uc3m | Universidad Carlos III de Madrid

Network Theory for the Social Sciences in Python*

Methods Workshop: Social Sciences PhD Program (2025/2026)

1. Introduction to Networks and Python

https://github.com/blas-ko/uc3m_networks_workshop_2025

Blas Kolic

blas.kolic@uc3m.es

**Inspired from the Social Network Analysis of the uc3m masters on Computational Social Science, by Esteban Moro and Blas Kolic*

Welcome to the networks workshop!

I'm **Blas Kolic,** a postdoc in Big Data in IBiDat (in this same building!).

I studied Physics (BsC) and Maths (PhD) with a focus on computational social science.

The main repository for this workshop is at:

https://github.com/blas-ko/uc3m_networks_workshop_2025

You can find the course syllabus at:

https://docs.google.com/document/d/1tbUFCsfmJMqQ1RelFVrBqFgI3F6eA9p0Fdt3vsGC9LE/edit?usp=sharing

Overview of the workshop

Sessions will be on **Wednesdays** from **15:00 to 18:00** in room **18.0.A04** or at Google Meet.

Sessions will be divided into a theoretical part (1.5h) and a practical part (1.5h). The practical part will be given in **Python** through **Jupyter Notebooks.**

Plan:

1. **19/Nov/25:** Introduction to Networks and Python
2. **26/Nov/25:** Network Models and Structural Analysis
3. **03/Dec/25:** Link Prediction and Network Inference
4. **10/Dec/25:** Dynamics on Networks

uc3m

Summary for today

- Part 1:
  - Why networks?
  - Mathematical definition of networks
  - Examples of networks
  - Types of networks
  - Representing networks

- Part2:
  - Key ideas, concepts and measures:
    - Degree
    - Degree distribution – Heavy tailed distribution
    - Degree correlation – Assortative & Dissortative
    - Clustering
    - Distance & Diameter – Small-world phenomenon
    - Connectedness

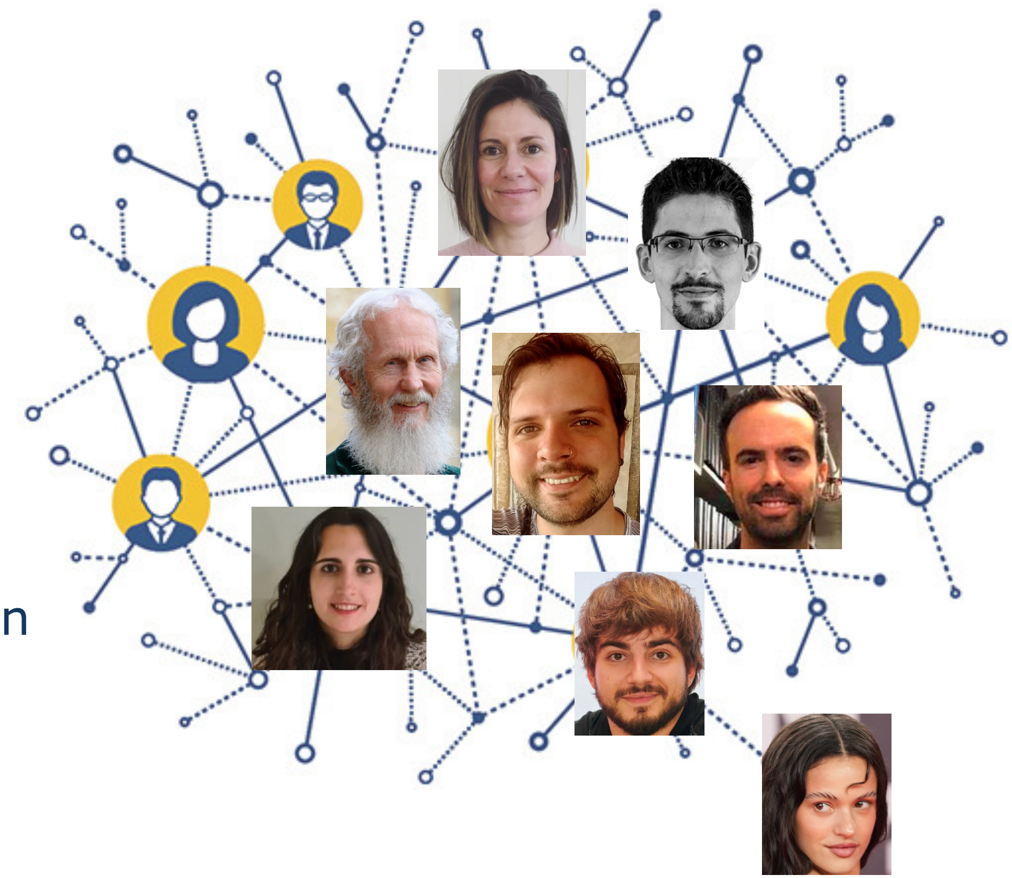
1 | Why networks?

uc3m

Why networks?

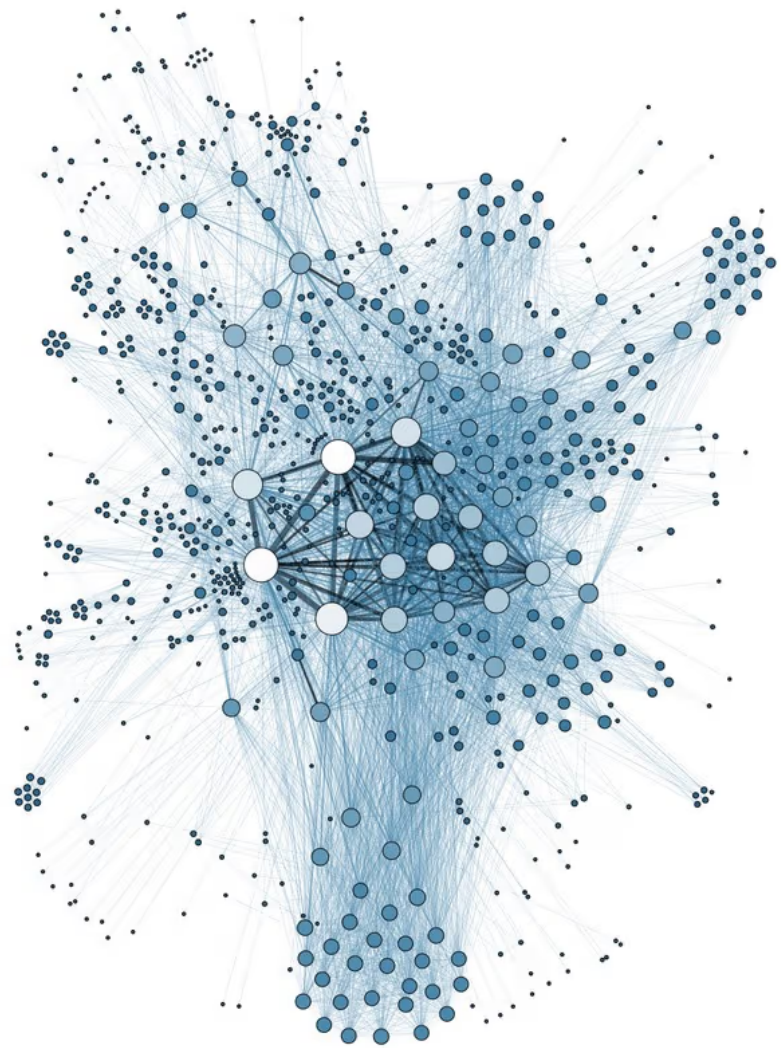
Individuals are interconnected

- A network is a framework to understand interconnected entities like people, institutions, events, places, etc.
- Thus, it makes sense to study our society globally from the perspective of a "network".
  - Example: instead of analyzing individual data on Facebook, we can study the network of friends of an individual



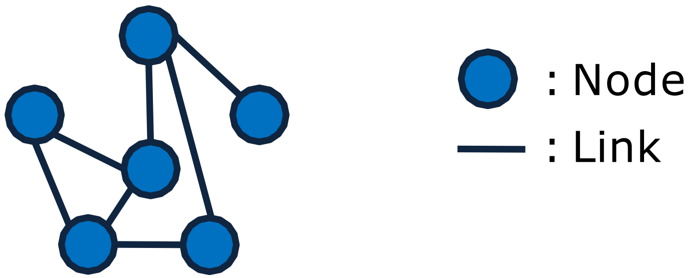
Why networks?

- Why studying those networks?
  - They allow to discover patterns that do not exist at the individual level.
  - Many of the daily-life or business process happen in networks:
    - Recommendation
    - Job searching
    - Task coordination
    - Information searching
    - Social media
    - Epidemics
- Without the understanding of the networks in which those process takes place we cannot understand those process.



Why networks?

- A simple **definition** of a network:
  - *A set of nodes connected by a set of links*



- What is a **node**?
  - Distinct objects/agents (e.g. persons, Webpages, neurons)
  - Also called **vertex**

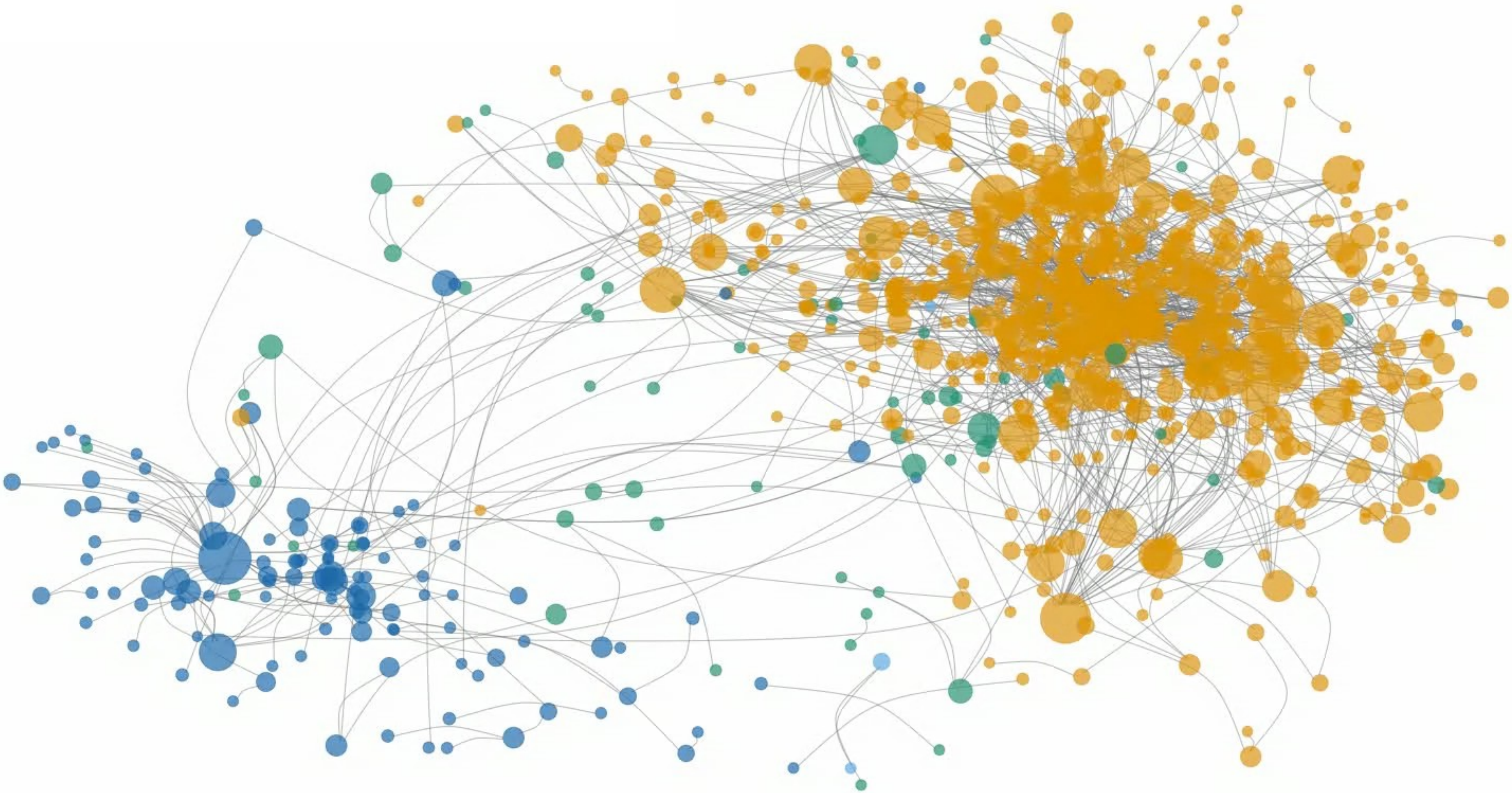
- What is a **link**?
  - Pairwise relations between two nodes (Ex. friendship, hyperlink, synaptic connection)
  - Also called **edge**

Social networks

Node: A person
Link: Friendship, collaborations, acquaintance.

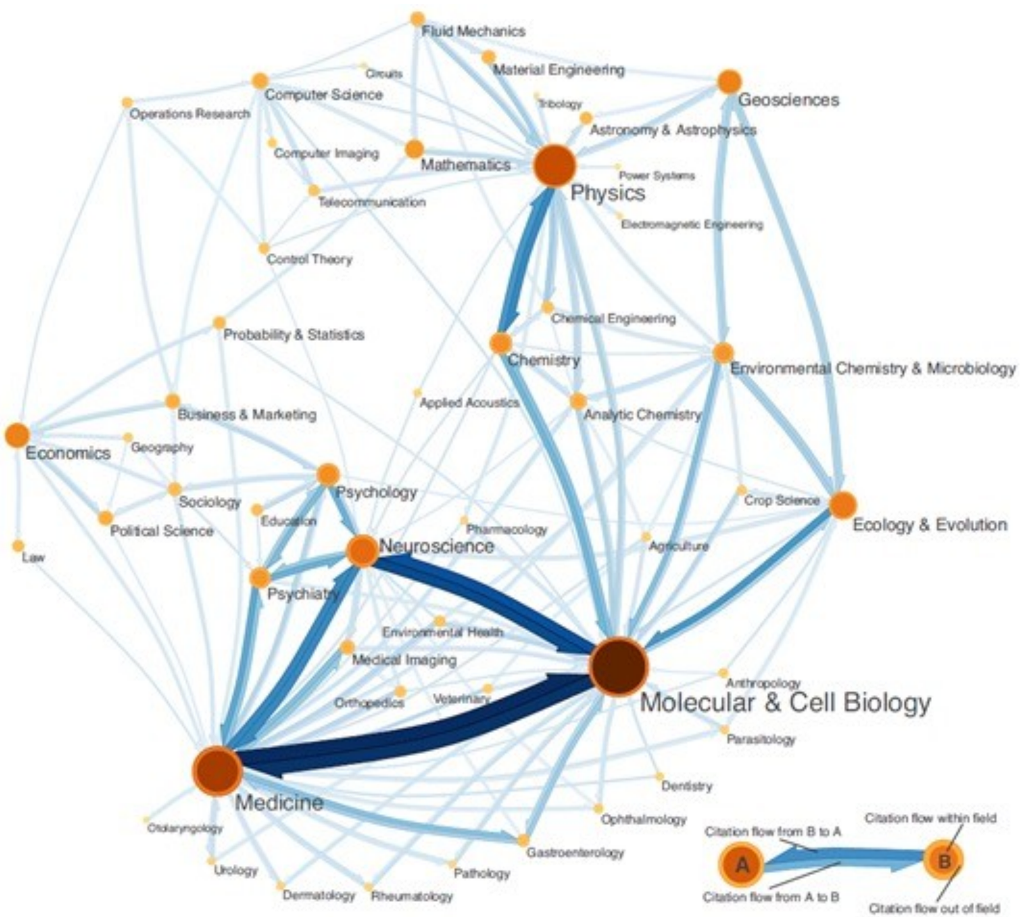
Social networks

Node: Twitter account
Link: Information Exchange, conversation (retweet, mention, etc.)

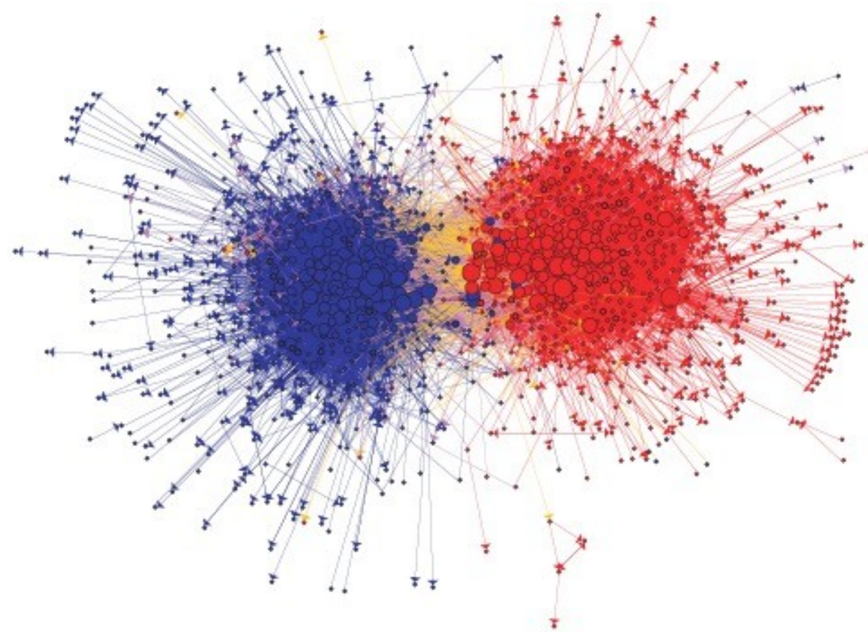


Information networks

Node: webpage, articles, books, etc.
Link: hyperlink, citation, etc



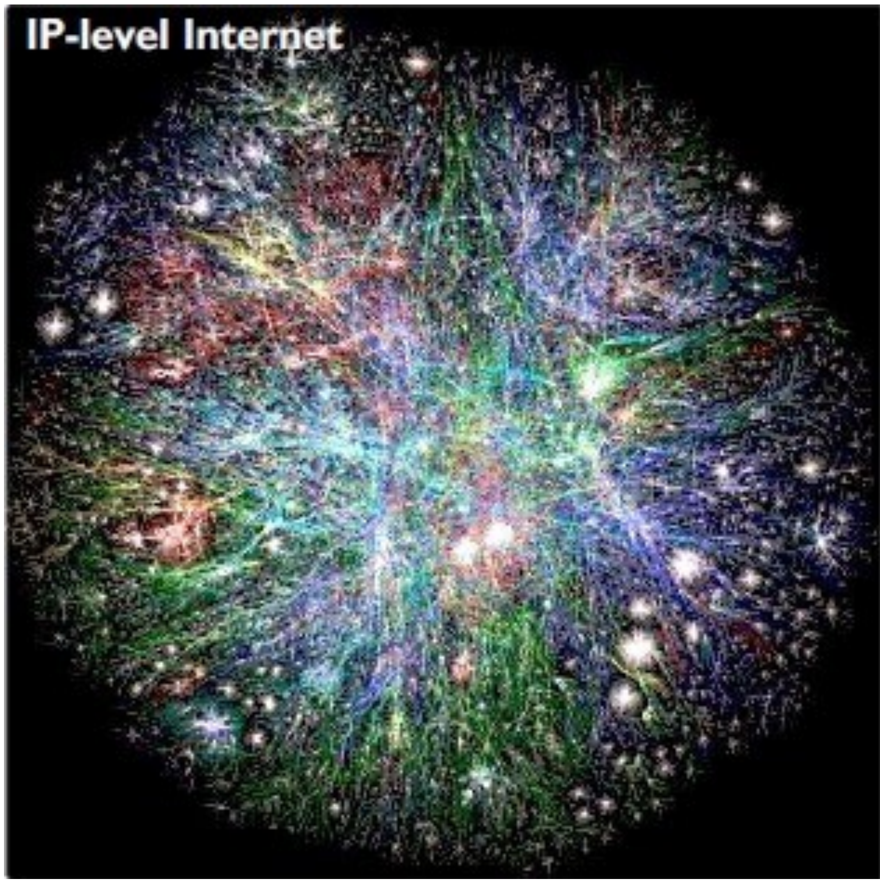
Citation network - M. Rosvall & C. T. Bergstrom (2008)

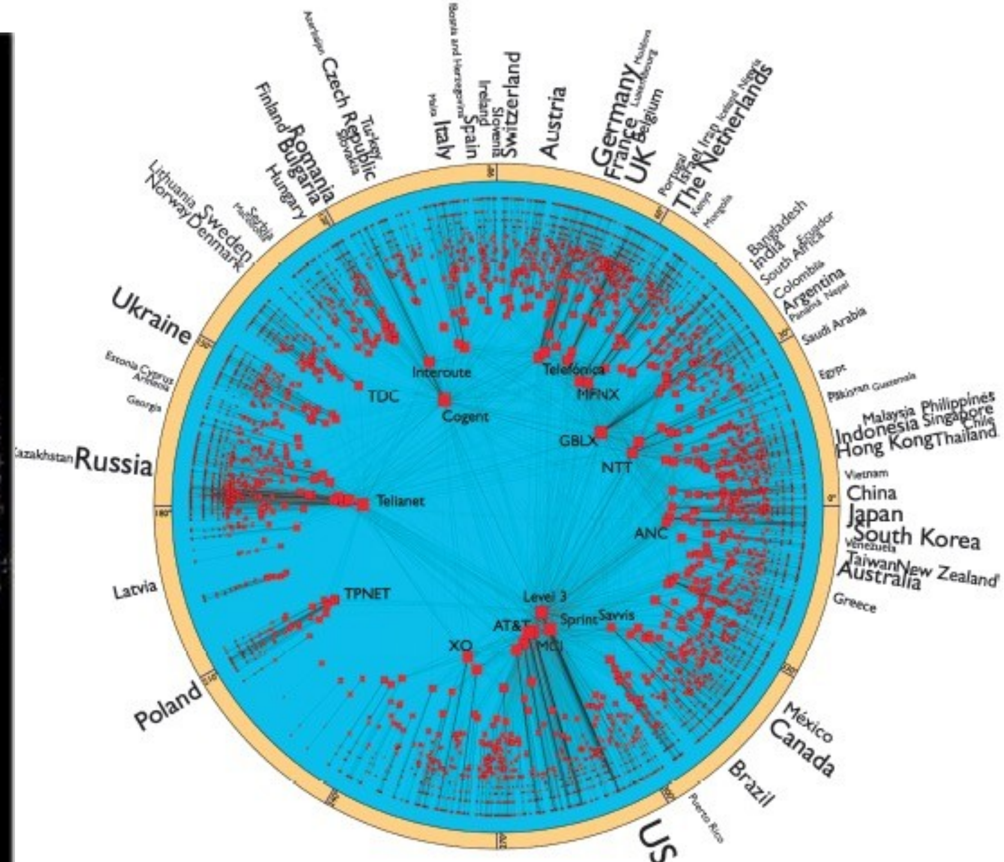


Blog network - L. Adamic & N. Glance (2005)

Communication networks

Node: Internet router, internet service provide, email address, mobile phone number
Link: Exchange of information, physical wire etc.



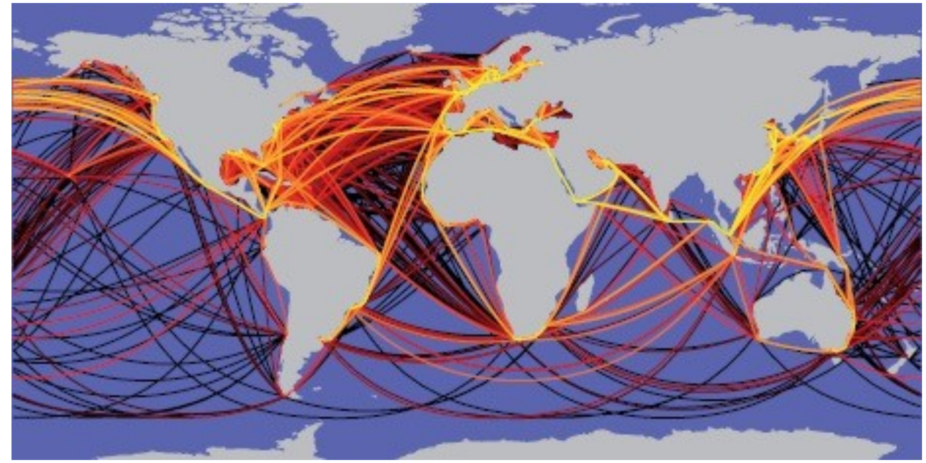


ISP network - M. Boguna, F. Papadopoulos & D. Krioukov (2010)

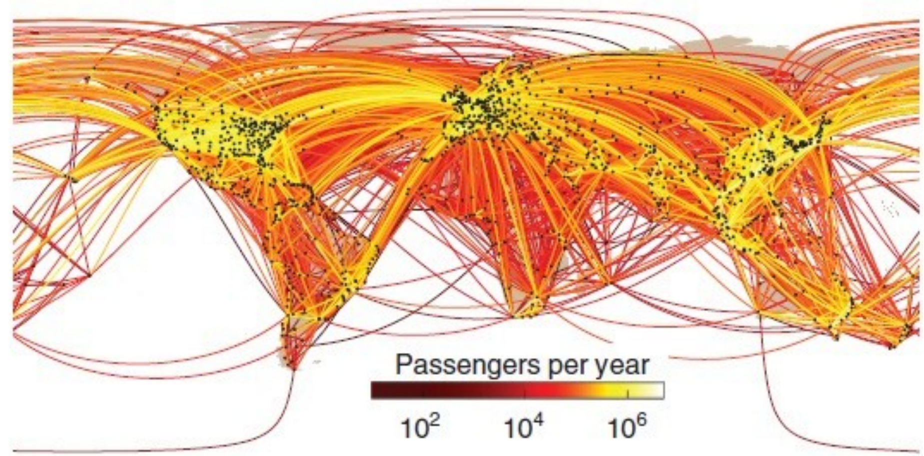
uc3m

Transportation networks

Node: Airports, cities, (railway or subway) station, ports
Link: Transportation of people or goods



Global shipping network - P. Kaluza, A. Kolzsch, M. T. Gastner & B. Blasius (2010)

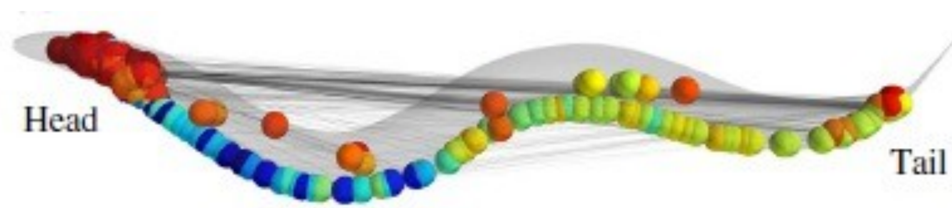


Air-transportation - D. Grady, C. Theimann, D. Brockmann (2012)

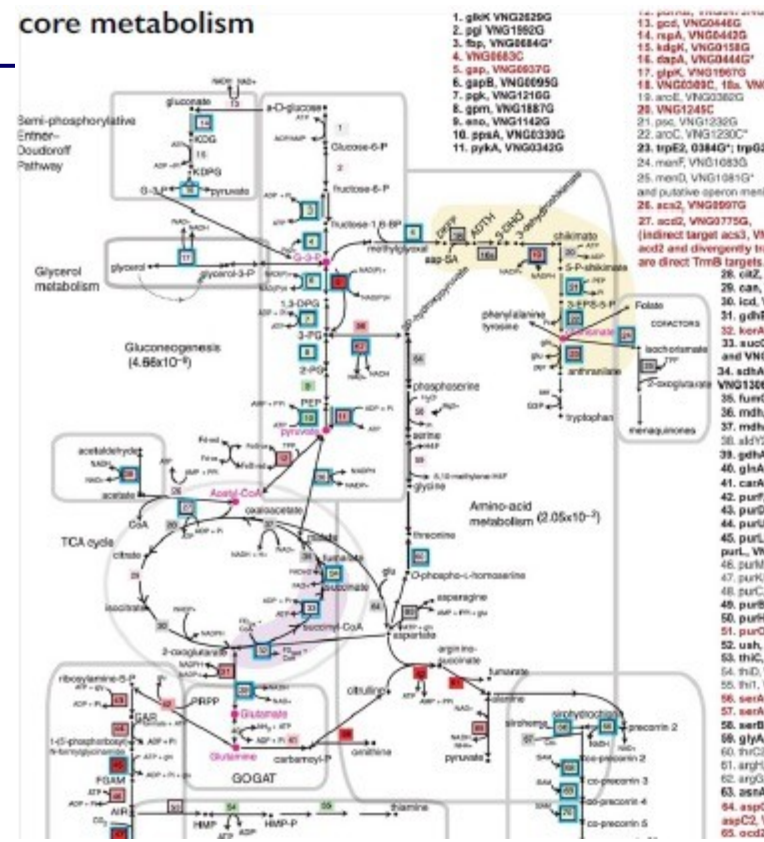
London tube

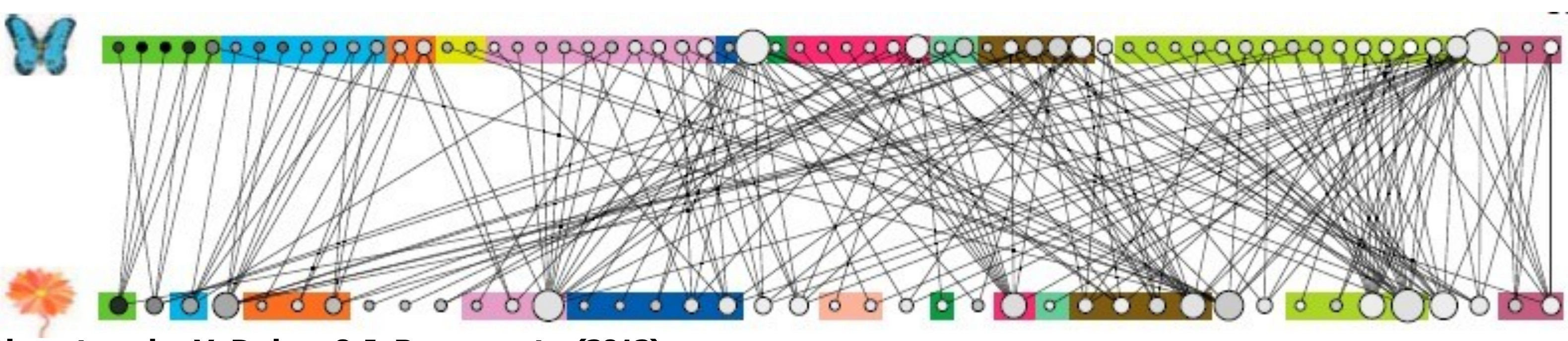
Biological networks

Node: Species, metabolites, proteins, genes, neurons
Link: Predation, mutualism, chemical reaction, binding, regulation, synaptic pulse.



Neural networks – M.E.J Newman & T. P. Peixoto (2015)

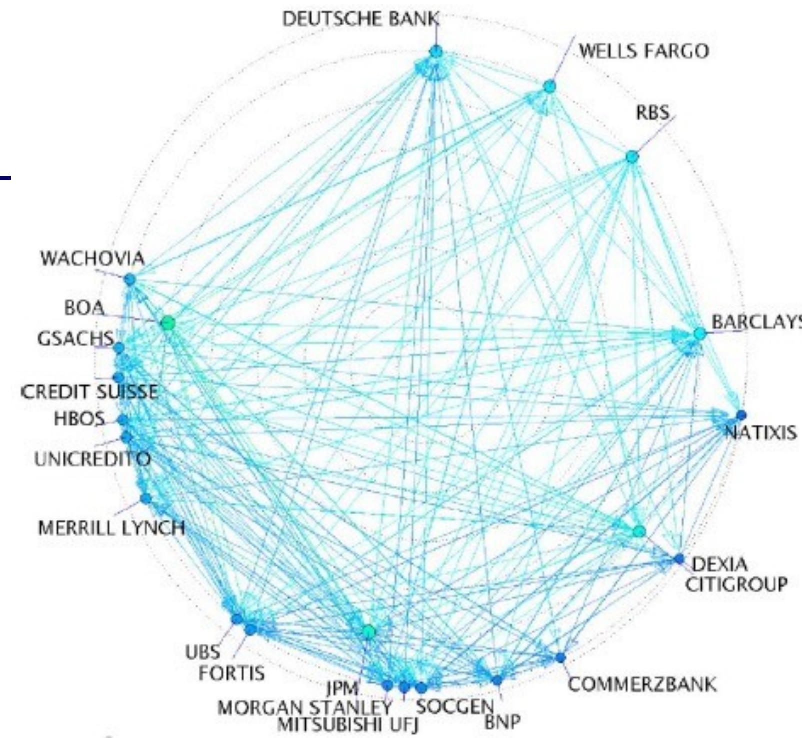




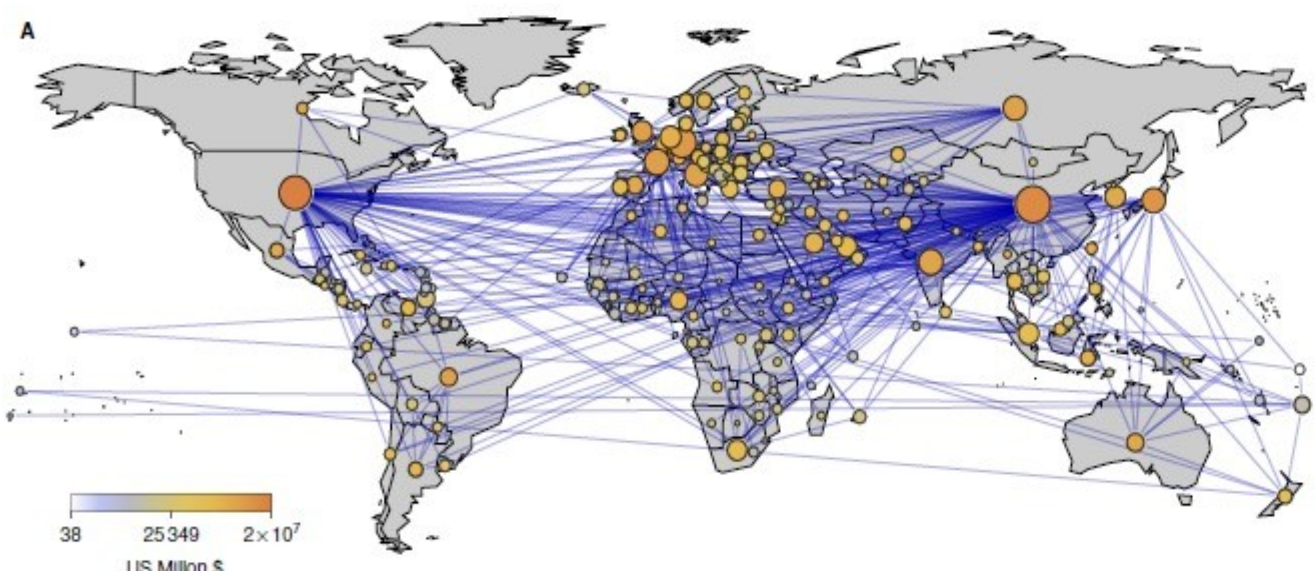
Mutualistic network - V. Dakos & J. Bascompte (2012)

Economical networks

Node: Country, bank, firm
Link: Trade, transaction, loan, control

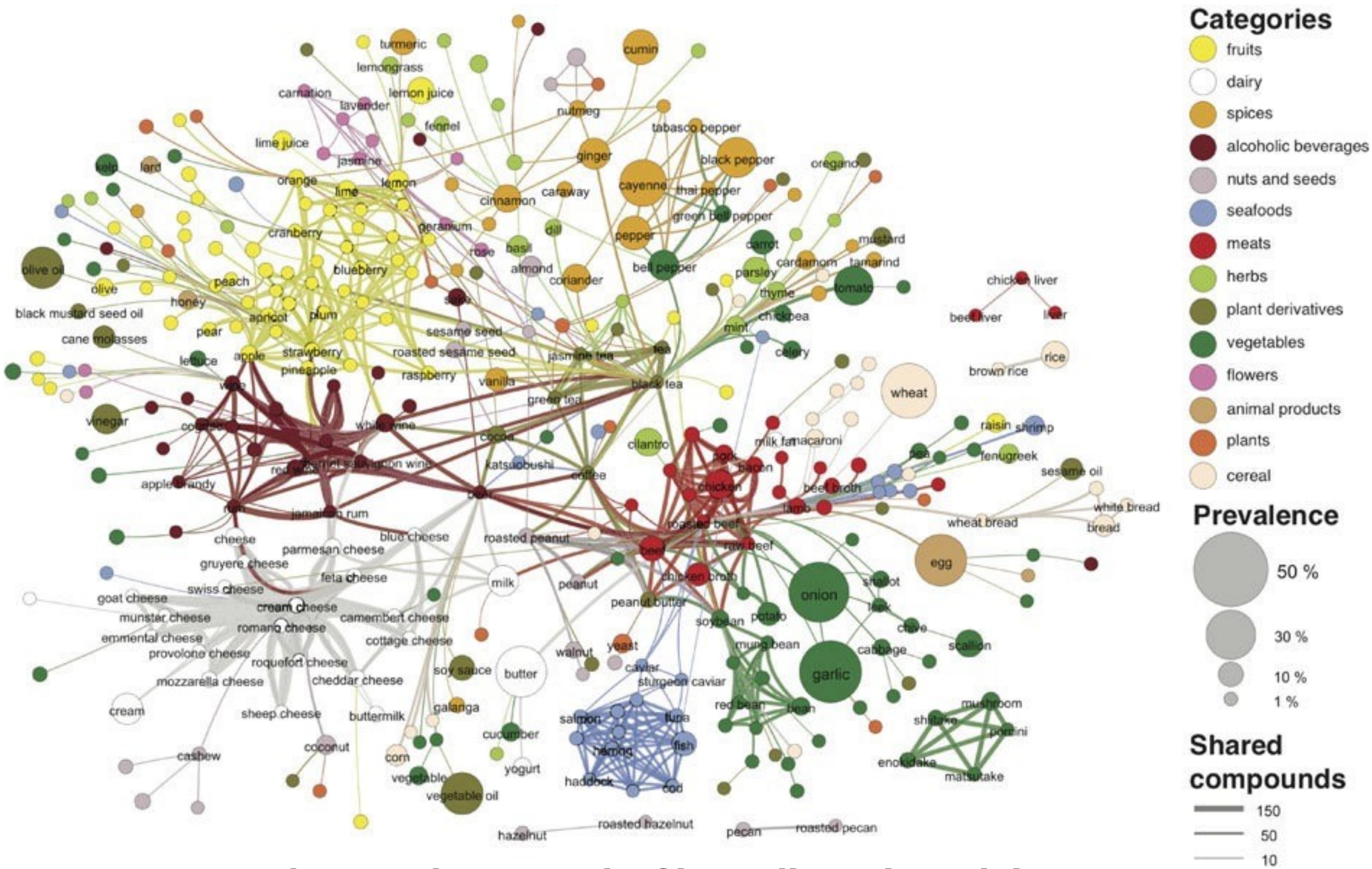


Bank networks – S. Battison et al. (2012)



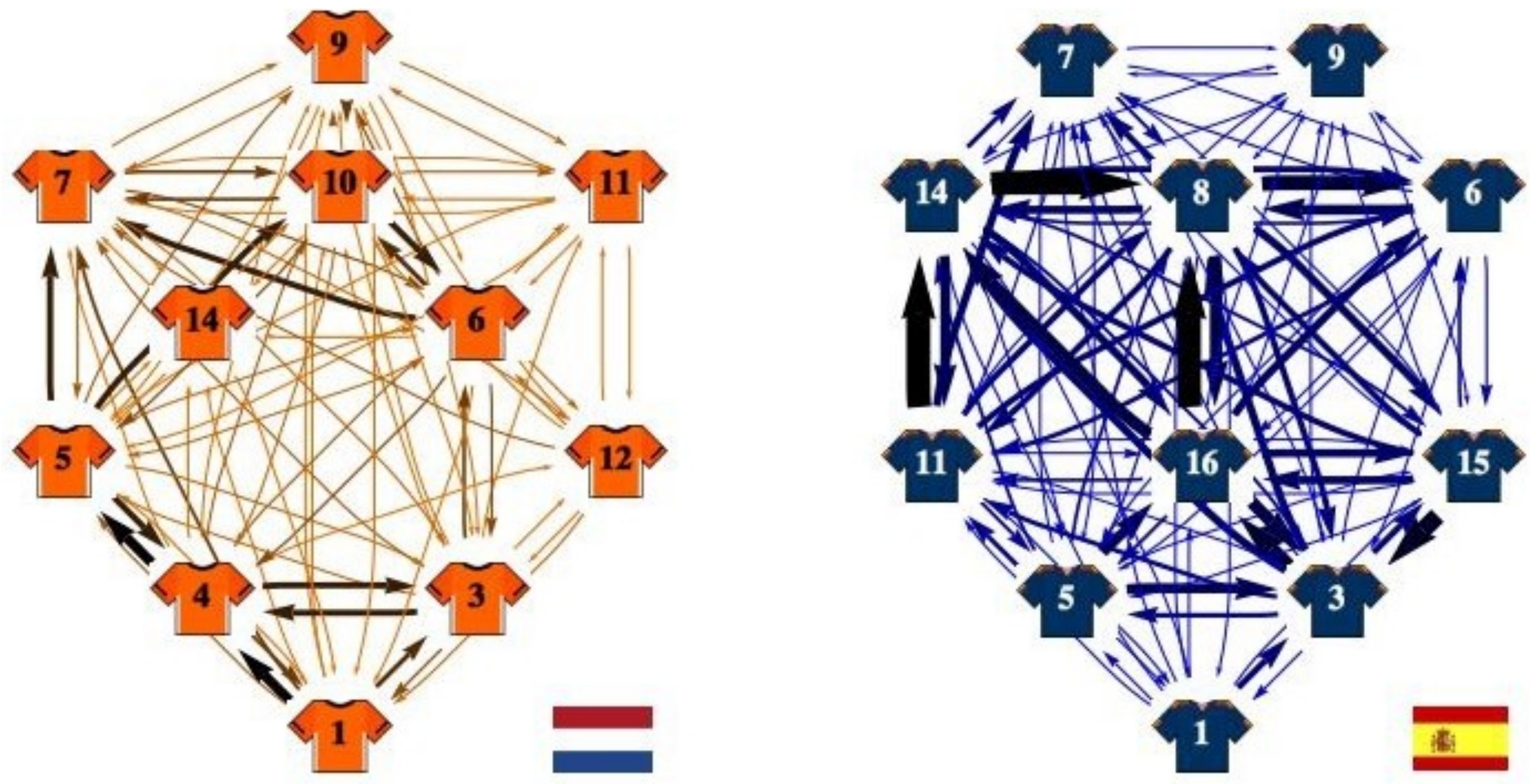
World trade web – G. Garcia-Perez, M. Boguna, A. Allard, M.A. Serrano (2015)

Other networks



Food network: Network of ingredients in recipies

Other networks



Network of passes between players in soccer

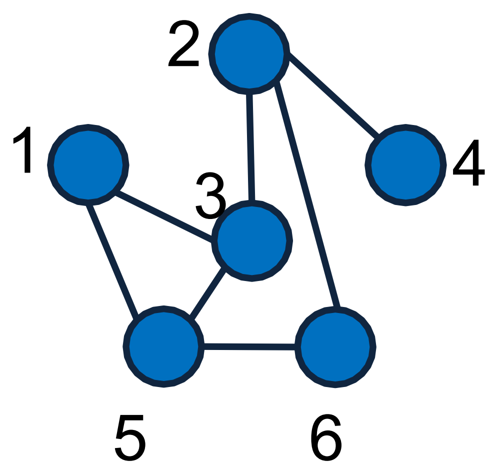
2 Defining networks

Basic definitions

A network/graph is a tuple of two sets: $G = (V, E)$

- Nodes or *vertices* (V)
- Links or *edges* (E)

- The **size** of the network is given by the *number of nodes* and the *number of links.*



$$V = \{1,2,3,4,5,6\}$$

$$E = \{(1,3), (1,5), (2,3), (2,4), (2,6), (3,5)\}$$

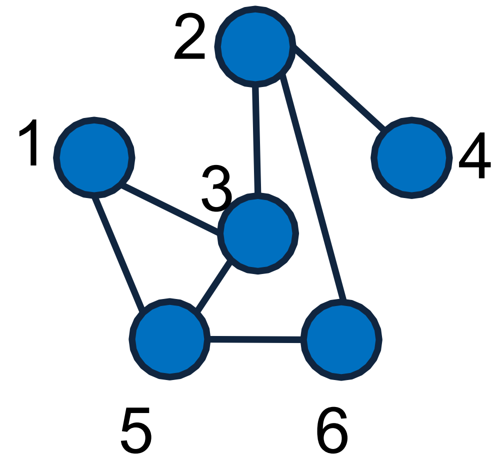
Number of nodes: $|V| = N = 6$

Number of links: $|E| = L = 7$

Basic definitions

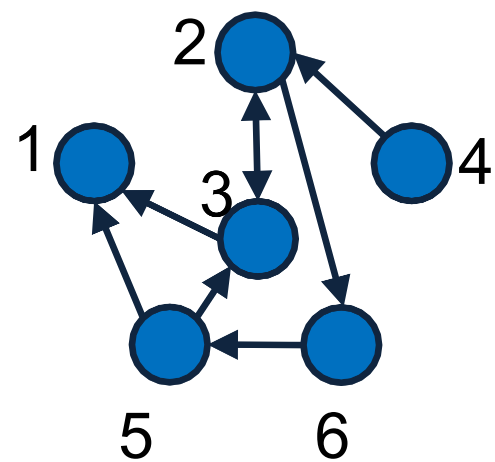
Networks are **undirected**
when all links are reciprocal

Example: Friends on Facebook



Networks are **directed**
when they are not

Example: Follows on Instagram

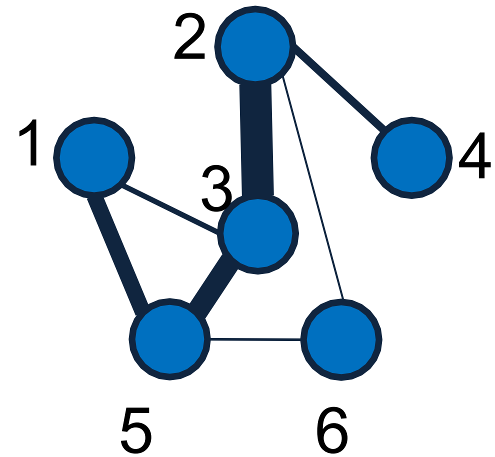


Basic definitions

Networks are **weighted**
if links have different intensities

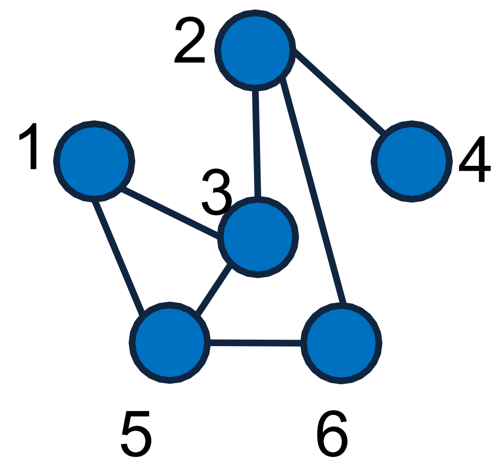
Example: Trade between countries

OBS: Here, each link has an associated *weight*



Networks are **unweighted**
if they do not

Example: Friends on Facebook



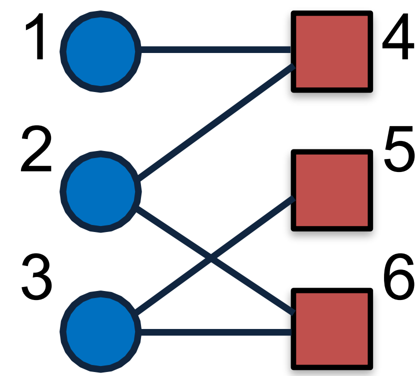
uc3m

Basic definitions

Networks can be **bipartite**
if two exclusive kind of nodes exist

Example: Movies–actors: who acted in which film

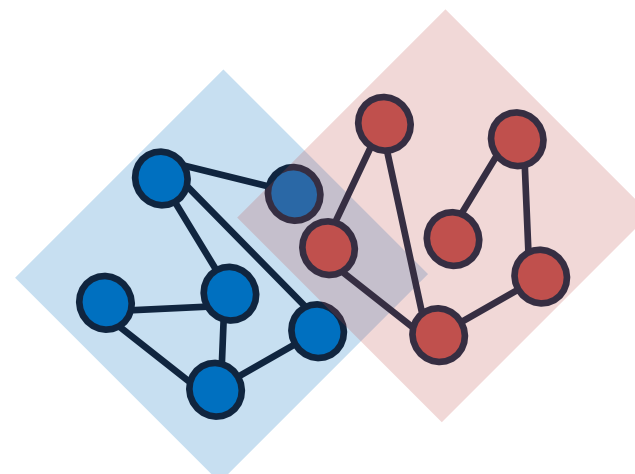
OBS: Here, links only happen between opposite kind of nodes



Networks can be **multilayered**
when connected layers of links exist

Example: City transport system (Bus + metro + bike lanes)

See: M. Kivela et al, *Multilayer networks*, J. of Complex network 2 (2014)



uc3m

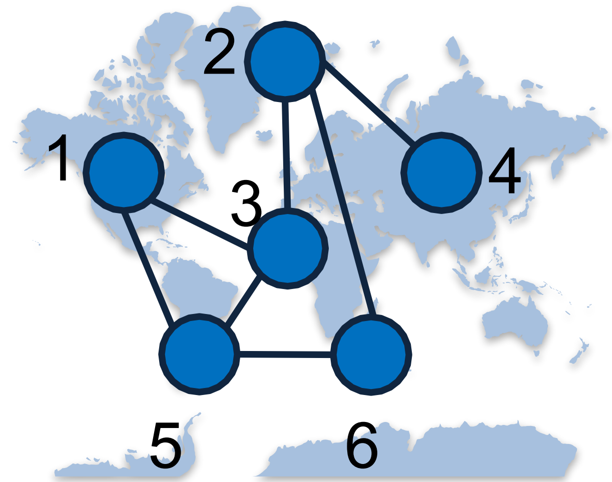
Basic definitions

Networks can be **spatial**
if nodes have a location

Example: Airports network

OBS: Here, nodes will have latitude
and longitude attributes.

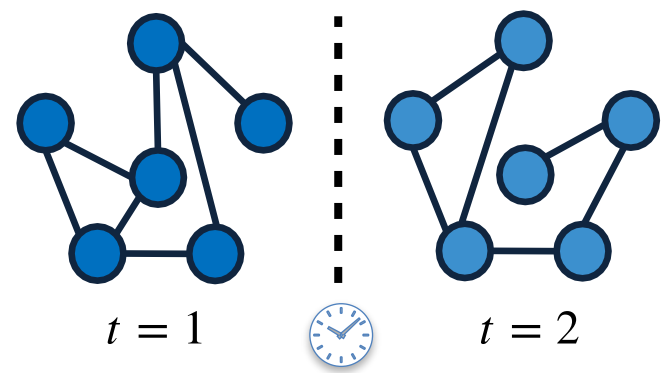
See: M. Barthelemy, *Spatial networks*, Physics Report 499 (2011) 1-101



Networks can be **temporal**
when nodes or links change over time

Example: Network of retweets

See P. Holme & J. Saramaki, *Temporal networks*, Physics Report 519 (2012) 97-125

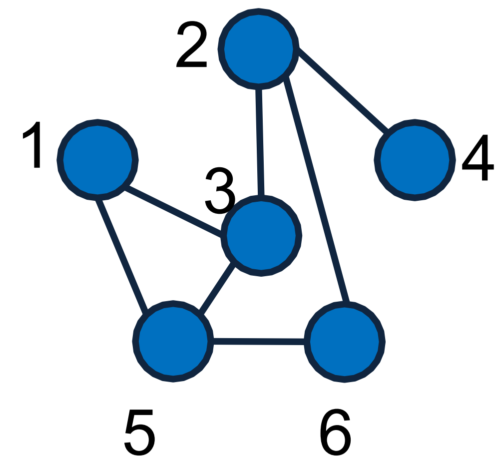


uc3m

Basic definitions

Moreover, **nodes** may have attributes too:

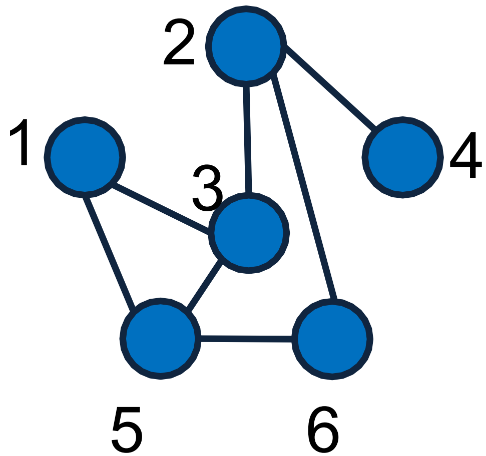
- age /gender
- name
- species
- size/color
- socioeconomic status



Id	Name	Gender	Age
1	Alice	F	23
2	Bob	M	27
3	Carla	F	31
4	David	M	22
5	Elena	F	29
6	Frank	M	35

3 Representing networks

Representing networks



$G=(V,E)$

1. The **node and edge sets:**

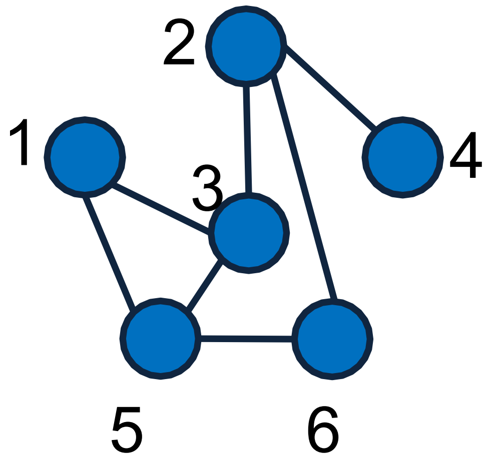
$$V=\{1,2,3,4,5,6\}$$

$$E=\{(1,3),(1,5),(2,3),(2,4),(2,6),(3,5)\}$$

Notes:

- If **undirected**, edge (1,3) is equivalent to (3,1) so no need to put both
- If **directed**, (1,3) and (3,1) are not equivalent
- If **weighted**, (1,3) → (1,3, w_{13}), where w_{13} is the weight of edge (1,3).

Representing networks



$G = (V, E)$

2. The **adjacency matrix:**

$A =$

	1	2	3	4	5	6
1	0	0	1	0	1	0
2	0	0	1	1	0	1
3	1	1	0	0	1	0
4	0	1	0	0	0	0
5	1	0	1	0	0	0
6	0	1	0	0	0	0

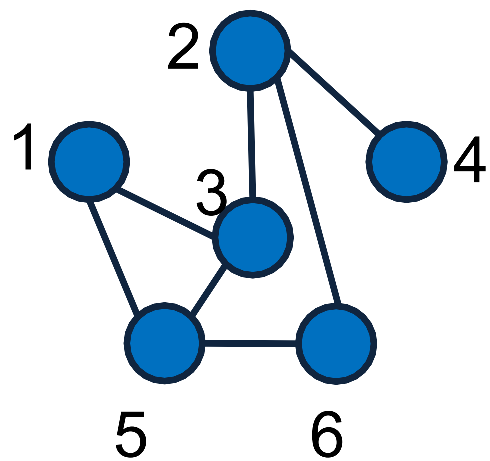
$A_{ij} = 1$ if node i is connected to node j

$A_{ij} = 0$ if not

Notes:

- If **undirected**, A is symmetrical ($A_{ij} = A_{ji}$).
- If **weighted**, put the weights instead of ones.
- If **temporal**, have one adjacency matrix per snapshot ($A_t, A_{t+1}, \ldots$).

Representing networks



$G = (V, E)$

3. The **edgelist:**

source	target
1	3
1	5
2	3
2	4
2	6
3	5

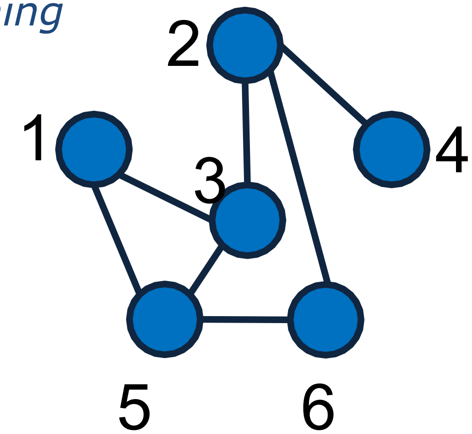
source: head of the link
target: tail of the link

Notes:

- The edgelist is a table, nice for data analysis.
- If **undirected**, a (source, target) pair indicates a *bidirectional* edge.
- If **weighted**, add *weight* as an extra column.
- If **temporal,** add *timestamp* as an extra column.

Representing networks

All of these represent the same thing



Node & edge sets

$$G = (V, E)$$

$V = \{1,2,3,4,5,6\}$

$E = \{(1,3), (1,5), (2,3), (2,4), (2,6), (3,5)\}$

Adjacency matrix

$A =$

	1	2	3	4	5	6
1	0	0	1	0	1	0
2	0	0	1	1	0	1
3	1	1	0	0	1	0
4	0	1	0	0	0	0
5	1	0	1	0	0	0
6	0	1	0	0	0	0

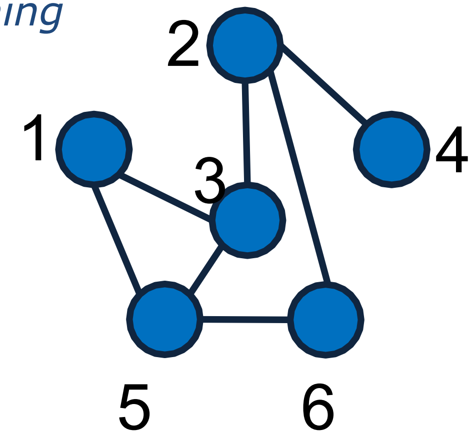
Edgelist

source	target
1	3
1	5
2	3
2	4
2	6
3	5

uc3m

Representing networks

All of these represent the same thing



Node & edge sets

$$G = (V, E)$$

$V = \{1,2,3,4,5,6\}$

$E = \{(1,3), (1,5), (2,3), (2,4), (2,6), (3,5)\}$

NetworkX

Network Analysis in Python

Adjacency matrix

$A =$

	1	2	3	4	5	6
1	0	0	1	0	1	0
2	0	0	1	1	0	1
3	1	1	0	0	1	0
4	0	1	0	0	0	0
5	1	0	1	0	0	0
6	0	1	0	0	0	0

NumPy

Edgelist

source	target
1	3
1	5
2	3
2	4
2	6
3	5

pandas

uc3m

4 Describing networks

Describing networks

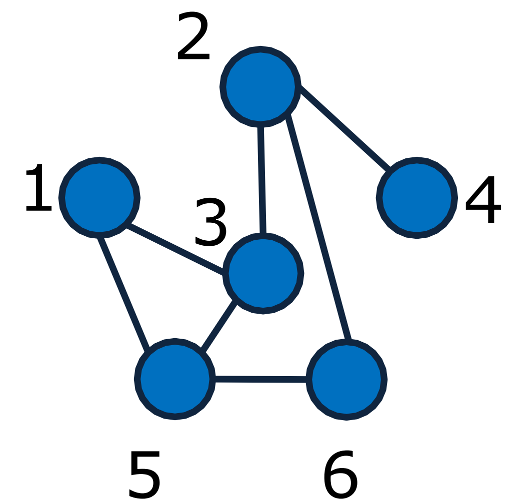
1. Degree, degree distribution
2. Degree distribution models
3. Network density
4. Assortativity (Degree correlation)
5. Clustering coefficient
6. Distance in networks, diameter, small world
7. Connectedness

Degree: the number of connections

We saw the two most basic network descriptors:
1. N: number of nodes
2. L: number of links

The **degree** is a key *node* descriptor, where
- k_i: number of connections of node i

degree sequence: collection of the degree for each node:
- In this case: $\mathbf{k} = (2, 3, 3, 1, 3, 2)$

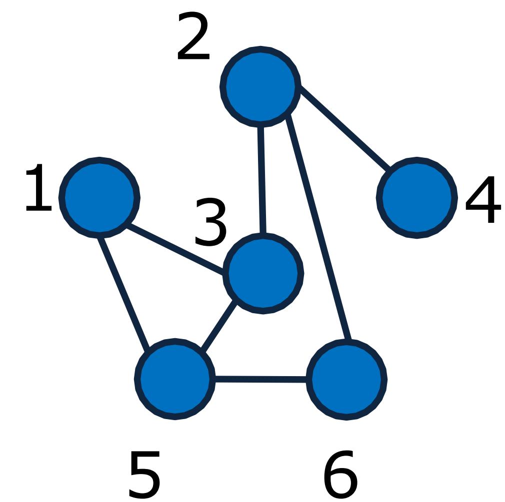


average degree

$$\langle k \rangle = \frac{1}{N} \sum_i k_i = \frac{2L}{N}$$

Degree: the number of connections

We saw the two most basic network descriptors:
1. N: number of nodes
2. L: number of links



The **degree** is a key *node* descriptor, where

- k_i: number of connections of node i

degree sequence: collection of the degree for each node:

- In this case: $\mathbf{k} = (2, 3, 3, 1, 3, 2)$

average degree

$$\langle k \rangle = \frac{1}{N} \sum_i k_i = \frac{2L}{N}$$

*Obs: In **directed** networks, we differentiate*

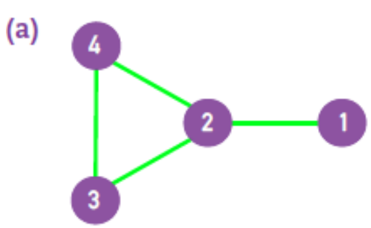
- ***in-degree:*** *incoming links* $\rightarrow \mathbf{k}_{in}$
- ***out-degree:*** *outgoing links* $\rightarrow \mathbf{k}_{out}$

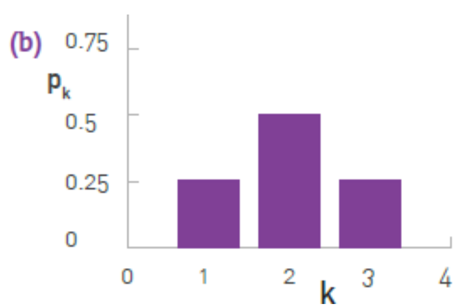
Degree distribution

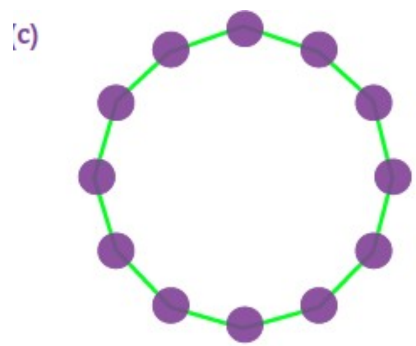
How are different
nodes connected
in the network?

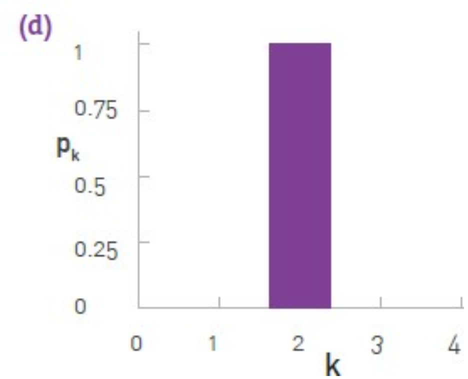
The **degree distribution** is the *normalised histogram* of the degree sequence $\mathbf{k}$, usually represented as $p(\mathbf{k})$ or p_k.

This tells us a great deal about how the network is connected *without* looking at the network itself.









$k_1 = 1,\ k_2 = 3,\ k_3 = k_4 = 2$

$p_{k=1} = 1/4,\ p_{k=2} = 2/4,\ p_{k=3} = 1/4$

$k_i = 2$

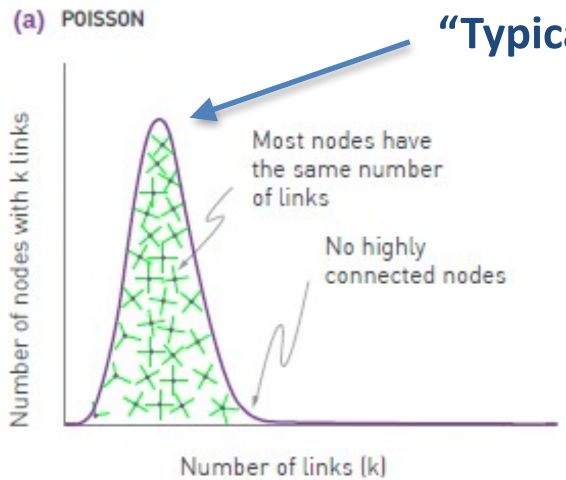
$p_{k=2} = 1$

Degree distribution: two models

Poisson distribution

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

"Typical" connectivity



Highway network

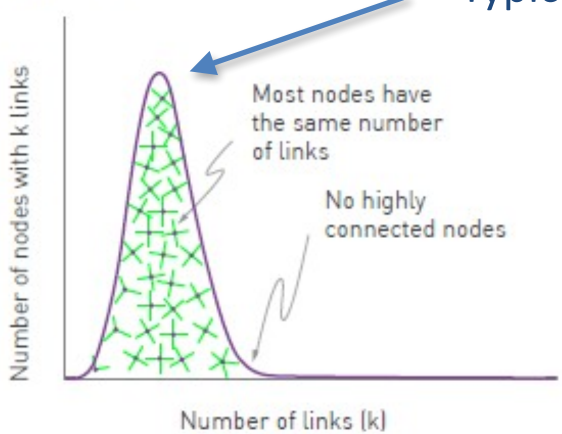
Degree distribution: two models

Poisson distribution

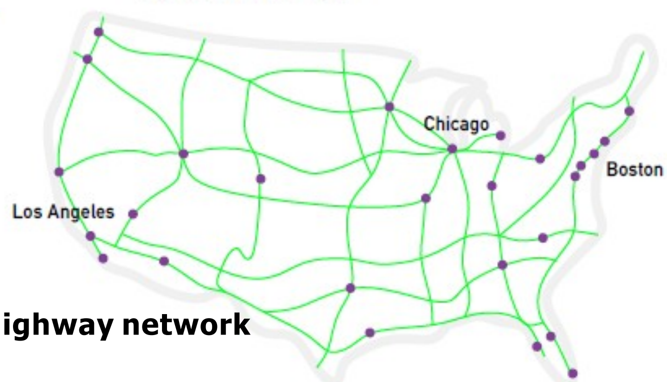
$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

(a) POISSON

"Typical" connectivity



(b)



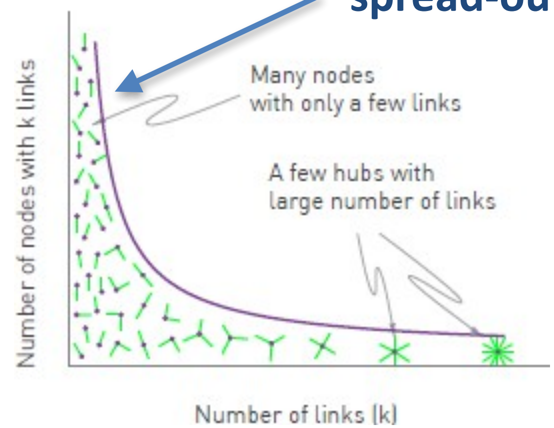
Highway network

Power-law distribution

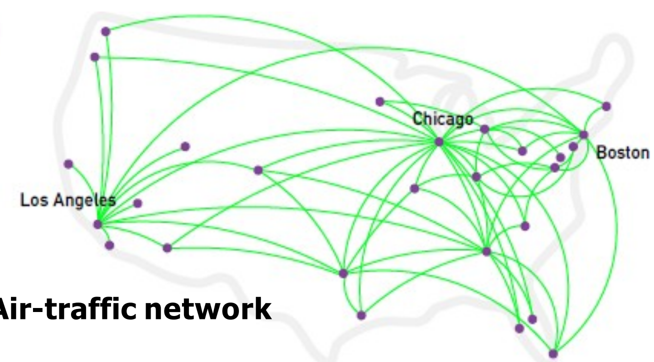
$$p_k = Ck^{-\gamma}.$$

(c) POWER LAW

"spread-out" connectivity



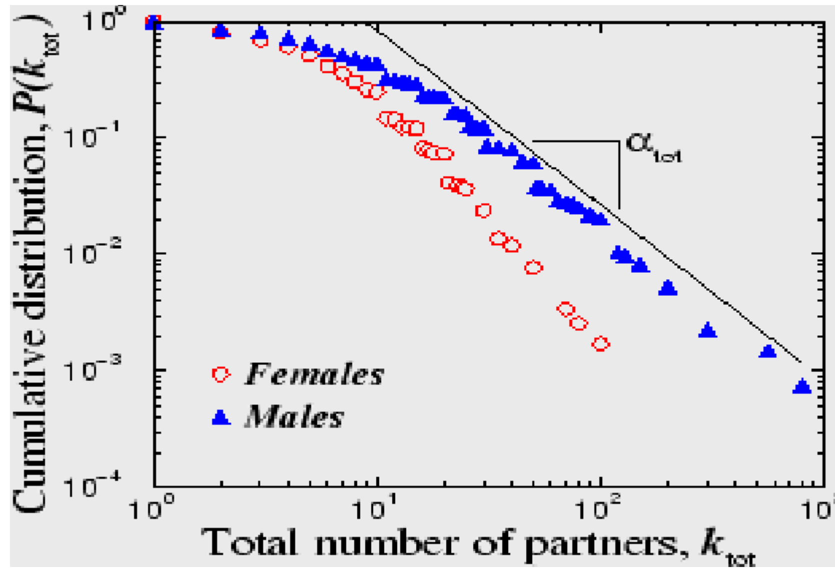
(d)



Air-traffic network

Degree distribution of real networks

2810 Swedish individuals (18-74)
Node: Person (Female;Male)
Link: Sexual relationship



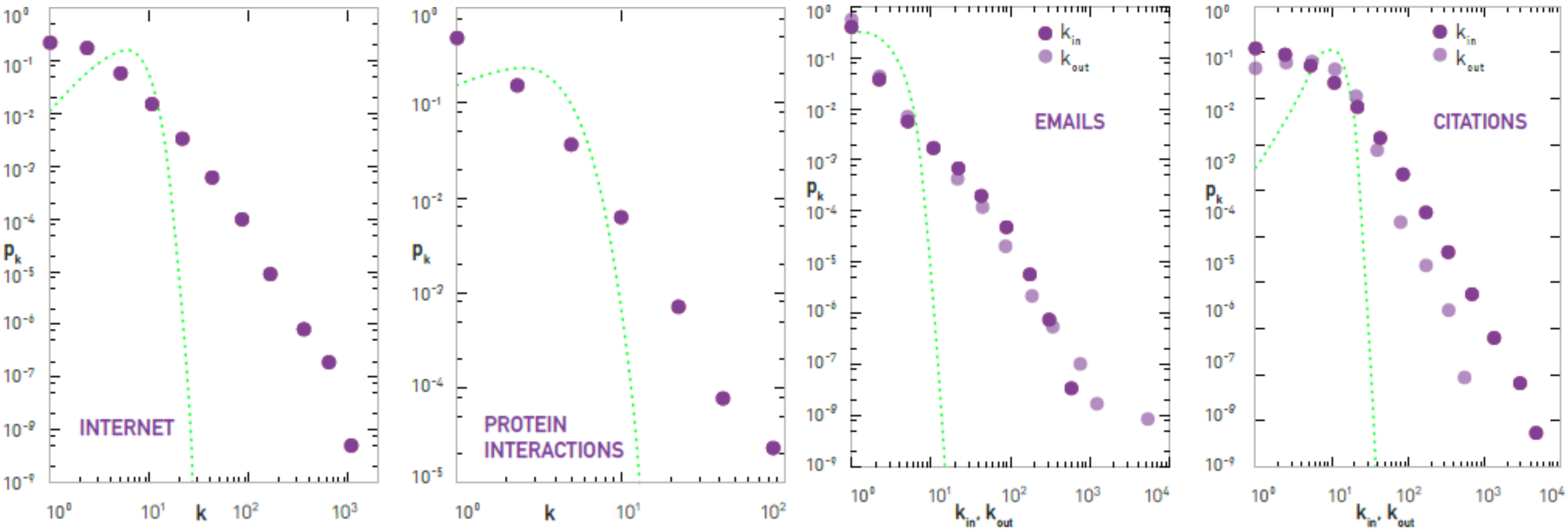
Can you see the '**hubs**'?

Most of people have few partners while some outliers have more than hundreds of partners

F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley (2001)

Degree distribution of real networks

Most real network distributions look like power laws ("heavy tails")



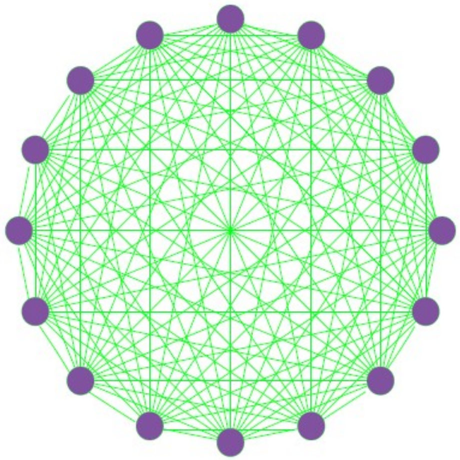
- Heavy tails → highly connected hubs nodes.
- Hubs → standard deviation σ_k is large. Average degree does not give us much information. There is no "typical" connectivity.

$$\sigma_k \gg \langle k \rangle$$

uc3m

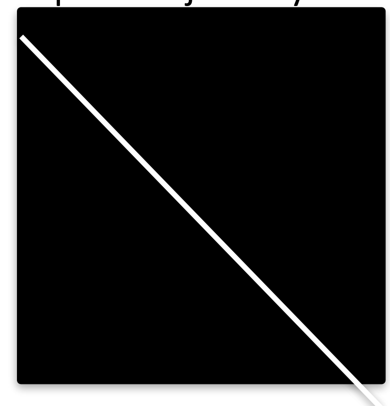
Real networks are sparse

- **Complete graph**: Each node connects to every other node



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

Complete adjacency matrix



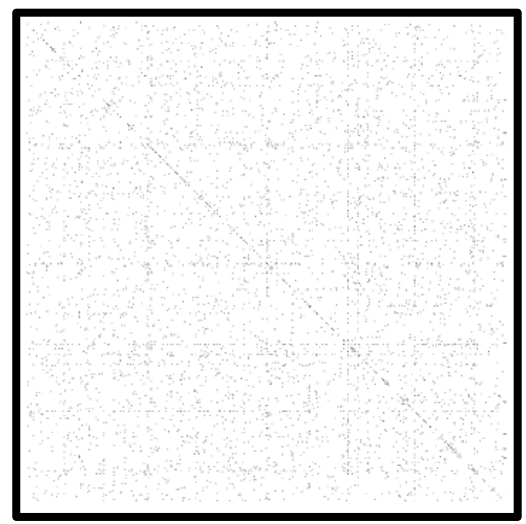
We call a network **sparse** if $L \ll L_{max}$
or, equivalently, if $\langle k \rangle \ll N$

→ *almost all entries in adjacency matrix are 0.*

Density: Proportion of existing links

$$\rho = \frac{L_{max}}{L} = \frac{N(N-1)}{2L}$$

Real network (sparse) adjacency



Protein-protein interaction network

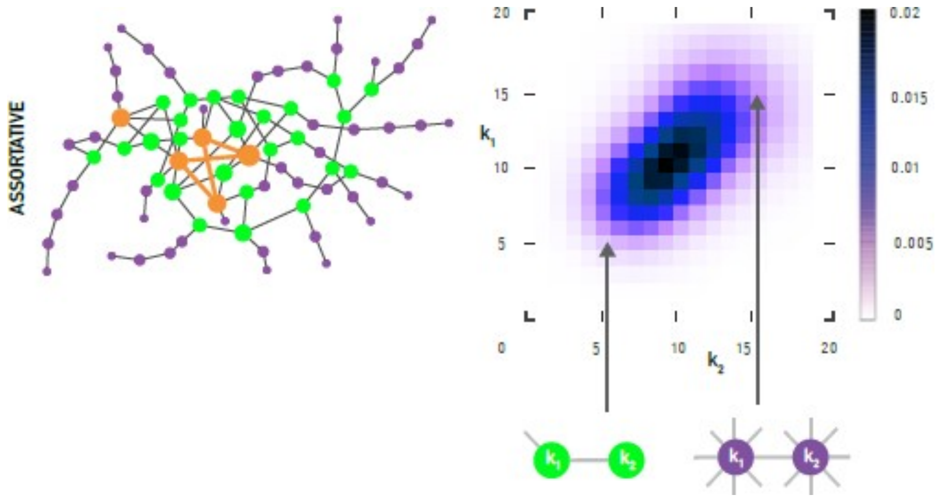
Real networks are sparse

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
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

$$\langle k \rangle \ll N$$

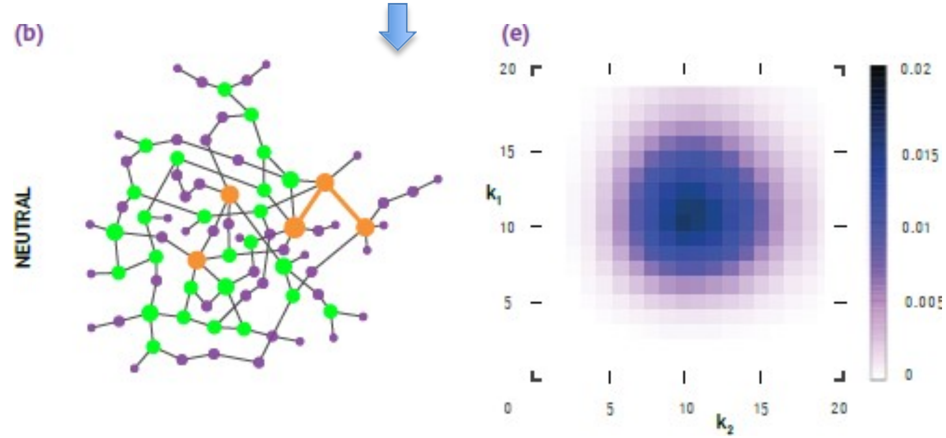
Degree correlation

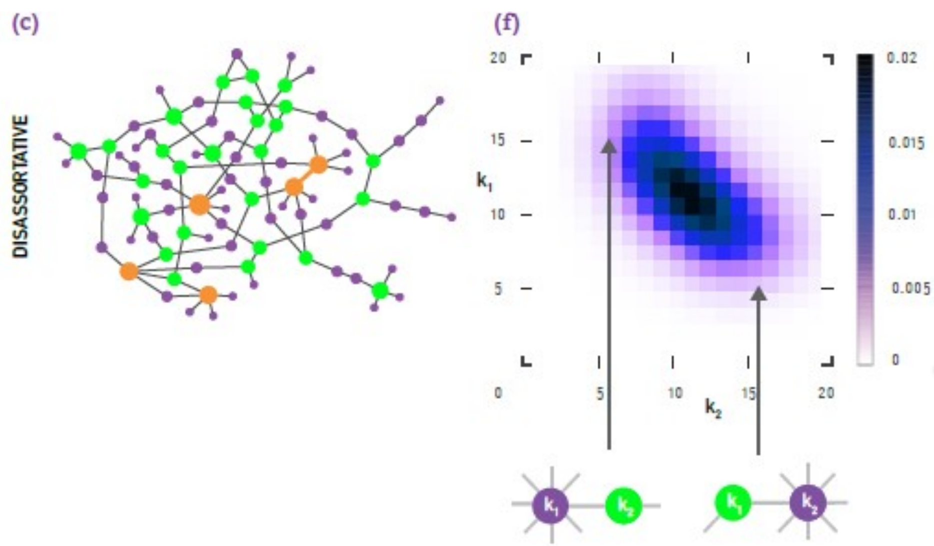
How do nodes with different degrees connect?



Assortative: High (low) degree nodes connect to high (low) degree nodes

Neutral





Dissortative: High (low) degree nodes connect to low (high) degree nodes

Degree correlation

How to measure this: **Assortativity coefficient** (Pearson correlation between the degrees of connected nodes).


assortative


disassortative

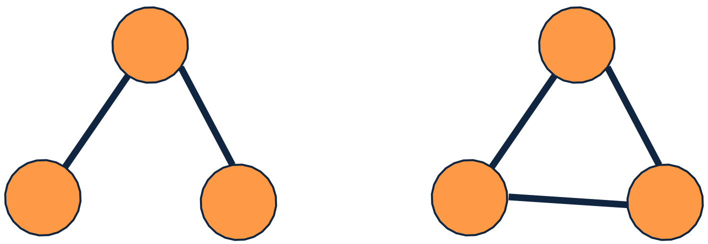
$$r \in [-1,1]$$

Network	n	r
Physics coauthorship (a)	52 909	0.363
Biology coauthorship (a)	1 520 251	0.127
Mathematics coauthorship (b)	253 339	0.120
Film actor collaborations (c)	449 913	0.208
Company directors (d)	7 673	0.276
Internet (e)	10 697	−0.189
World-Wide Web (f)	269 504	−0.065
Protein interactions (g)	2 115	−0.156
Neural network (h)	307	−0.163
Marine food web (i)	134	−0.247
Freshwater food web (j)	92	−0.276
Random graph (u)		0
Callaway *et al.* (v)		$\delta/(1+2\delta)$
Barabási and Albert (w)		0

- **Rich-club effect**: nodes with higher connectivity are connected between themselves
- **Social networks tend to be assortative**, while technological networks are disassortative. Social interactions tend to be "redundant." Technological relationships tend to be less frequent to optimize for cost, efficiency, etc.

Clustering coefficient

- Nodes tend to share mutual relationships: **your friends are likely friends between themselves too**.

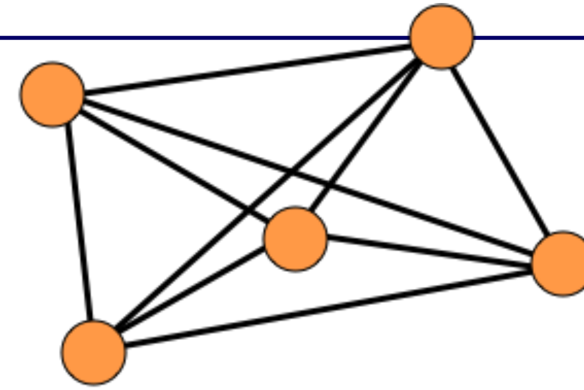


Open triangle **Closed triangle**

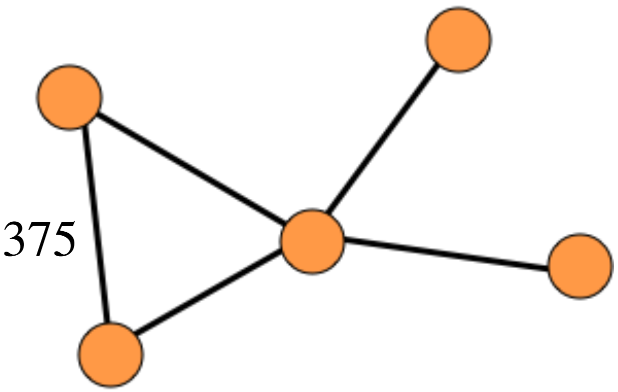
- This "clusterization" of links is measured with the **clustering coefficient** (or transitivity):

$$T = \frac{3 \times |\triangle|}{|\wedge|}$$

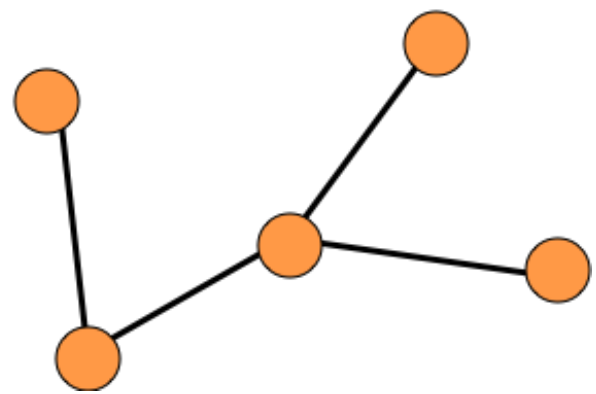
$T = 1$



$T = 3/8 = 0{,}375$



$T = 0$



Clustering in real networks

Real networks are way more clustered than random networks

Clustering coeff

	Network	Type	N	L	$\langle k \rangle$	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
Social	film actors	undirected	449 913	25 516 482	113.43	0.20	0.78	0.208	[20, 415]
	company directors	undirected	7 673	55 392	14.44	0.59	0.88	0.276	[105, 322]
	math coauthorship	undirected	253 339	496 489	3.92	0.15	0.34	0.120	[107, 181]
	physics coauthorship	undirected	52 909	245 300	9.27	0.45	0.56	0.363	[310, 312]
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	0.088	0.60	0.127	[310, 312]
	telephone call graph	undirected	47 000 000	80 000 000	3.16				[8, 9]
	email messages	directed	59 912	86 300	1.44		0.16		[136]
	email address books	directed	16 881	57 029	3.38	0.17	0.13	0.092	[320]
	student relationships	undirected	573	477	1.66	0.005	0.001	−0.029	[45]
	sexual contacts	undirected	2 810						[264, 265]
Information	WWW nd.edu	directed	269 504	1 497 135	5.55	0.11	0.29	−0.067	[14, 34]
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46				[74]
	citation network	directed	783 339	6 716 198	8.57				[350]
	Roget's Thesaurus	directed	1 022	5 103	4.99	0.13	0.15	0.157	[243]
	word co-occurrence	undirected	460 902	17 000 000	70.13		0.44		[119, 157]
Technological	Internet	undirected	10 697	31 992	5.98	0.035	0.39	−0.189	[86, 148]
	power grid	undirected	4 941	6 594	2.67	0.10	0.080	−0.003	[415]
	train routes	undirected	587	19 603	66.79		0.69	−0.033	[365]
	software packages	directed	1 439	1 723	1.20	0.070	0.082	−0.016	[317]
	software classes	directed	1 377	2 213	1.61	0.033	0.012	−0.119	[394]
	electronic circuits	undirected	24 097	53 248	4.34	0.010	0.030	−0.154	[155]
	peer-to-peer network	undirected	880	1 296	1.47	0.012	0.011	−0.366	[6, 353]
Biological	metabolic network	undirected	765	3 686	9.64	0.090	0.67	−0.240	[213]
	protein interactions	undirected	2 115	2 240	2.12	0.072	0.071	−0.156	[211]
	marine food web	directed	135	598	4.43	0.16	0.23	−0.263	[203]
	freshwater food web	directed	92	997	10.84	0.40	0.48	−0.326	[271]
	neural network	directed	307	2 359	7.68	0.18	0.28	−0.226	[415, 420]

Mean degree

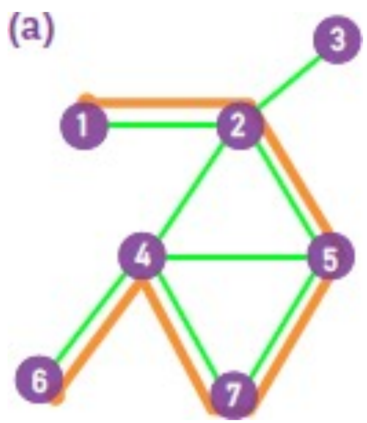
Distance in networks

•What is the distance between two webpages in WWW, or between two individual who do not know each other in social network?

•Two webpages can be connected a single hyperlink although their servers are located in different continents.
➔ Physical distance is not useful to define "distance" in networks

• In networks physical distance is replaced by **path length**

• A **path** is a route that runs along the links of the networks
• A **path's length** represents the number of links the path contains.

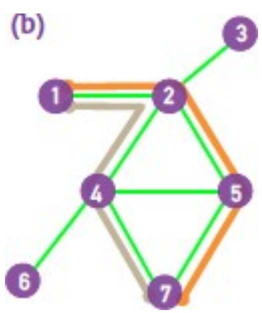


Orange line is a **path** from node 1 to node 6 (length = 5)

uc3m

Distance in networks

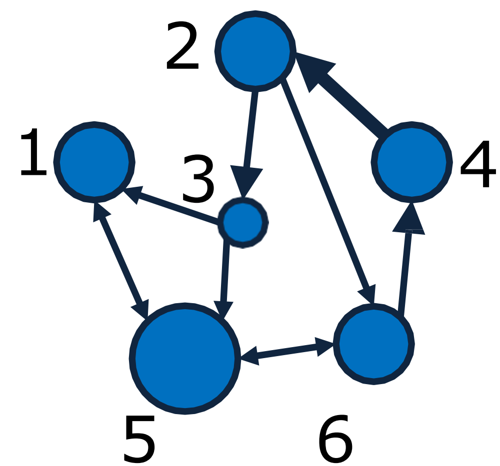
- **Shortest path** between node i and j is the path with the fewest number of links.
- **Distance** d_{ij}**:** *shortest path length* between node i and node j
- **diameter:** maximum distance of a network



-Orange line and grey line are the shortest paths from node 1 to node 7.
- The distance is 3.
- $d_{ij} = d_{ji}$ in undirected networks.

- Directed networks: $d_{ij} \neq d_{ji}$

$$d_{2\to 5} = 2$$
$$d_{5\to 2} = 3$$



Diameter of real networks

- **Small world phenomenon**: Networks have a tiny diameters
- Mathematically, small-world property is defined by

$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}$$

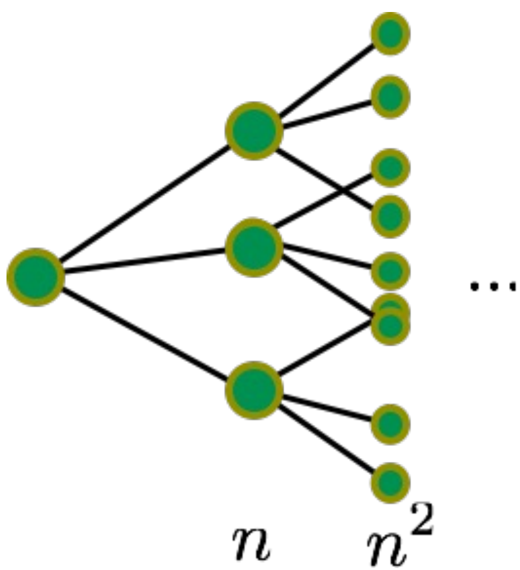
- Many real networks show the small-world properties

NETWORK	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\frac{\ln N}{\ln \langle k \rangle}$
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,439	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83.71	3.91	14	3.04
Citation Network	449,673	4,707,958	10.43	11.21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

uc3m

Diameter of real networks

•Small world experiment: Milgram and Travers (1977) sent letters between persons in the US.
•They found that, on average, a letter took 6 hops between persons to reach their destination (6 degrees of separation)

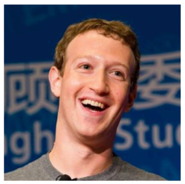


$$43^6 \simeq 6 \times 10^9$$

Diameter of real networks

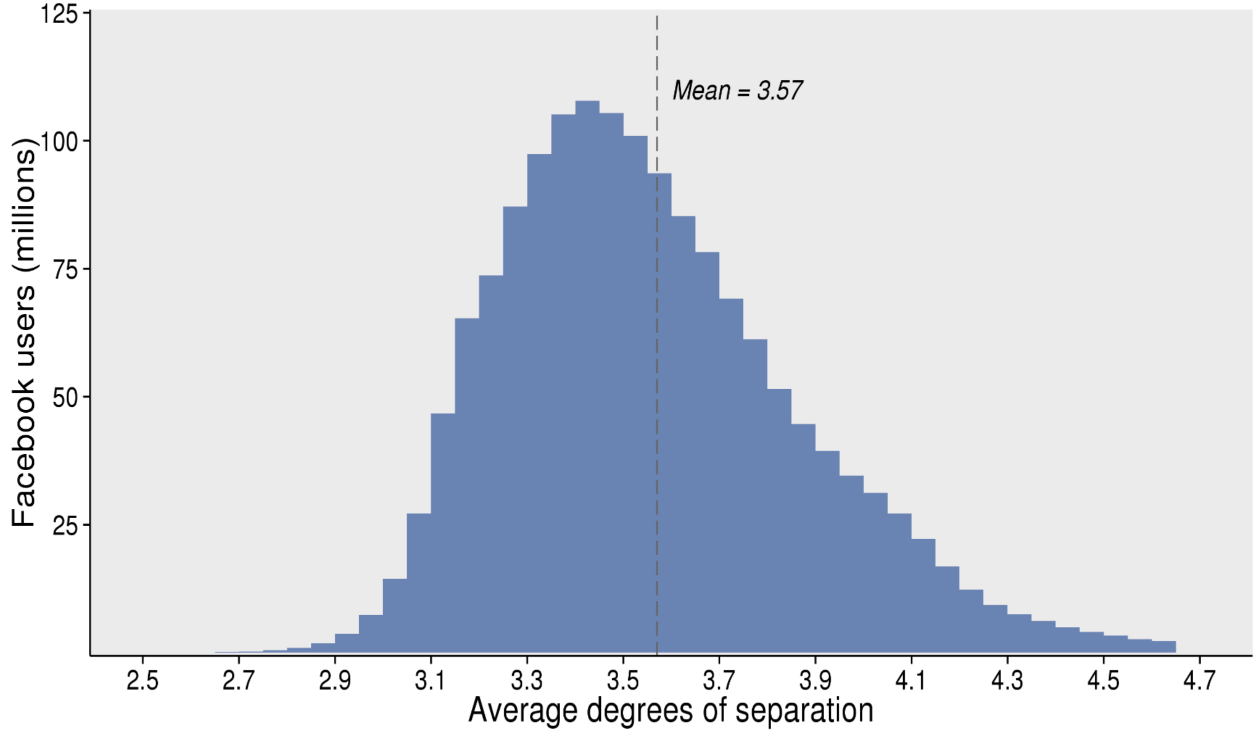
- In Facebook, the average distance to any other user is *very small*

Some Facebook employees



Mark Zuckerberg
3.17 degrees of separation

Sheryl Sandberg
2.92 degrees of separation



- https://research.facebook.com/blog/2016/2/three-and-a-half-degrees-of-separation/

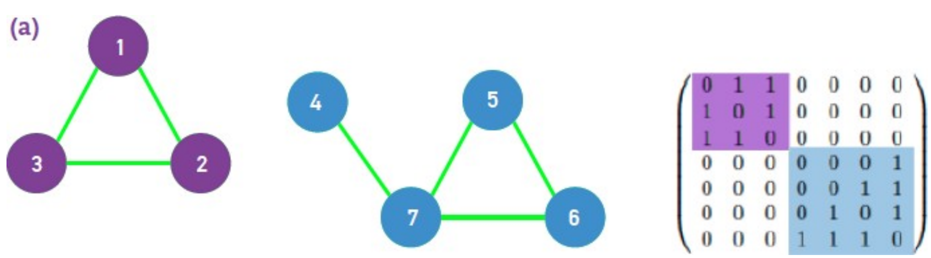
Connectedness

A network is **connected** if there is a path between all pairs of nodes.

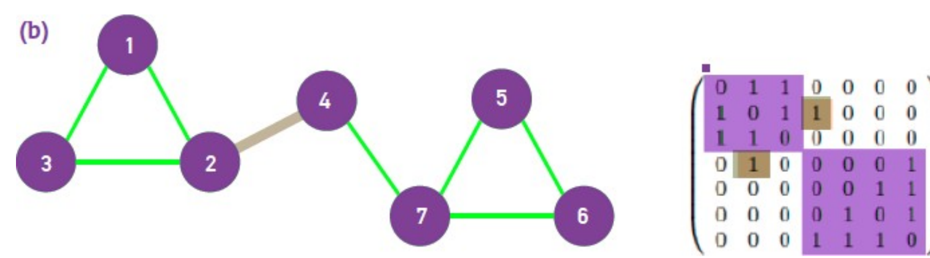
A **network component** is a subset of nodes in a network, so that there is a path between any two nodes that belong to the component to the component.

A **bridge** is any link that, if cut, disconnects the network-

Disconnected network



Connected network



5 Summary

Summary of session 1

- Networks are everywhere. We need to understand their structure to model data, processes.
- Degree, average degree, degree distribution:
  - Most real networks have heavy-tailed distribution
- Degree correlation:
  - Asssortative / Dissortative / Neutral
- Clustering:
  - Real networks are highly clustered → a lot of closed triangles!
- Path, Distance & Diameter
  - Real networks are "small-world"