Network Analysis

Computational International Relations - Week 5

Pouria Mirelmi
5 March 2025
Leiden University

What are Networks?

What are Networks?

- Networks are maps or scaffolds of complex systems.
- They provide a structural representation of interesting data, which makes it well-suited for mathematical analysis.
- This enables the study of structure at different scales.
- Examples: Network of brain cells, friendships, Instagram users, trade networks.
- Other examples?

What are Networks?

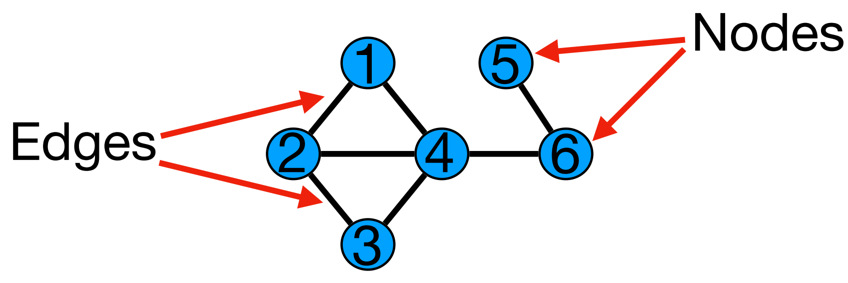
- Networks consist of two main ingredients:

    1. Nodes (or Vertices): Entities or actors

    2. Edges (or Links, or Ties): Connections between nodes

- Different networks can be defined on the same set of nodes! —> Co-authorships

- Nodes can have different attributes: gender, age, location, etc.



Friendship

What are Networks?

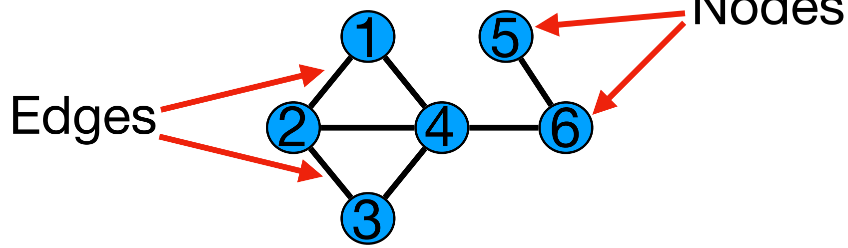
- Networks consist of two main ingredients:

  1. Nodes (or Vertices): Entities or actors

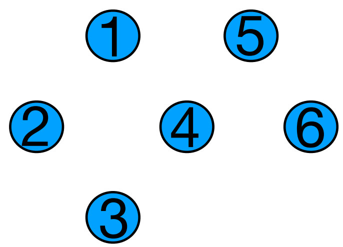
  2. Edges (or Links, or Ties): Connections between nodes

- Different networks can be defined on the same set of nodes! —> Co-authorships

- Nodes can have different attributes: gender, age, location, etc.



Friendship



Cheating (isolated nodes)

What are Networks?

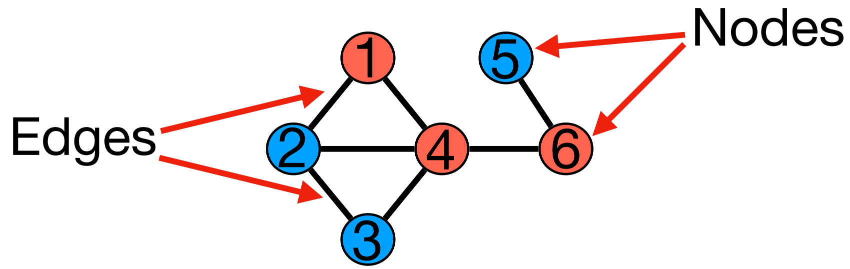
- Networks consist of two main ingredients:

    1. Nodes (or Vertices): Entities or actors

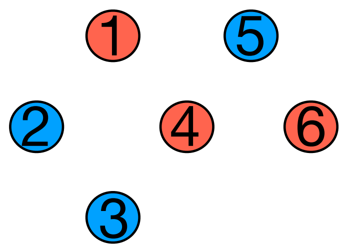
    2. Edges (or Links, or Ties): Connections between nodes

- Different networks can be defined on the same set of nodes! —> Co-authorships

- Nodes can have different attributes: gender, age, location, etc.



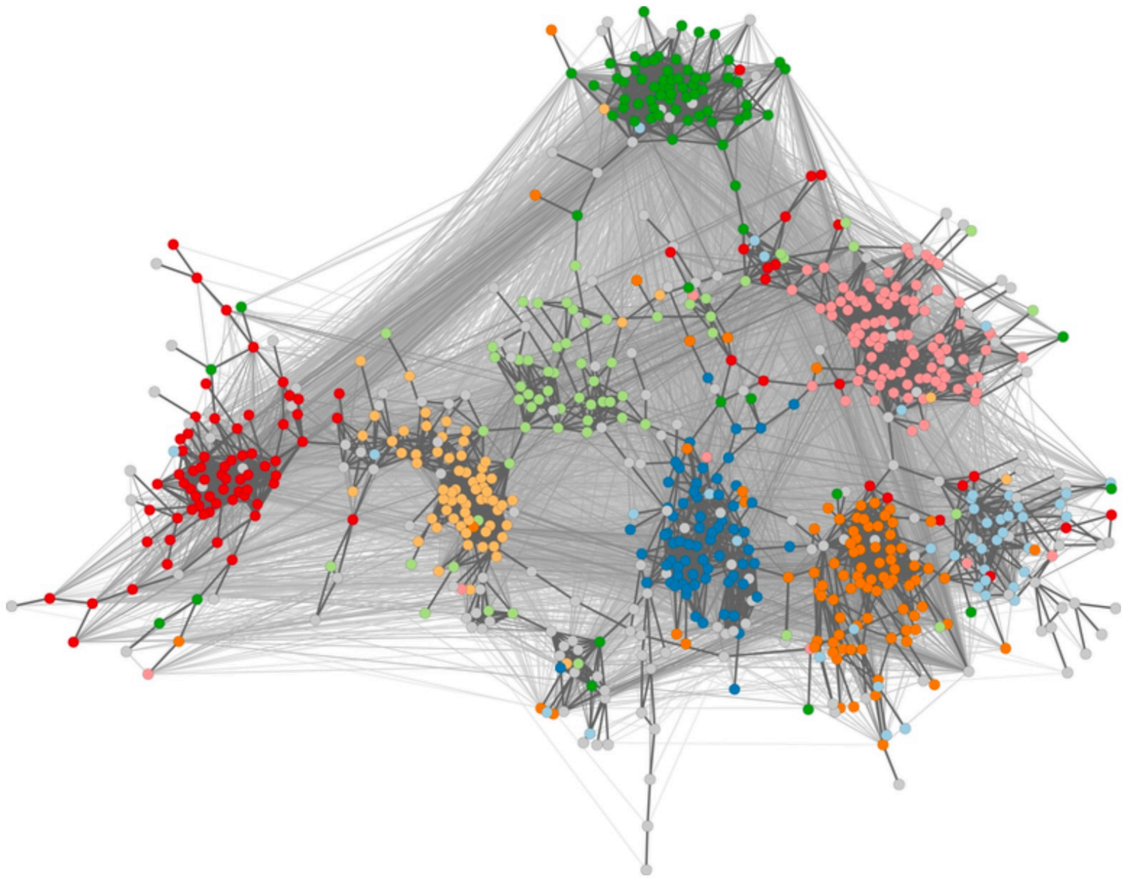
Friendship



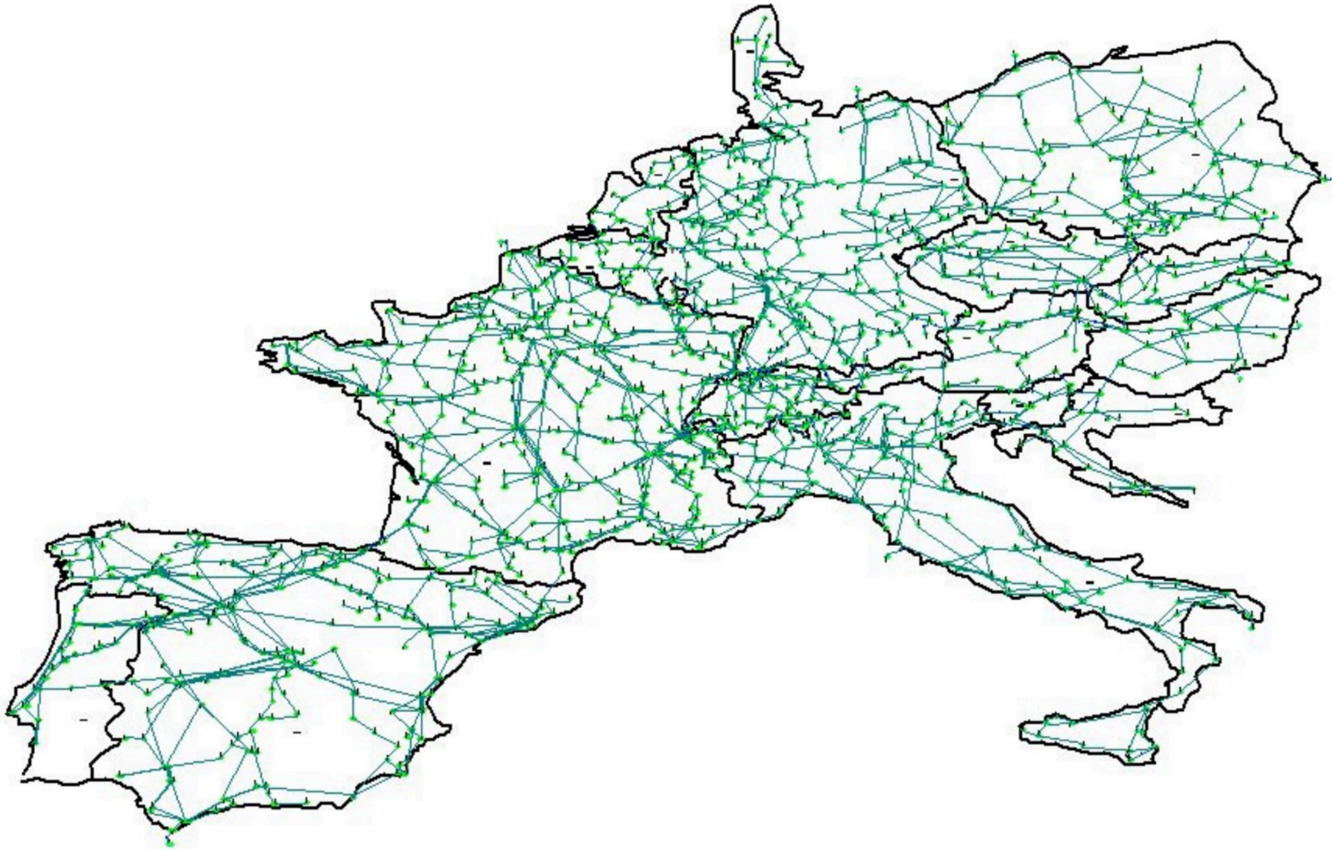
Cheating (isolated nodes)

Example: Friendships in a University Dorm

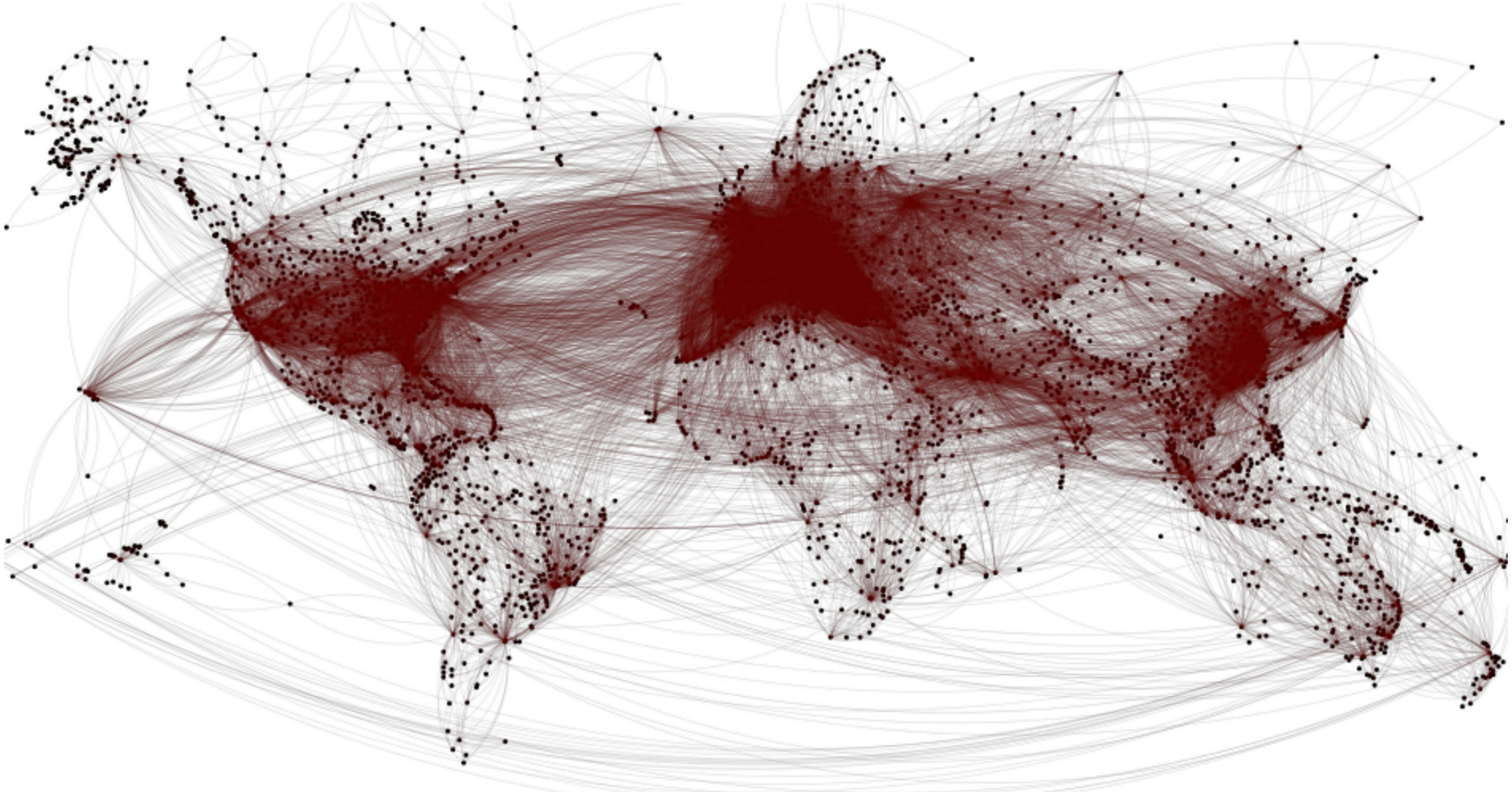
- 769 nodes (students)
- 16600 edges (symmetric friendship ties)
- Colours encode the dorm variable (grey for missing values)



Example: European Power Grid (Simplified)



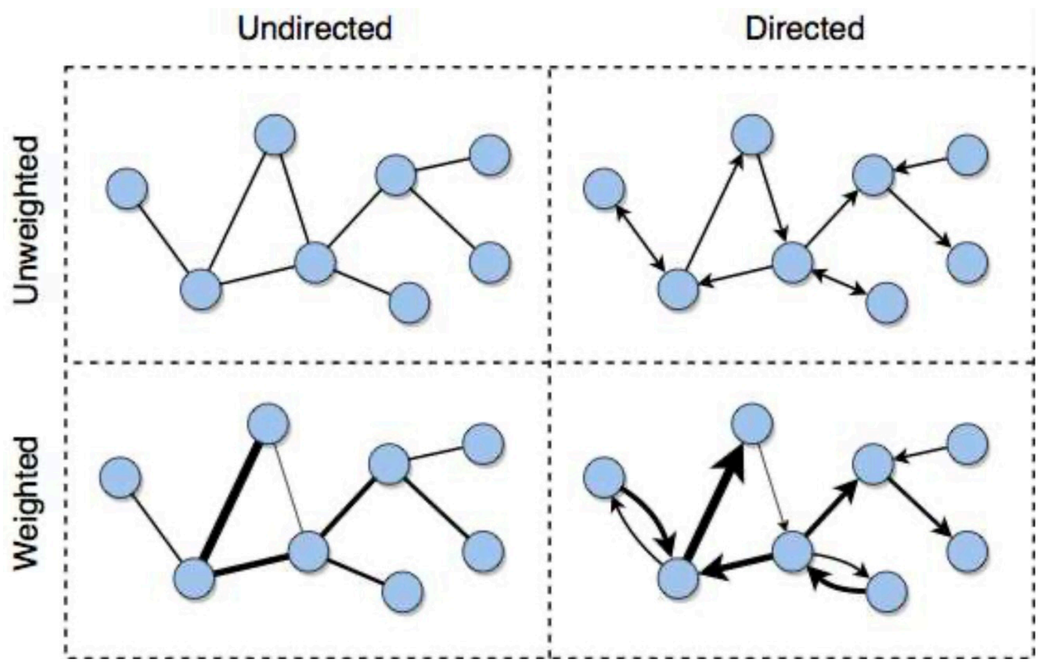
Example: Global Air Traffic Network



More Definitions

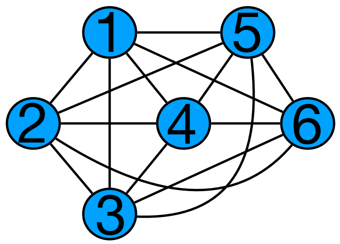
- A network can be **undirected** or **directed.** A directed network is also called a **digraph.** In directed networks, links are called **directed links.**

- A network can be **unweighted** or **weighted.** In a weighted network, links have associated weights: **the weighted link between nodes i and j will have weight w.** A network can be both directed and weighted, in which case it has directed weighted links.



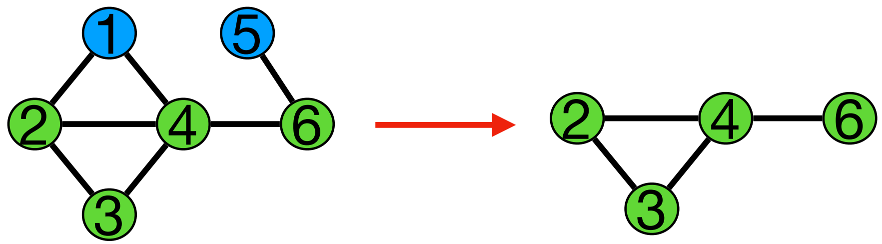
More Definitions

- A **complete graph/network** is an undirected graph/network in which every pair of distinct vertices is connected by a unique edge.



N1: A complete graph

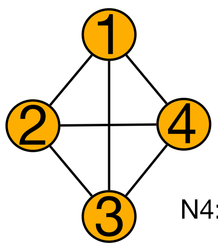
- A **subnetwork** is a network obtained by selecting a subset of the nodes and all of the links **among these nodes.**



N2: The friendship network

N3: A subnetwork of N2

- A **clique** is a complete subnetwork



N4: A clique of N1

Basic Measures

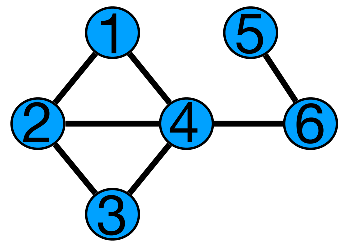
- Network size N = number of nodes
- L = number of links
- Density: $d = \frac{L}{L_{max}} = \frac{2L}{N(N-1)}$
- The network is **sparse** if d<<1
- Large networks are usually very sparse
- Facebook: $d \approx 0.000001$

Table 1.1 Basic statistics of network examples. Network types can be (D)irected and/or (W)eighted. When there is no label the network is undirected and unweighted. For directed networks, we provide the average in-degree (which coincides with the average out-degree).

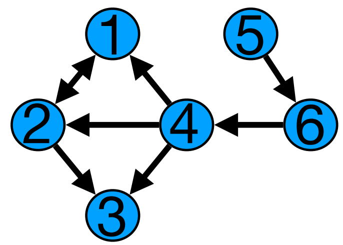
Network	Type	Nodes (N)	Links (L)	Density (d)	Average degree ($\langle k \rangle$)
Facebook Northwestern Univ.		10,567	488,337	0.009	92.4
IMDB movies and stars		563,443	921,160	0.000006	3.3
IMDB co-stars	W	252,999	1,015,187	0.00003	8.0
Twitter US politics	DW	18,470	48,365	0.0001	2.6
Enron Email	DW	87,273	321,918	0.00004	3.7
Wikipedia math	D	15,220	194,103	0.0008	12.8
Internet routers		190,914	607,610	0.00003	6.4
US air transportation		546	2,781	0.02	10.2
World air transportation		3,179	18,617	0.004	11.7
Yeast protein interactions		1,870	2,277	0.001	2.4
C. elegans brain	DW	297	2,345	0.03	7.9
Everglades ecological food web	DW	69	916	0.2	13.3

Basic Measures

- The **degree** k_i of a node is its number of links, or neighbours
  - Most basic metric of a node, many others build on it

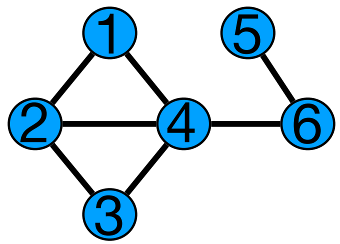


- In directed networks
  - **In-degree:** number of incoming links k_i^{in}
  - **Out-degree:** number of outgoing links k_i^{out}



Basic Measures

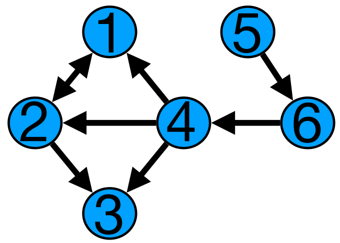
- The **degree** k_i of a node is its number of links, or neighbours
    - Most basic metric of a node, many others build on it



$k_4 = ?$

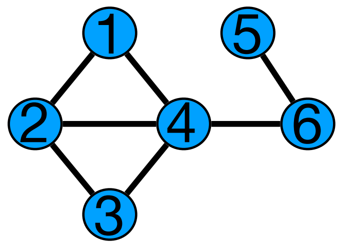
$k_5 = ?$

- In directed networks
    - **In-degree:** number of incoming links k_i^{in}
    - **Out-degree:** number of outgoing links k_i^{out}



Basic Measures

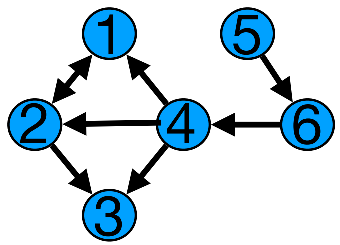
- The **degree** k_i of a node is its number of links, or neighbours
    - Most basic metric of a node, many others build on it



$k_4 = 4$

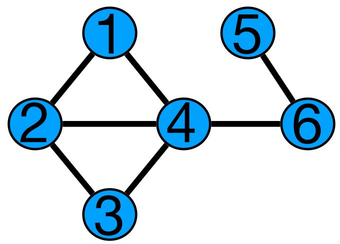
$k_5 = 1$

- In directed networks
    - **In-degree:** number of incoming links k_i^{in}
    - **Out-degree:** number of outgoing links k_i^{out}



Basic Measures

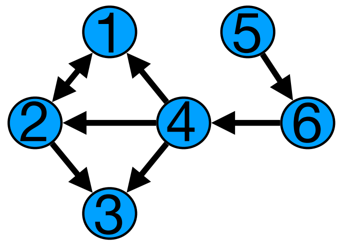
- The **degree** k_i of a node is its number of links, or neighbours
  - Most basic metric of a node, many others build on it



$k_4 = 4$

$k_5 = 1$

- In directed networks
  - **In-degree:** number of incoming links k_i^{in}
  - **Out-degree:** number of outgoing links k_i^{out}

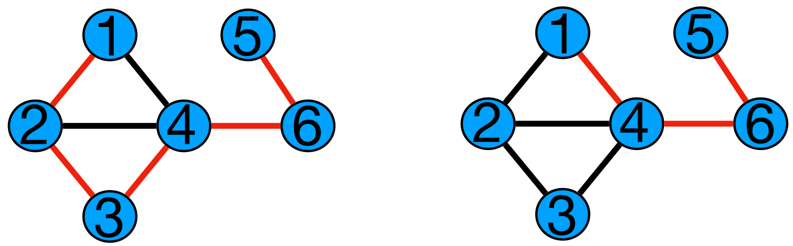


$k_4^{in} = 1$

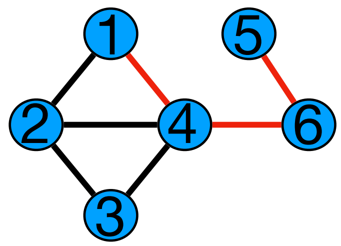
$k_4^{out} = 3$

Paths and Distances

- A **path** between two nodes is a sequence of edges that connects them, passing through intermediate nodes if necessary.
- A **path length** is the number of edges in the sequence.
- A **shortest path** is the path with the fewest number of edges connecting two nodes.



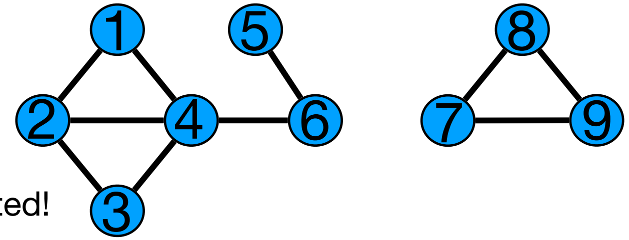
Two different paths between 1 and 5



The shortest path with length 3

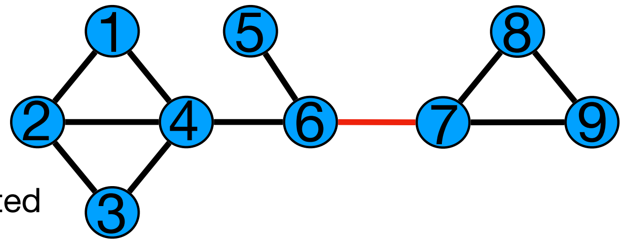
Connectedness and Components

- A network is **connected** if there is a path between every pair of its nodes.



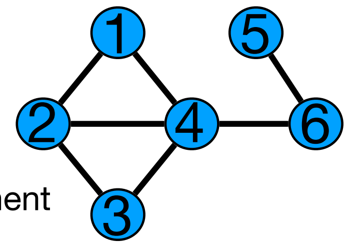
Not connected!

- Connected components



Connected

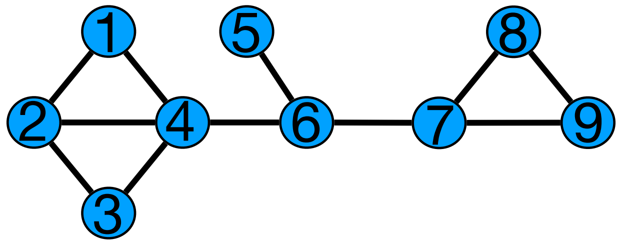
- Largest connected component (Giant component)



Giant component

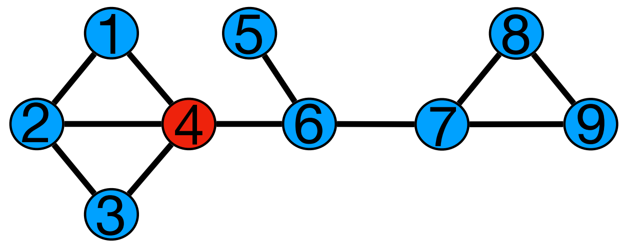
Node Importance: Centrality Measures

- Node centrality
  - A measure to quantify the importance of a node within a network.
  - It helps identify key actors who have the most influence, best access to information, or play crucial roles in connectivity.
- Different types of centrality measures
  - Capturing different aspects of a node's importance.
- What aspects...?



Degree and Eigenvector Centralities

- Degree centrality
  - Importance based on the **number of direct connections (links)**
  - Nodes with high degree centrality: many direct connections, can spread information quickly
  - Applications: Influencers, super-spreaders in decease transmission, opinion leaders in viral marketing

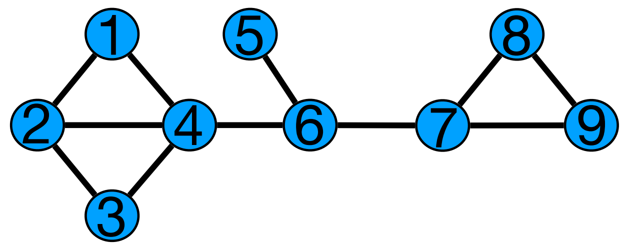


- Eigenvector centrality
  - Quantity AND quality of the neighbouring nodes
  - Example: A lobbyist with multiple connections to important US senators

$$x_i = \frac{1}{\lambda} \sum_j A_{ij} x_j$$

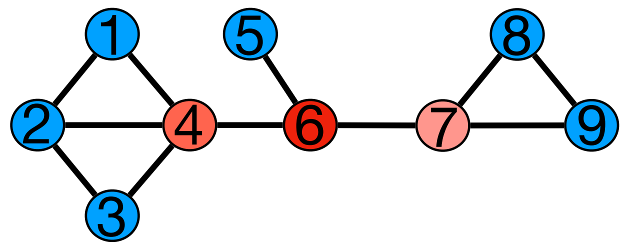
Betweenness Centrality

- Betweenness centrality
  - Importance based on **a node's role as a bridge between other nodes.**
  - Quantifies how often a node appears on the shortest paths between pairs of other nodes in the network.
  - High betweenness centrality: significant control over information flow in the network.
  - Removal of nodes with high betweenness can disrupt communication between different parts of the network.



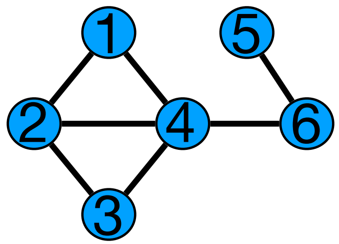
Betweenness Centrality

- Betweenness centrality
  - Importance based on **a node's role as a bridge between other nodes.**
  - Quantifies how often a node appears on the shortest paths between pairs of other nodes in the network.
  - High betweenness centrality: significant control over information flow in the network.
  - Removal of nodes with high betweenness can disrupt communication between different parts of the network.



Closeness centrality

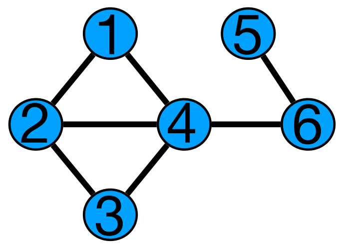
- Closeness centrality
  - Importance based on **how close a node is to all other nodes in the network.**
  - High closeness: The ability to quickly reach many others in the network
  - Low closeness: Takes longer to reach others in the network
  - Applications in transportation and logistics
  - Closeness of nodes 4 and 5??



$$C(v) = \frac{1}{\sum_{u \neq v} d(v, u)}$$

Network Representation

- Edge Lists
  - Very efficient
  - Representing edges of the network
  - Stored as pairs of nodes
  - Can be used to represent directed and weighted networks as well



$$\{[1,2], [1,4], [2,4], [2,3], [3,4], [4,6], [5,6]\}$$

- In directed networks —> order matters
- In weighted networks —> weights added:

$$\{[1,2,1], [1,4,4], \ldots\}$$

Homophily

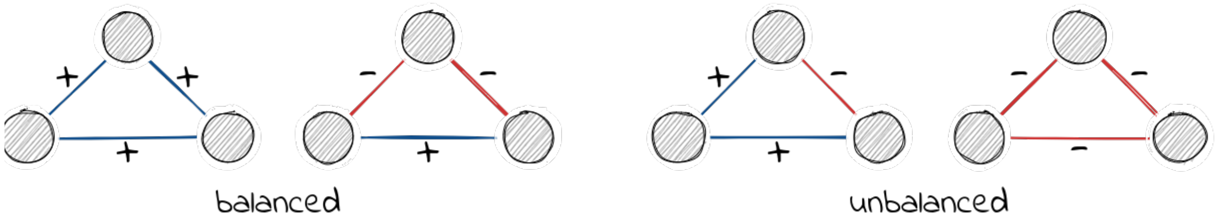
- **Homophily** is the tendency of individuals to associate and bond with others who are similar to them in characteristics such as age, gender, ethnicity, education, occupation, social status, or interests.

- It is a fundamental principle in social network theory and plays a significant role in shaping relationships, communities, and information flow.

- Examples?

Structural Balance Theory

- **Structural Balance Theory** explains how relationships in a network tend to evolve toward a state of **psychological stability.**
- It is based on the idea that **triads** (groups of three nodes) in a network should maintain balance
  - A friend of my friend is my friend
  - An enemy of my enemy is my friend
  - The product of the edge signs should be positive



R Hands On: The Train Bombing Network

- Two R scripts: train1.R and train2.R
- One edgelist file: train_edgelist.txt
- The Train Bombing Network
  - Undirected network, containing contacts between suspected terrorists involved in the train bombing of Madrid on March 11, 2004.
  - Reconstructed from newspapers
  - Nodes: terrorists, edges: contacts between two terrorists
  - Edge weights: how strong a connection was. Including friendship or co-participating in training camps
  - Weights are ignored in today's exercise

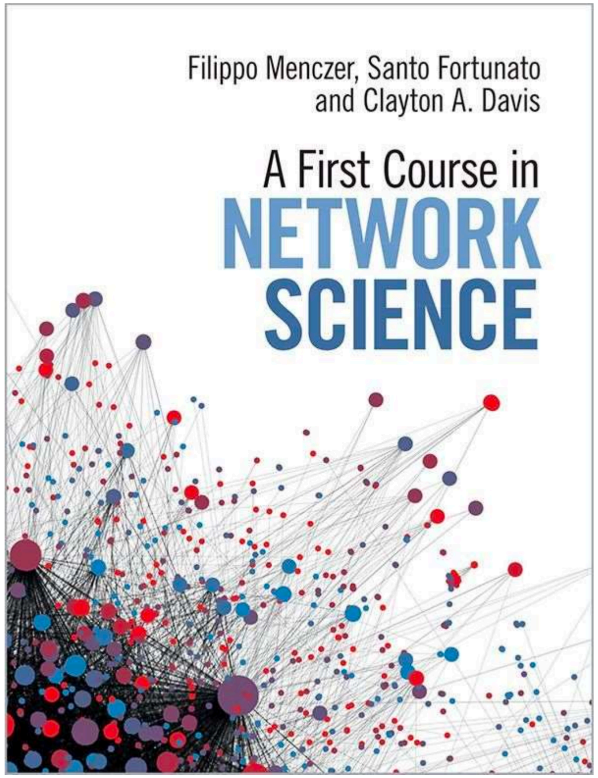
Web Scraping in R

A self-explanatory introductory lesson:

https://statsandr.com/blog/web-scraping-in-r/

References

- “A First Course in Network Science” by Menczer et al.
- “Social Networks” course by Prof. Dr. Markus Strohmaier, Summer 2021, RWTH Aachen University.
- https://github.com/joaopn/teaching_networks_2023



Thank you!