Graph Theory: Basics

Introduction to network Science

Instructor: Michele Starnini — https://github.com/chatox/networks-science-course

Universitat Pompeu Fabra Barcelona

Network Science or Graph Theory?

	Field	When	What	How
Graphs	mathematics (computer science)	1960-70	Theory	Structural properties
Networks	physics	2000-present	Applications	Complex systems, Dynamical processes

Nobody cares! Completely equivalent

What is (modern) Network Science

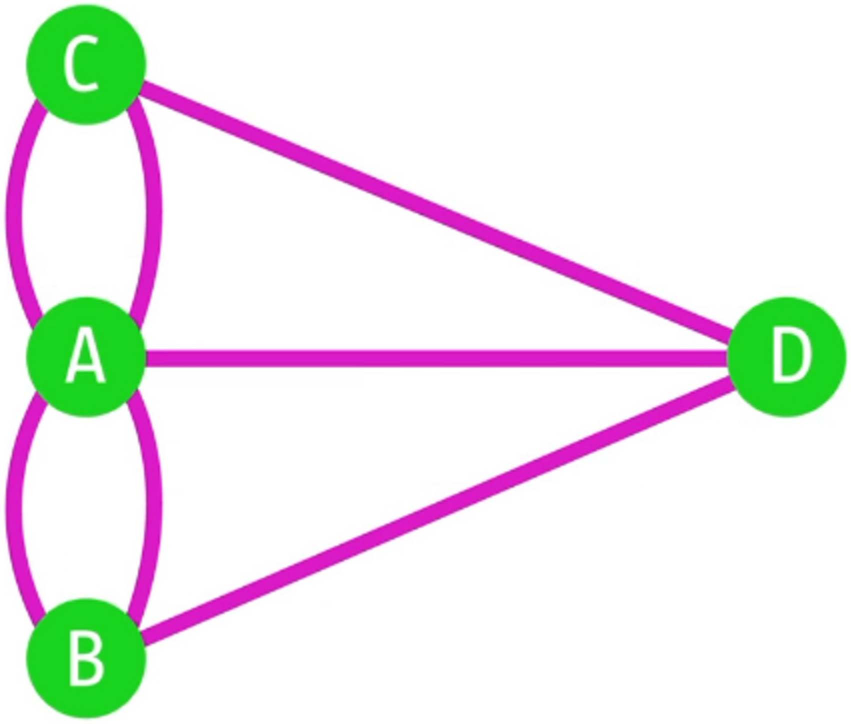
- **Mathematical formalism** from graph theory
- New insights from **statistical physics**
- Recent availability of very **large databases** on networked systems
- **Computational resources** to analyze them

Contents

- Directed, weighted, … many kind of graphs!
- Density & Sparsity
- Degrees
- Adjacency matrices

Notation for a graph

- G = (V,E)
  - V: nodes or vertices
  - E: links or edges
- |V| = N size of graph
- |E| = L (or E) number of links



Subgraph

- Given G = (V,E)
- A **subgraph** induced by a nodeset S is the graph G'=(S,F) defined by:
  - nodes in S
  - edges in F = { (u,v) ∊ E s.t. u ∊ S and v ∊ S }

Directed vs undirected graphs

- In an undirected graph
  - E is a symmetric relation

$$(u, v) \in E \Rightarrow (v, u) \in E$$

- In a directed graph, also known as “digraph”
  - E is not a symmetric relation

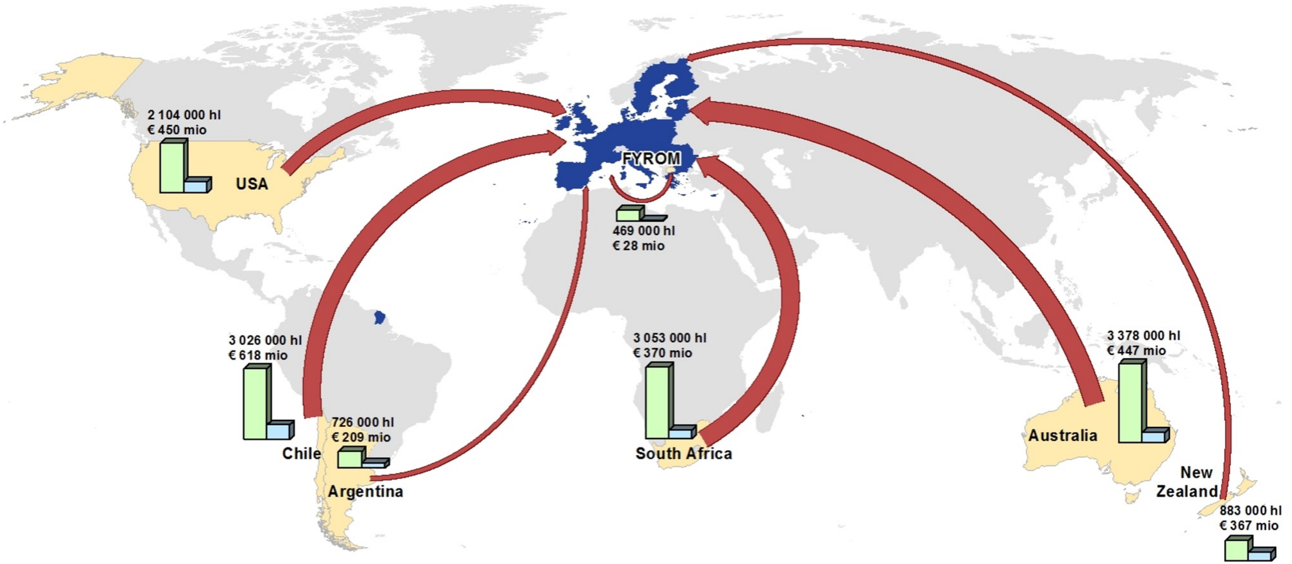
$$(u, v) \in E \not\Rightarrow (v, u) \in E$$

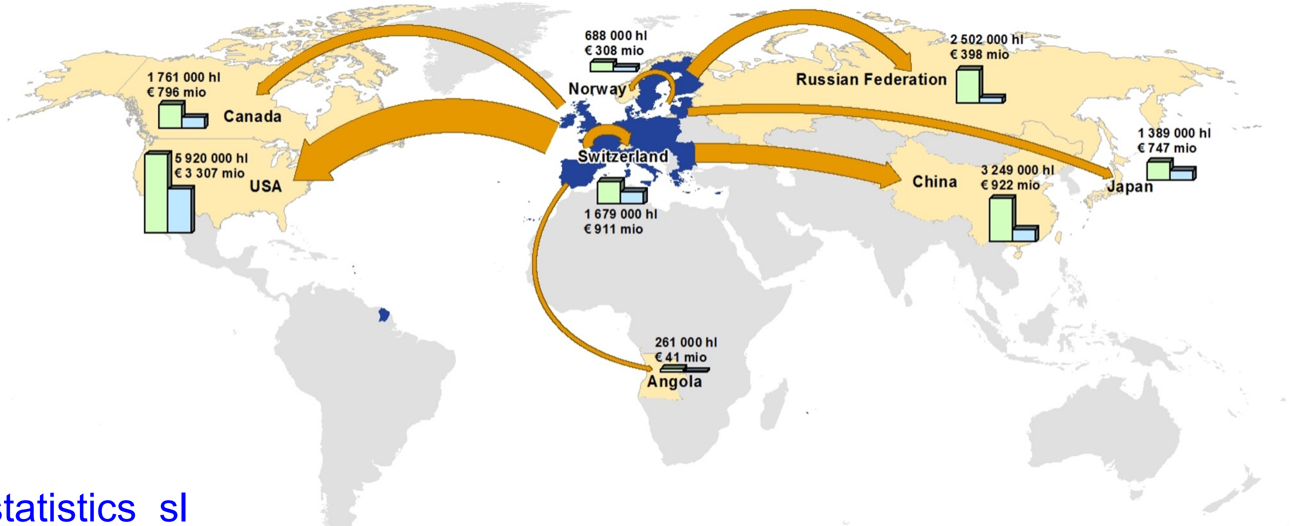
Weighted vs unweighted graphs

- In a weighted graph edges have **weights** denoting the strength or importance of a connection
- When drawing, typically larger weights are drawn with ticker lines

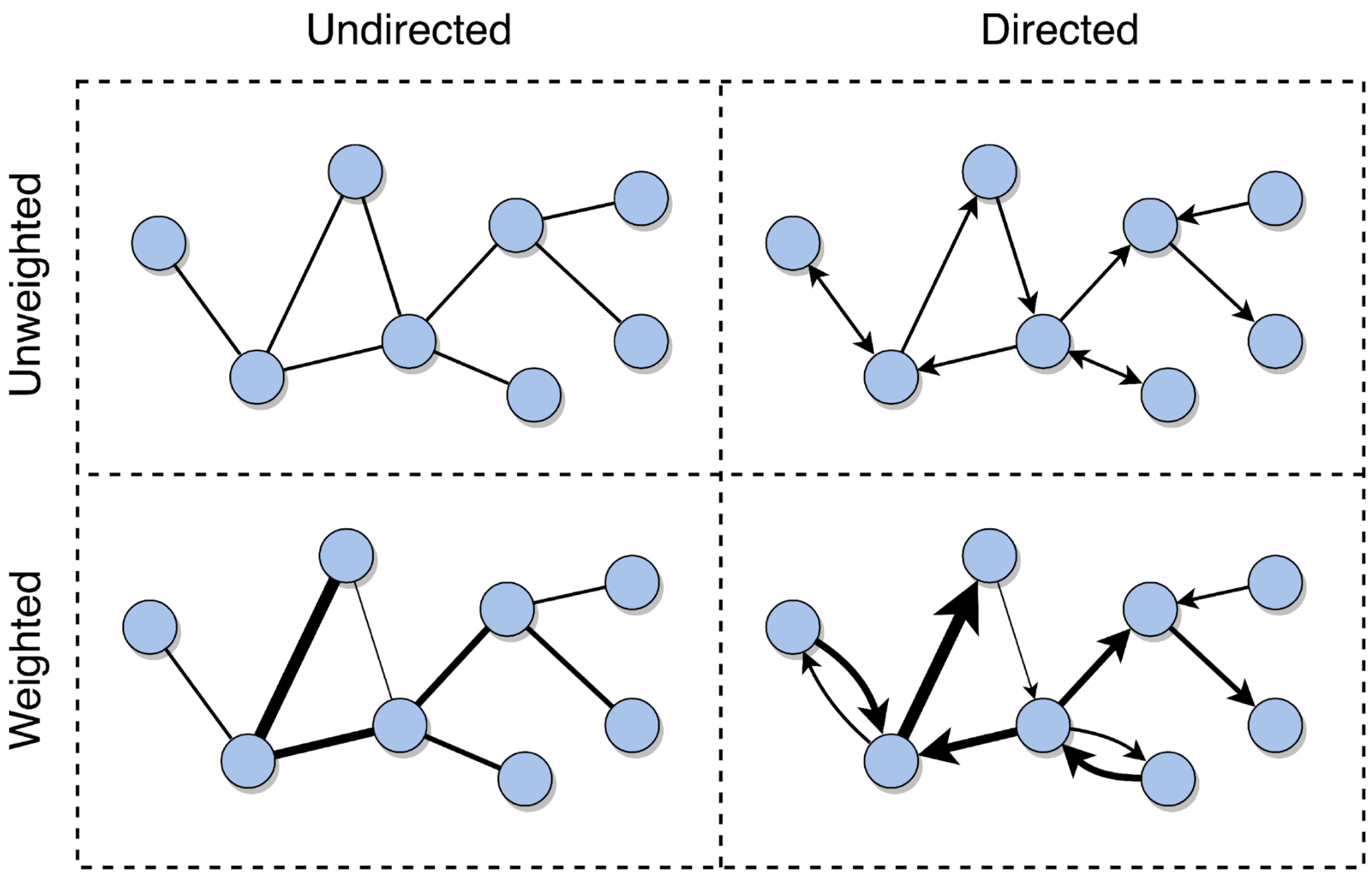
Weighted networks

EU imports (top)
and exports (bottom)
of wine





https://ec.europa.eu/agriculture/wine/statistics_sl



Source: Menczer, Fortunato, David: A First Course on Networks Science. Cambridge, 2020.

Example graphs we will use

Network	**N**	**E**
Zachary’s Karate Club (karate.gml)	34	78
Game of Thrones (got-relationships.csv)	84	216
US companies ownership	1351	6721
Marvel comics (hero-network.csv)	6K	167K

https://github.com/chatox/networks-science-course/tree/master/practicum/data

Density and sparsity

- Network size N = number of nodes
- L = number of links
- Maximum possible number of links:

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

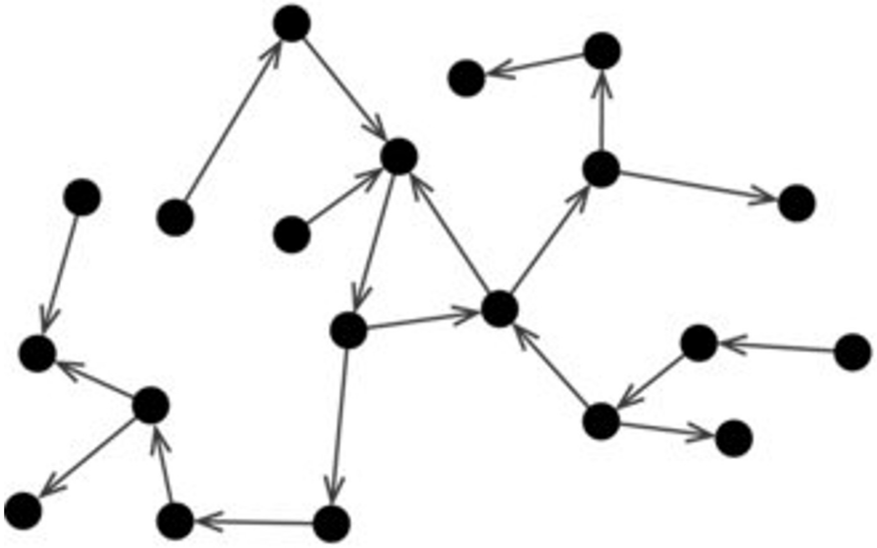
- Density:

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

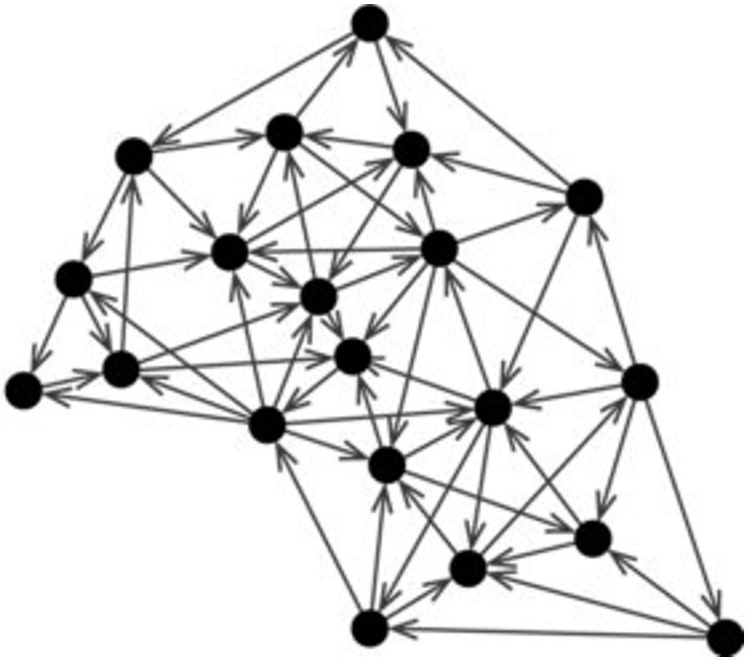
- The network is **sparse** if $d << 1$

Source: Menczer, Fortunato, David: A First Course on Networks Science. Cambridge, 2020.

Sparse network



Dense network



[Source]

Example: Facebook

Rough orders-of-magnitude approximations:

- $N \approx 10$
- $L \approx 10^3 \times N$
- $d \approx L / N^2 \approx 10^3 N / N^2 \approx 10^3 / 10^9 = 10^{-6}$
- Most (but not all) real-world networks are similarly sparse because the number of links scales proportionally to N, whereas the maximum scales with N^2

Source: Menczer, Fortunato, David: A First Course on Networks Science. Cambridge, 2020.

Complex networks are sparse

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

- Most real networks are sparse, i.e., $$L \ll L_{\max}$$

L is the number of links in the network, N is the number of nodes on it

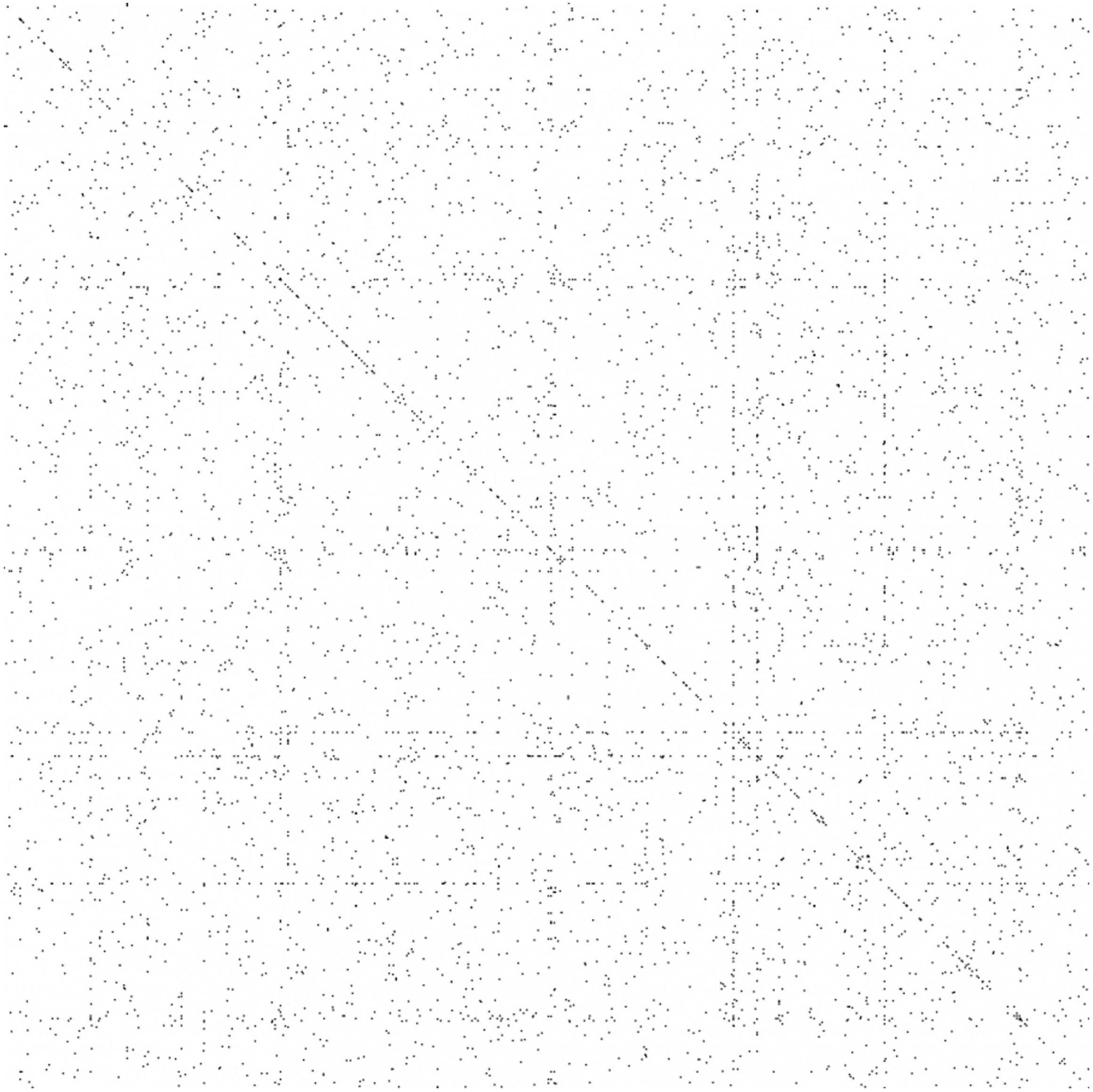
How sparse are some networks?

Network	|V|	|E|	Max |E|
Zachary's Karate Club	34	78	561
Game of Thrones	84	216	3496
US companies ownership	1351	6721	911K
Marvel comics	6K	167K	17M

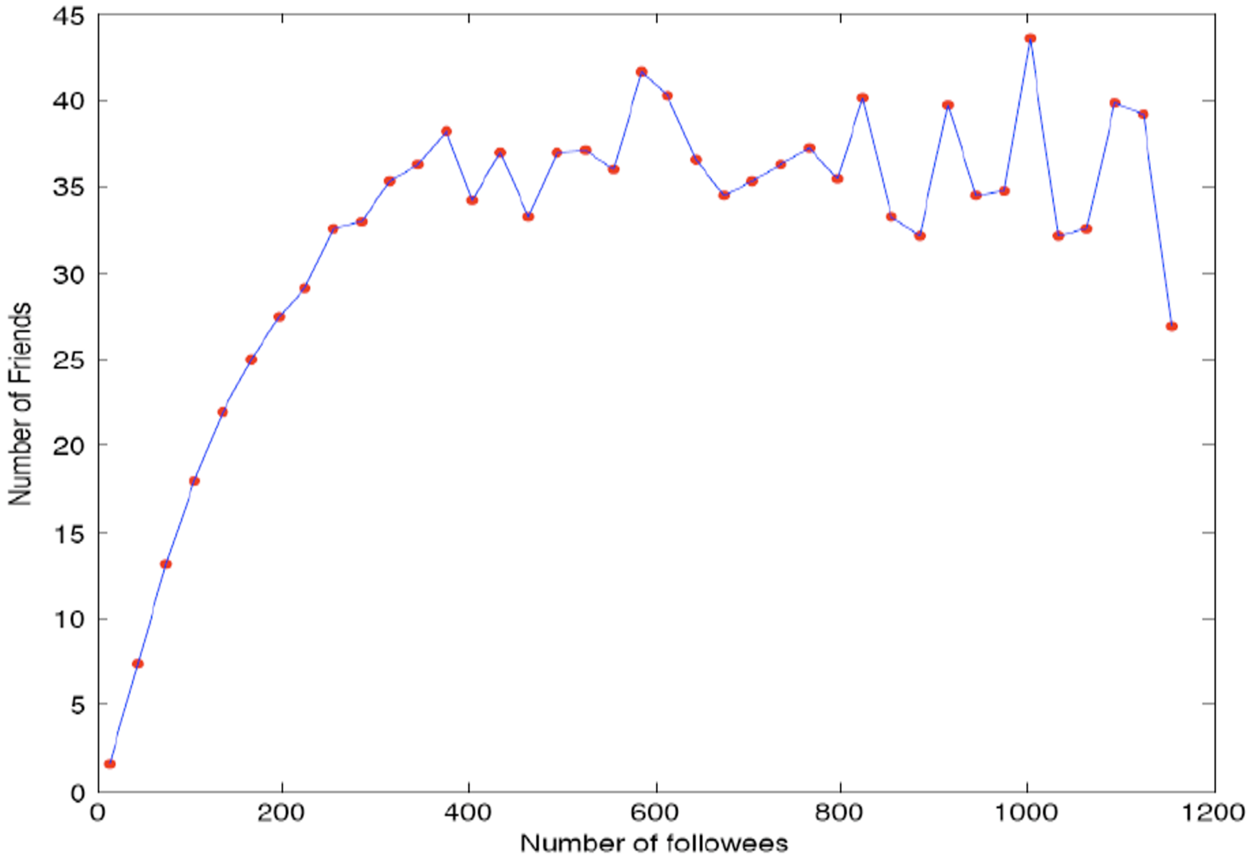
https://github.com/chatox/networks-science-course/tree/master/practicum/data

Example: protein interaction network

(N=2K, L=3K)



Example: people you follow on Twitter (followees) vs people you have sent at least two messages to ("friends")

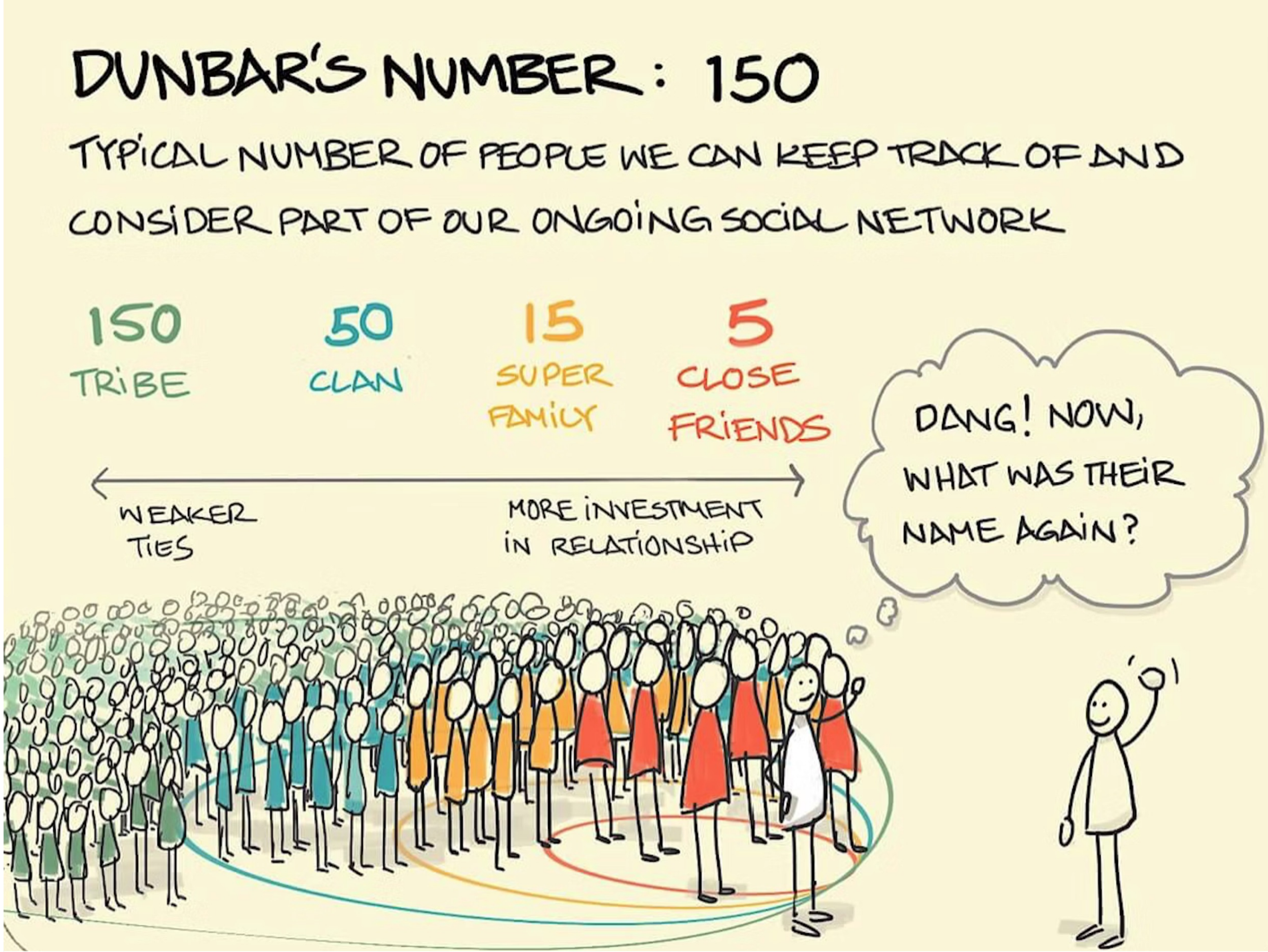


Huberman, B., Romero, D. M., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. First Monday, 14(1).

Why are networks sparse?

- Different mechanisms, think about it from the node perspective:
  - How many items **could** the node be connected to
  - Would it be **realistic** to connect to a large fraction of them?
- In social networks, Dunbar's number ($\simeq$150)

https://en.wikipedia.org/wiki/Dunbar's_number



[Sketchplanations]

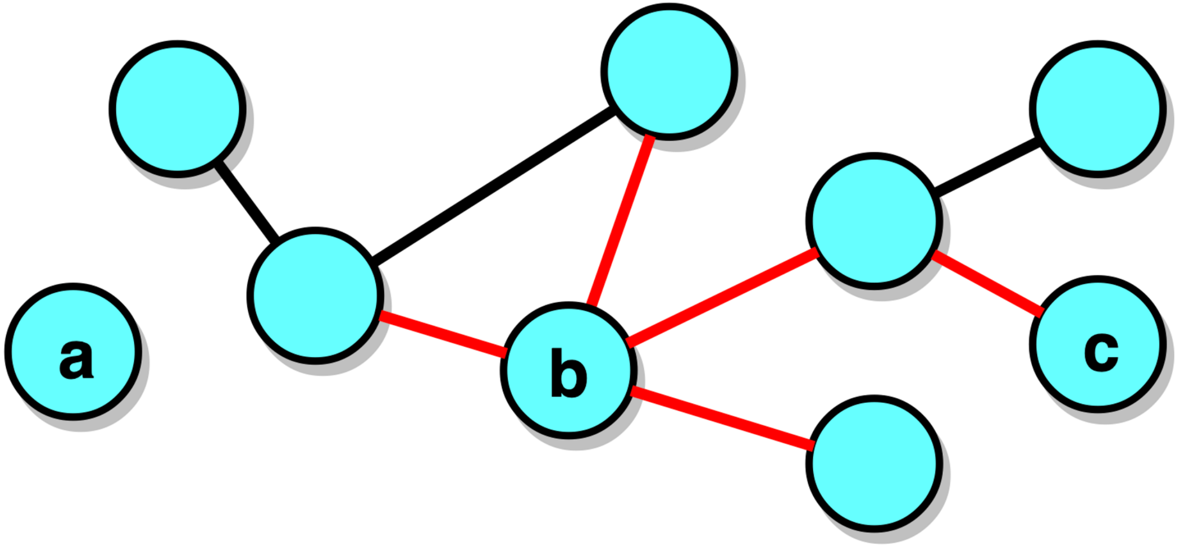
Degree

Node i has degree k_i

This is the number of links incident on this node

The total number of links L is given by

$$L = \frac{1}{2} \sum_{i=1}^{N} k_i$$



Average degree

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N}$$

Average degree & density

$$d = \frac{\langle k \rangle}{N-1} = \frac{\langle k \rangle}{k_{max}}$$

- Let us revisit the Facebook example
- Since *N* is very large, *N* – 1 can be approximated by *N*

$$d = \frac{\langle k \rangle}{N-1} \approx \frac{\langle k \rangle}{N} \approx \frac{10^3}{10^9} = 10^{-6}$$

In directed graphs

We distinguish in-degree from out-degree

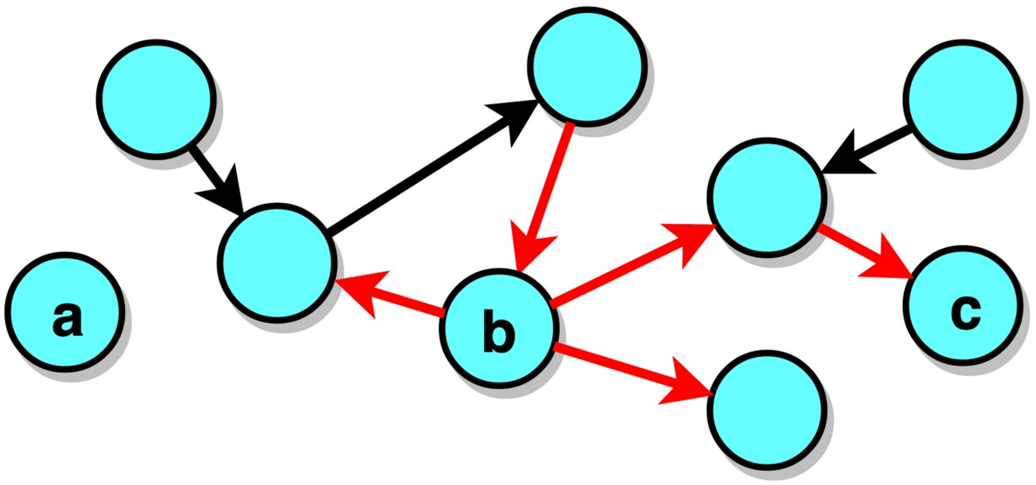
Incoming and outgoing links, respectively

Degree is the sum of both

$$k_i = k_i^{\text{in}} + k_i^{\text{out}}$$

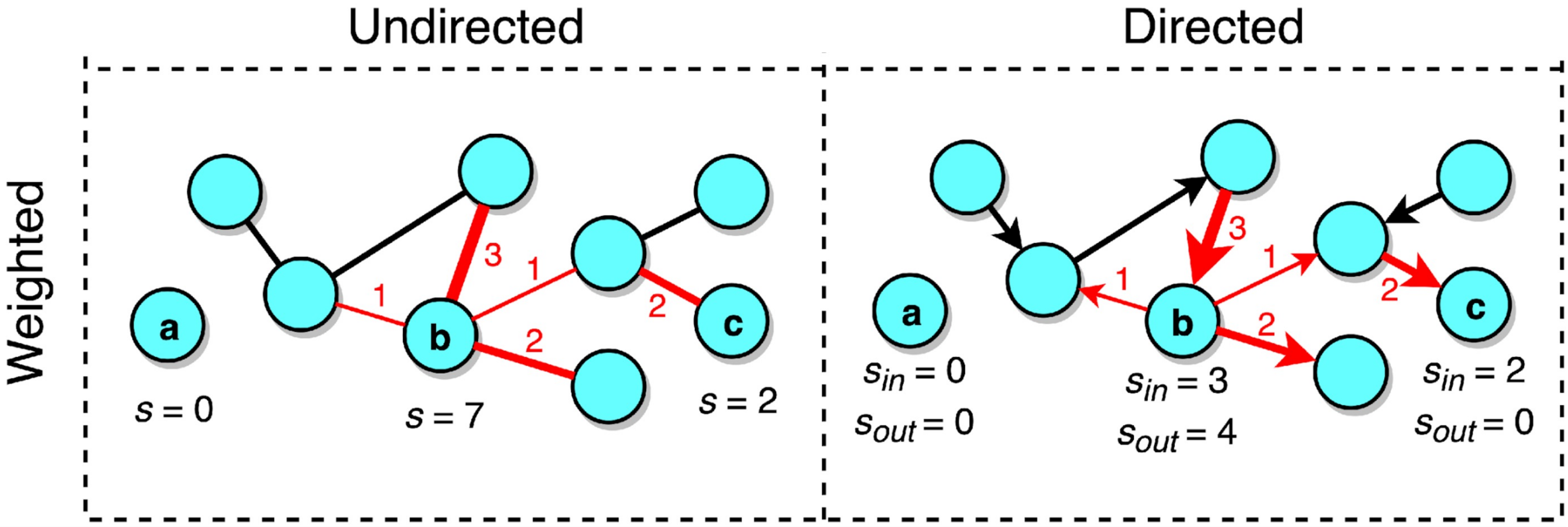
Counting total number of links:

$$L = \sum_{i=1}^{N} k_i^{\text{in}} = \sum_{i=1}^{N} k_i^{\text{out}}$$



In weighted graphs

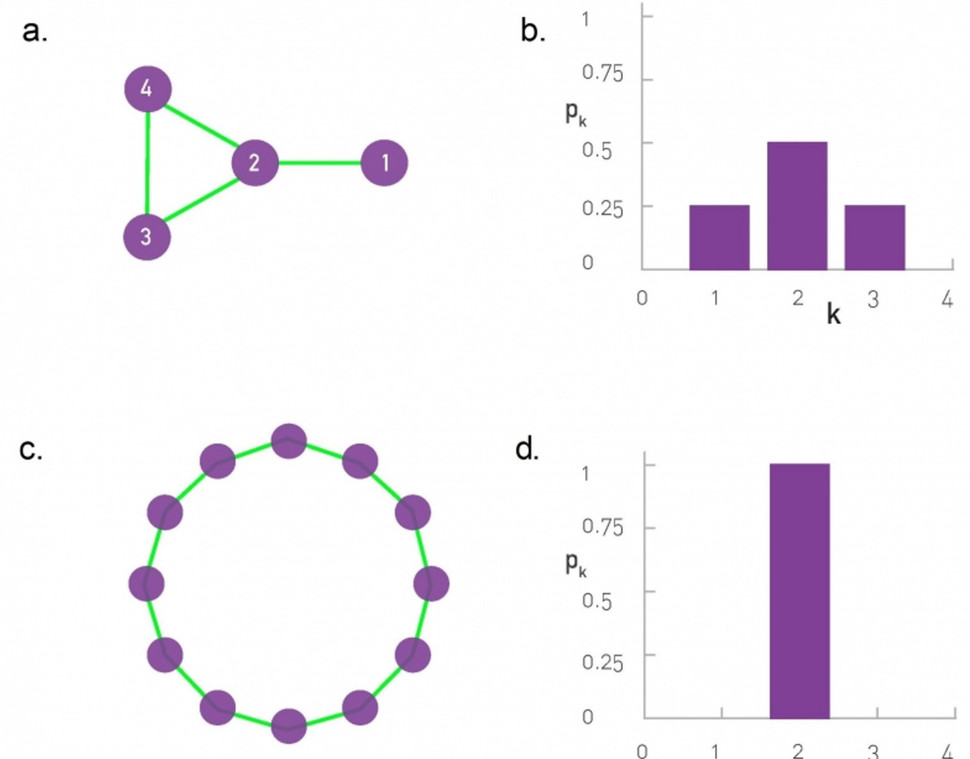
- We speak of “weighted degree” or “strength”



Degree distribution

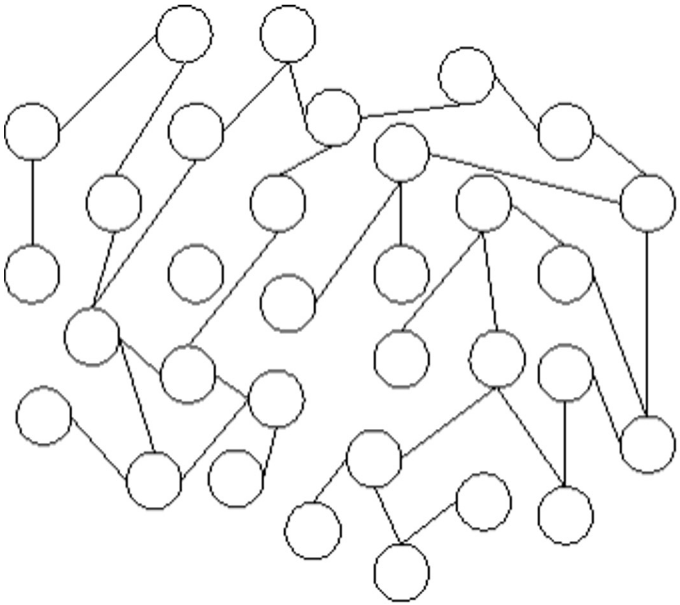
- If there are N_k nodes with degree k
- The degree distribution is given by $$p_k = \frac{N_k}{N}$$
- The average degree is then $$\langle k \rangle = \sum_{k=0}^{\infty} k p_k$$

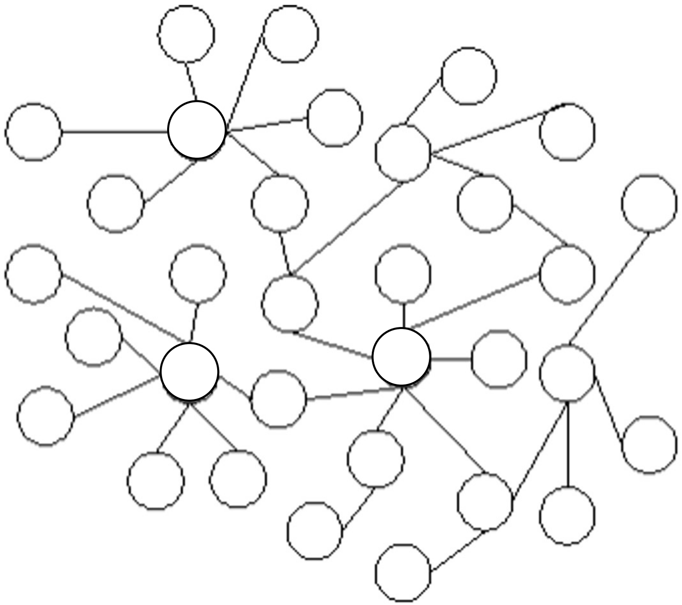
Degree distribution; two toy graphs



Exercise

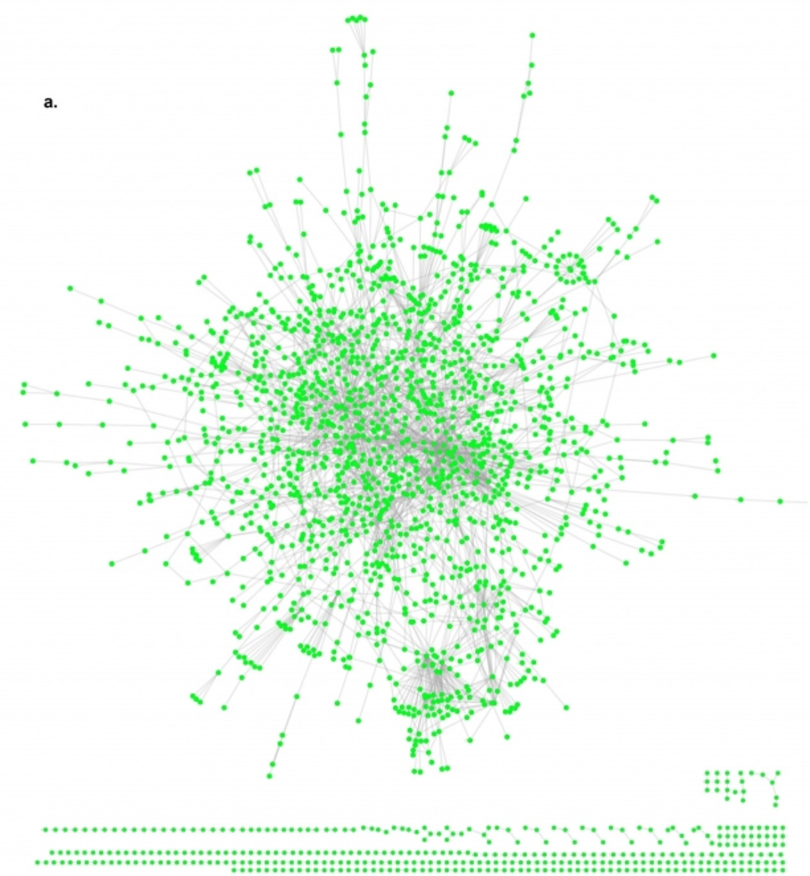
Draw the degree distribution of these graphs

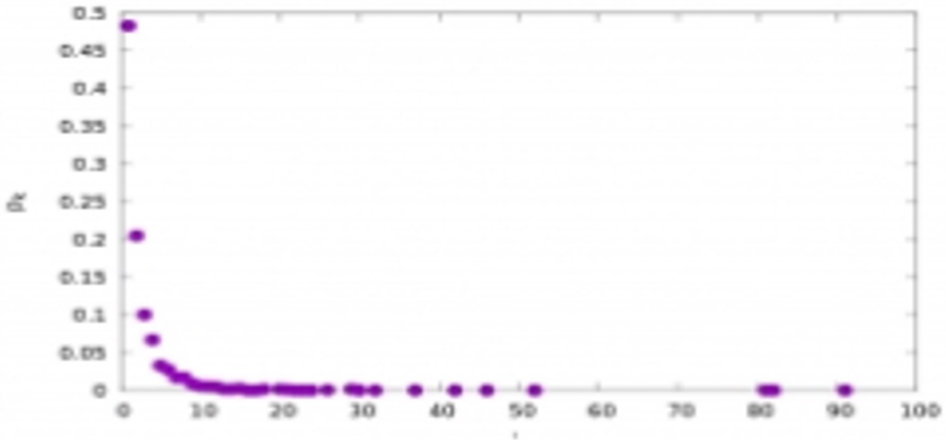




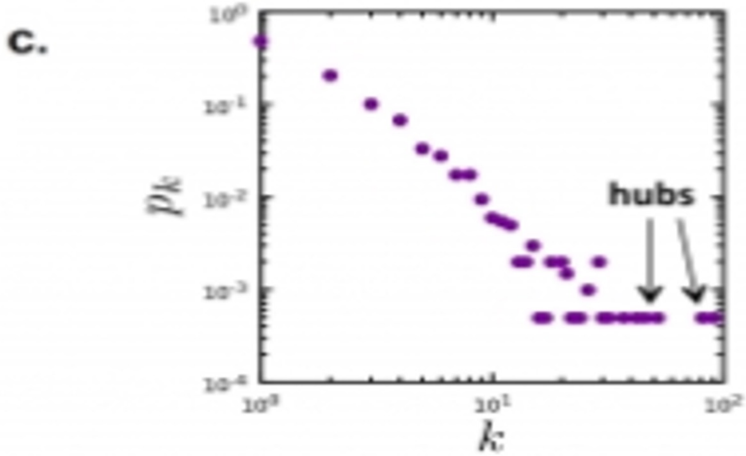
Spreadsheet links: https://upfbarcelona.padlet.org/chato/shyq9m6f2g2dh1bw

Real graphs





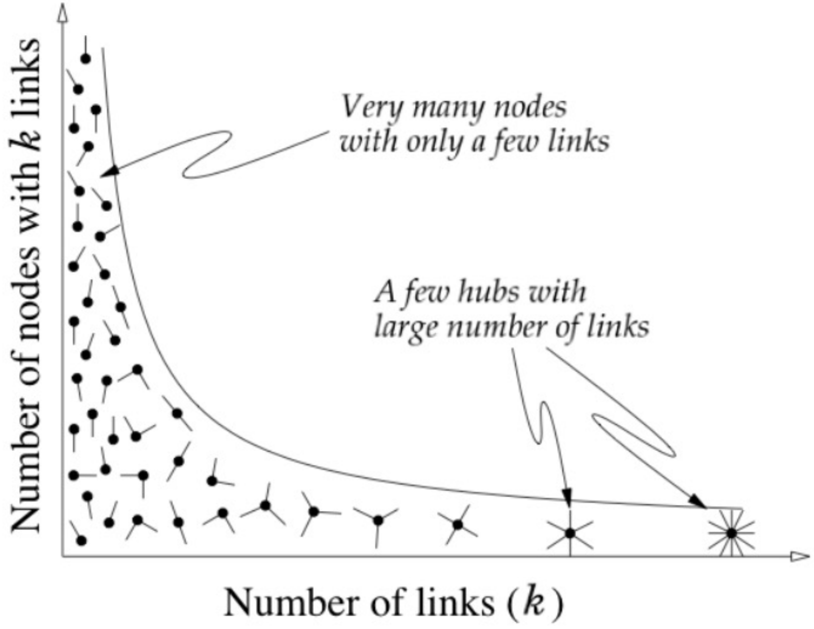
Linear scale



Log-log scale

Degree distribution P(k)

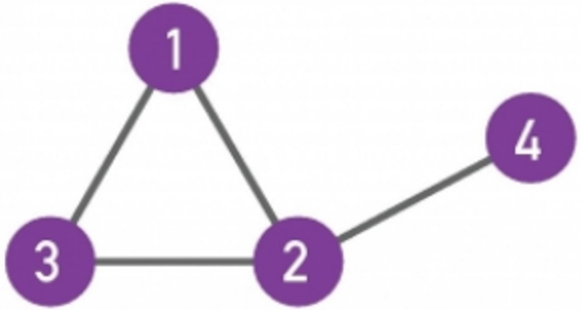
- Most important statistical property of networks
- Nodes with large degree = hubs, nodes with small degree = leaves
- When hubs are present, the P(k) is heterogeneous
- Most real (complex) networks are heterogeneous



Adjacency matrix

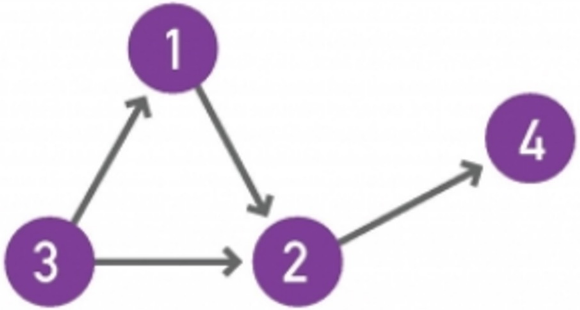
- A is the adjacency matrix of $G = (V, E)$ iff:
  - A has $|V|$ rows and $|V|$ columns
  - $A_{ij} = 1$ if $(i,j) \in E$
  - $A_{ij} = 0$ if $(i,j) \notin E$
- **A_{ij} always means row i, column j**
  - Sometimes Barabási's book has this wrong

Examples



Undirected graph

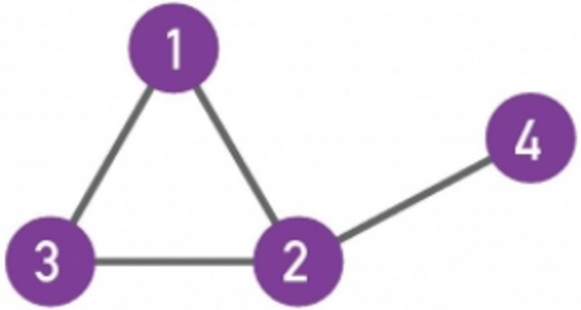
$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$



Directed graph

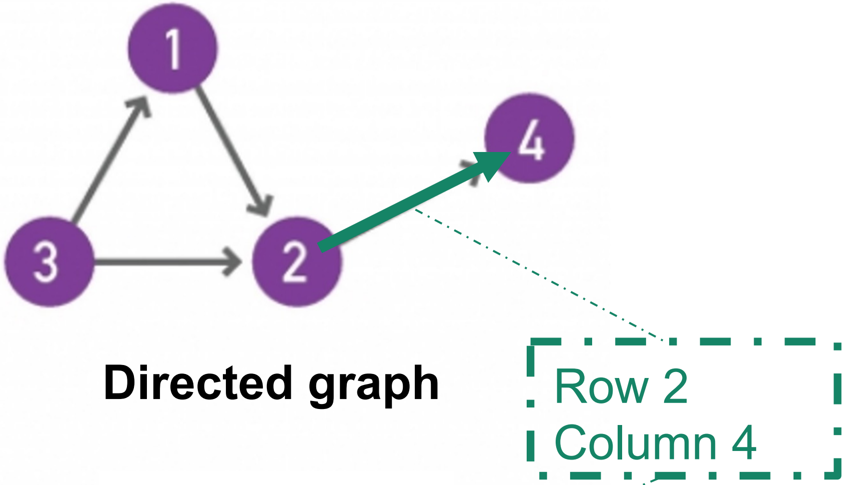
$$A_{ij} = \begin{matrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

A_{ij} always means row i, column j



Undirected graph

$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$



Directed graph

$$A_{ij} = \begin{matrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

Properties of adjacency matrices

- G is undirected ⇔ A is symmetric
- G has a self-loop
  ⇔ A has a non-zero element in the diagonal
- G is complete ⇔ $A_{ij} \neq 0$ (except if i=j)

Quick Exercise

- In terms of A, what is the expression for:

$$k_i^{\text{in}} =$$

$$k_i^{\text{out}} =$$

Weighted networks

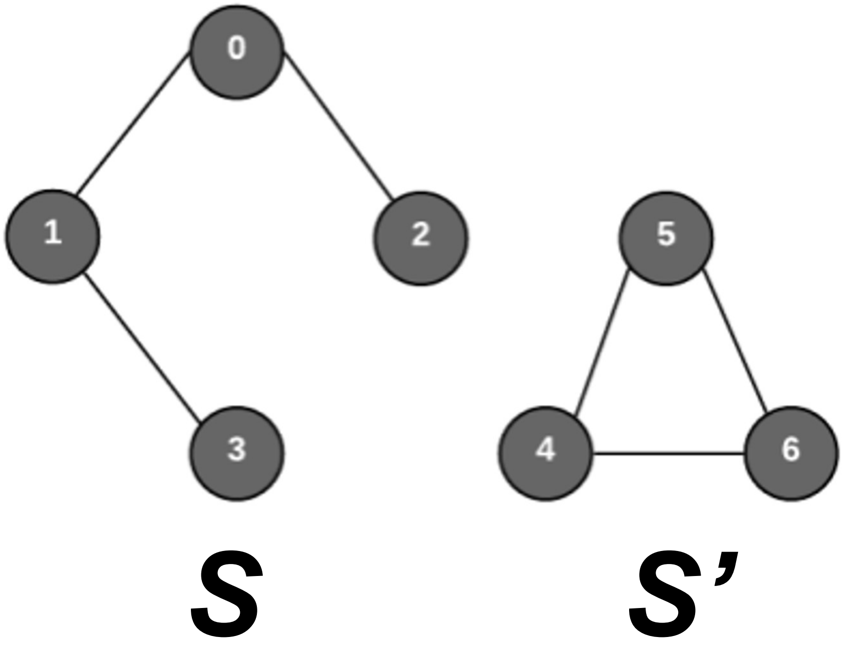
- Element w_{ij} indicates the **weight of the link** from node i to node j
- The link from node B to node C has a weight of 2
- Out-strenght (weighted out-degree): $s_i^{out} = \sum_j A_{ij}$
- In-strenght (weighted in-degree): $s_i^{in} = \sum_j A_{ji}$
- Total weight: $W = \sum_i s_i^{in} = \sum_i s_i^{out} = \sum_{ij} A_{ij}$

$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \end{array}
\begin{array}{c}
\begin{array}{cccccc} A & B & C & D & E & F \end{array} \\
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
2 & 1 & 3 & 1 & 1 & 0
\end{pmatrix}
\end{array}.
$$

If a graph is disconnected

Disconnected graphs have adjacency matrices with **block structure**

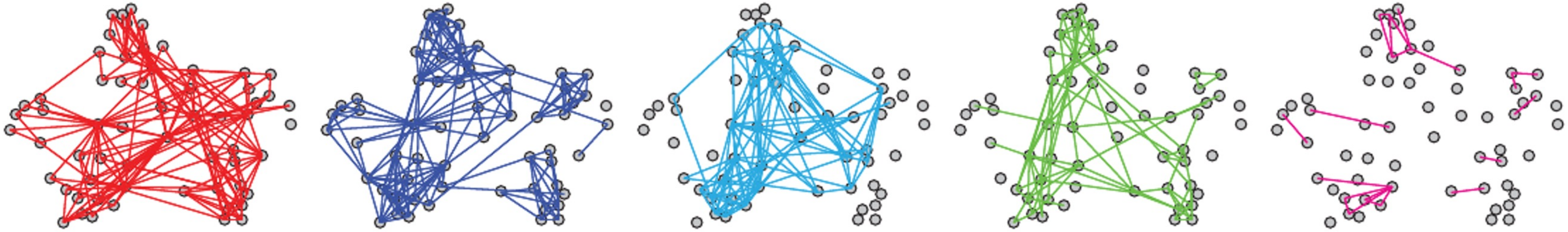
$$A = \begin{bmatrix} S & 0 \\ 0 & S' \end{bmatrix}$$



Beyond simple graphs

Some networks are multi-layer

- Multi-layer graphs have different edges **over the same nodes**
- Example graph of 61 employees of a university department, with links indicating: coworking, having lunch together, being connected on Facebook, being actual friends, or being co-authors

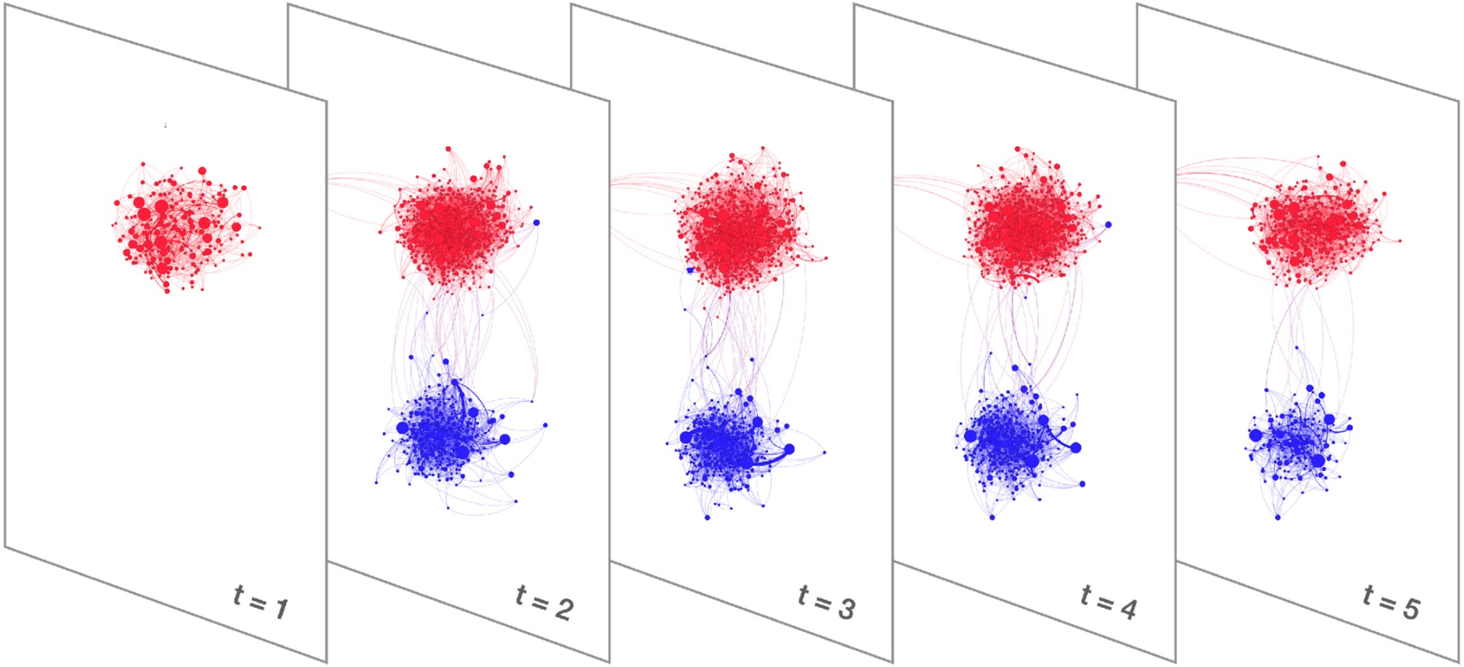


(a) Work. (b) Lunch. (c) Facebook. (d) Friend. (e) Coauthor.

https://dl.acm.org/doi/abs/10.1145/2898362

Some networks are time-evolving

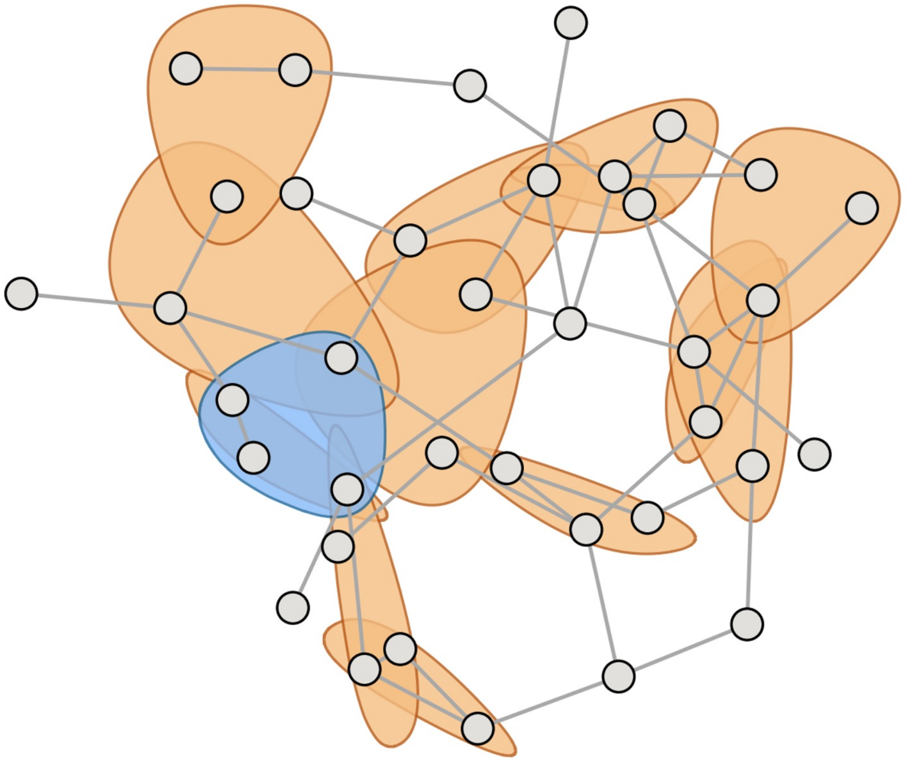
Temporal, or "time-evolving" networks
At each timestep there are new nodes and/or edges (and/or deletions)



Source: Menczer, Fortunato, David: A First Course on Networks Science. Cambridge, 2020.

Some networks are higher-order

Higher-order networks, or “hypergraphs”:
Hyper-links involve more than two nodes

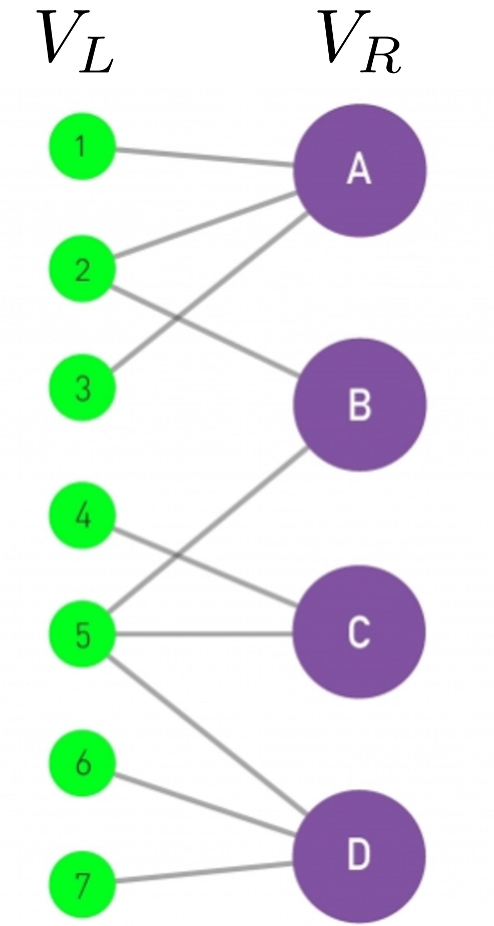


Source: Menczer, Fortunato, David: A First Course on Networks Science. Cambridge, 2020.

Some networks are **bi-partite**

- A **bipartite** graph is a graph
- G = (V,E) such that

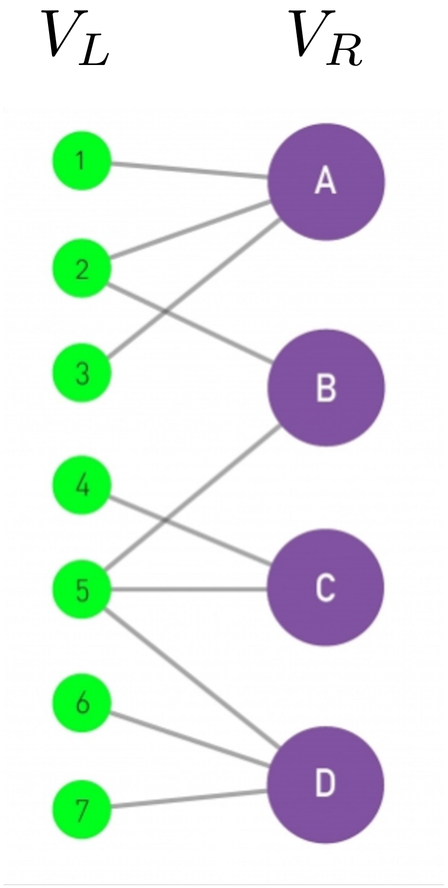
$$V = V_L \cup V_R, V_L \cap V_R = \emptyset, E \subseteq V_L \times V_R$$



Exercise: project a bipartite network

?

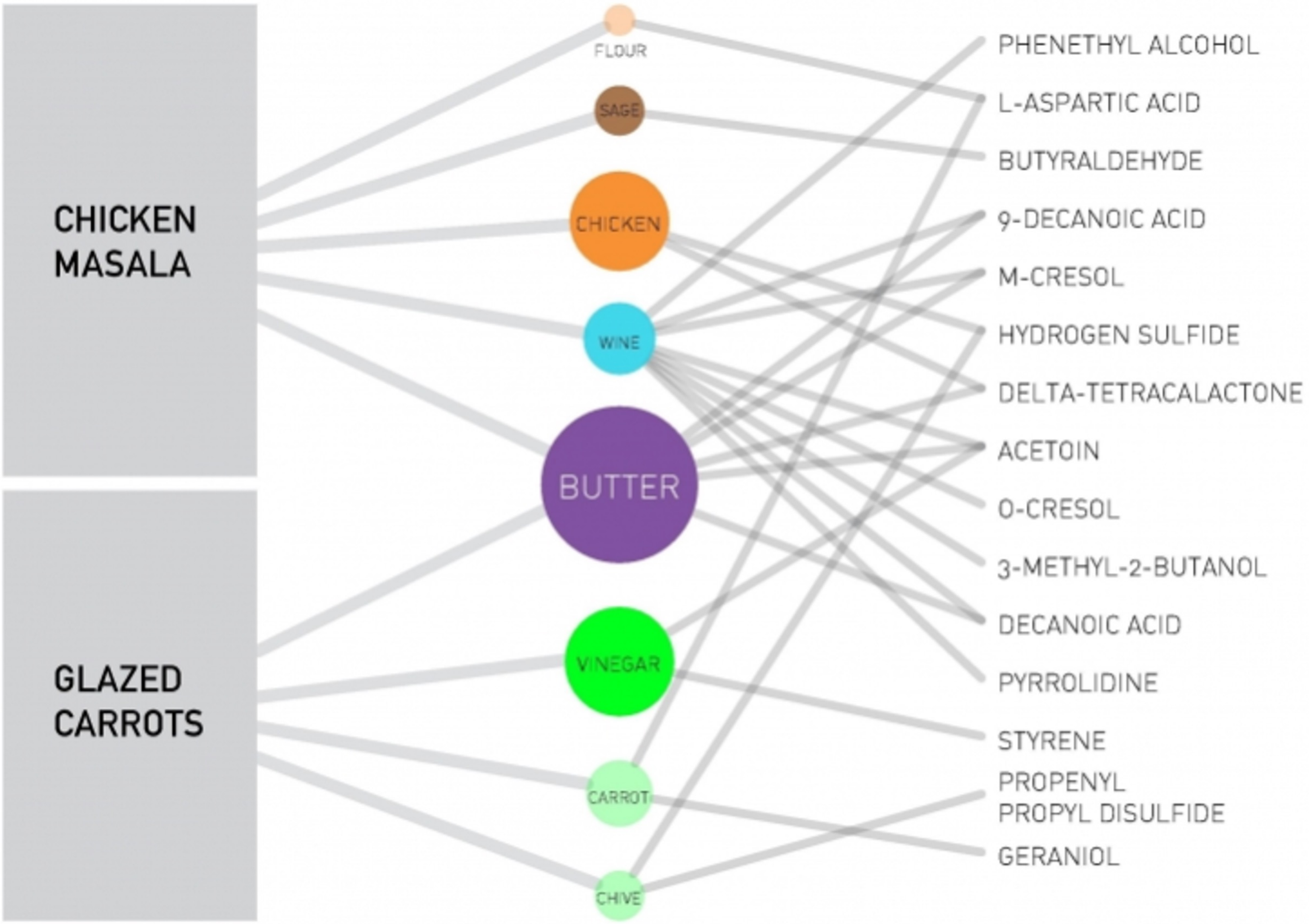
Left projection: graph where nodes are 1, 2, …, 7 and nodes are connected if they share a neighbor



?

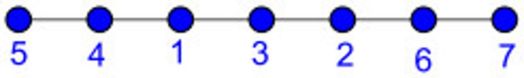
Right projection: graph where nodes are A, B, …, D and nodes are connected if they share a neighbor

Tripartite network

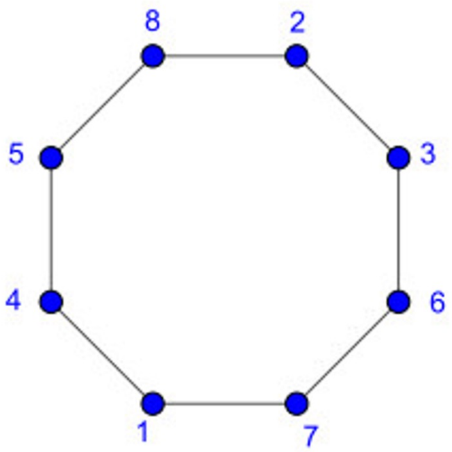


Some graphs have a name

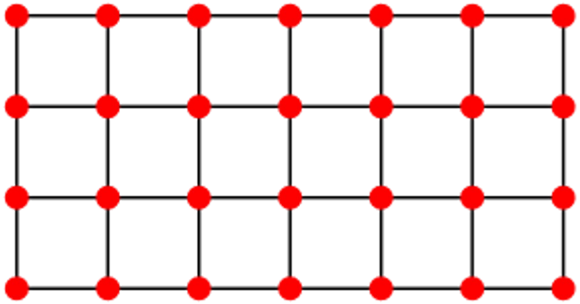
Line



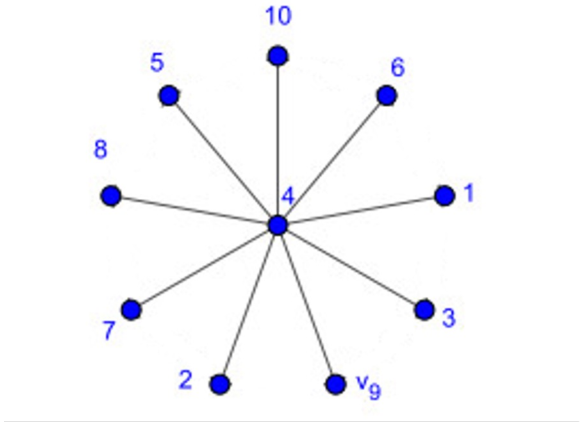
Cycle



Lattice



Star



These are not complex networks!

Clique and Bi-partite clique

- A **clique** is a complete (sub)g $E = (V \times V)$
- An **n-clique** is a complete graph of n nodes
- A **bi-partite clique** is such that

$$V = V_1 \cup V_2, V_1 \cap V_2 = \emptyset, E = (V_1 \times V_2)$$

- A **(n_1, n_2)-clique** is a bipartite clique such that

$$|V_1| = n_1, |V_2| = n_2$$

Summary

Things to remember

- Definitions
  - degree, in-degree, out-degree, strength
  - time-evolving graph, multi-layer graph
  - line graph, cycle graph, star graph, lattice, bi-partite graph, clique
- Writing the adjacency matrix of a graph, and drawing a graph given its adjacency matrix
- Plotting the degree distribution of a graph
- Projecting a bi-partite graph

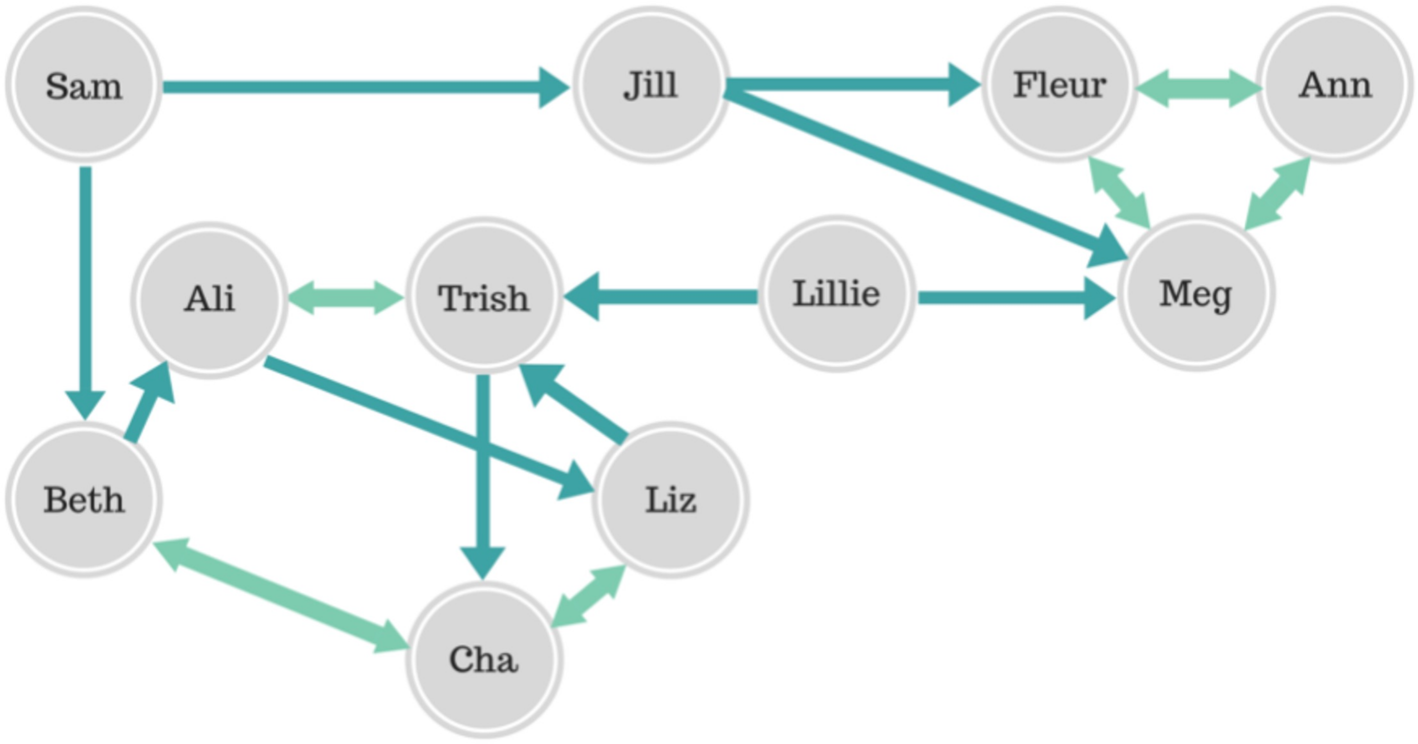
Sources

- A. L. Barabási (2016). Network Science – Chapter 02
- URLs cited in the footer of specific slides

Practice on your own

Draw the indegree, outdegree, degree distribution

Write the adjacency matrix



https://www.6seconds.org/2017/07/03/20125/

Practice on your own

How do you call the sub-graph induced by nodesets:

- {H, A, B}
- {G, H, D}
- {B, D, E, G}
- {A, B, D, E}