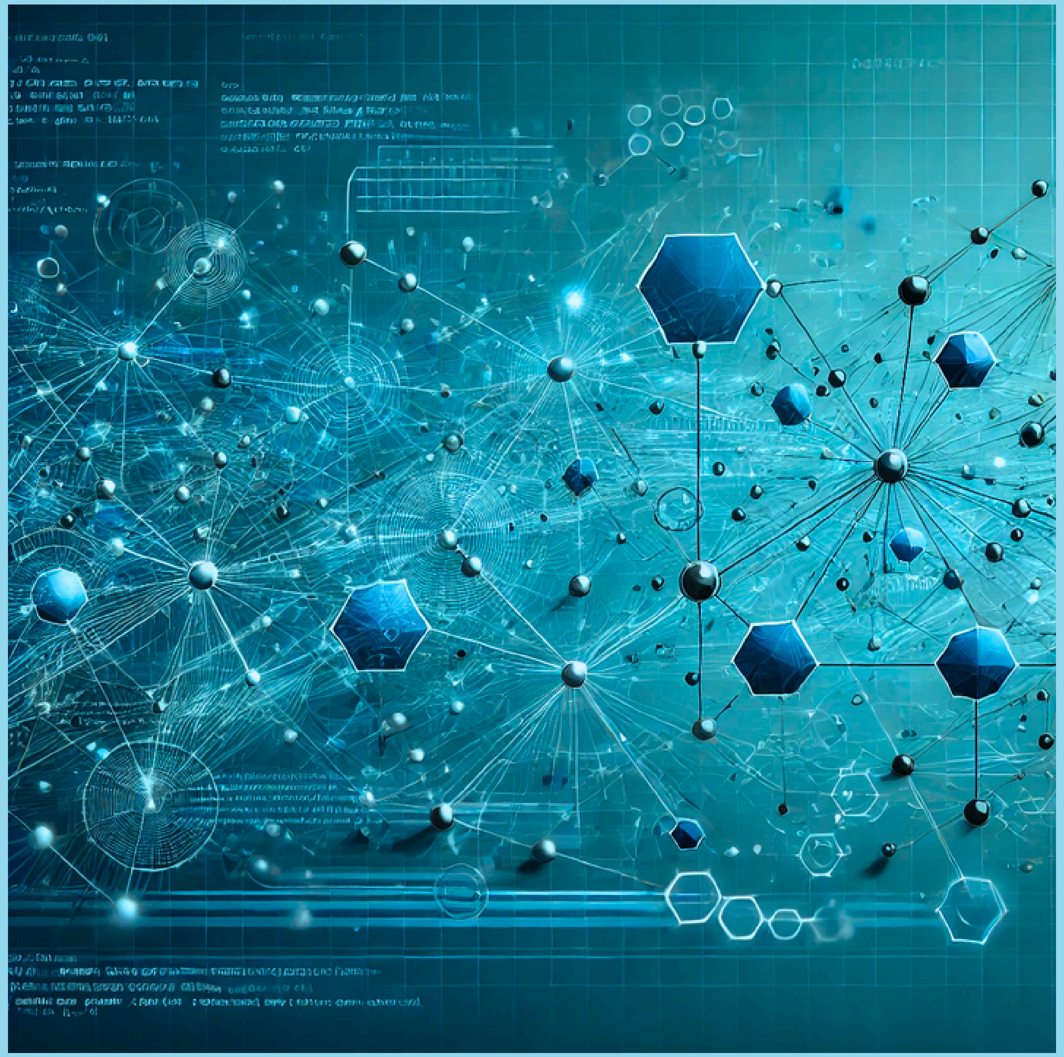


Universität Konstanz

UNIVERSITÄT KONSTANZ

Network Theory Basics

Network Science of Socio-Economic Systems

Giordano De Marzo

Outline

1. Networks Basics
2. Measuring Networks
3. Real World Networks
4. The Value of Networks

Networks Basics

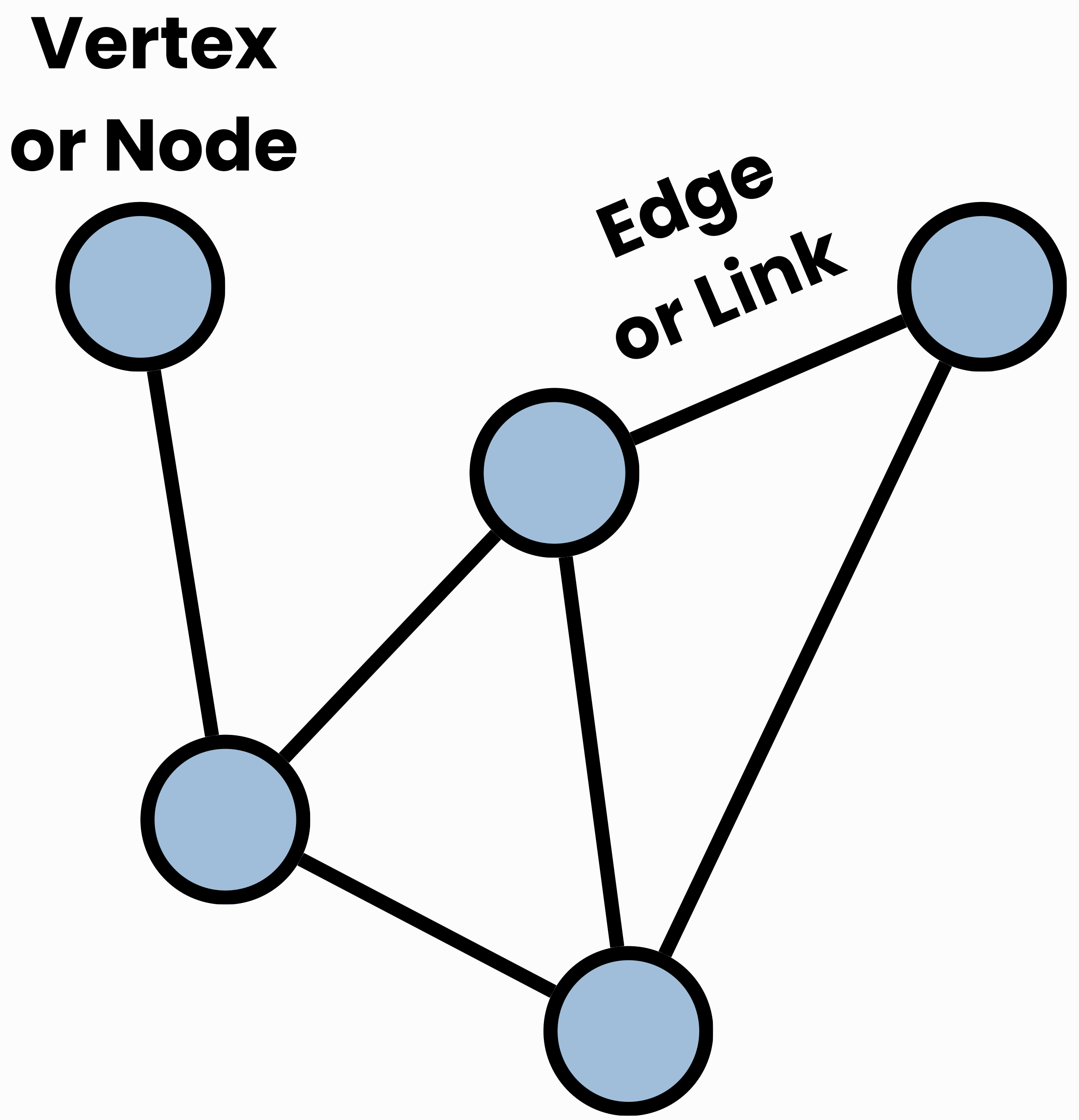
What is a **Network?**

A Network or Graph G(V, E) is a set of vertices or nodes V and and edges or links E

- nodes represent entities in the system (eg. people on a social network)
- edges represent connections among the nodes (eg. friendship in a social network)

We denote by

- N the number of nodes
- E the number of links



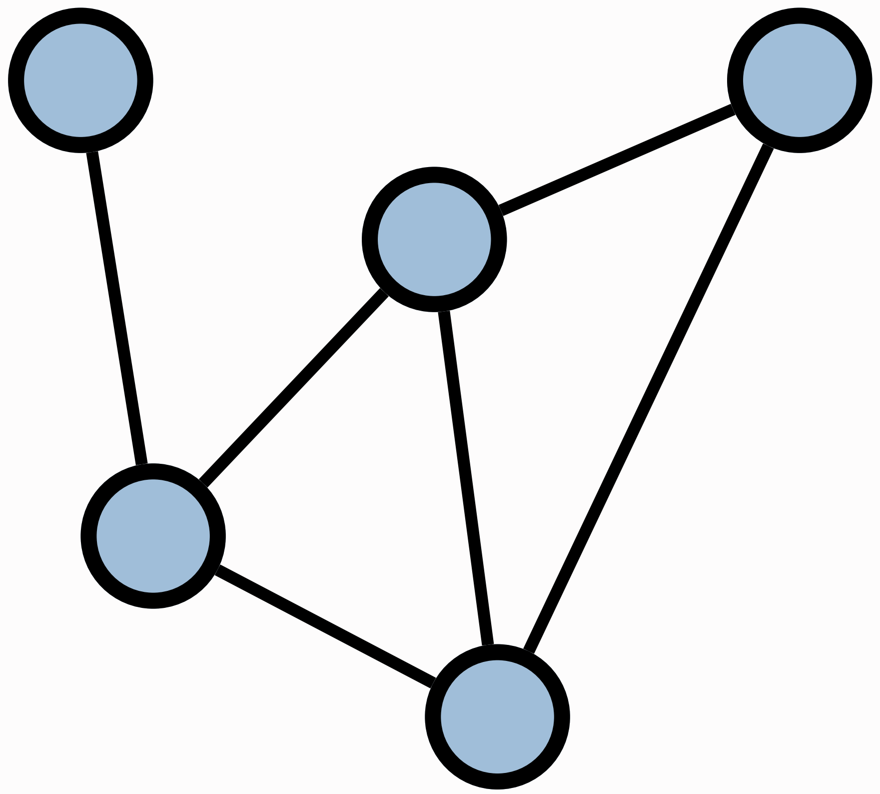
Network **Types**

There are three main types of graphs

Undirected

Links are bidirectional

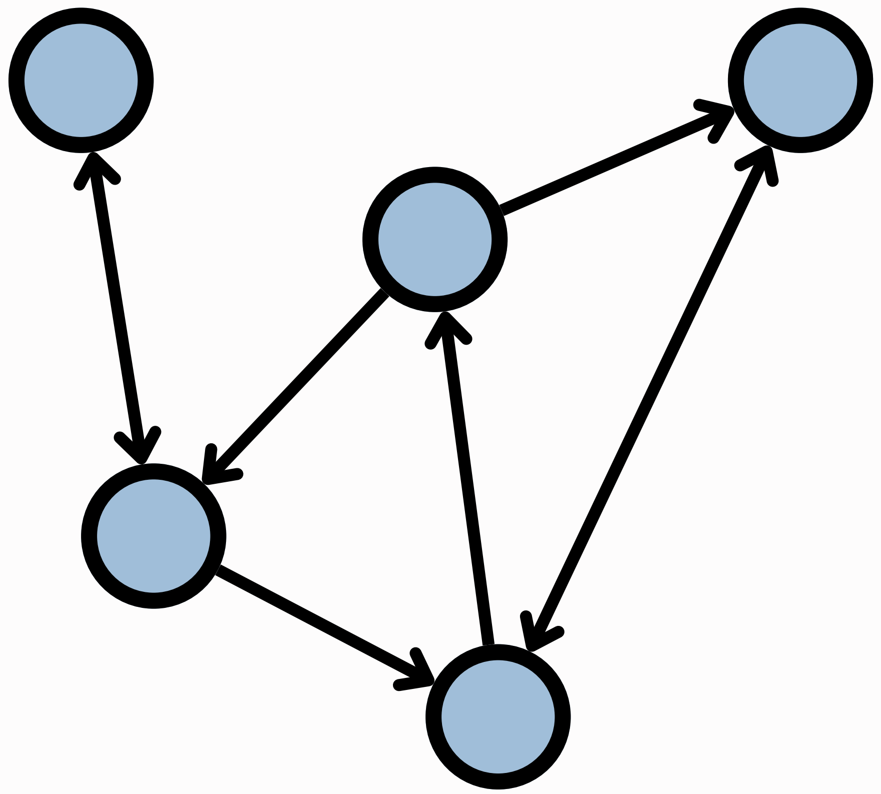
E.g. Facebook



Directed

Links are directional

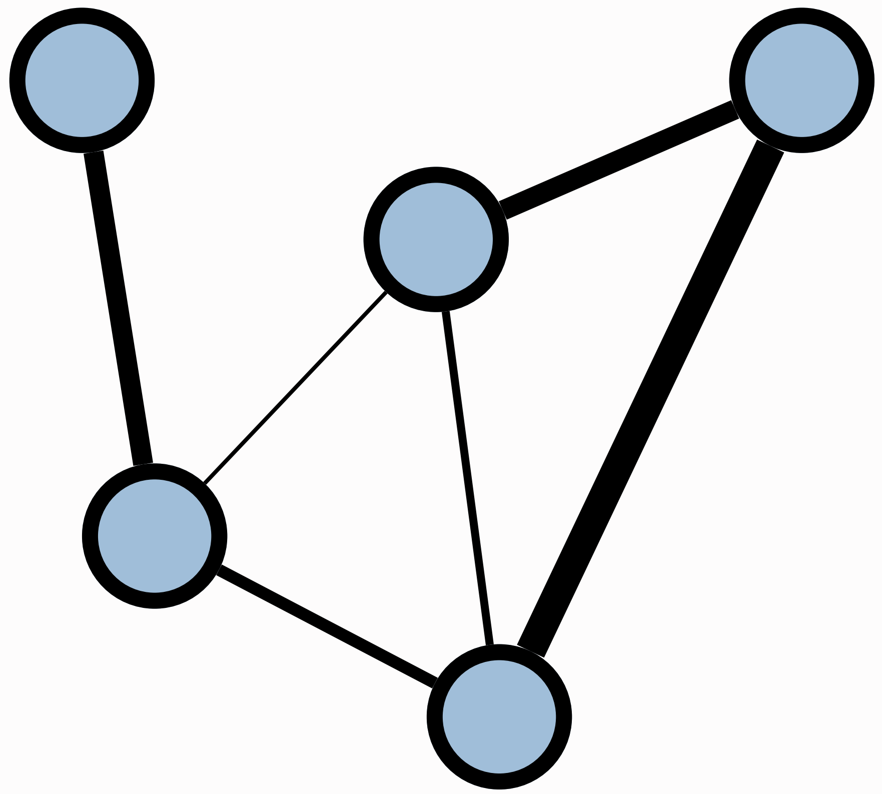
E.g. Twitter



Weighted

Links are weighted

E.g. Road Network

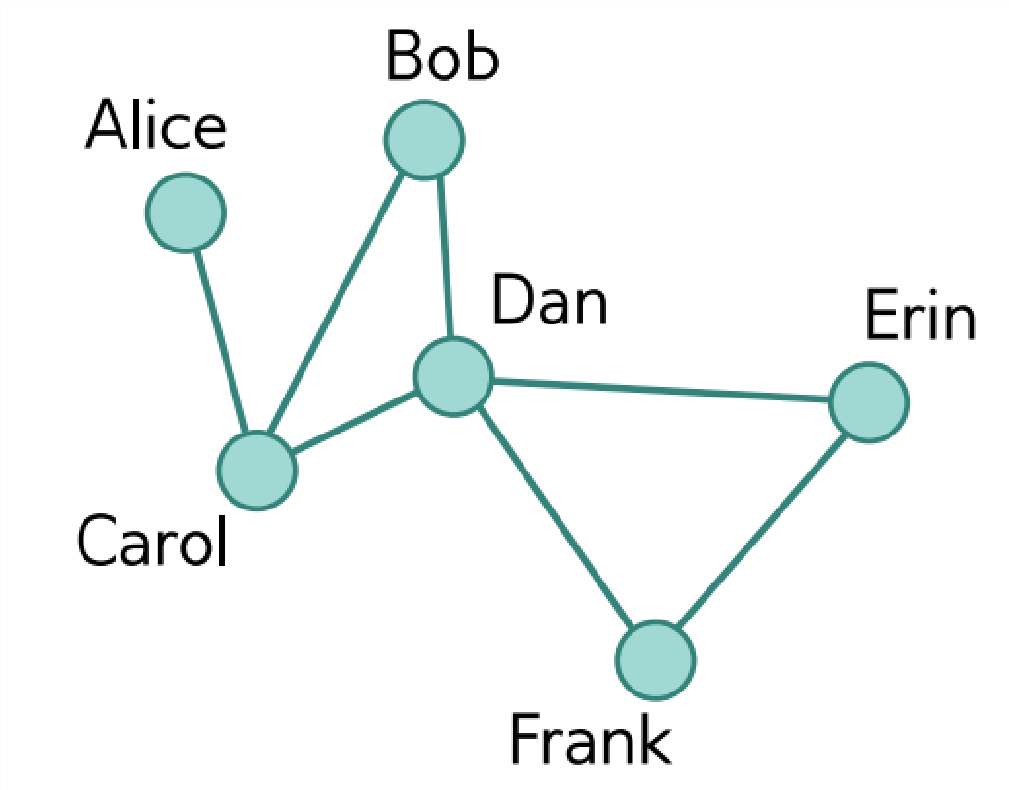


Examples of Networks

Graphs are everywhere, many systems can be described using this formalism

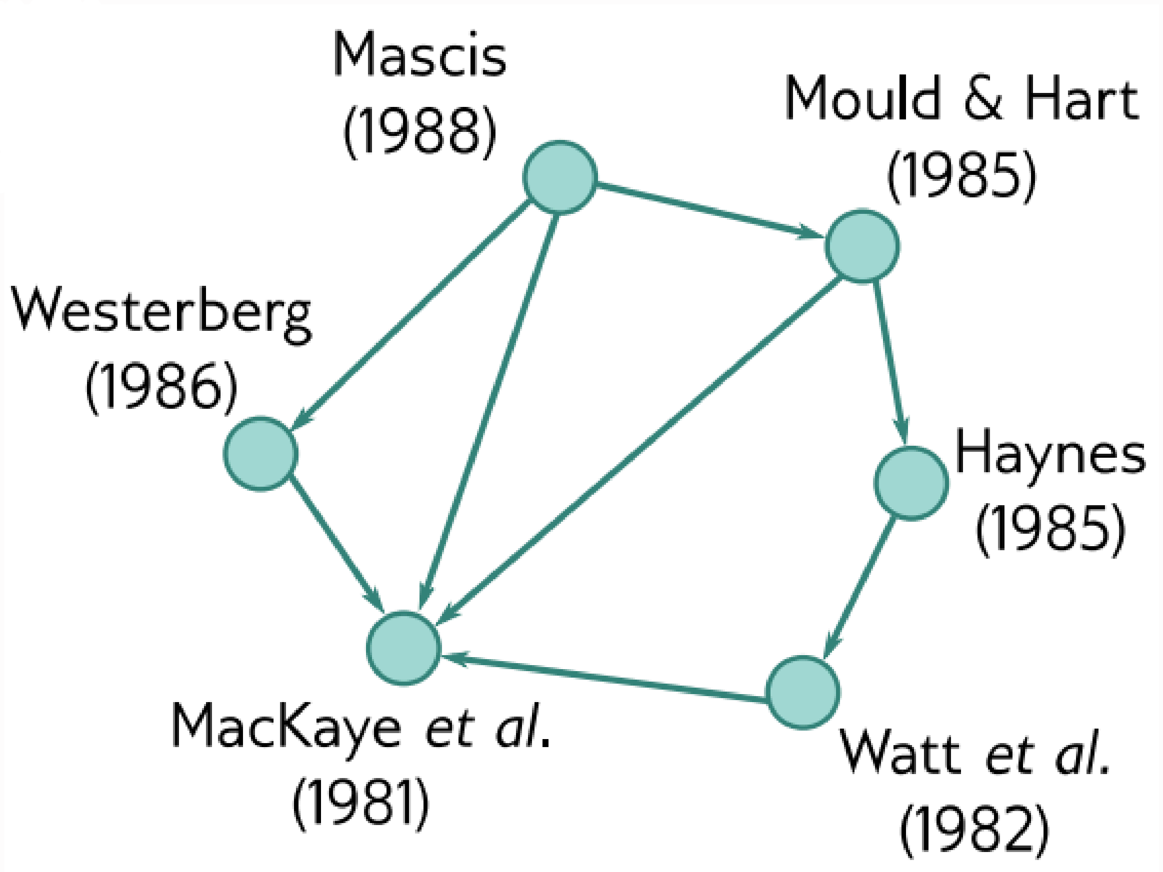
Undirected

Friendship Networks



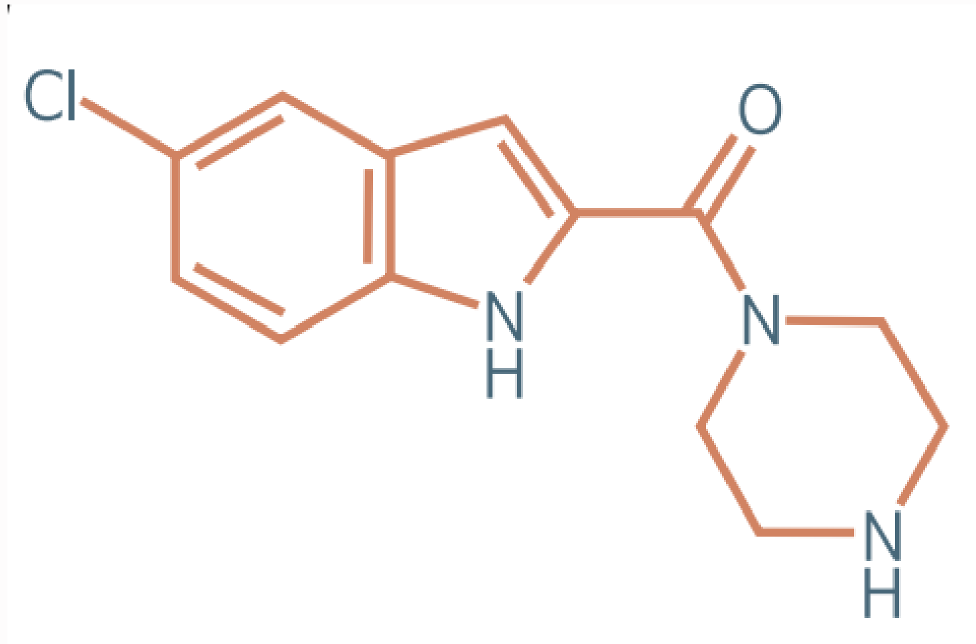
Directed

Citation Networks



Weighted

Molecules



Examples of Simple Networks

Among all networks, the most simple are lattices, trees and random networks

Lattice

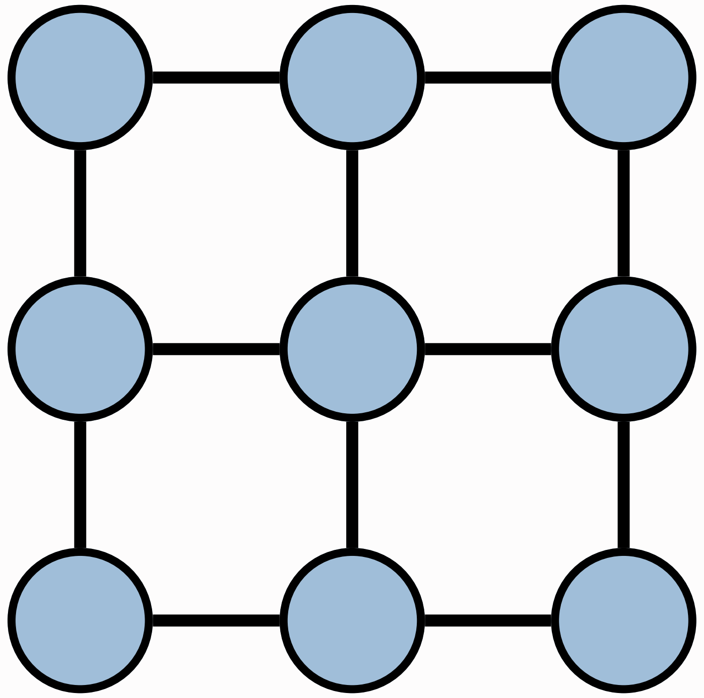
All nodes have the same degree and the structure is homogeneous

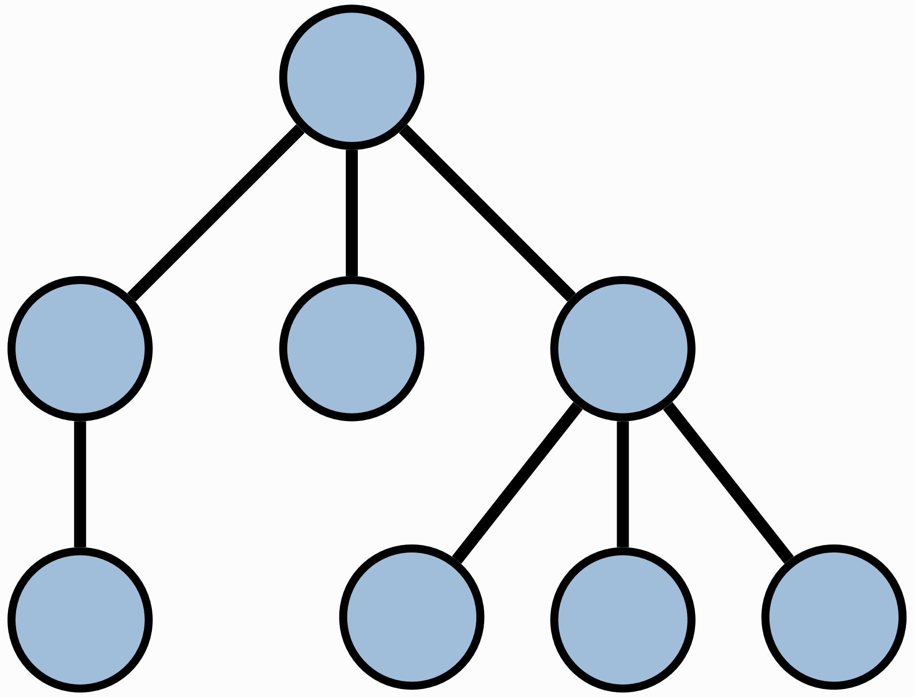
Tree

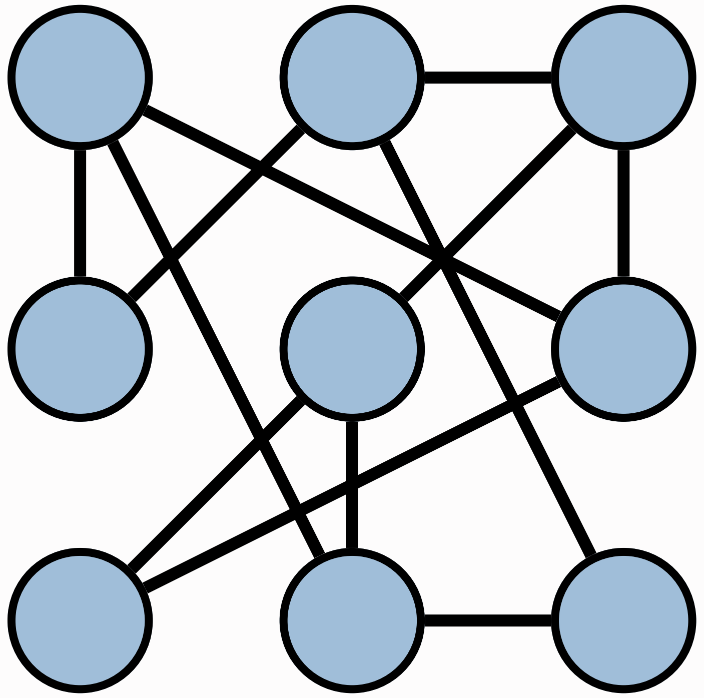
Characterized by the absence of loops

Random

Links are formed at random, nodes have the same degree on average



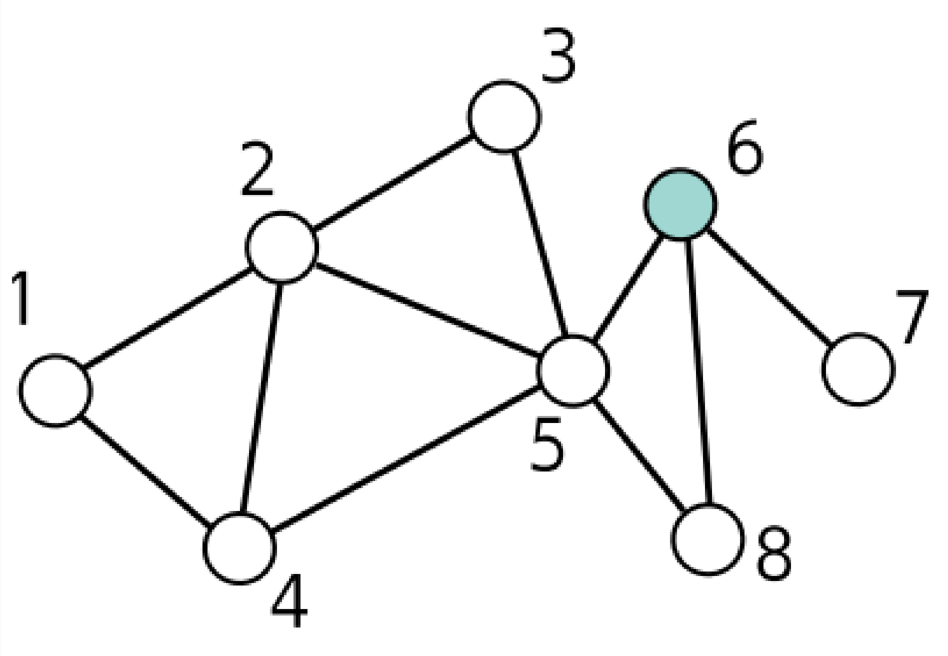




The Adjacency Matrix

A network is mathematically represented by its adjacency matrix A

- it's an NxN matrix
- the element A_{ij} of the matrix is different from zero if there is a link going from i to j
    - $A_{ij}=1$ in unweighted graphs
    - $A_{ij}=w_{ij}$ in weighted graphs



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

Properties of the Adjacency Matrix

Many of the characteristics of a network can be inferred by the structure of its adjacency matrix

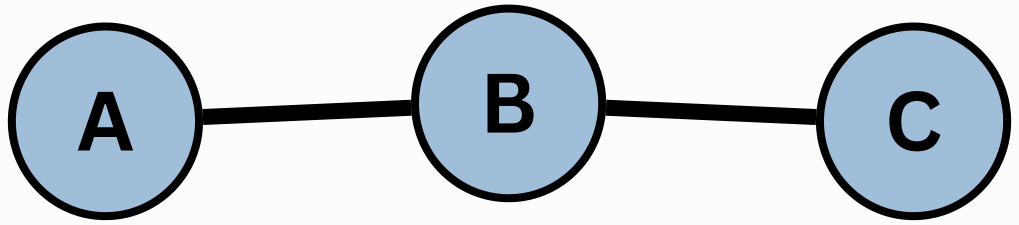
- if the graphs is undirected the adjacency matrix is simmetric $A_{ij}=A_{ji}$
- the elements on the diagonal are self-loops (which are typically absent)

Note that the same network can be described by several adjacency matrix

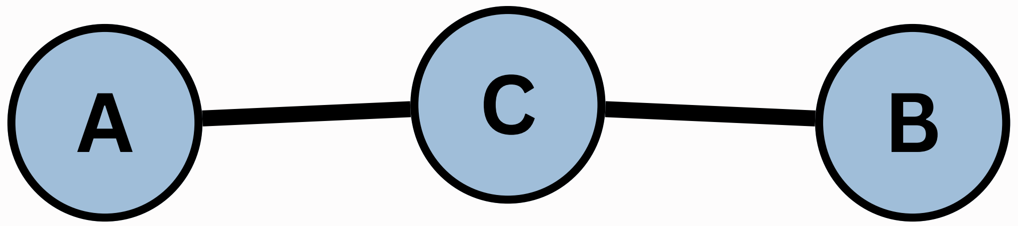
- two networks are isomorphic if one can be transformed into the other by relabeling nodes
- the adjacency matrices will appear different, but they are describing the same network

Example: Isomorphic Networks

The two networks below are isomorphic, they only difference is how we label and order the nodes



	A	B	C
A	0	1	0
B	1	0	1
C	0	1	0



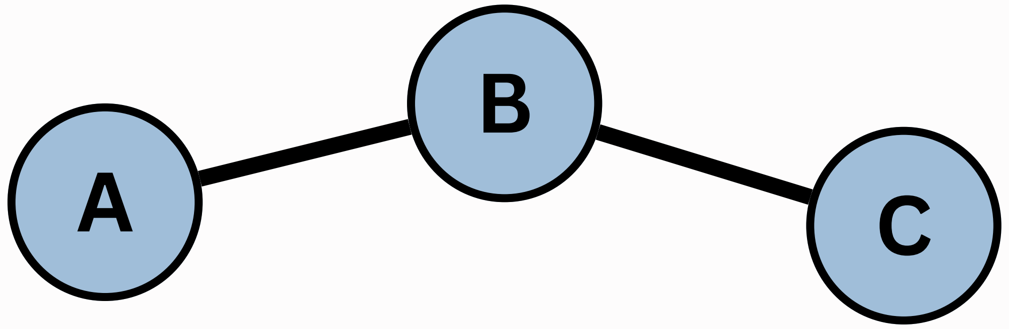
	A	B	C
A	0	0	1
B	0	0	1
C	1	1	0

Network Representations

The Adjacency matrix is not the only way of represent a network. Each representation has advantages and disadvantages and the choice depends on the particular task. There are 3 possibilities

- **Adjacency Matrix:** Represents the network as a square matrix where rows and columns correspond to nodes, and entries indicate edges between nodes.
- **Adjacency List:** Represents the network as a list or dictionary where each node is associated with a list of adjacent nodes.
- **Edge List:** Represents the network as a list of edges, where each edge is a pair of nodes.

Example: Network Representations



Adjacency Matrix

0	1	0
1	0	1
0	1	0

Adjacency List

A: B
B: A, C
C: B

Edge List

(A, B)
(B, C)

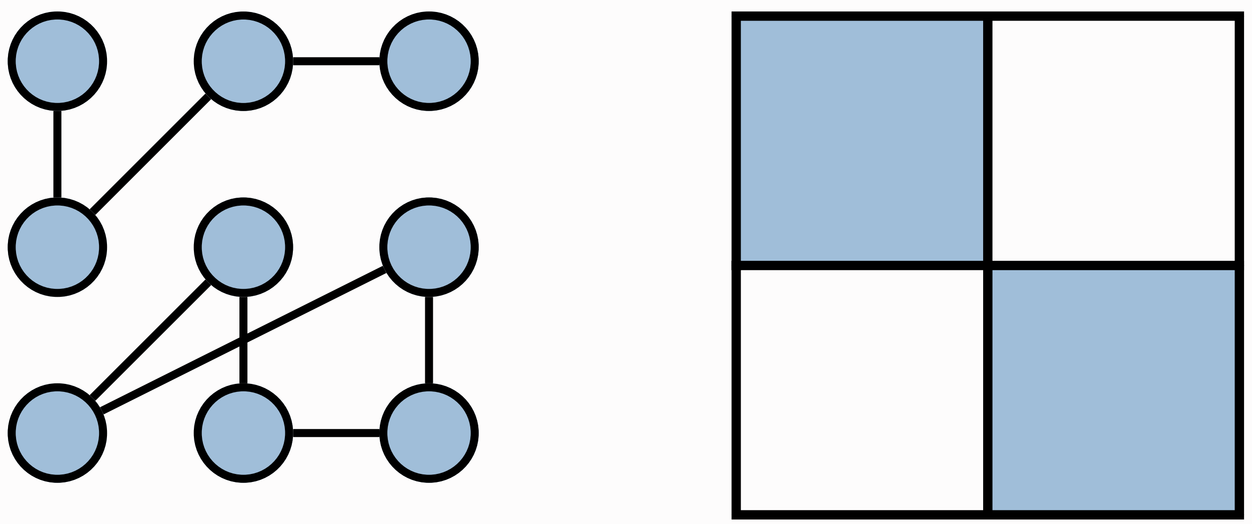
Network Representations

Representation	Memory Usage	Best for	Limitations
Adjacency Matrix	$O(N^2)$	Networks with many links Algebraic operations	Memory intensive
Adjacency List	$O(N+E)$	Networks with few links Neighbors Retrieval	Inefficient when there are many links
Edge List	$O(E)$	Networks with few links	Slow neighbor look-up

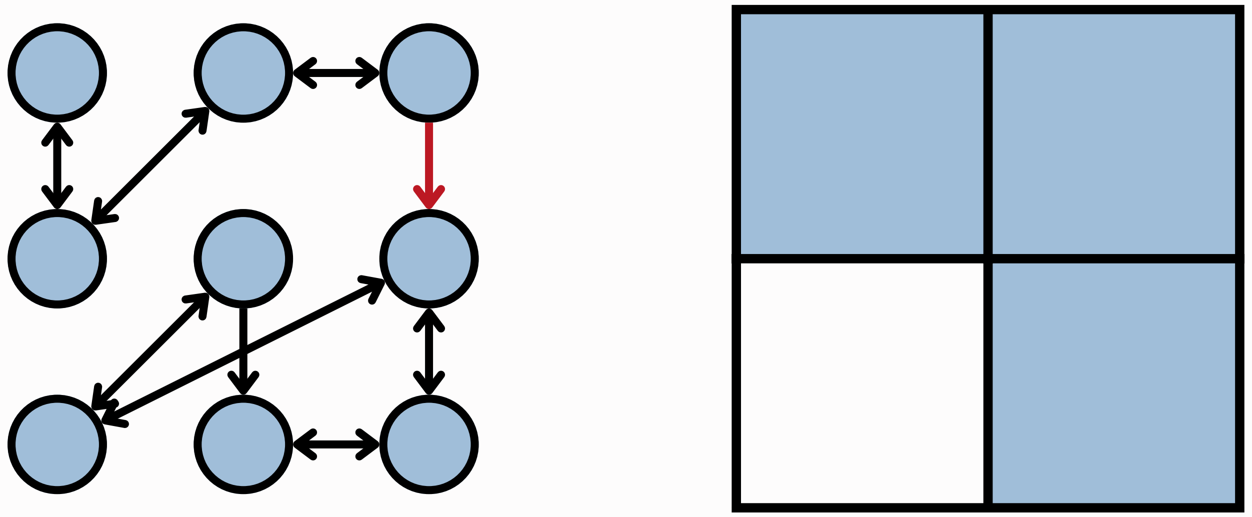
Network **Components**

A **component** is a group of nodes that are all linked together. For directed graphs:

- **Strongly Connected Component**: all nodes in this group can "reach" each other, even following specific directions, like one-way streets
- **Weakly Connected Component:** Nodes are connected when you ignore directions



Undirected network with 2 connected components



Weakly connected directed network

Measuring Networks

Properties of Networks

Given two networks, how can we understand if they are similar or very different?

- we need a set of properties or measures to characterize networks
- these properties must be invariant under isomorphism/node reordering

The most relevant properties we cover are

- diameter and average path length (size of a network)
- degree distribution (connectivity)
- clustering and assortativity (local structure)

Distance on a Network

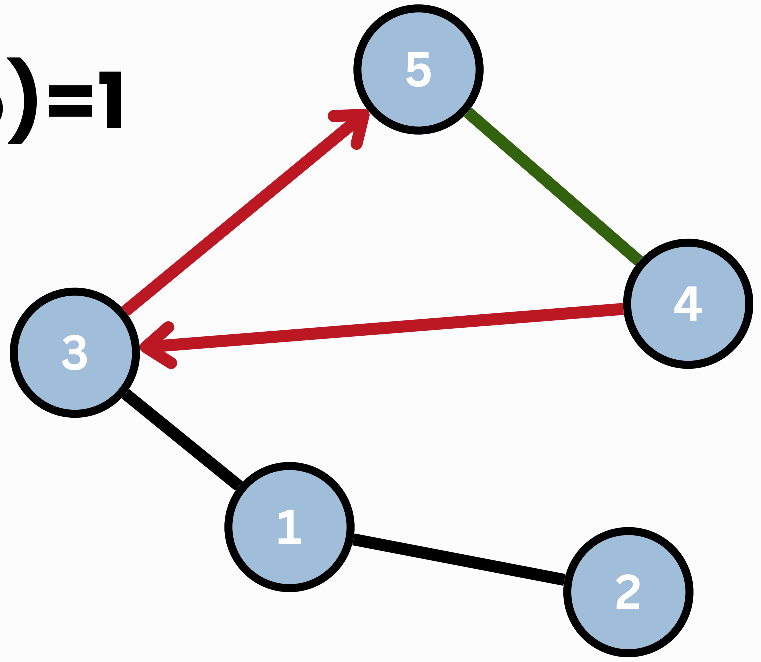
A path on a network is a sequence of links connecting a series of nodes

- the length of a path is the number of links it contains
- the distance d(i,j) between two nodes i, j defined as the length of the shortest path connecting them

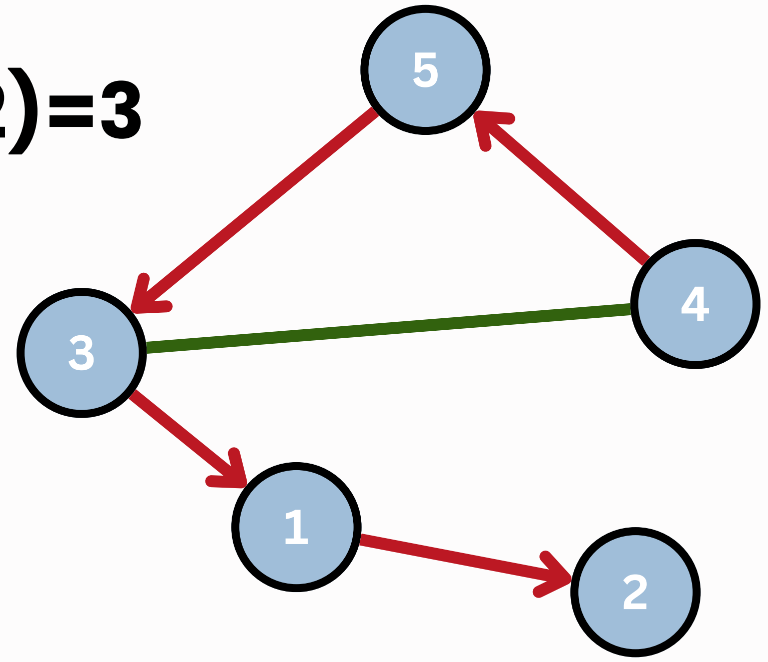
We can find the number of paths between two nodes using the adjacency matrix

- A_{ij} gives the number of length 1 paths from i to j
- $(\mathbf{A}^2)_{ij}$ gives the number of length 2 paths from i to j

d(4,5)=1



d(4,2)=3



Diameter of a Network

A network is different from a typical metrical space like a lattice. However also in this case we can define the dimension or diameter of a network

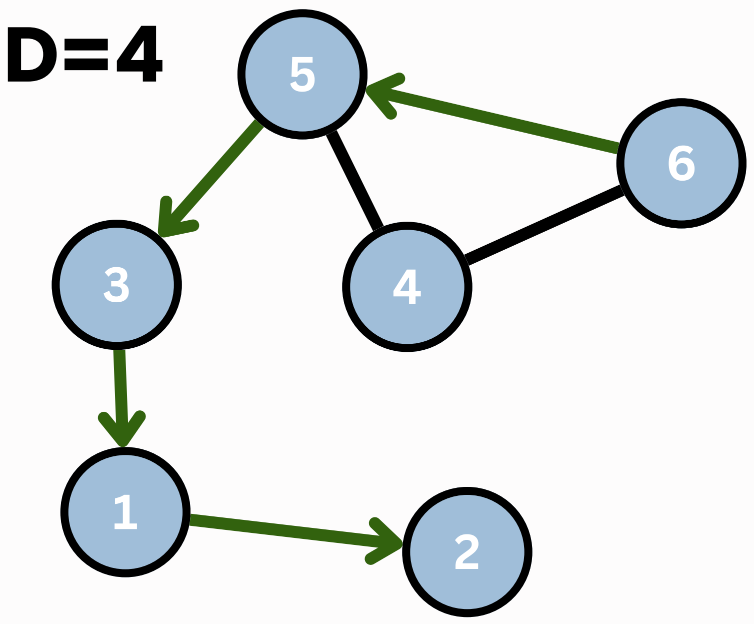
- we take all possible distances between any pair of nodes in the network
- the diameter is defined as the largest of them

$$D = \max[d_{ij}]$$

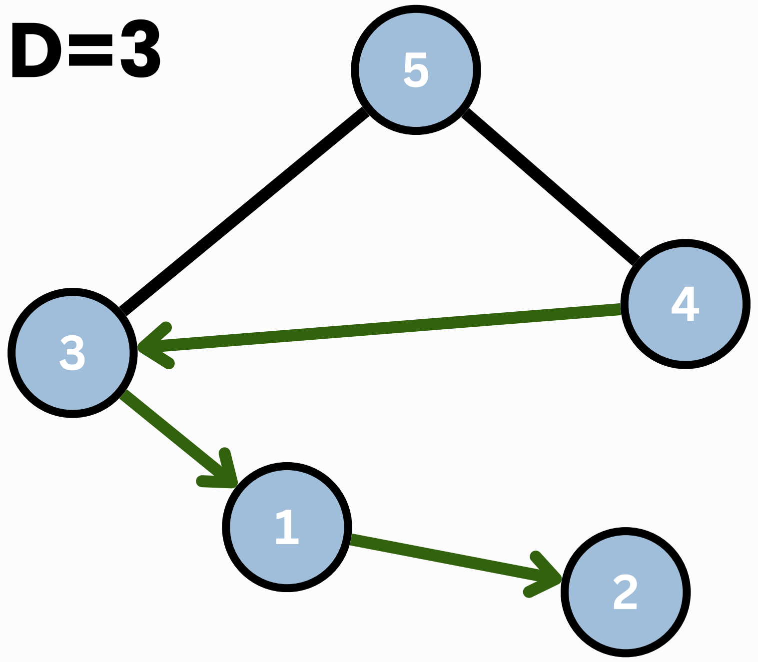
Another useful measure is the average distance L

$$L = \frac{1}{N(N-1)} \sum_{i,j} d_{ij}$$

D=4



D=3



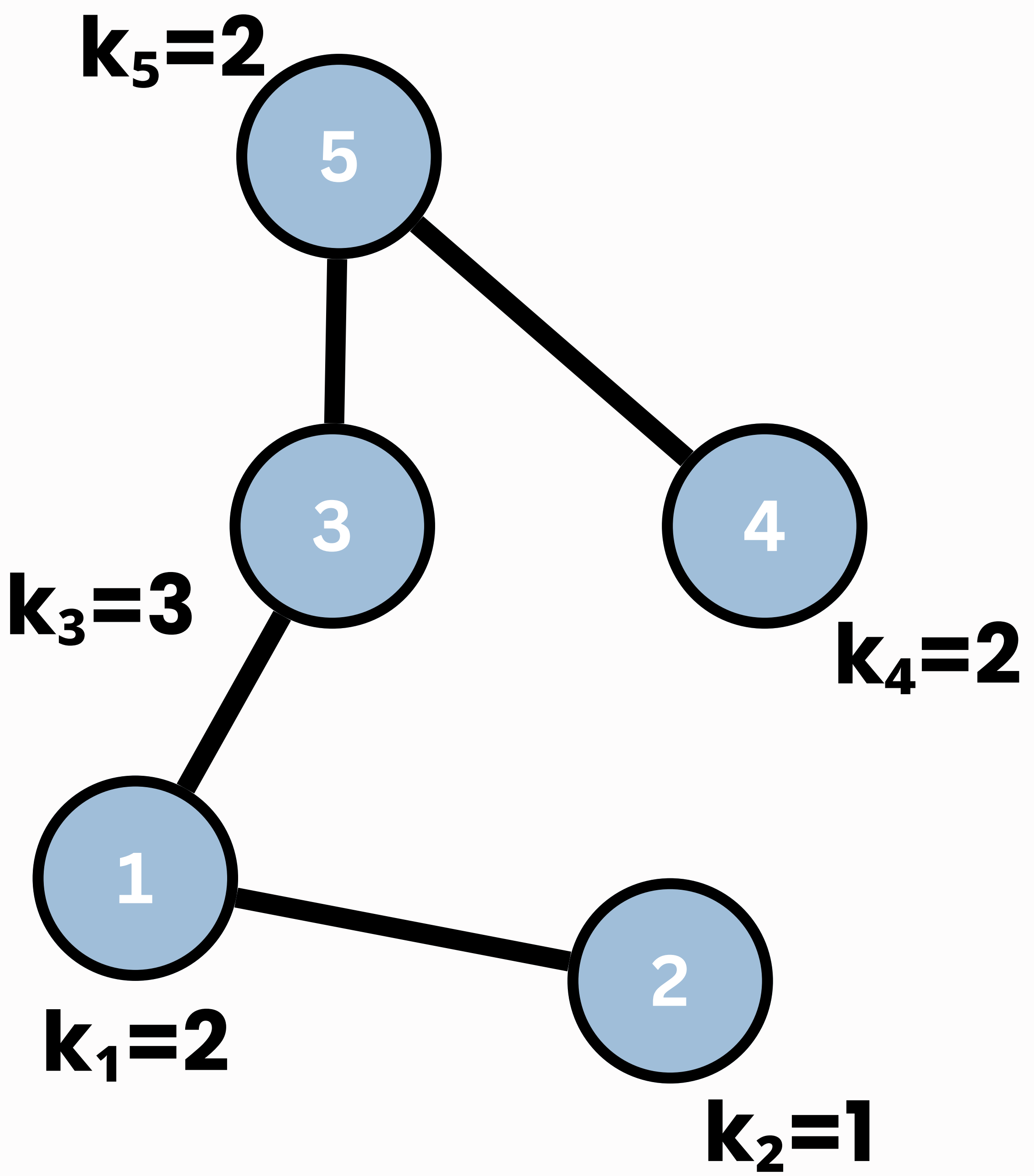
Degree and Strength

We define the degree of a node as the number of links this node has.

- For directed networks
  - in-degree (number of incoming links)
  - out-degree (number of outgoing links)
- For a weighted networks
  - the degree becomes the strength, defined as the sum of all weighted links

The degree/strength k_i of node i can be computed using the adjacency matrix as

$$k_i = \sum_j A_{ij}$$



Degree Distribution

In general two similar networks will have a different degree sequence

- the degree sequence is the list of all degrees
- it is often too strong as a property
- instead we look at the degree distribution

The degree distribution $P(k)$

- is the probability distribution of the degrees
- it is the histogram of all the degrees
- gives the probability of selecting a node at random and observing a degree k

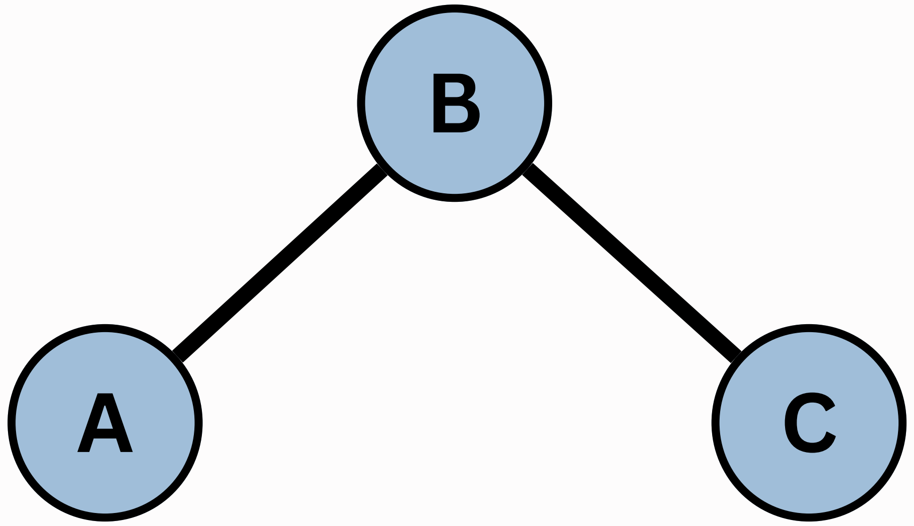
Triangles in Networks

In social networks often "friends of friends are friends"

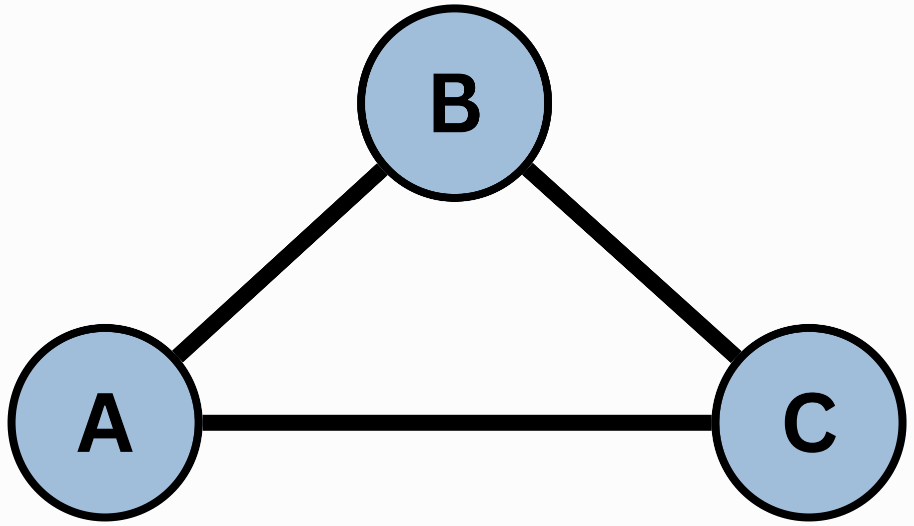
- technically speaking we say that in social networks there are many closed triangles
- if B is friend with both A and C, it is likely that also A and C are friends (or will become friends in future)

Triangles are related to the tendency to form communities and tightly connected groups

- this property is completely absent in random networks



Open Triangle



Closed Triangle

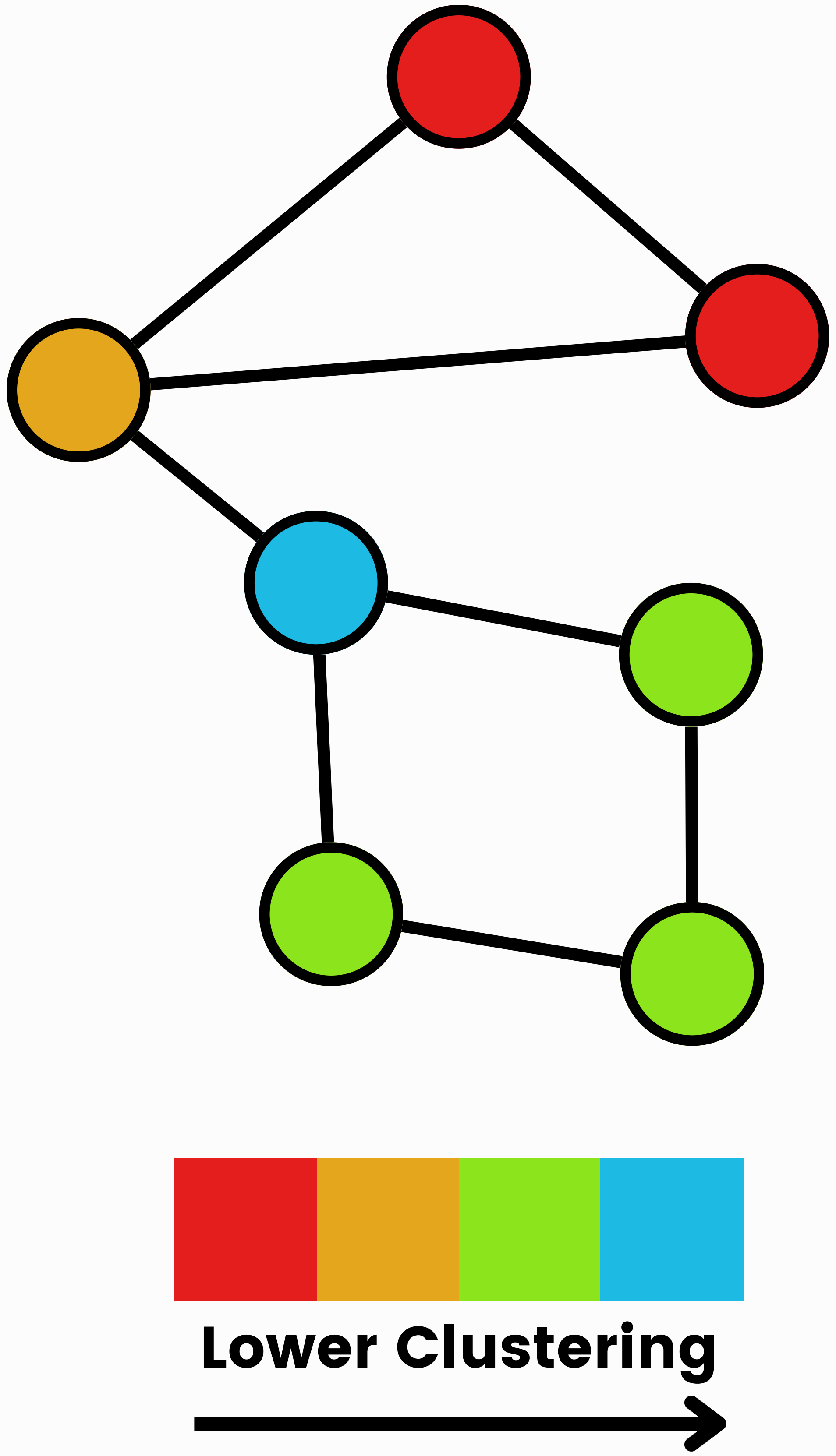
Local and Global Clustering

The tendency of a network to form triangles is measured using the clustering coefficients

- **Local Clustering Coefficient:** For a single node, it is the ratio between the number of existing and potential triangles

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

- **Global and Average Clustering Coefficient:** The first is the average of the local clustering coefficients of all nodes. The second is defined as the ratio between all existing and all potential triangles.



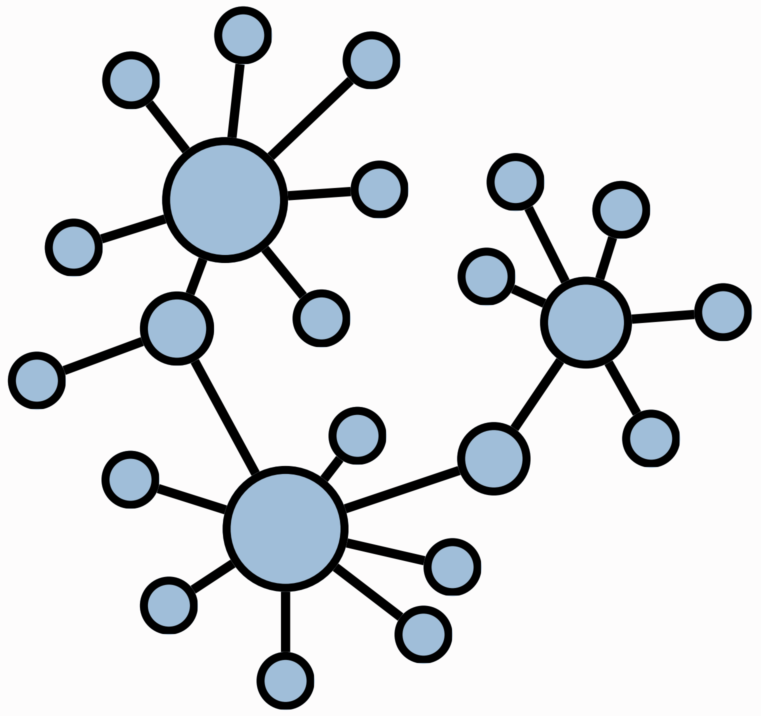
Assortativity

In many networks nodes tend to connect to other similar nodes. This tendency is called assortativity or homophily

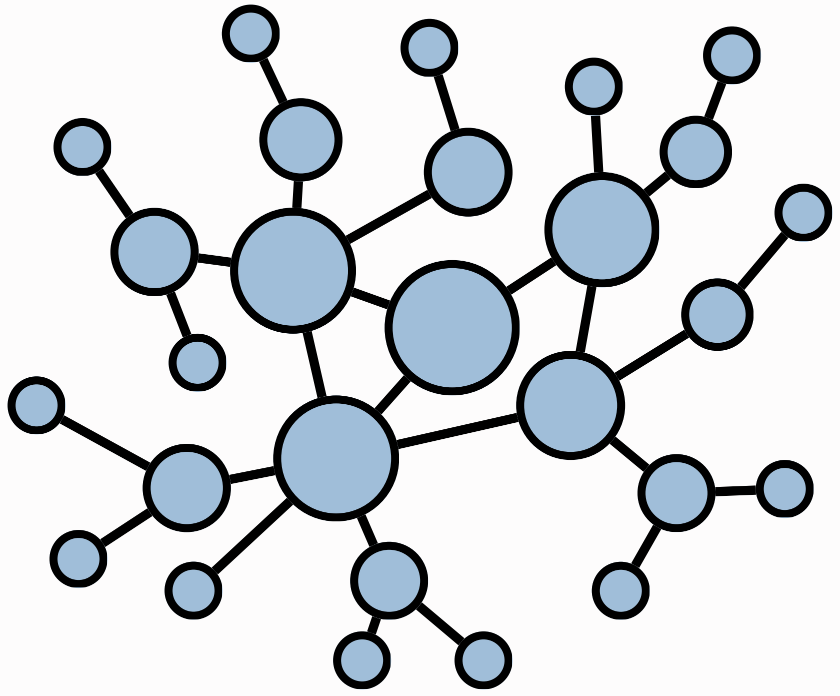
- it can involve external features such as gender or ethnicity
- it can involve network properties such as the degree of nodes

Social networks are generally characterized by a degree assortativity

- we tend to connect to people with a similar degree (level of popularity)



Disassortative



Assortative

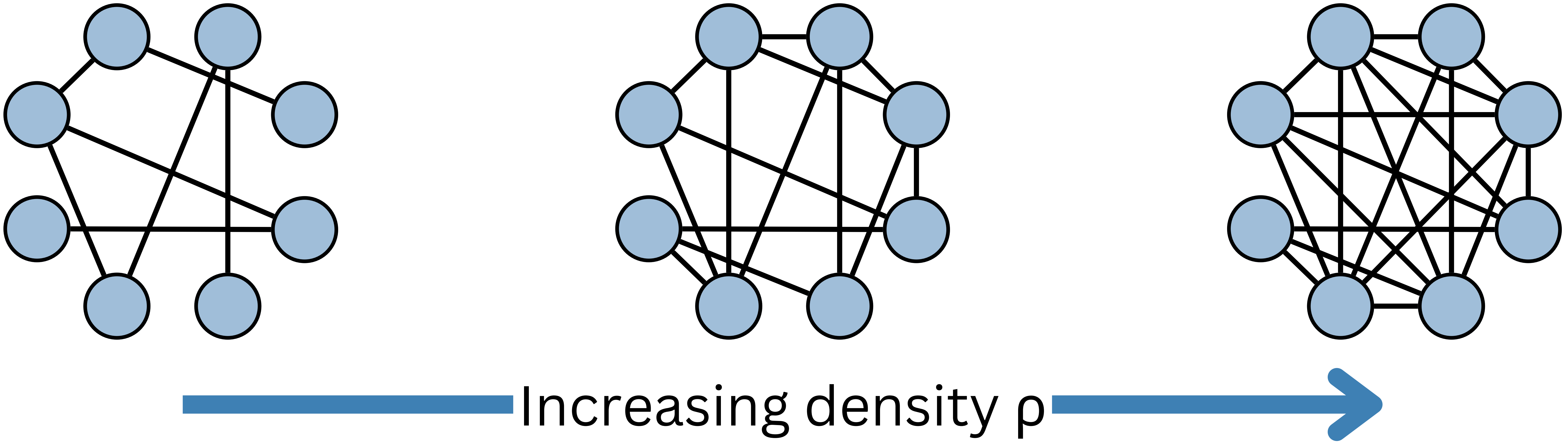
Real World Networks

Sparsity

Most real world networks are **sparse.** This means that they have a low density ρ

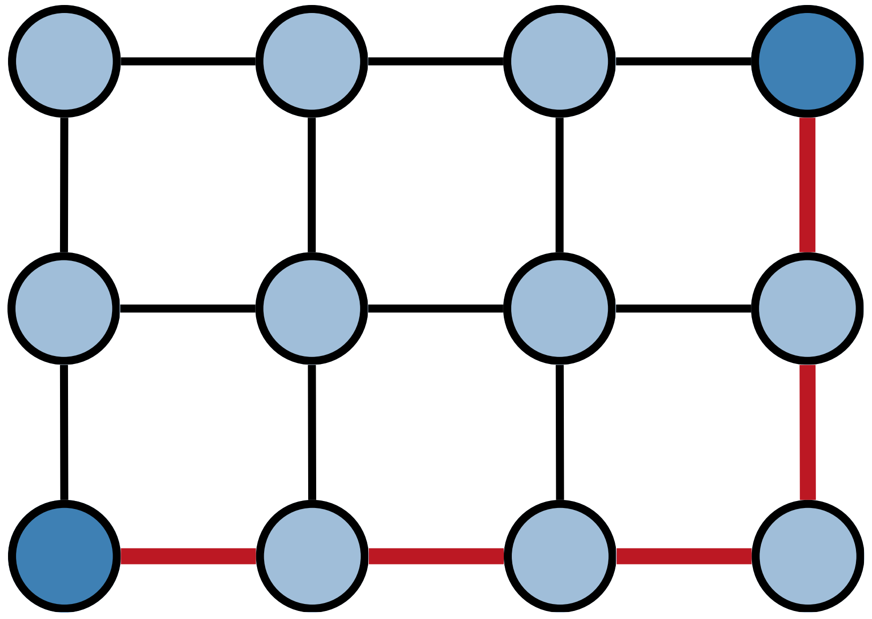
$$\rho = \frac{\text{N. of edges}}{\text{max N. of edges}} = \frac{E}{\frac{1}{2}N(N-1)}$$

Sparse Network **Dense Network**



The Small World Property

Despite networks can be huge, often the path connecting any two elements in the network can be surprisingly short. This phenomenon is often summarized by the popular notion of "six degrees of separation".

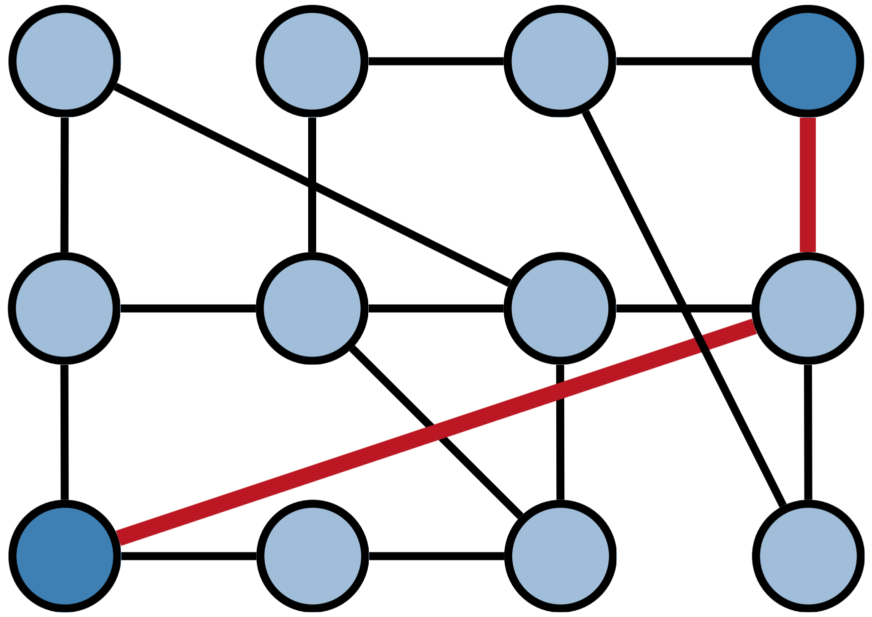


This property is mathematically expressed in terms of the average path length L and it is called small world property

$$L \sim \log(N)$$

Note that this is not true for lattices, for instance in D=2

$$L \sim \sqrt{N}$$



High Clustering

Many real world networks are characterized by two apparently opposite properties:

- high clustering C (nodes tend to form triangles)
- small world (the average path length L is small)

These two properties are apparently in conflict since

- random networks have small L but also small C (no local structure)
- regular networks can have high C, but also high L (too much local structure)

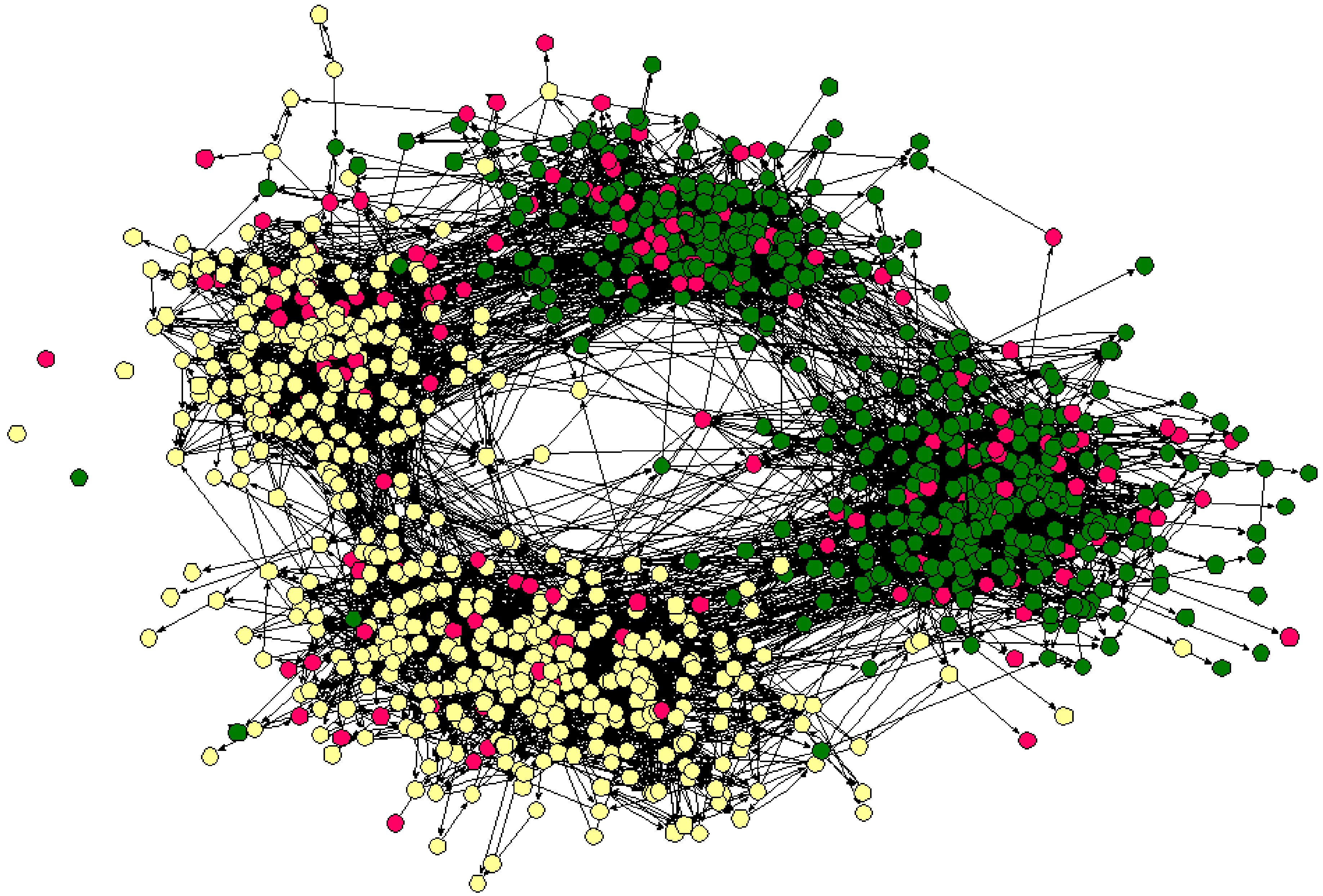
Network	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Assortativity and Homophily

Networks from the social domain tend to exhibit an high homophily based on features such as

- ethnicity
- political leaning
- gender

This leads to block structures that can favor the emergence of polarization



"Race, school integration, and friendship segregation in America," American Journal of Sociology 107, 679-716 (2001).

Scale-free Networks

Many real world networks are characterized by a power law distribution of degrees. We call such graphs scale-free networks. In a scale-free network there are many nodes with few connections, but also few nodes with an enormous number of links.

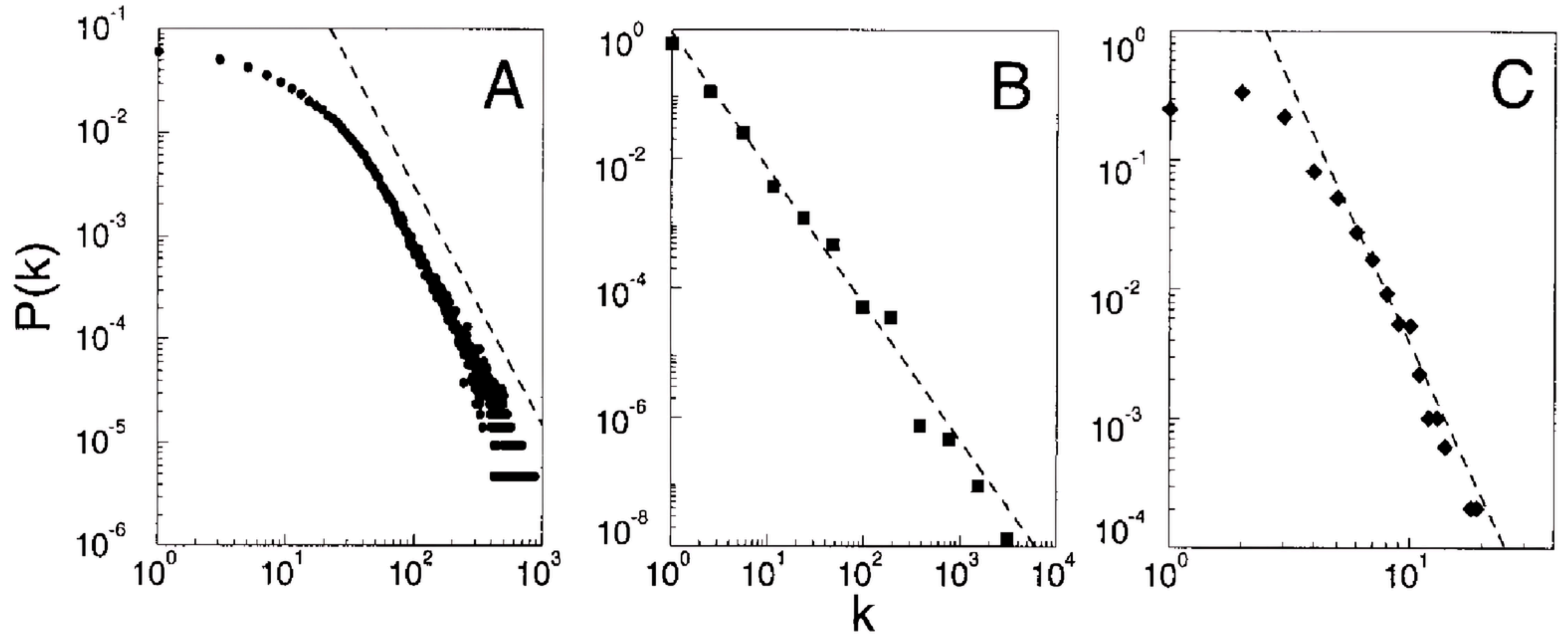


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (*6*). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

Research Questions

We saw that real world networks are characterized by a number of common properties. We want to understand what are the factors and the mechanisms that are making these features emerge. In particular:

- how can we have both high clustering and the small world property?
- which are the generative mechanisms producing these features?
- how can we get a scale-free network from a simple model?
- can we get realistic networks only using local mechanisms?

The Value of Networks

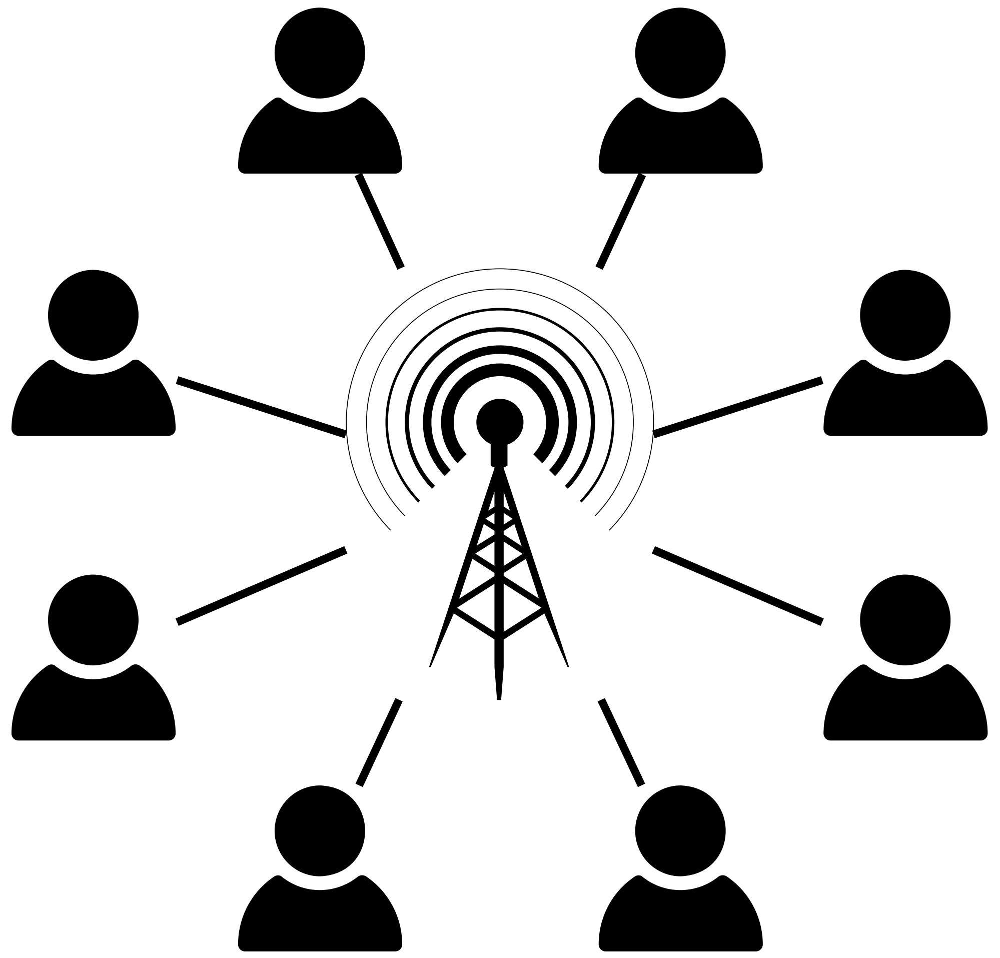
What is the Value of a Network?

Our society is pervaded by networks and many of them are the backbone of billion dollar companies

- broadcasting networks
  - Fox Corporation 18B
  - RTL Group 5B
- telecommunication networks
  - T-Mobile US 263B
  - China Mobile 205B
- social networks
  - Meta Platforms 1238B
  - Tencent Holdings Limited 506B

What makes these companies so valuable?

Sarnoff's Law



A broadcasting network is like a star

- there is a central broadcaster
- all the users are linked to this central node

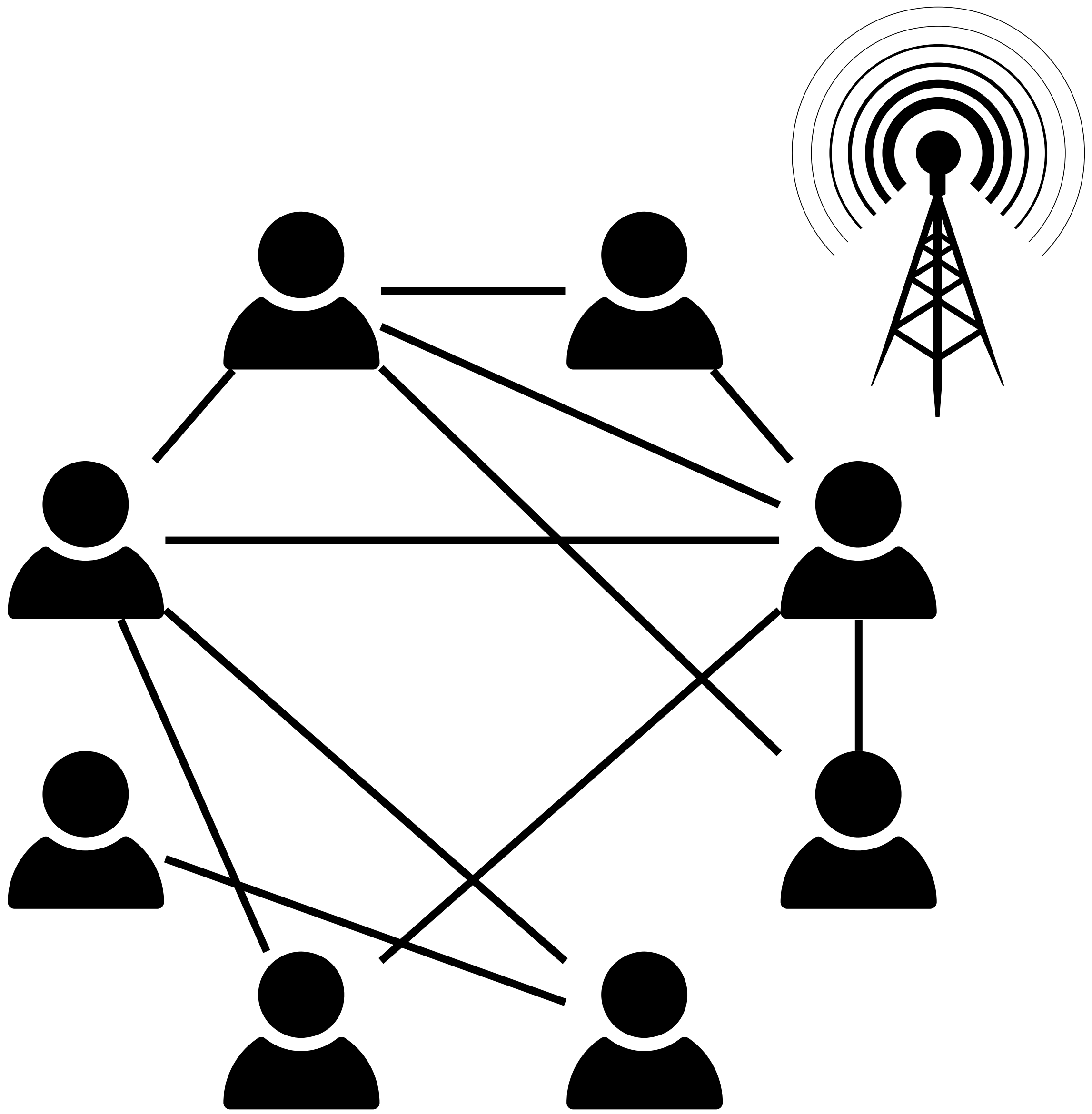
In this situation revenue R is proportional to the number of users N

$$R \propto N$$

- the same adds/products are provided to all users
- each user can buy a single subscription

This is known as **Sarnoff's law**

Metcalfe's Law



In a telecommunication network

- users call/communicate with each other
- profit is made for each call

In this situation revenue R is proportional to the number of links, that grow with the square of the users

$$R \propto N^2$$

- this is an approximation because the networks can be sparse

This is known as **Metcalfe's law**

How Social Networks Make Money

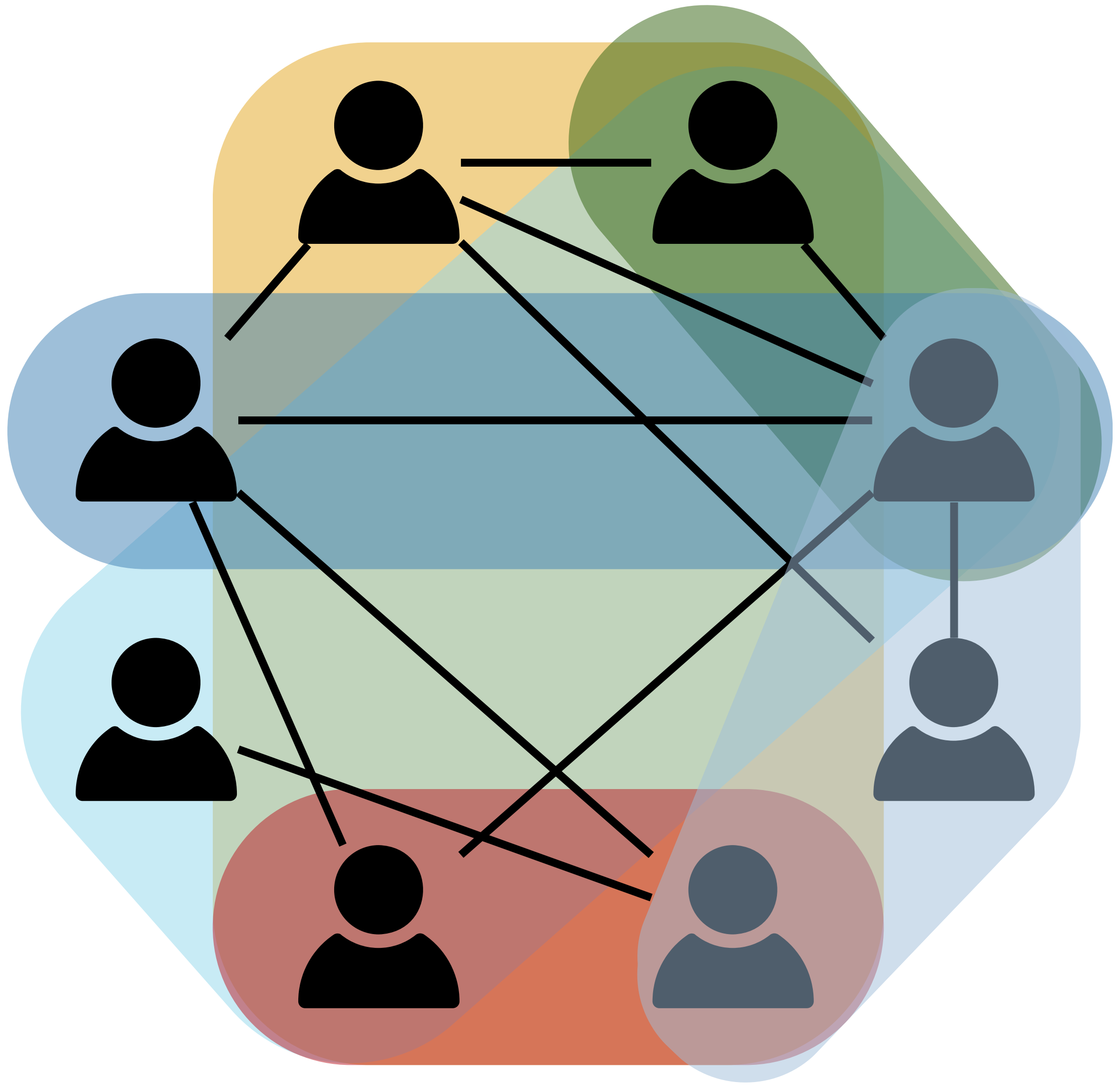
Social networks mainly profit from **Targeted Advertising**

- social networks collect detailed user data
- these include interests, behaviors, and connections
- in this way platforms can deliver highly personalized ads
- companies pay social networks for placing these adds

This is completely different from the previous business models, revenue can scale much faster!

https://adsmanager.facebook.com/adsmanager

Reed’s and Nivi’s Laws



Social networks profit from Targeted Advertising

- each user can be assigned to several advertising campaigns
- the number of combinations grow exponentially with the number of users

$$R \propto e^{cN}$$

This is know as Reed’s law, but also slower growth has been hypothesized (Nivi’s law)

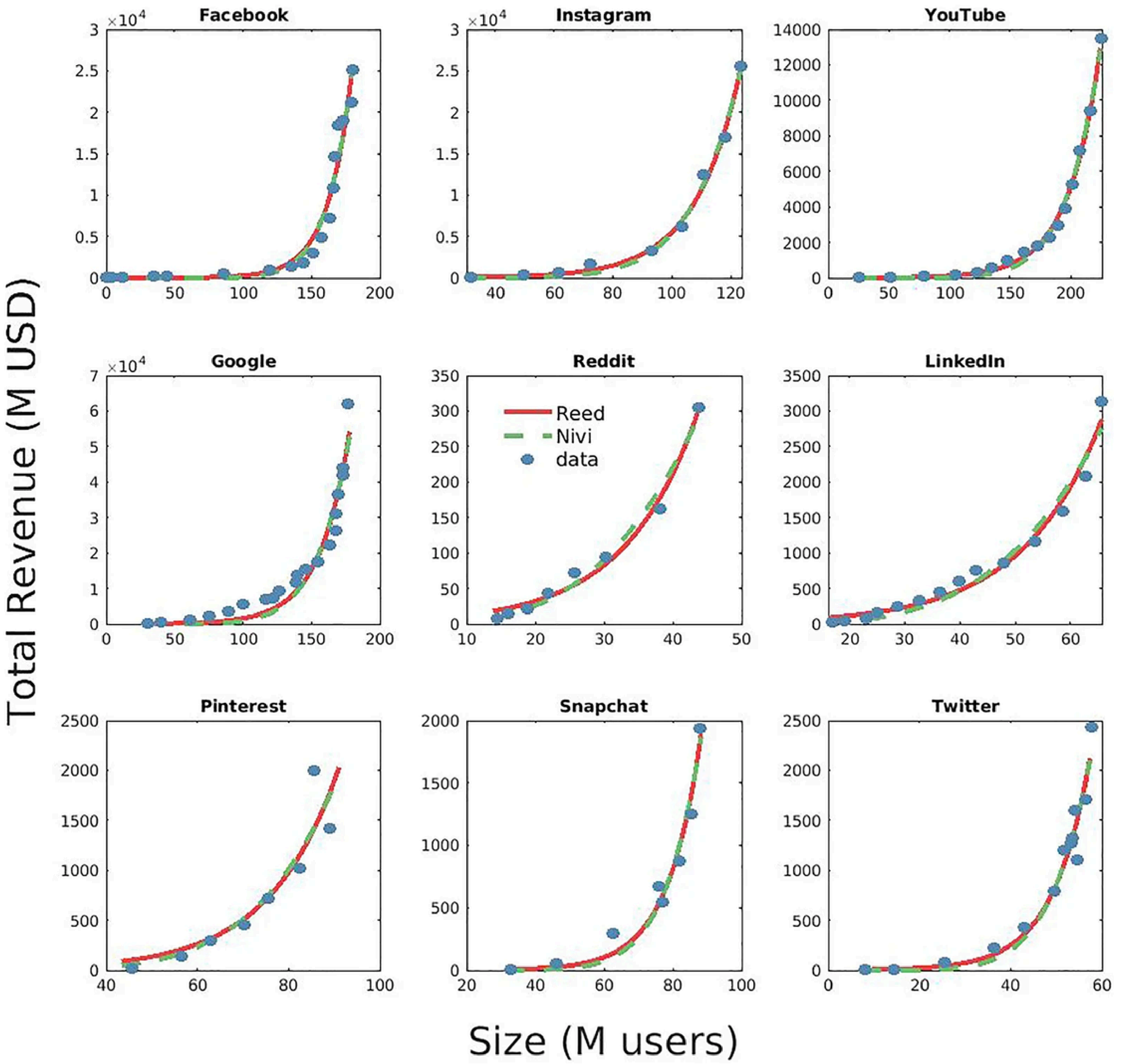
$$R \propto N^{\gamma}$$

Testing the Laws

The growth of social networks revenue shown an explosive trend

- both Reed' and Nivi's law well fit the data
- the growth is much faster than a simple square

This is observed in most social networks across different sizes

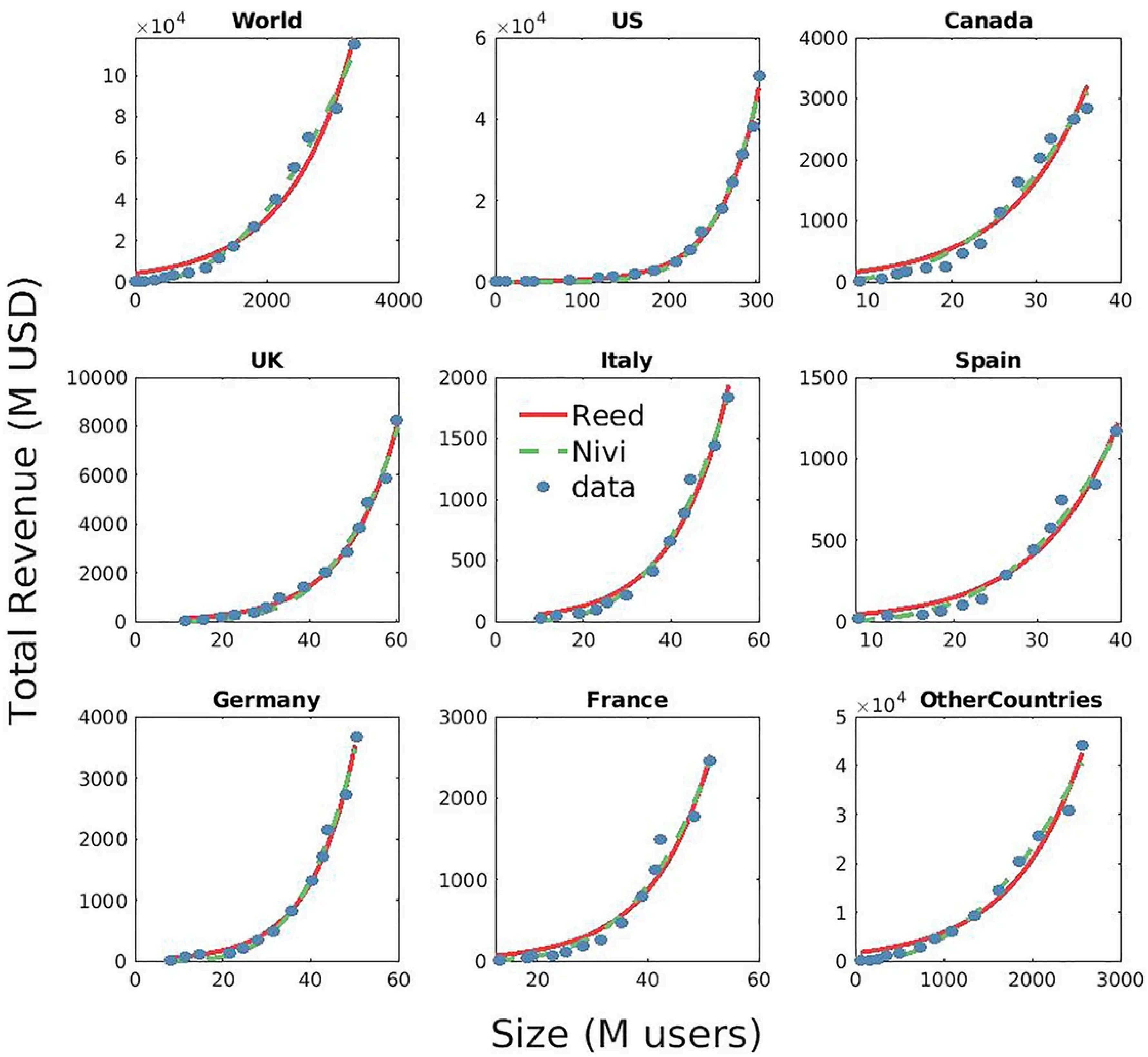


Testing the Laws

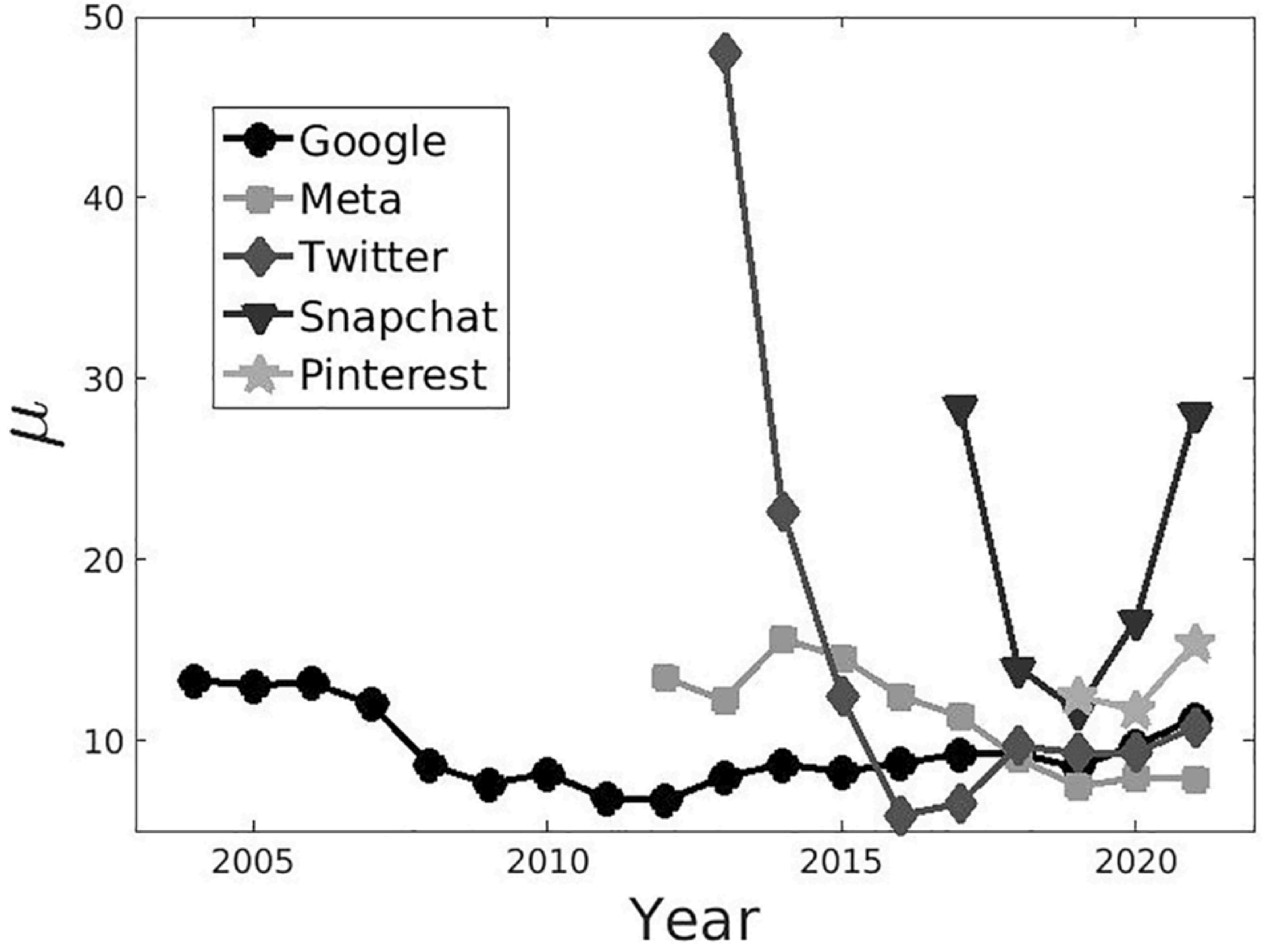
The growth of social networks revenue shown an explosive trend

- both Reed’ and Nivi’s law well fit the data
- the growth is much faster than a simple square

This is also observed analyzing a single social network in different countries



Economic Value vs Financial Value



The revenues of social networks also determines their financial value (market capitalization)

- μ is the ratio between the market cap and the revenue
- it oscillates around a value of 10 for many social networks

Conclusions

Networks Basics

We introduced the basics concepts of network science and the discussed the different representations of networks

Measuring Networks

Networks can be characterized in terms of graph invariant properties such as diameter, clustering, degree distribution, assortativity

Real World Networks

Many real world networks are characterized by similar properties: small world, high clustering, scale free degree distribution, homophily, sparsity

The Value of Networks

Social networks revenue increase very fast with the number of users, much faster than a square. This is thanks to the targeted advertisement

Quiz

- What type of network is a Reddit discussion?
- Can you provide examples of weighted, directed and undirected networks from your everyday experience?
- Which are the mechanism that lead to high clustering in social networks?
- Which law is Bitcoin network expected to follow and why?
- How much do you think social network earns selling your data?
- From your personal experience, which social network does targeted advertisement the best?