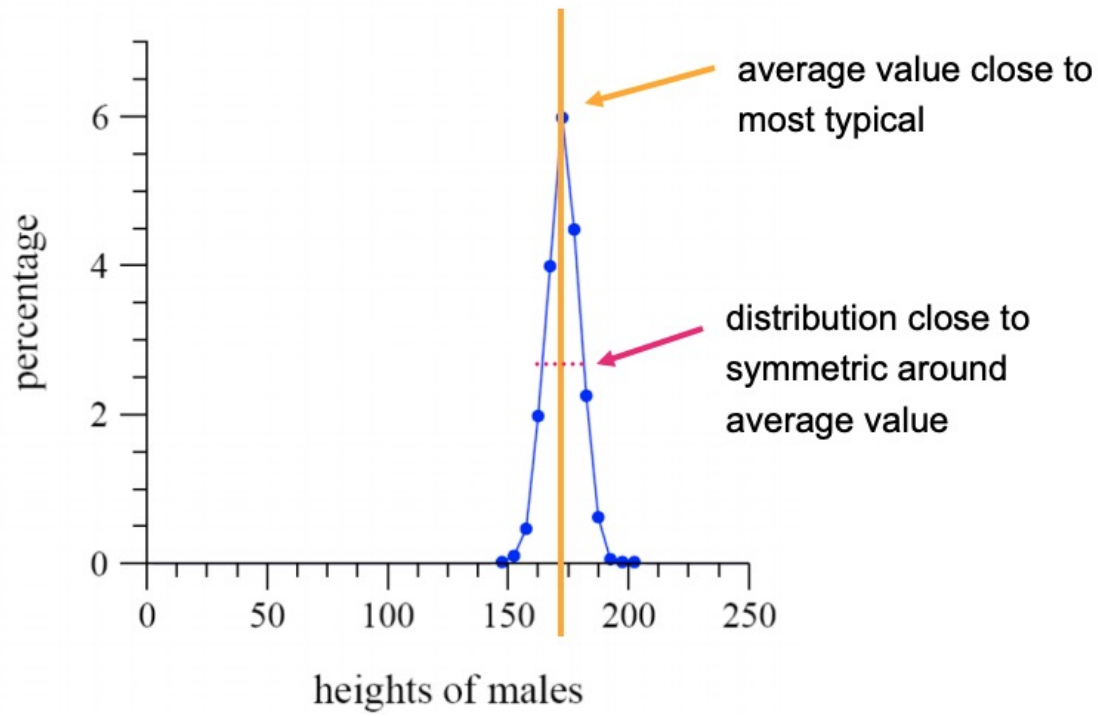
- **Weakly connected components** — … undirected path …
- **Strongly connected components** — … two directed paths in both directions …



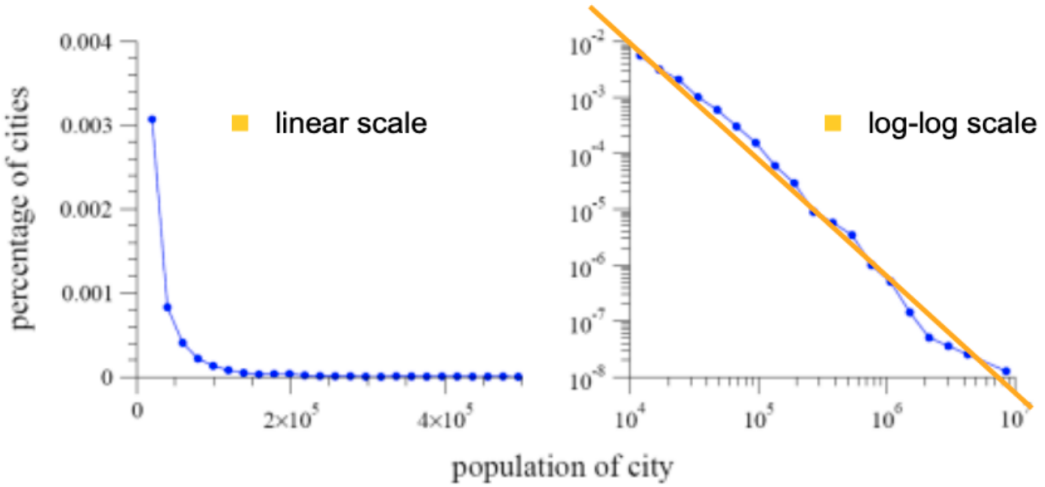
# Complex networks properties

1. **Scale-free property**: some nodes are hubs, power law node degree distribution.
2. **Transitivity property**: hight clustering coefficient.
3. **Small world property:** small diameter and average path length.

# Typical normal distributions



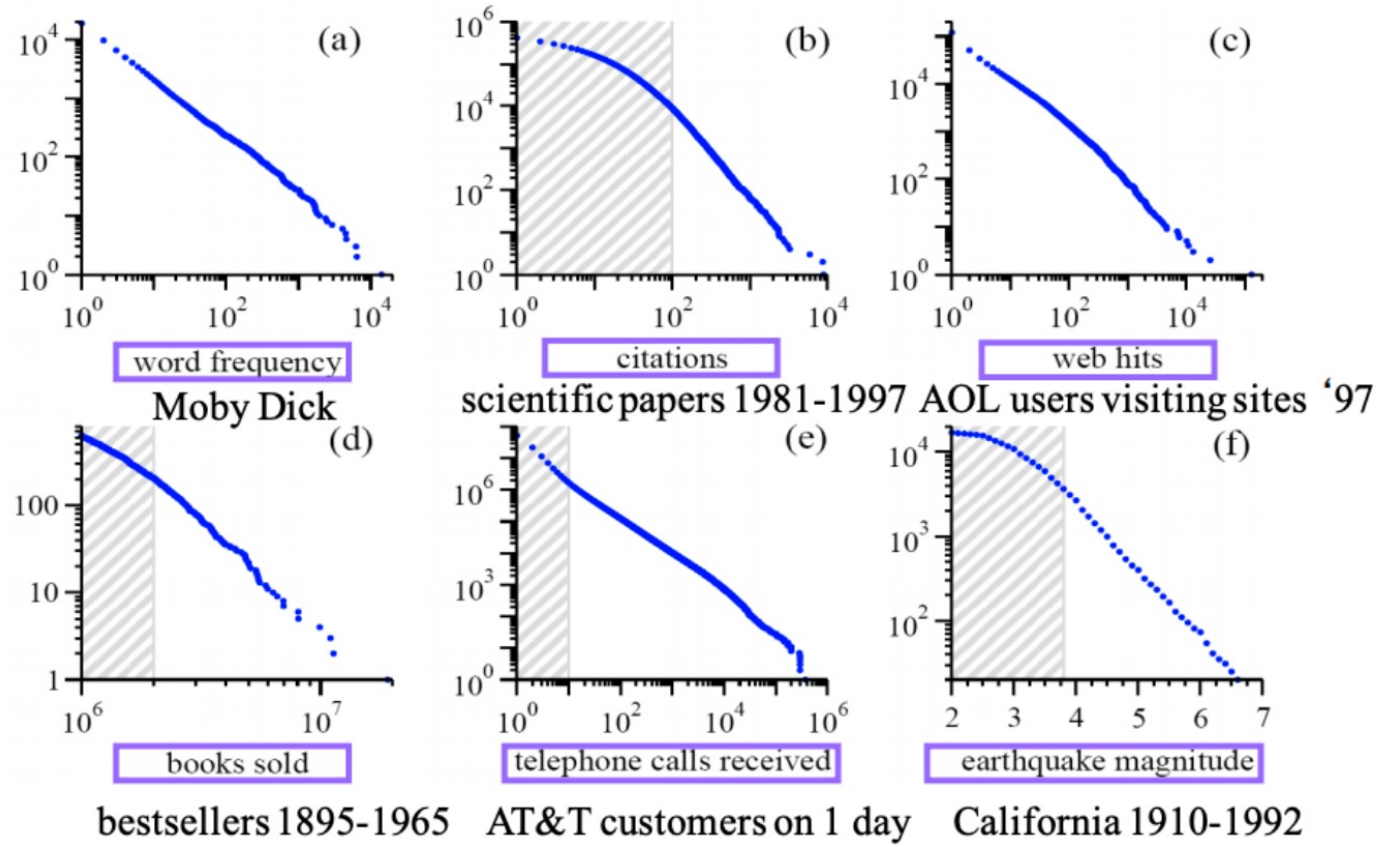
# Power law distributions




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

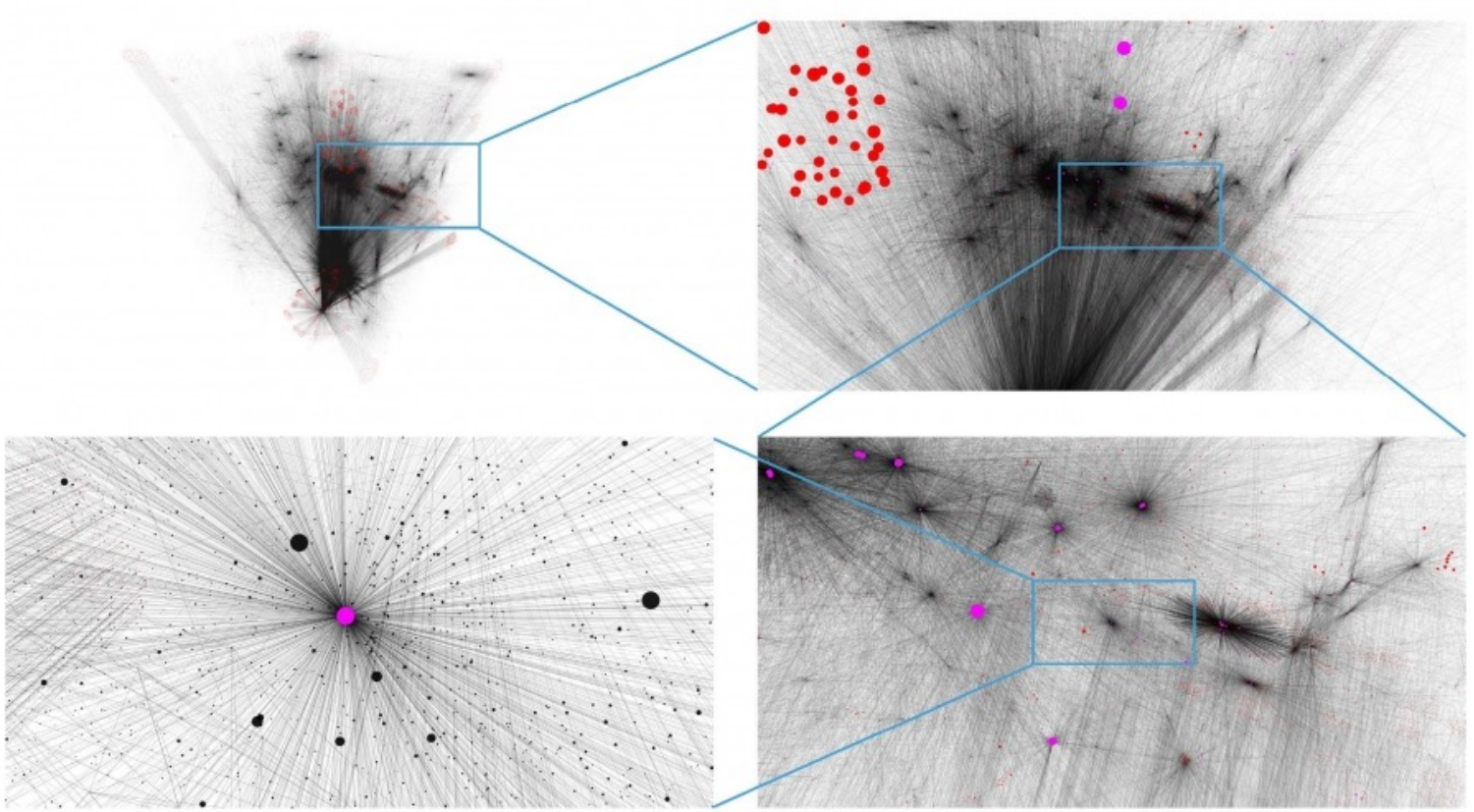
$$\log f(k) = log\ \ C - \gamma log\ \ k$$

# Power law distributions



# Power law node degree distribution

# Hubs in social networks

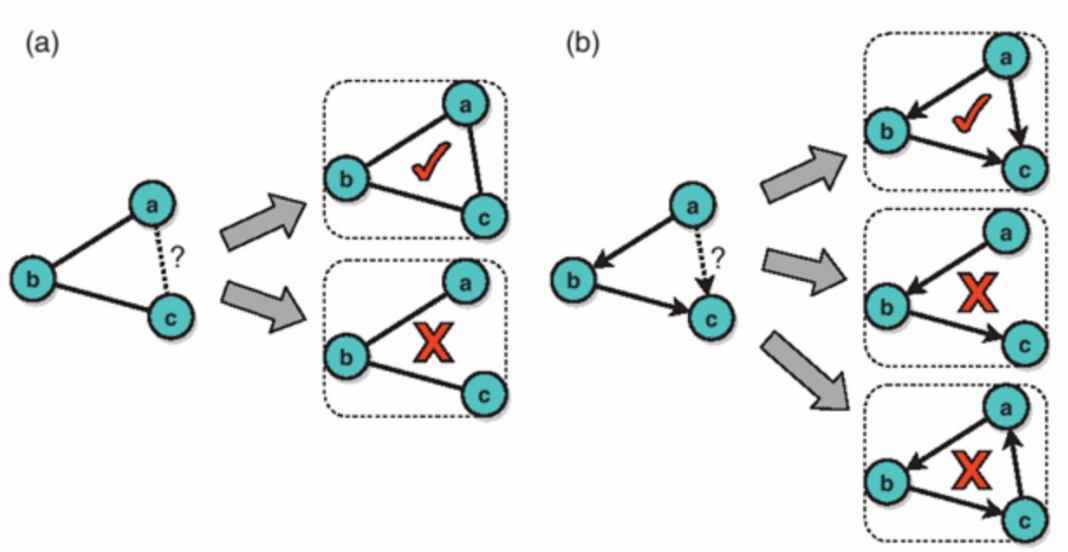


The Topology of the World Wide Web

# Friends of friends

In a social network, if Alice and Bob are both friends of Charlie' s, they are also likely to be friends of each other

In other words, there is a good chance that a friend of my friend is also my friend.

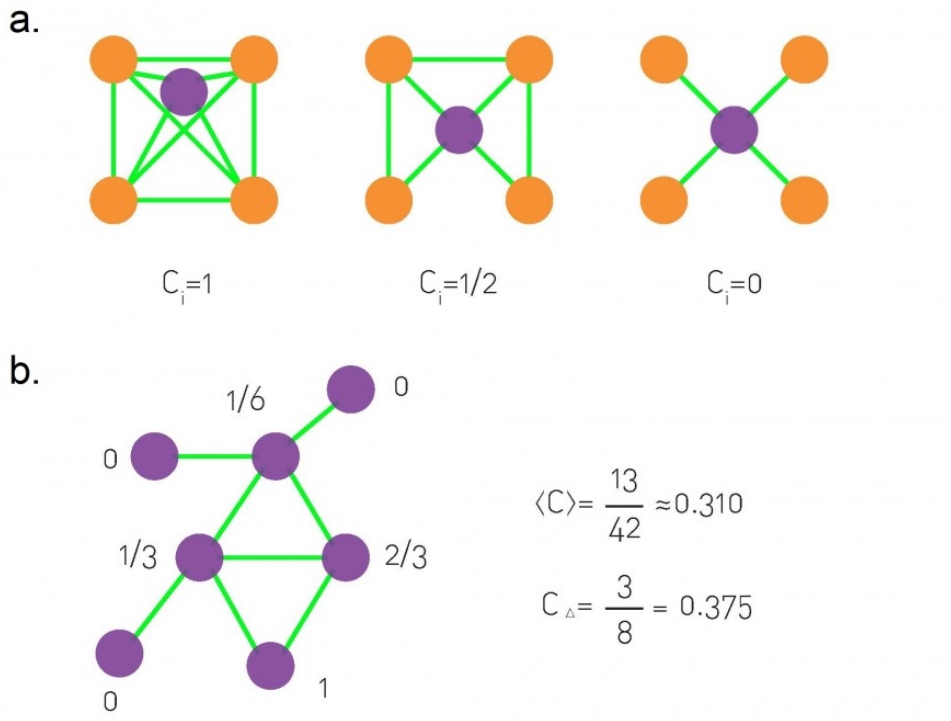


# Clustering coefficient

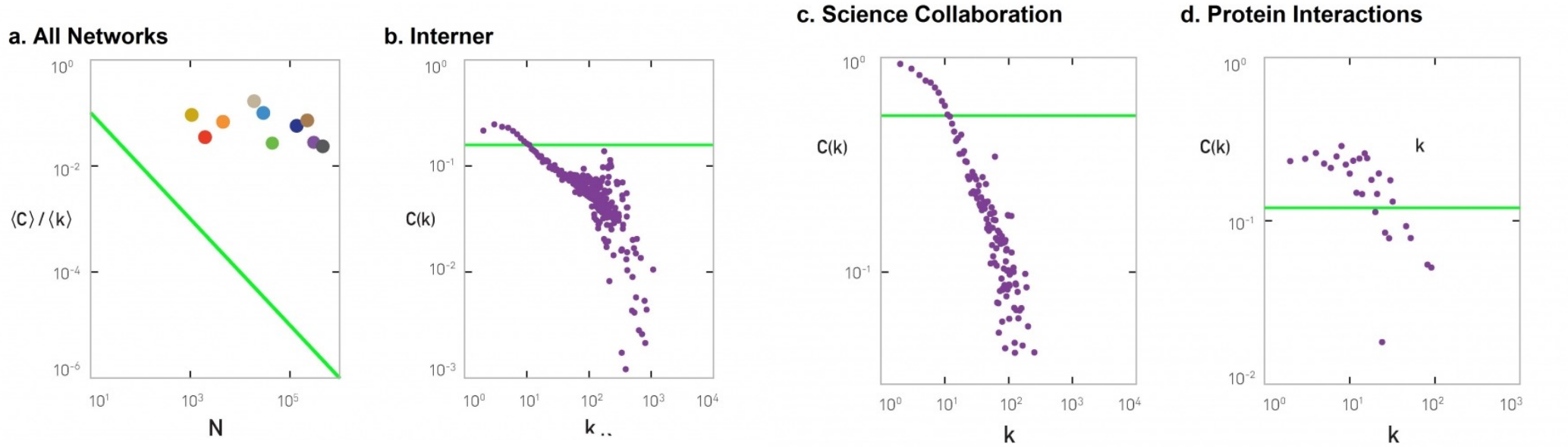
$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

where $L_i$ represents the number of links between the $k_i$ neighbours of node $i$.

- $C_i = 0$ if none of the neighbours of node $i$ link to each other
- $C_i = 1$ if the neighbours of node $i$ form a complete graph, i.e. they all link to each other



# High clustering coefficient



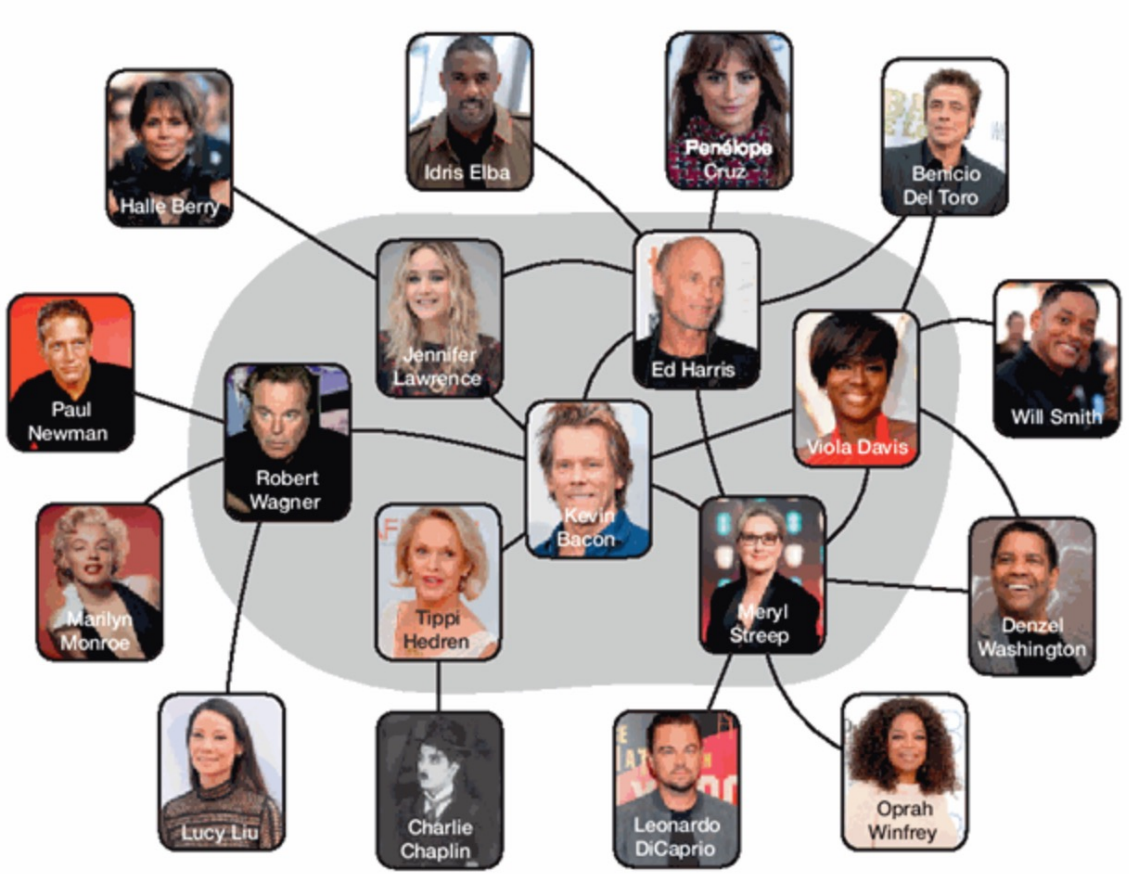
# Small world: six degrees of separation

Avarage path length is small if

$$\langle l \rangle \approx \log N$$

World population has reached 8 billion people on 15 November 2022

$$\log(8 \cdot 10^9) \approx 9.9$$



# Small world experiment

Jeffrey Travers, Stanley Milgram, 1969

Starting persons:

- 296 volunteers, 217 sent
- 196 in Nebraska
- 100 in Boston

Target person — Boston stockbroker

Given information: target name, address, occupation, place of employment, college, hometown

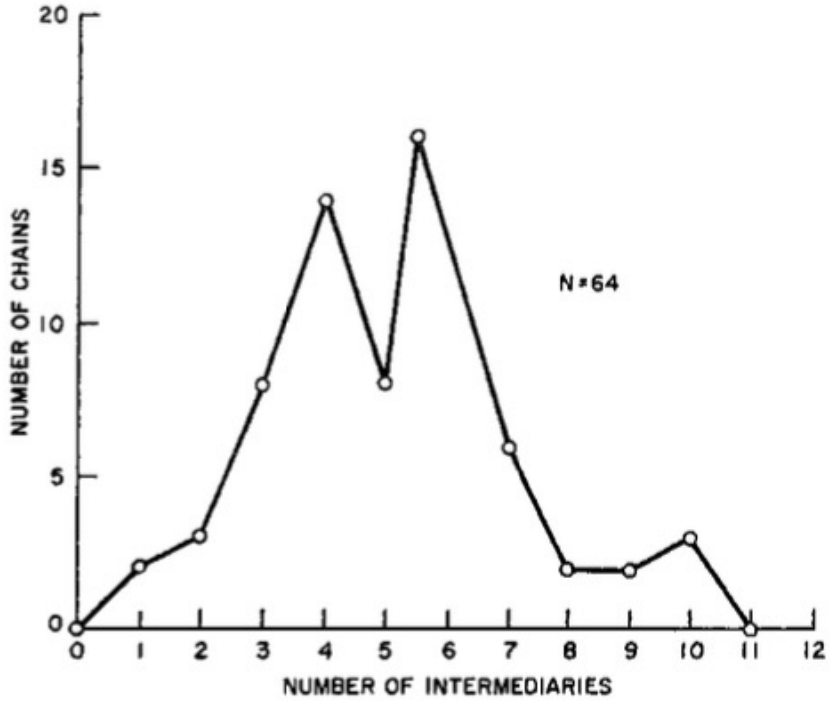
# Small world experiment: results

Reached the target N = 64(29%)

Average chain length ⟨L⟩ = 5.2

Channels:

- hometown ⟨L⟩ = 6.1
- business contacts ⟨L⟩ = 4.6
- from Boston ⟨L⟩ = 4.4
- from Nebraska ⟨L⟩ = 5.7



# Average path length and clustering coefficient in real networks

| Network | Nodes ($N$) | Links ($L$) | Average path length ($\langle \ell \rangle$) | Clustering coefficient ($C$) |
|---|---|---|---|---|
| Facebook Northwestern Univ. | 10,567 | 488,337 | 2.7 | 0.24 |
| IMDB movies and stars | 563,443 | 921,160 | 12.1 | 0 |
| IMDB co-stars | 252,999 | 1,015,187 | 6.8 | 0.67 |
| Twitter US politics | 18,470 | 48,365 | 5.6 | 0.03 |
| Enron email | 87,273 | 321,918 | 3.6 | 0.12 |
| Wikipedia math | 15,220 | 194,103 | 3.9 | 0.31 |
| Internet routers | 190,914 | 607,610 | 7.0 | 0.16 |
| US air transportation | 546 | 2,781 | 3.2 | 0.49 |
| World air transportation | 3,179 | 18,617 | 4.0 | 0.49 |
| Yeast protein interactions | 1,870 | 2,277 | 6.8 | 0.07 |
| *C. elegans* brain | 297 | 2,345 | 4.0 | 0.29 |
| Everglades ecological food web | 69 | 916 | 2.2 | 0.55 |

# References


- Leonid E. Zhukov, Lecture Notes on Social Network Analysis and Linguistic Networks. HSE University, 2022.
- Menczer, Filippo, Santo Fortunato, and Clayton A. Davis. A first course in network science. Cambridge University Press, 2020.
- Barabási, Albert-László. "Network science." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371.1987 (2013): 20120375.
- Travers, Jeffrey, and Stanley Milgram. "An experimental study of the small world problem." Social networks. Academic Press, 1977. 179-197.
- Adamic, Lada A., and Natalie Glance. "The political blogosphere and the 2004 US election: divided they blog." Proceedings of the 3rd international workshop on Link discovery. 2005.
- Labatut, Vincent, and Xavier Bost. "Extraction and analysis of fictional character networks: A survey." ACM Computing Surveys (CSUR) 52.5 (2019): 1-40.
- Shao, Chengcheng, et al. "The spread of low-credibility content by social bots." Nature communications 9.1 (2018): 1-9.
- Normile, Dennis. "The Metropole, superspreaders, and other mysteries." (2013): 1272-1273.
- Bearman, Peter S., James Moody, and Katherine Stovel. "Chains of affection: The structure of adolescent romantic and sexual networks." American journal of sociology 110.1 (2004): 44-91.


HSE UNIVERSITY
EST 1992