

# Graph theory basics

Introduction to Network Science

Carlos Castillo

Topic 02



Universitat  
Pompeu Fabra  
*Barcelona*

# Contents

- Degree
- Sparsity
- Bi-partite networks
- Connectedness

# Sources

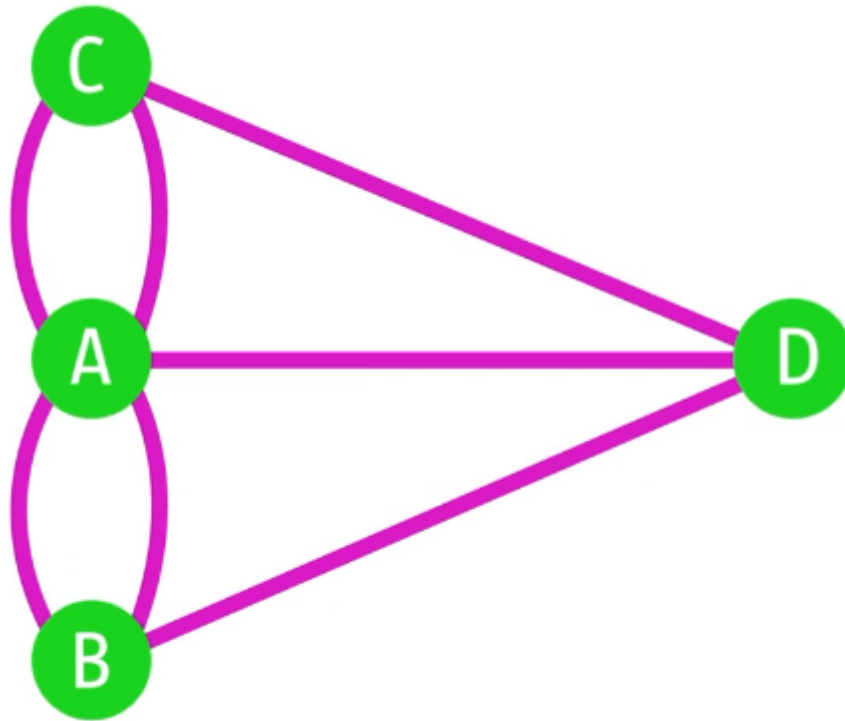
- Albert-László Barabási: Network Science. Cambridge University Press, 2016.
  - Follows almost section-by-section chapter 02
- URLs cited in the footer of specific slides

# The seven bridges of Königsberg



<http://networksciencebook.com/images/ch-02/video-2-1.m4v>

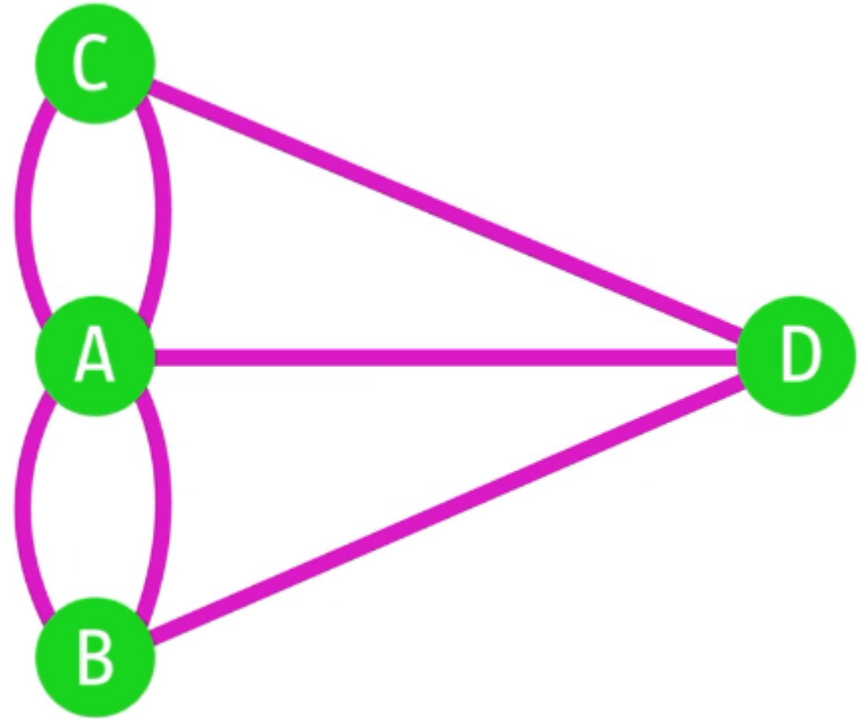
Can one walk across the 7 bridges without crossing the same bridge twice?



# Basic concepts

# Notation for a graph

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



# Typical notation variations

- You may find that  $G$  is denoted by  $(N, A)$ , this is typical of directed graphs
- You may find that
  - $|V|$  is denoted by  $n$  or  $N$
  - $|E|$  is denoted by  $m$ ,  $M$ , or  $L$



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

# Example graphs we will use

Network	$ V $	$ E $
Zachary's Karate Club (karate.gml)	34	78
Les Misérables (lesmiserables.gml)	77	254
E-mail exchanges (email-eu-core.csv)	868	25K
US companies ownership	1351	6721
Marvel comics (hero-network.csv)	6K	570K

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

# In directed networks

- 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:

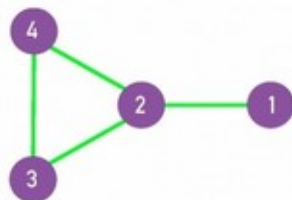
$$L = \sum_{i=1}^N k_i^{\text{in}} = \sum_{i=1}^N k_i^{\text{out}}$$

# Degree distribution

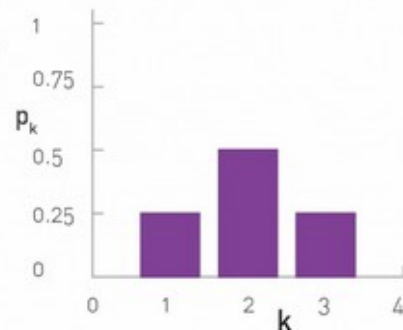
- If there are  $N_k$  nodes with degree  $k$
- The degree distribution is given by  $p_k = \frac{N_k}{N}$
- The average degree is then  $\langle k \rangle = \sum_{k=0}^{\infty} k p_k$

# Degree distribution; two toy graphs

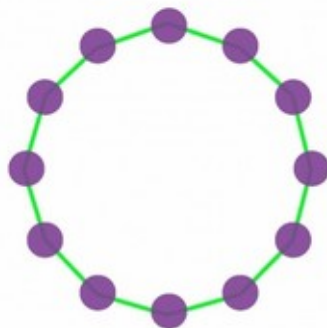
a.



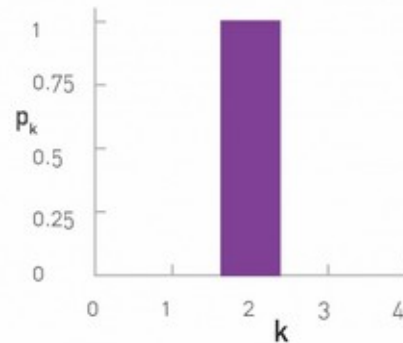
b.



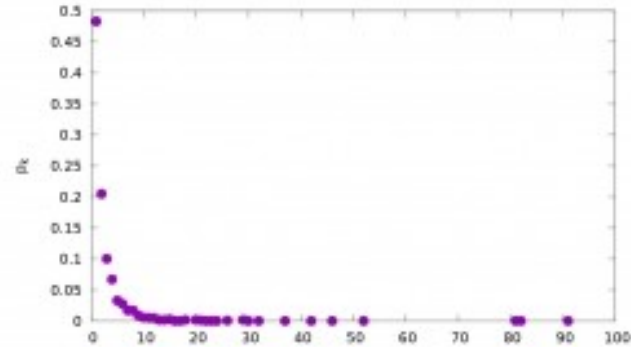
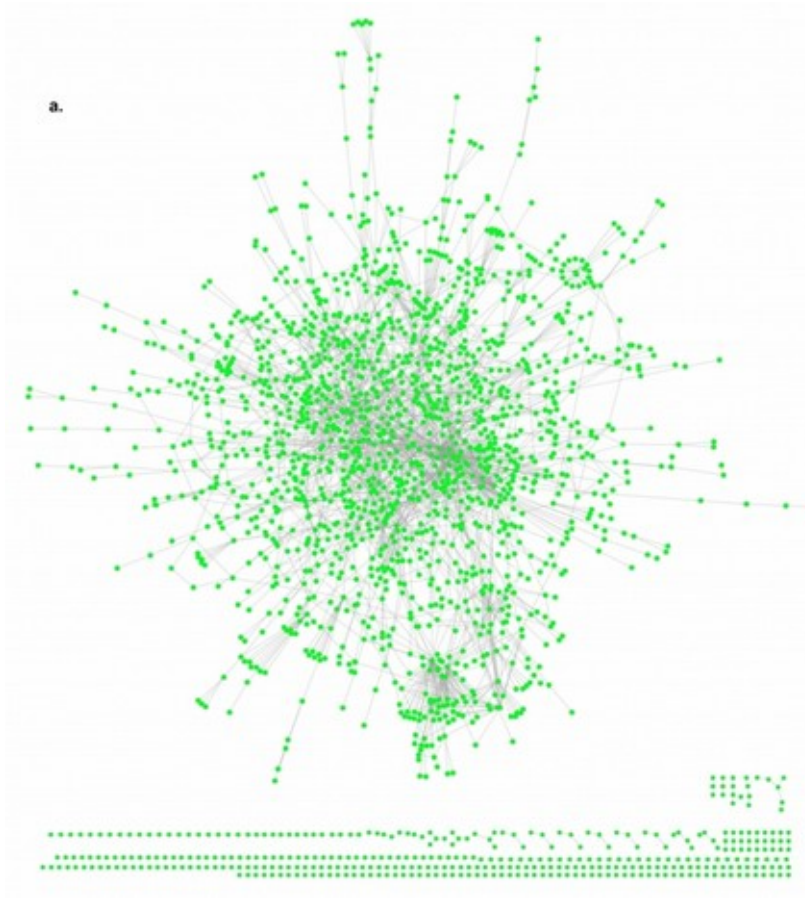
c.



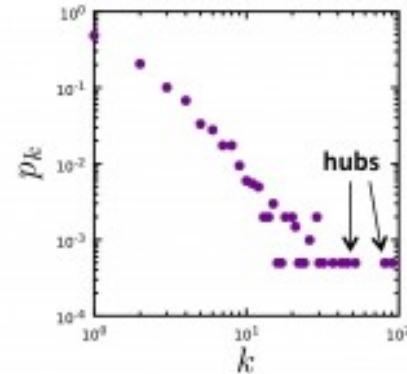
d.



# Degree distribution; real graph



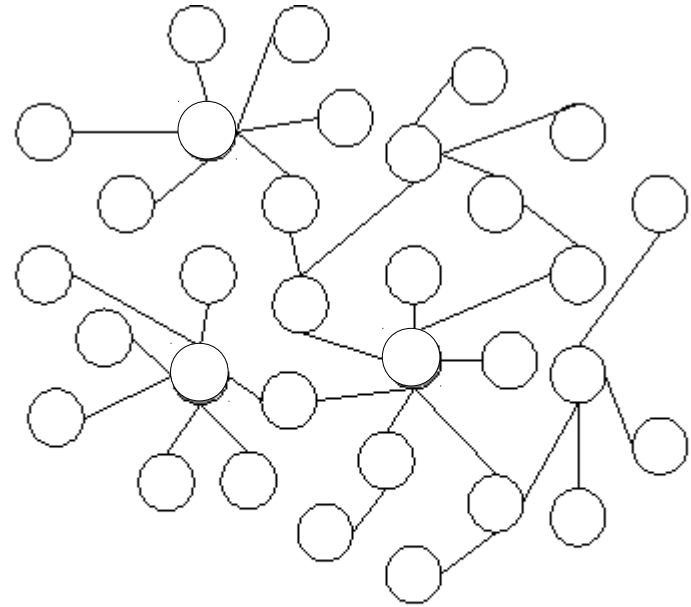
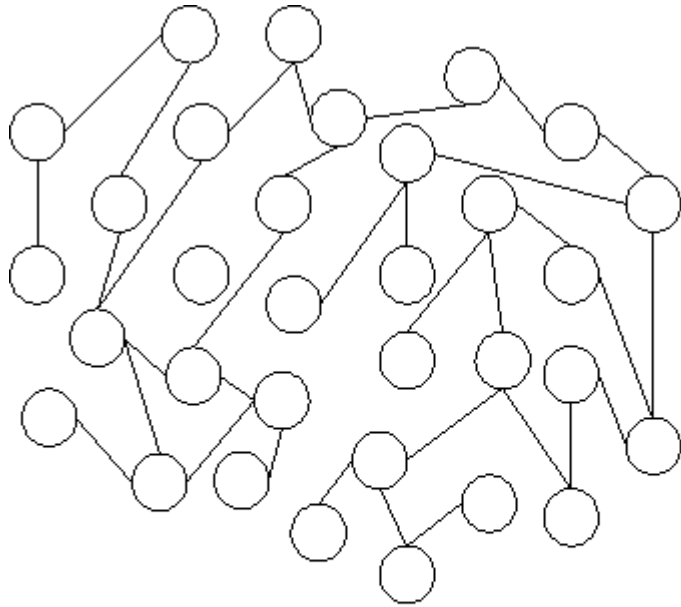
Linear  
scale



Log-log  
scale

# Exercise

- Draw the degree distribution of these graphs



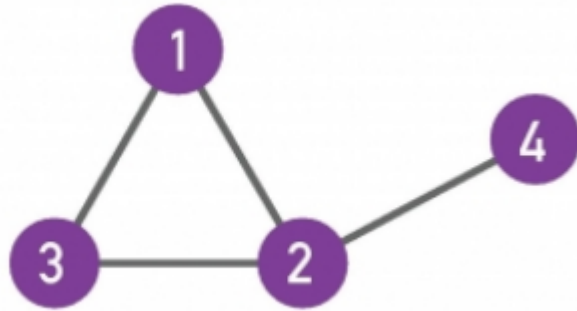


# Adjacency matrix

# What is an adjacency matrix

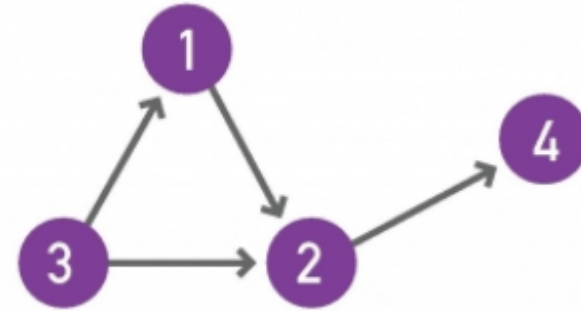
- $A$  is the adjacency matrix of  $G = (V, E)$  iff:
  - $A$  has  $|V|$  rows and  $|V|$  columns
  - $A_{ij} = 1$  if  $(i,j) \in E$
  - $A_{ij} = 0$  if  $(i,j) \notin E$

# Examples



Undirected graph

$$A_{ij} = \begin{matrix} & 1 & 2 & 3 & 4 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix} \end{matrix}$$



Directed graph

$$A_{ij} = \begin{matrix} & 1 & 2 & 3 & 4 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{matrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} \end{matrix}$$

# Quick exercise

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

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

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

# Some “graphology” ...

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

# Real networks are sparse

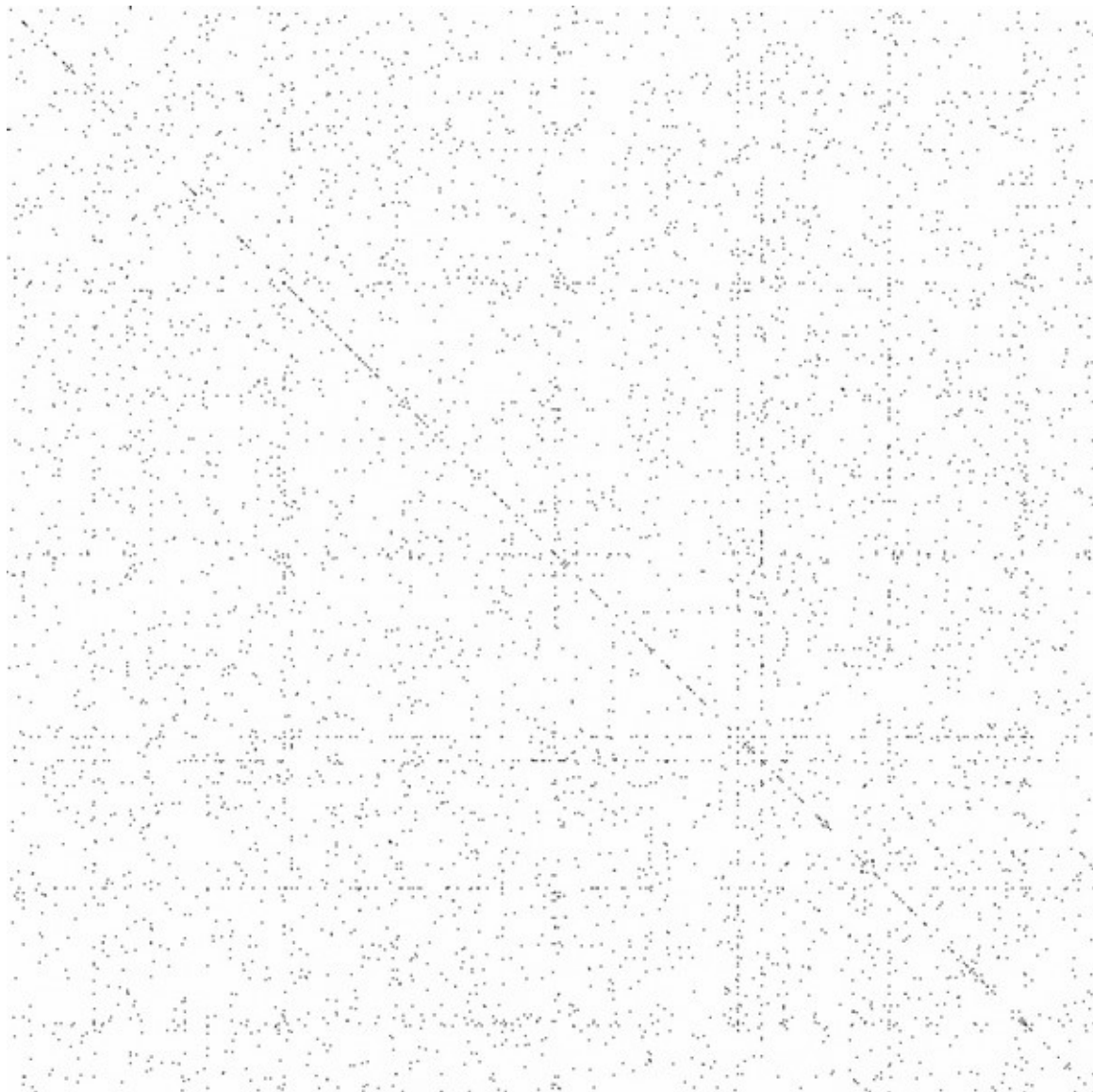
- Theoretically  $L_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$
- Most real networks are sparse, i.e.,  $L \ll L_{\max}$

# 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	570K	17M

# Example: protein interaction network

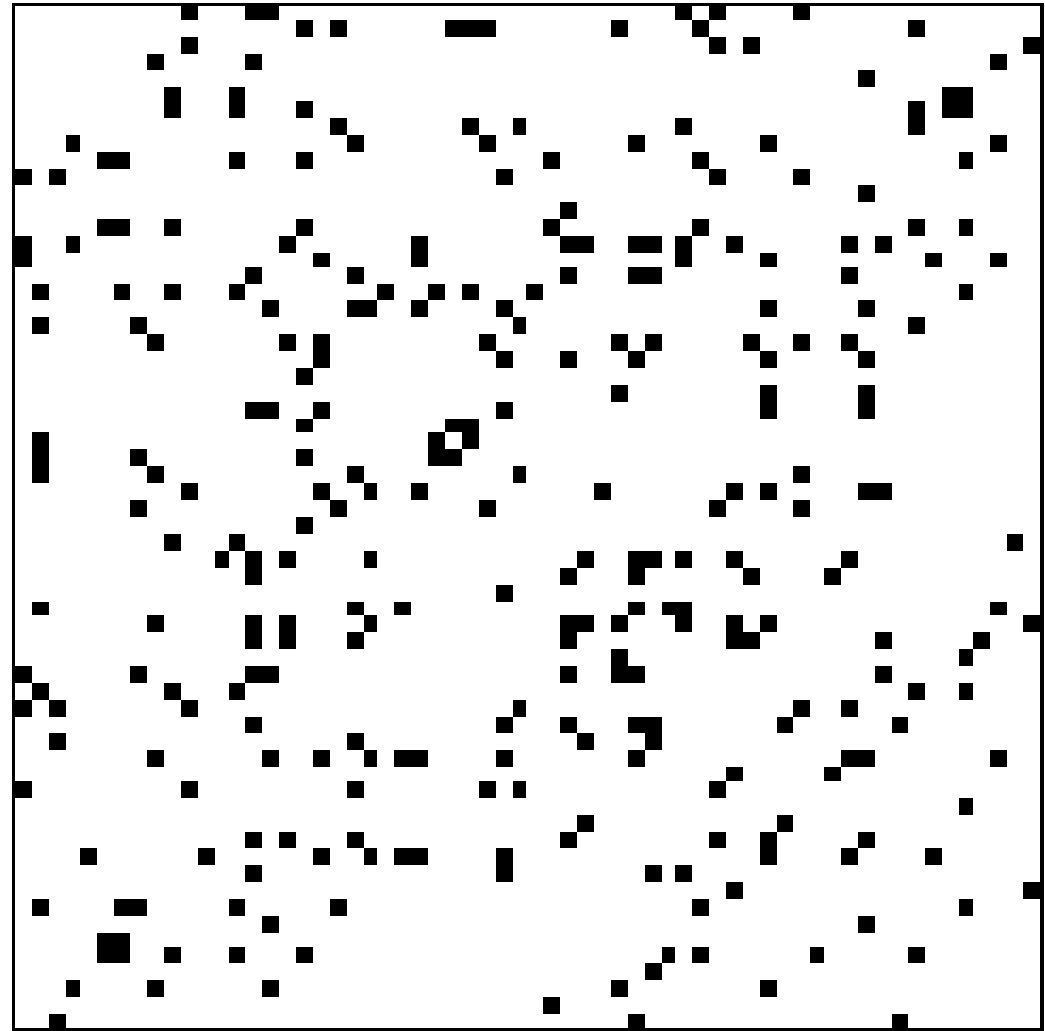
( $N=2K$ ,  $L=3K$ )





Example:  
dolphins

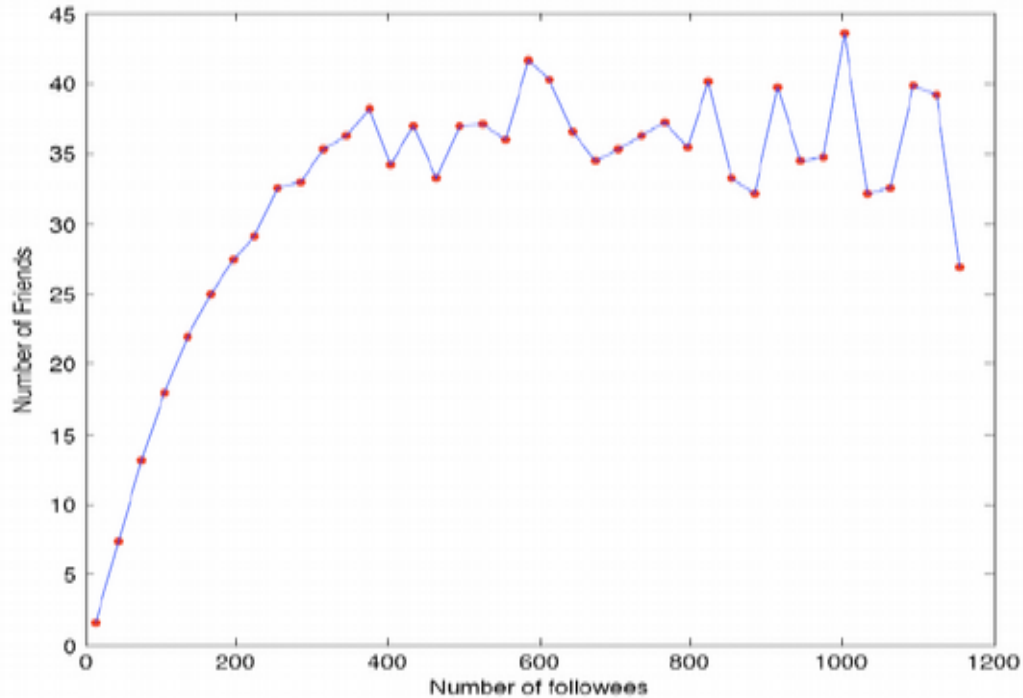
( $N=62$ ,  $L=318$ )



# 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 ( $\approx 150$ )

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

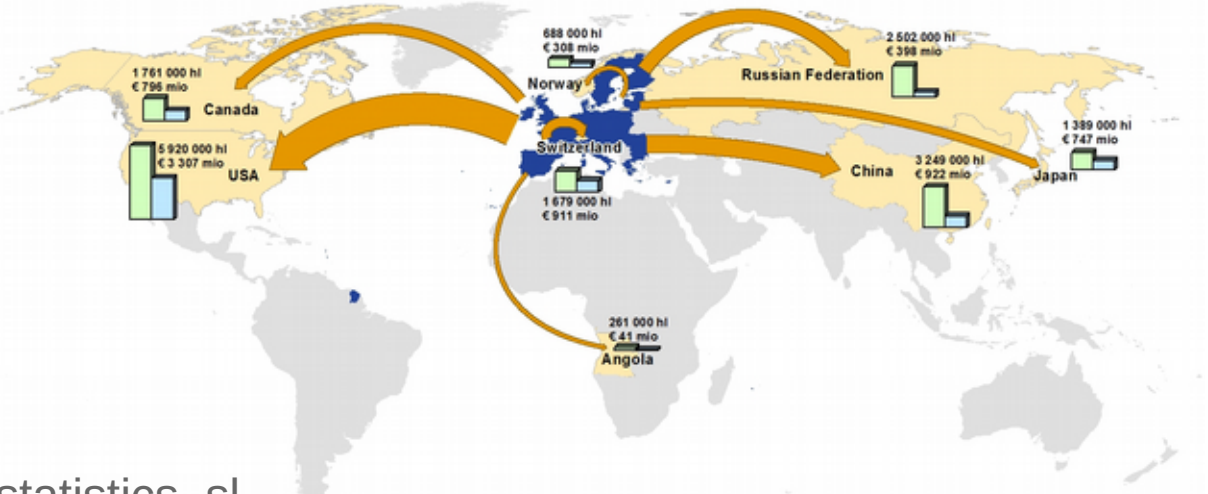
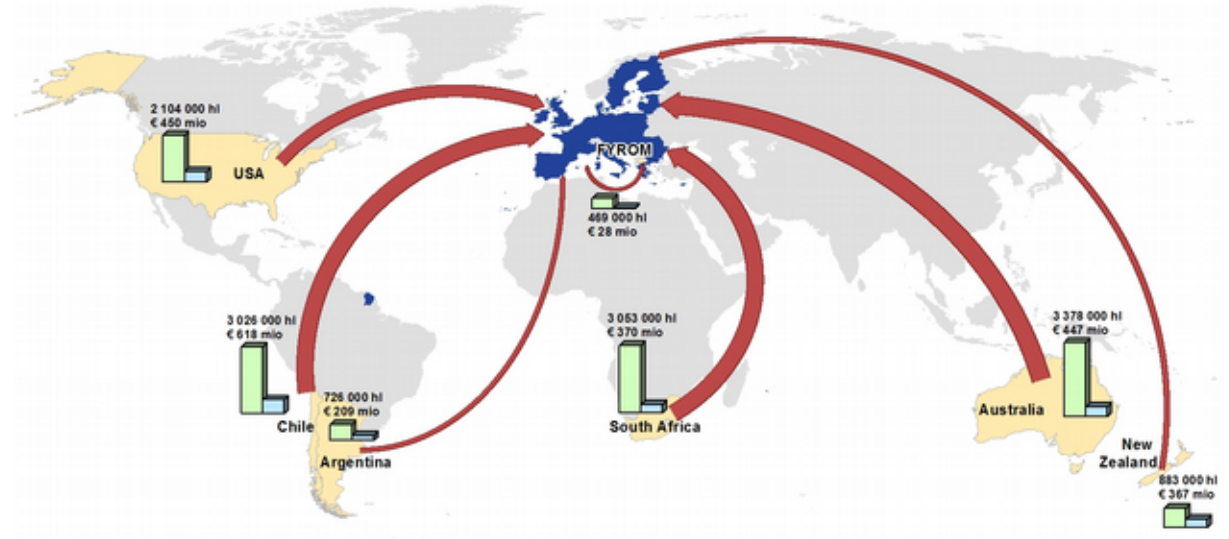


# Weighted networks

- 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

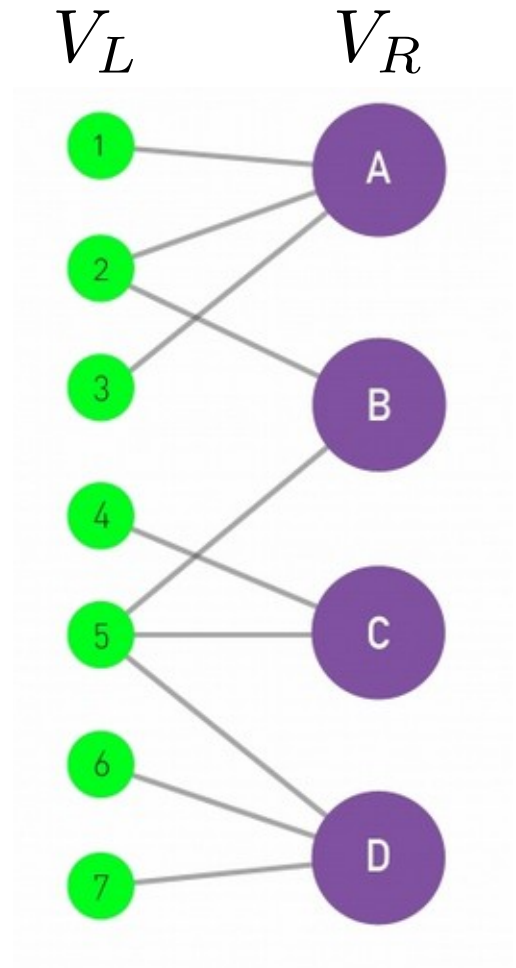
EU imports (top) and  
exports (bottom) of wine



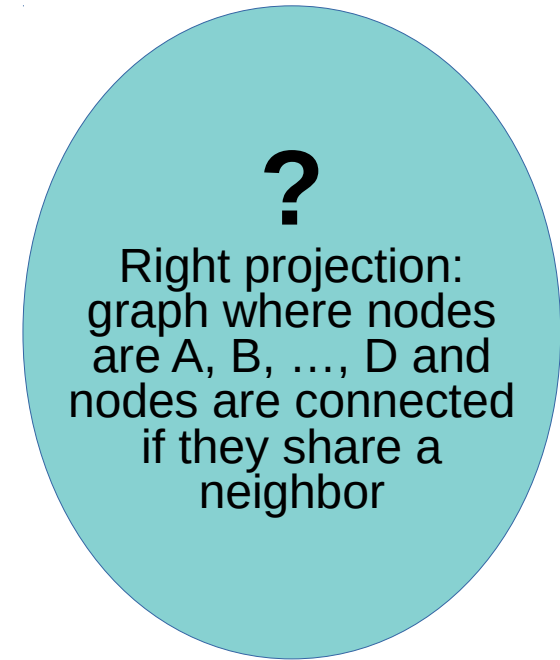
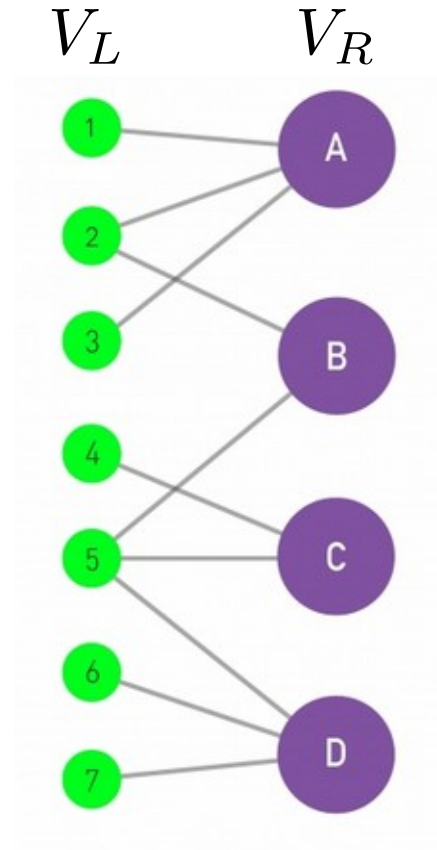
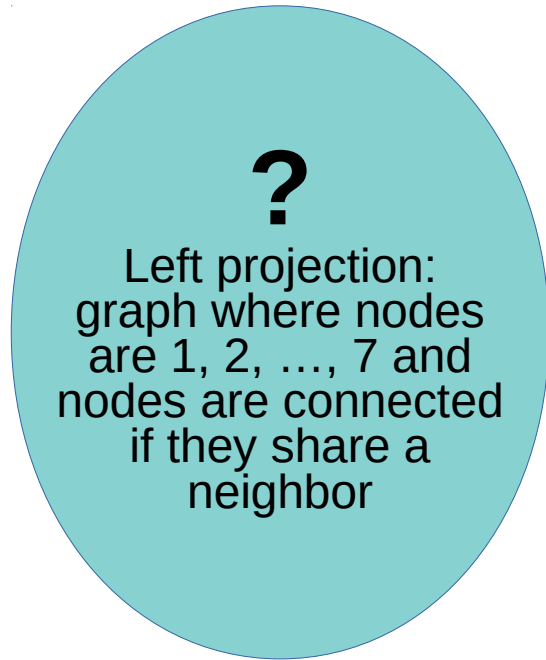
# Bipartite networks

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

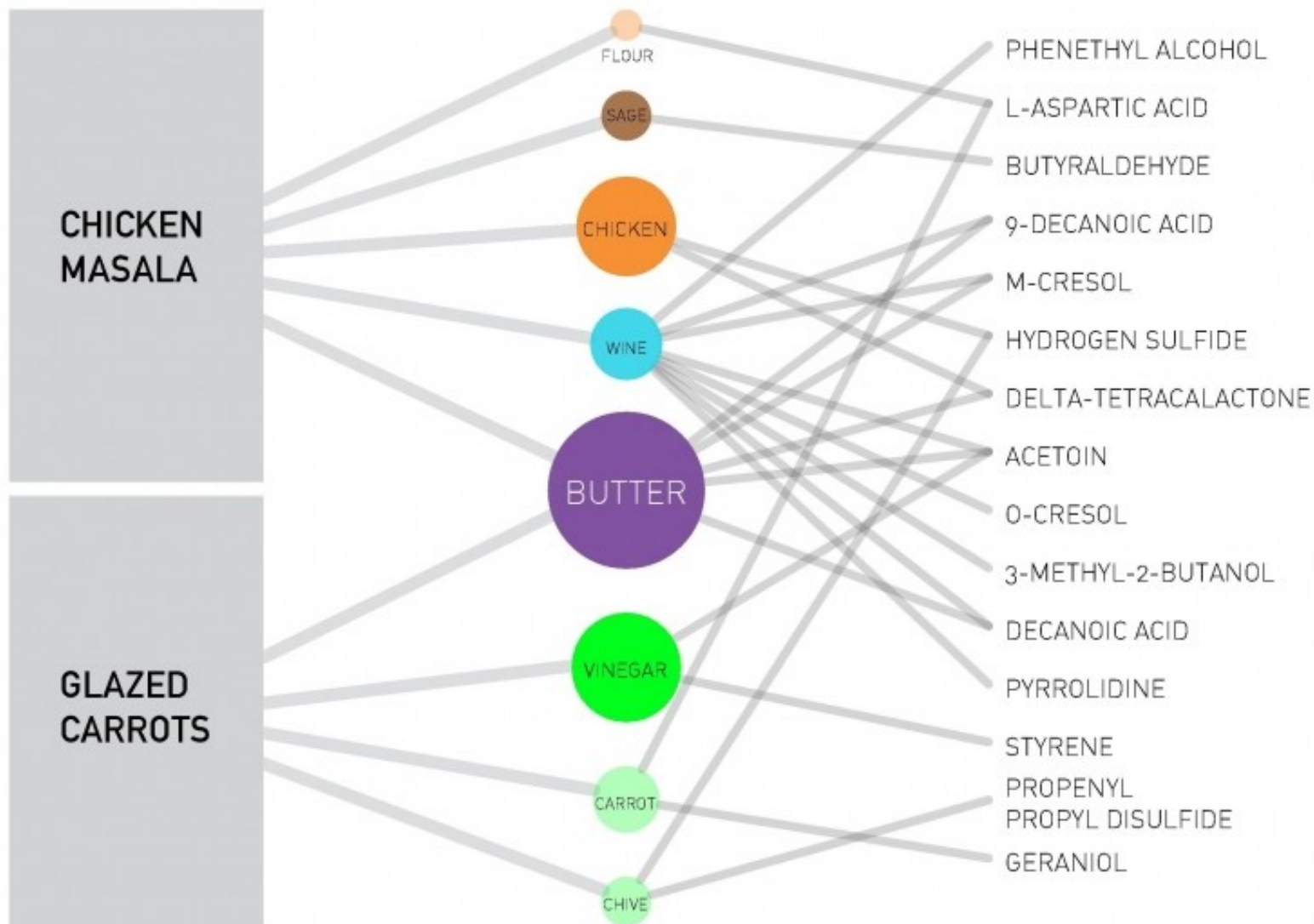
$$V = V_L \cup V_R, V_L \cap V_R = \emptyset, E \subseteq V_L \times V_R$$



# Projecting a bipartite network



# 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

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

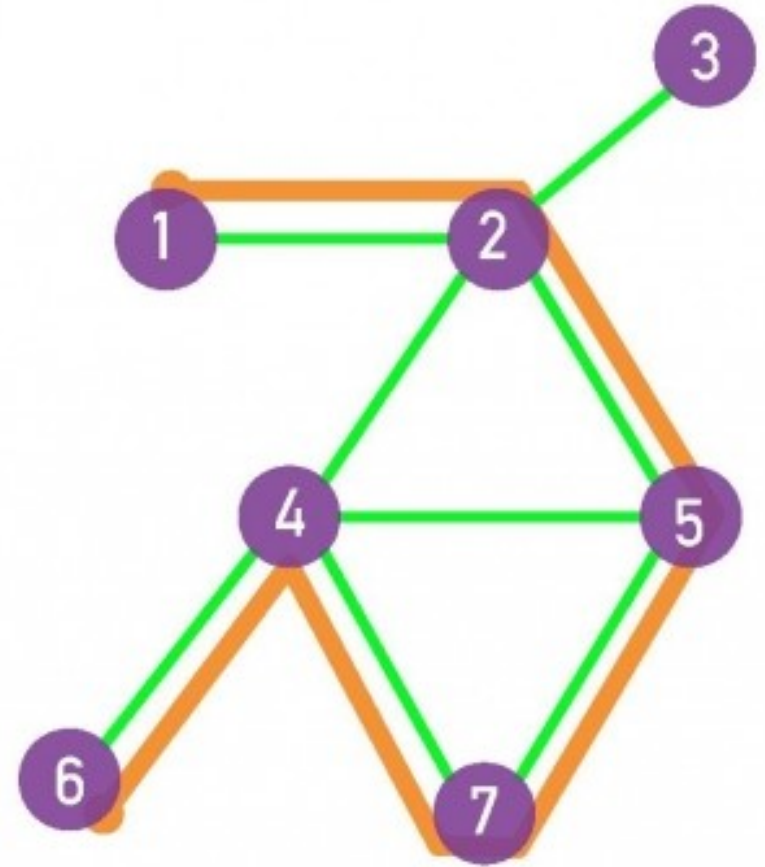
- A **(n<sub>1</sub>, n<sub>2</sub>)-clique** is a bipartite clique such that

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

# 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



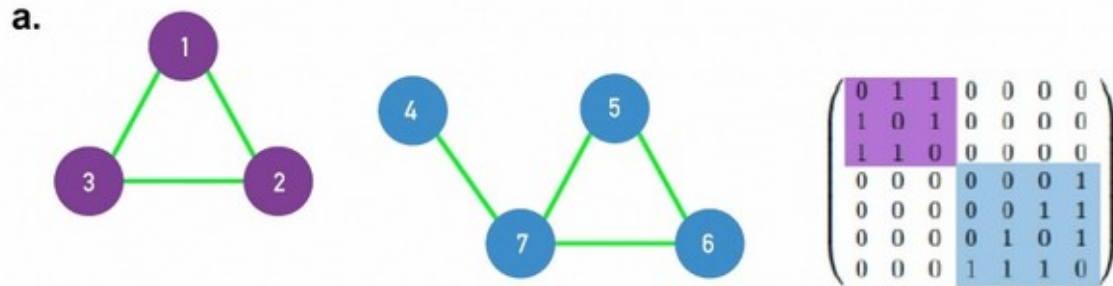
# Connectedness

- 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**

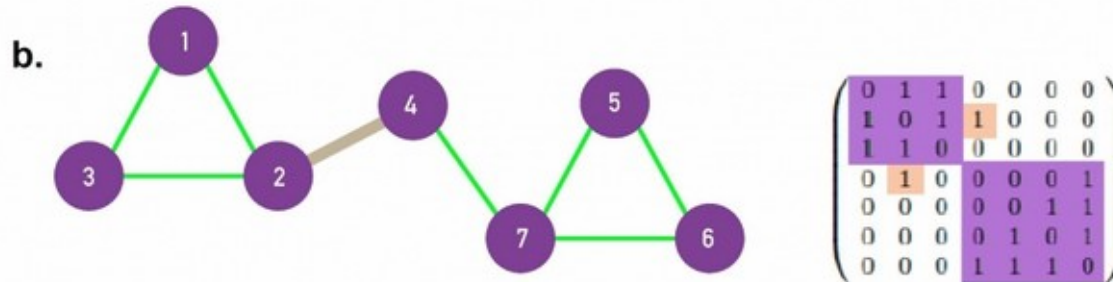
# Connected graphs

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  $\langle d \rangle$ , and is measured only for nodes that are in the same connected component

# Local clustering coefficient

- The **local clustering coefficient**  $C_i$  is a property of a node  $i$
- Let  $L_i$  represent the number of links among neighbors of node  $i$

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



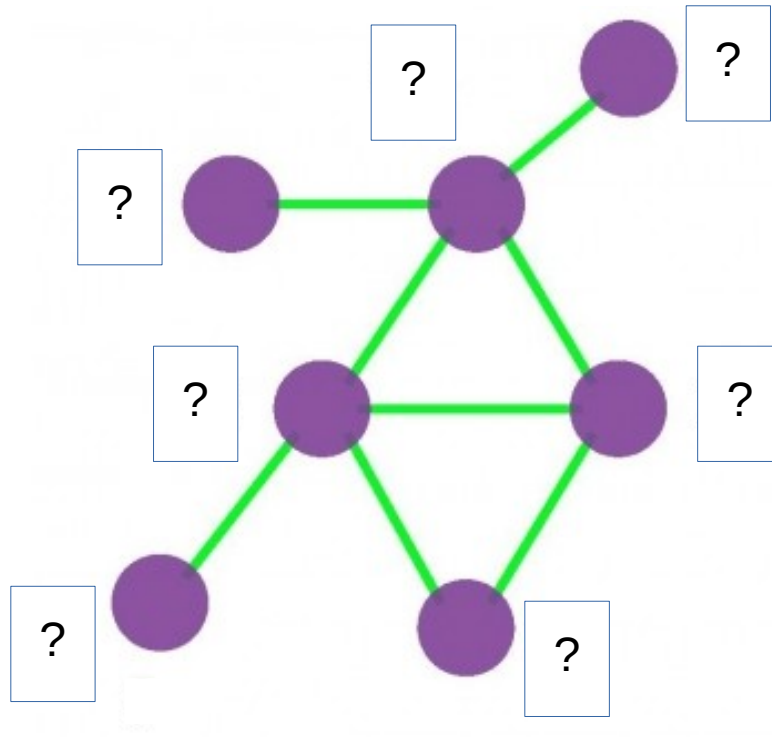
# Average clustering coefficient

- The **average clustering coefficient** is a property of the entire graph

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$$

# Try it!

What is the local clustering coefficient of each node?



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

# Practice on your own

- Obtain a degree distribution
- Write the adjacency matrix of a graph
- Measure the sparsity of a graph
- Compute the distance between two nodes
- Compute the diameter of a graph
- Identify connected components
- Calculate local clustering coefficient