Sparsity and Connectivity

Introduction to Network Science

Carlos Castillo

Topic 05

Universitat Pompeu Fabra Barcelona

Contents

- Degree
- Sparsity
- Bi-partite networks
- Connectedness

Sources

- Albert László Barabási: Network Science. Cambridge University Press, 2016.
  - Follows almost section-by-section chapter 02
- URLs cited in the footer of specific slides

Real 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

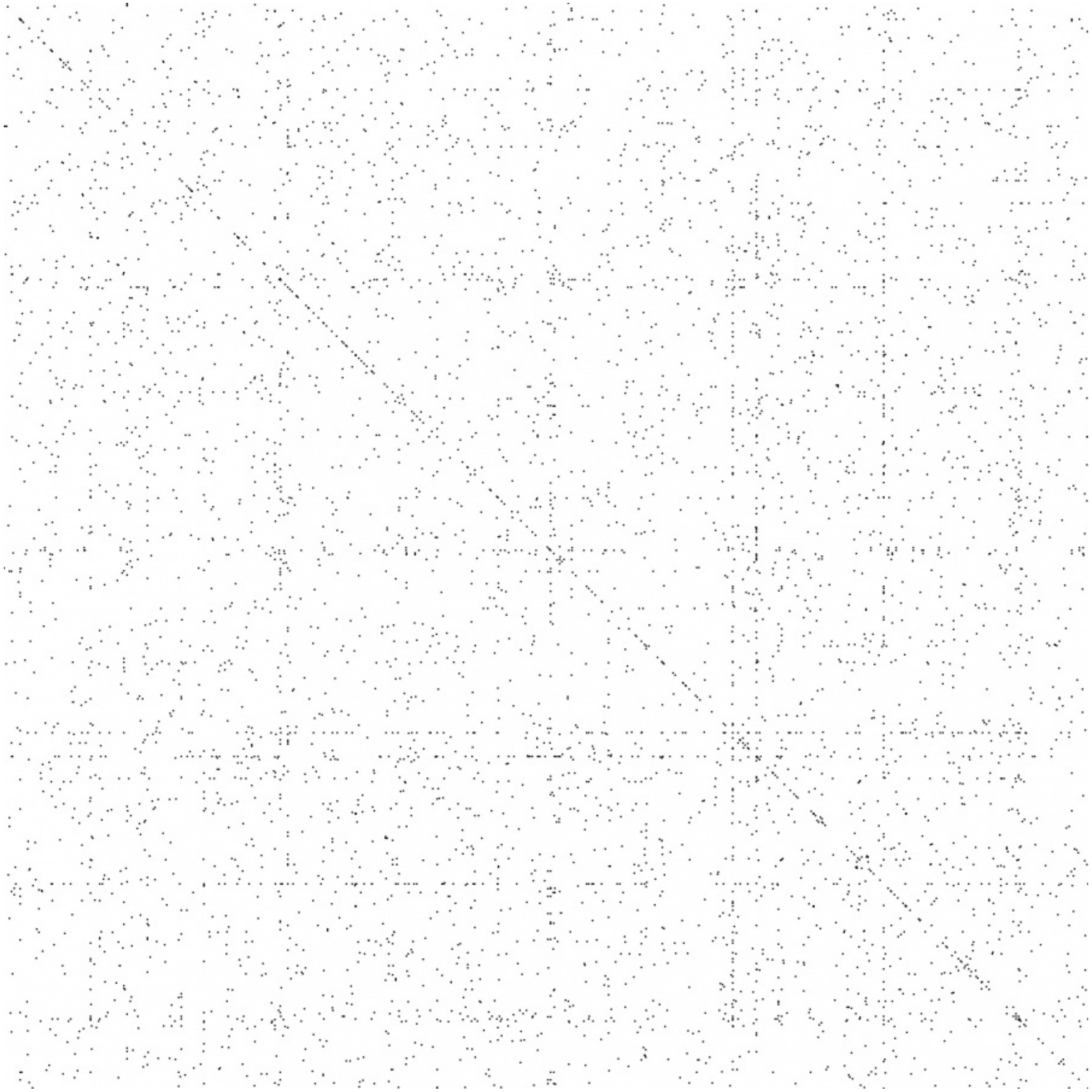
How sparse are some networks?

Network	|V|	|E|	Max |E|
Zachary’s Karate Club	34	78	561
Les Misérables	77	254	2962
E-mail exchanges	868	25K	376K
US companies ownership	1351	6721	911K
Marvel comics	6K	167K	17M

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

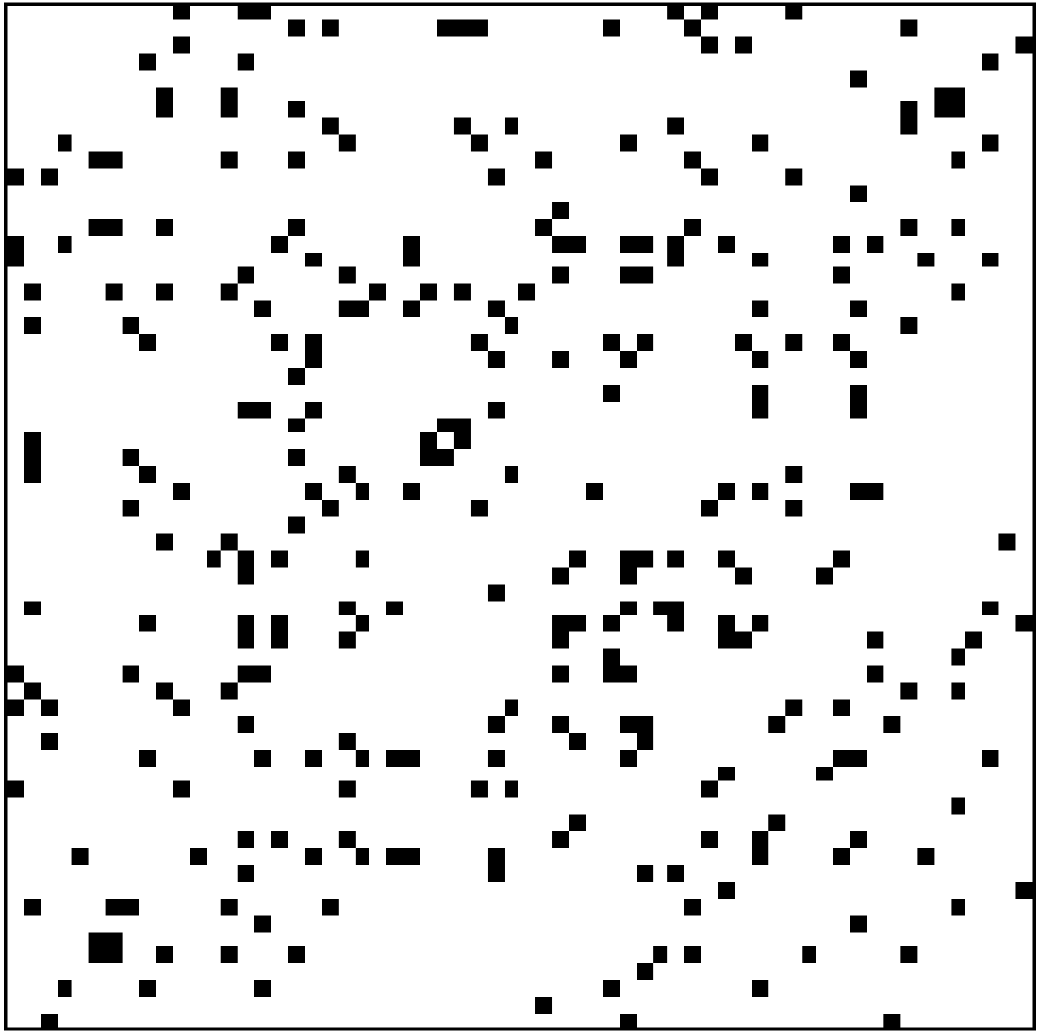
Example: protein interaction network

(N=2K, L=3K)



Example: dolphins

(N=62, L=318)



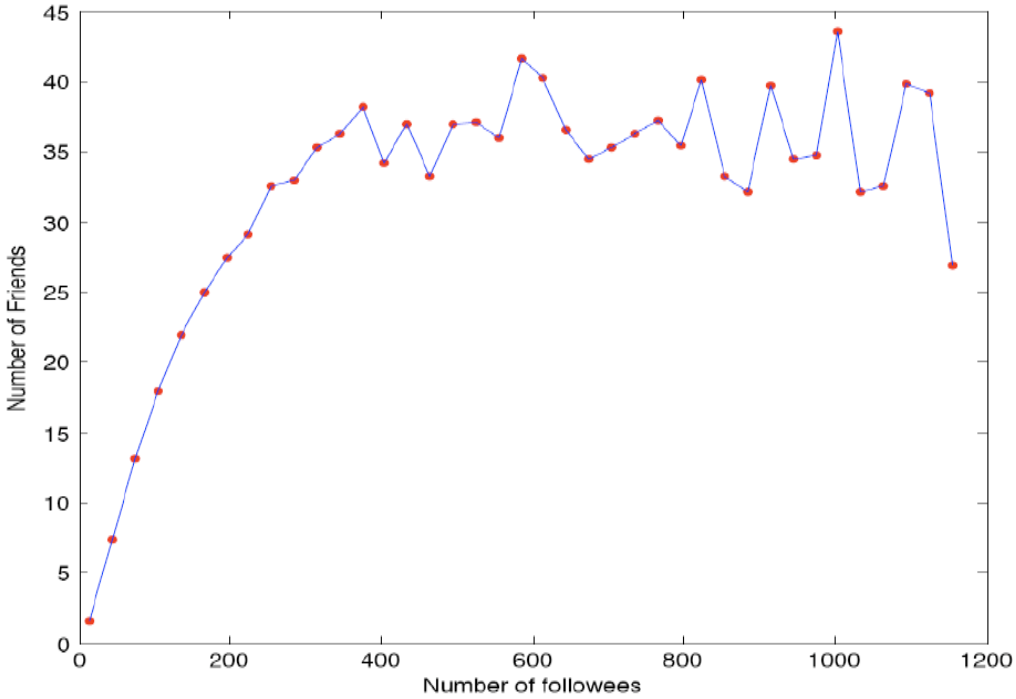
https://www.cise.ufl.edu/research/sparse/matrices/Newman/dolphins.html

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 (≃150)

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

Example: actual friends in Twitter vs people you follow in Twitter

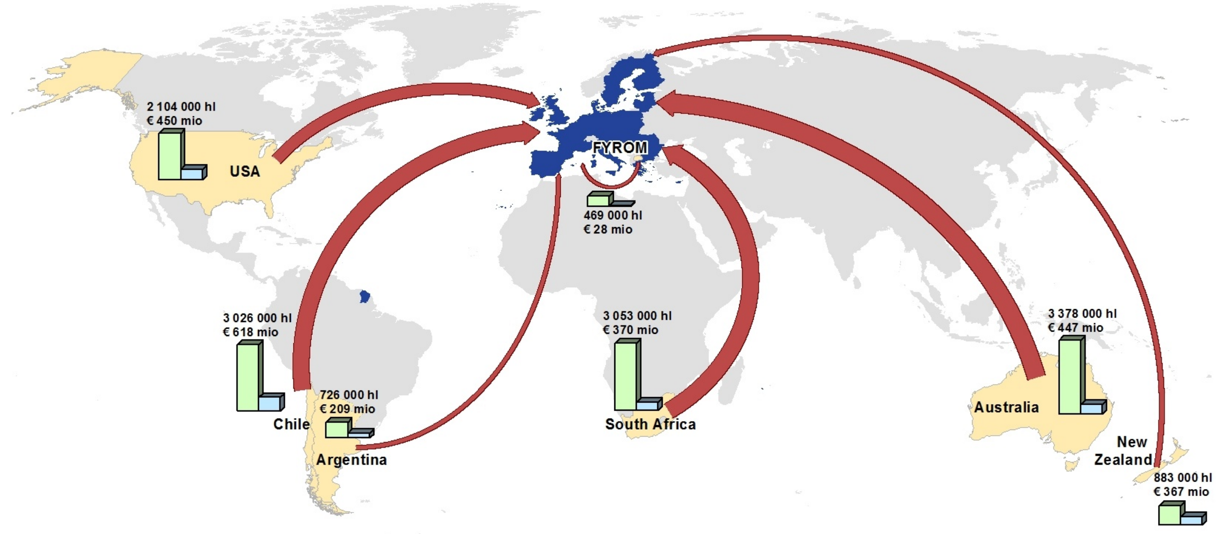


Huberman, B., Romero, D. M., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. First Monday, 14(1).

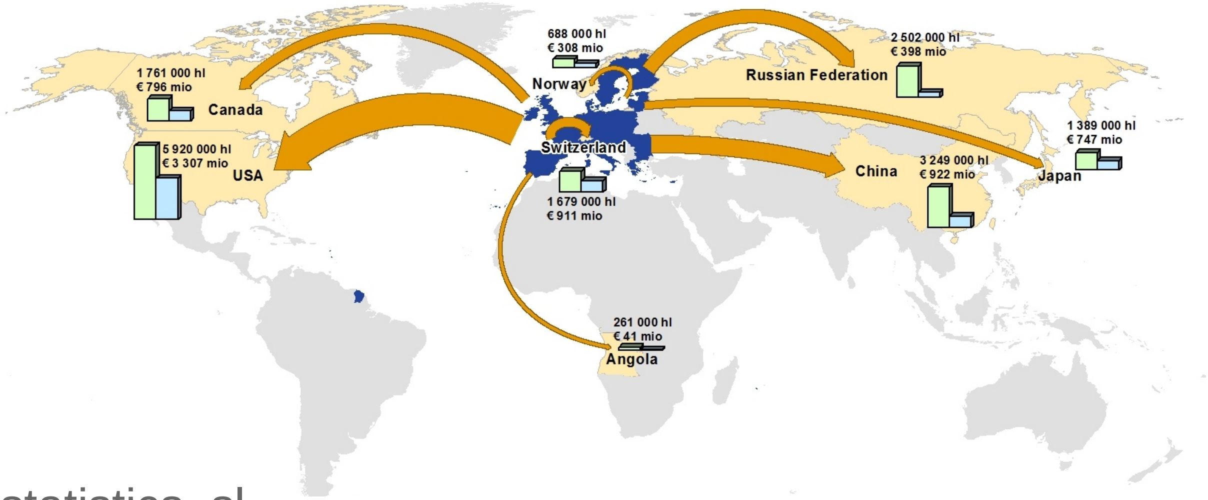
Weighted networks

- In weighted networks, instead of $A_{ij} \in [0, 1]$
- We have that $A_{ij} \in \mathbb{R}$
- Weights may represent different tie strengths

Example: weighted networks



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

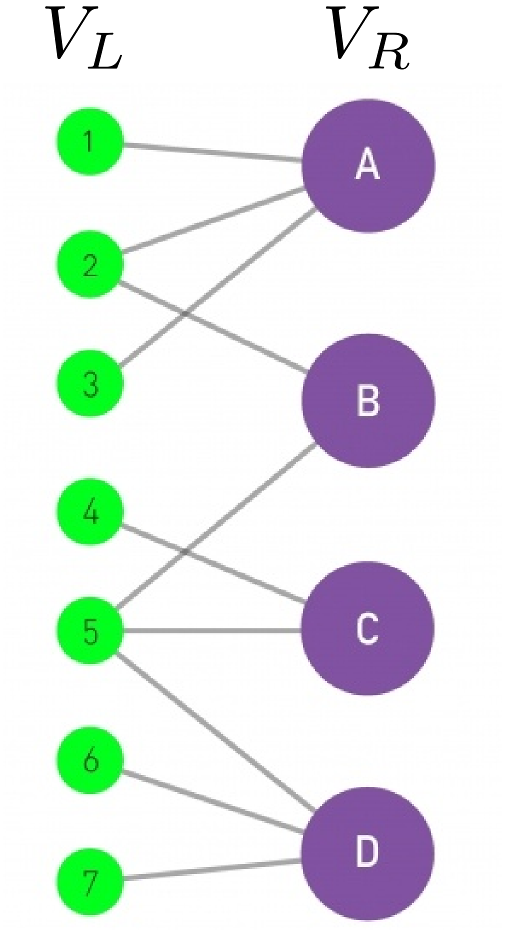


https://ec.europa.eu/agriculture/wine/statistics_sl

Bipartite networks

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

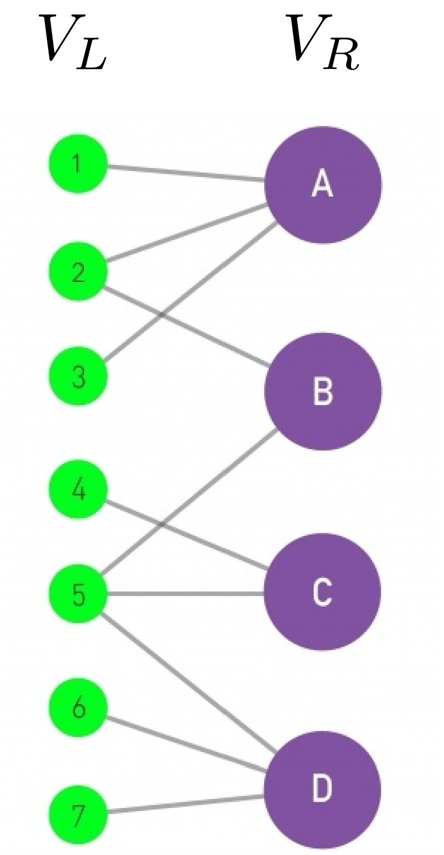
$$V = V_L \cup V_R, V_L \cap V_R = \emptyset, E \subseteq V_L \times V_R$$



Exercise: project a bipartite network

?

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

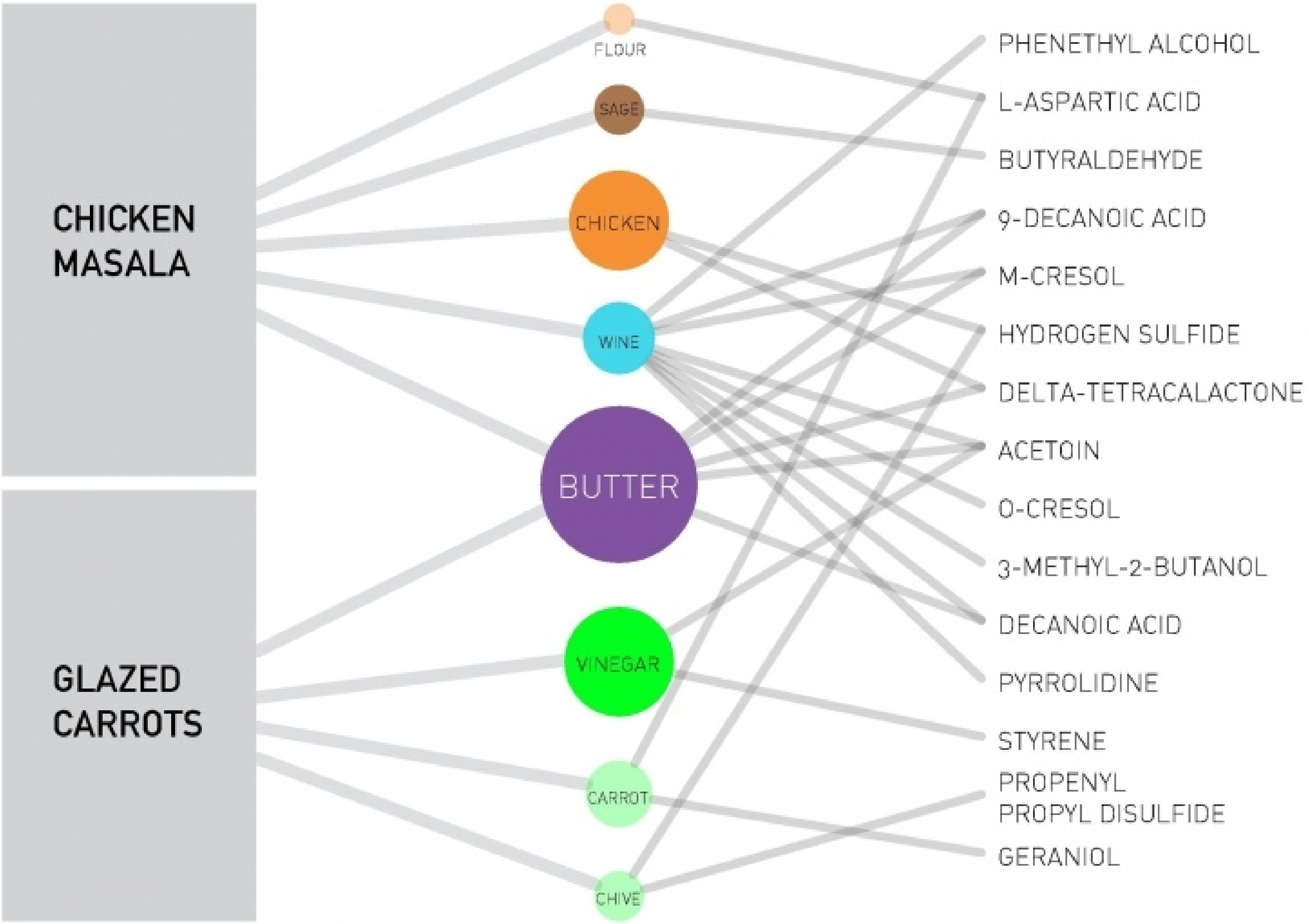


?

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

Draw in Nearpod Collaborate
https://nearpod.com/student/
Code to be given during class

Tripartite network



Clique and Bi-partite clique

- A **clique** is a complete (sub)graph: $E = (V \times V)$
- 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

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

The word “clique” in popular culture

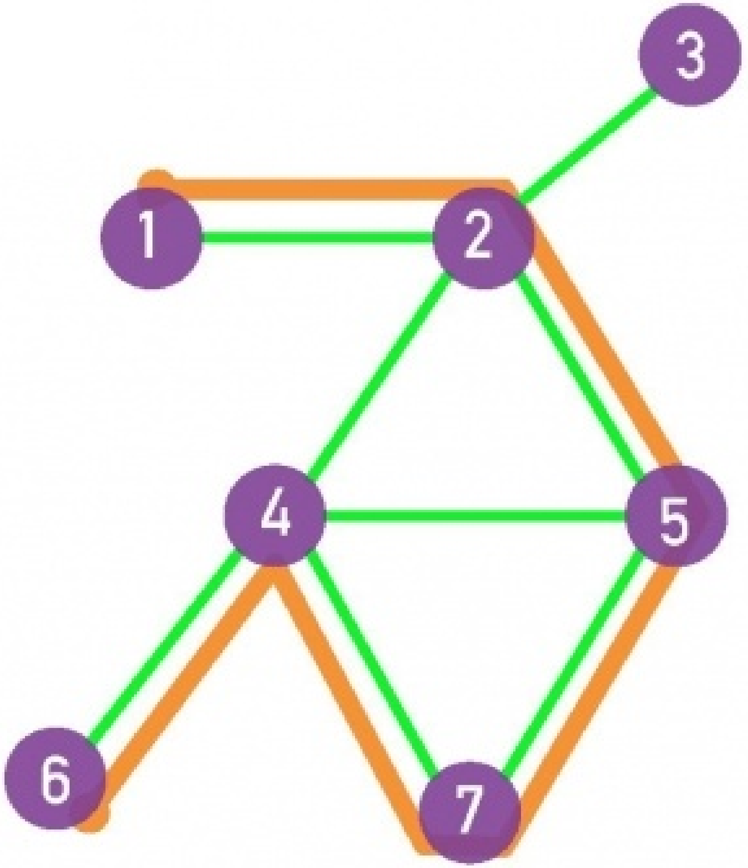
In some parts of Latin America, a “***clika***” or “***clica***” means a close group of friends, sometimes a gang

Photo credit: @astro_jr

Paths and distances

Paths

- A path is a sequence of edges from E
- The destination of each edge is the origin of the next edge
- The length of the path is the number of edges on it
- Example: a path marked in orange, having length 5



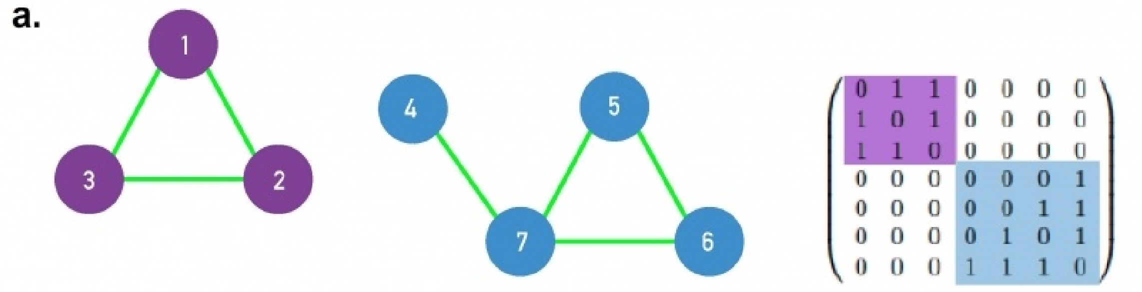
Connectedness

- If a path exists between two nodes i, j:
  - those nodes are part of the same **connected component**
- A graph that has only one connected component is called a **connected graph**

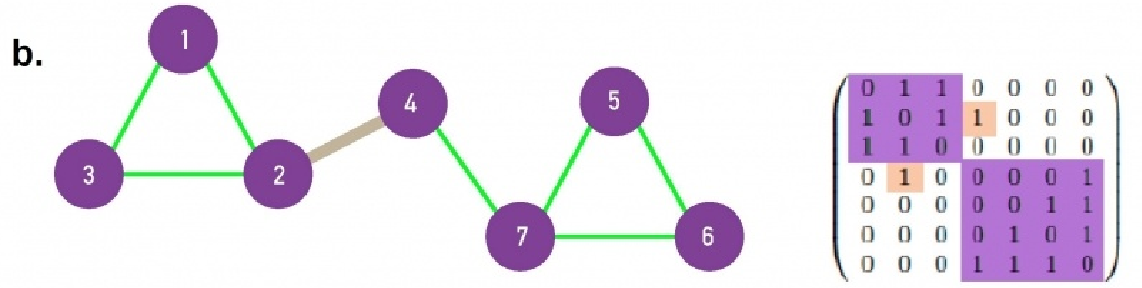
Connected graphs

A disconnected graph has an adjacency matrix that can be arranged in block diagonal form

a. disconnected



b. connected



Distance

- If two nodes i, j are in the same connected component:
  - the **distance** between i and j, denoted by d_{ij} is the **length of the shortest path** between them

Diameter

- The **diameter** of a network is the maximum distance between two nodes on it, d_{max}
- The **effective diameter** (or effective-90% diameter) is a number d such that 90% of the pairs of nodes (i,j) are at a distance smaller than d
- The **average distance** is <d>, and is measured only for nodes that are in the same connected component

Summary

Things to remember

- Definitions:
  - Degree, in-degree, out-degree
  - Bi-partite graph, clique
  - Sparse vs dense graph
- Distance, diameter, effective diameter
- Connected components

Practice on your own

- Measure the sparsity of a graph $L/L_{\max}$
- Compute the distance between two nodes
- Compute the diameter of a graph
- Identify connected components