Graph Theory Basics

Social Networks Analysis and Graph Algorithms

Prof. Carlos "ChaTo" Castillo — https://chato.cl/teach

Contents

- Notation for graphs

- Degree distributions

- Adjacency matrices

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

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
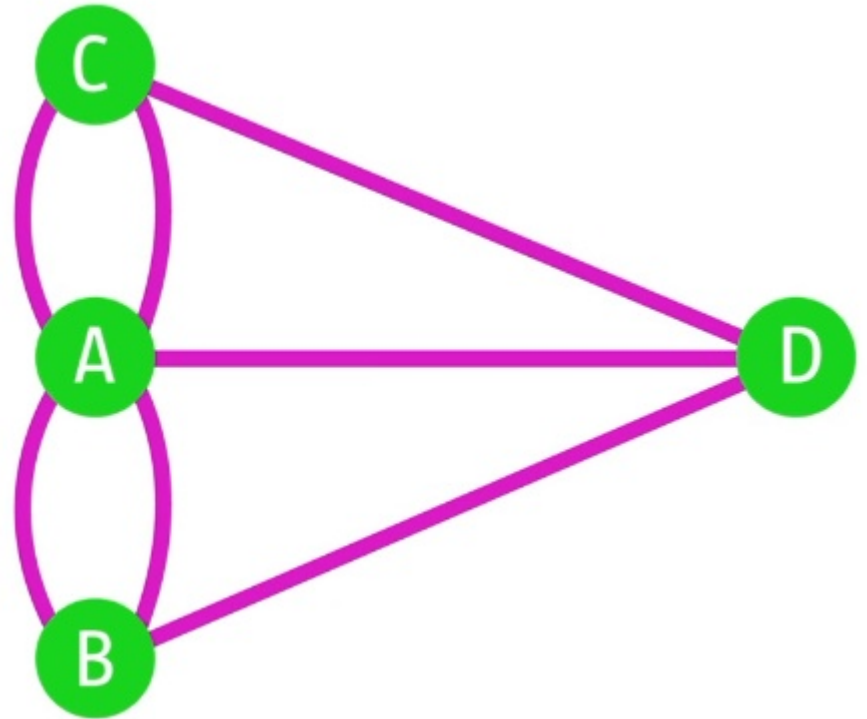
Subgraph

- Given $G = (V,E)$

- A **subgraph** induced by a nodeset S is the graph $G=(S,F)$ defined by all of the nodes in S and

  - $F = \{$ edges $(u,v) \in E$ s.t. $u \in S$ and $v \in S \}$

Typical notation variations

- You may find that G is denoted by (N, A), this is typical of directed graphs, means "*nodes, arcs*"

- You may find that
  - |V| is denoted by n or N
  - |E| is denoted by m, M, or L

Example graphs we will use

Network	|V|	|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

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

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

Weighted vs unweighted graphs

- In a weighted graphs edges have **weights** denoting the strength or importance of a connection

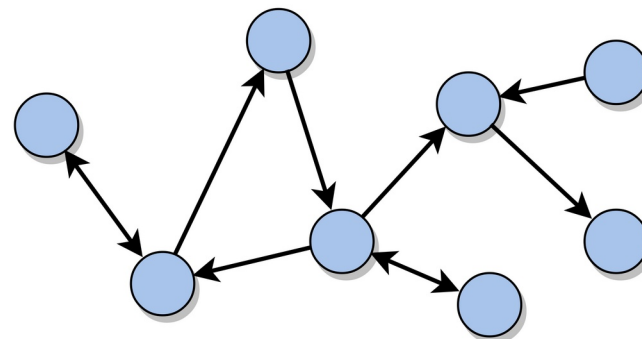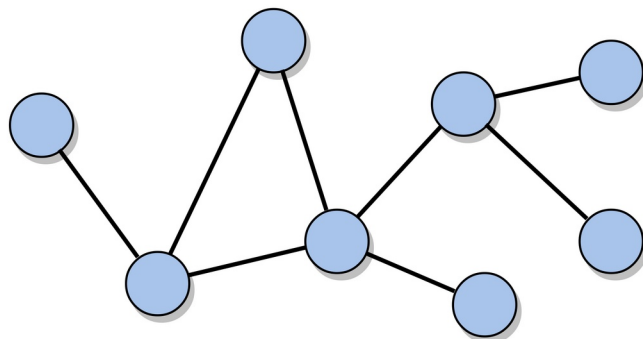- When drawing, typically larger weights are drawn with ticker lines

Example: weighted networks

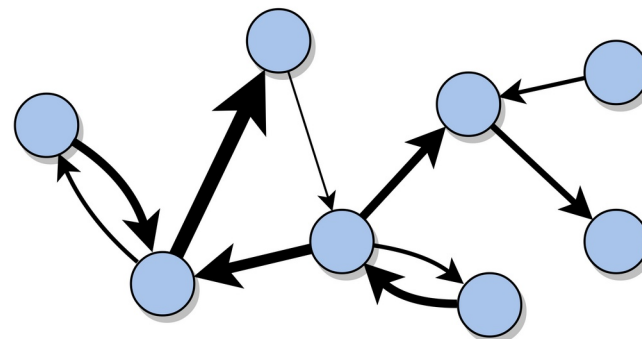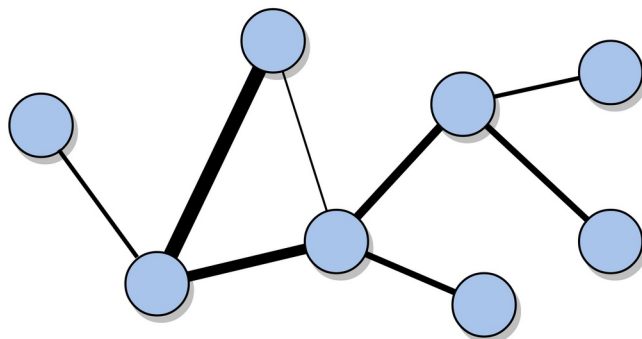**EU imports (top) and exports (bottom) of wine**
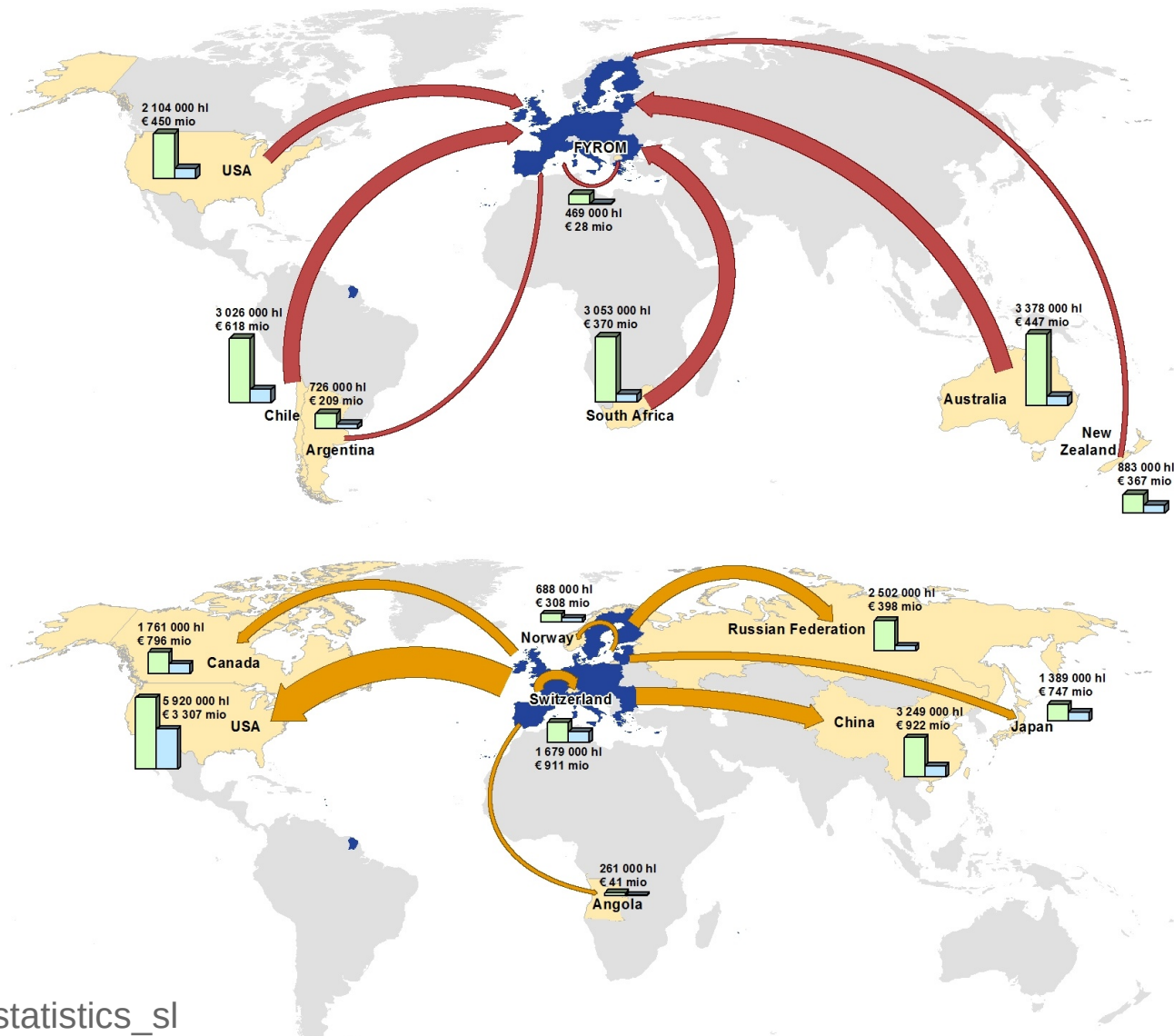
Undirected / Directed; Unweighted / Weighted

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

11/38

Degree

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

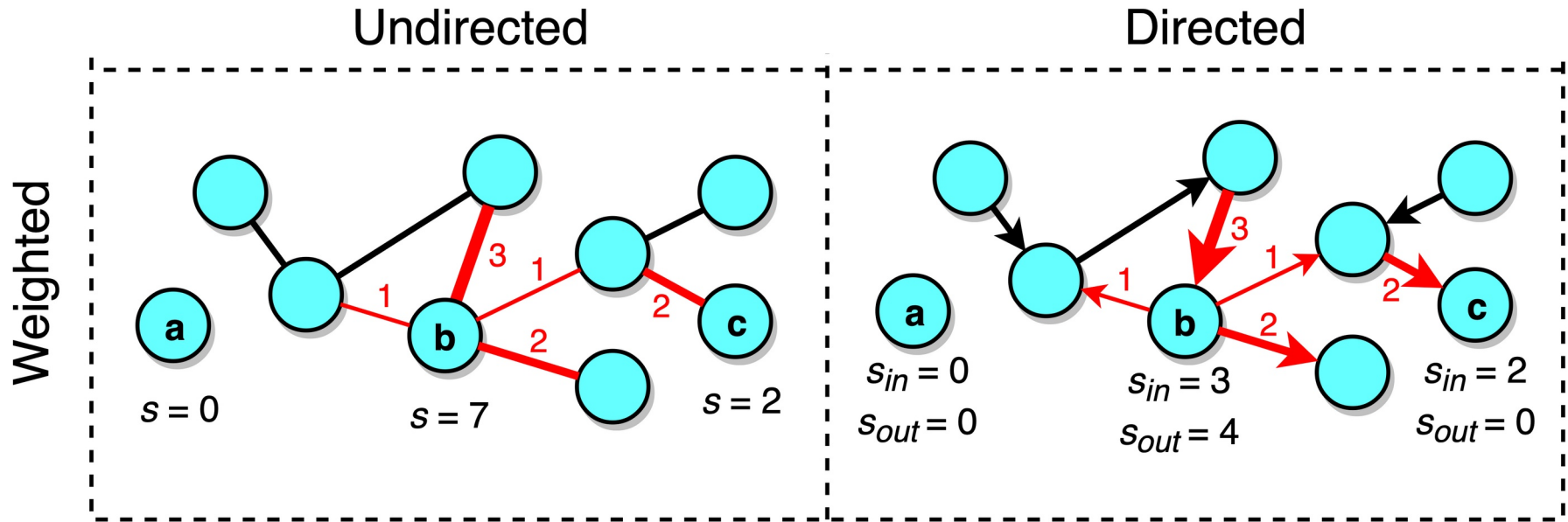$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N}$$

In directed graphs

- We distinguish in-degree from out-degree

  - Incoming and outgoing links, respectively

- Degree is the sum of both $\quad k_i = k_i^{\mathrm{in}} + k_i^{\mathrm{out}}$

- Counting total number of links:

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

In weighted graphs

We speak of "weighted degree" or "strength"

Undirected · Directed

Weighted

Undirected:
$s = 0$ (a), $s = 7$ (b), $s = 2$ (c)

Directed:
$s_{in} = 0$, $s_{out} = 0$ (a)
$s_{in} = 3$, $s_{out} = 4$ (b)
$s_{in} = 2$, $s_{out} = 0$ (c)

Degree distribution

- If there are N_k nodes with degree k

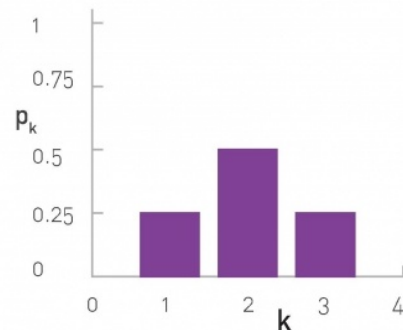- The degree distribution is given by
$$p_k = \frac{N_k}{N}$$

- The average degree is then
$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k$$

Degree distribution; two toy graphs
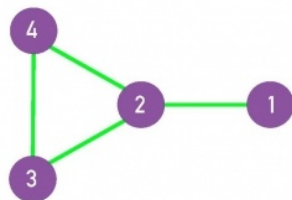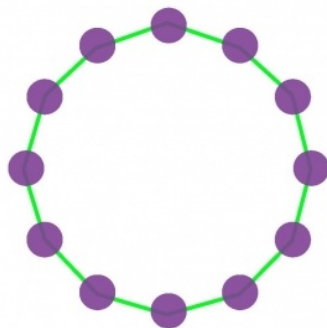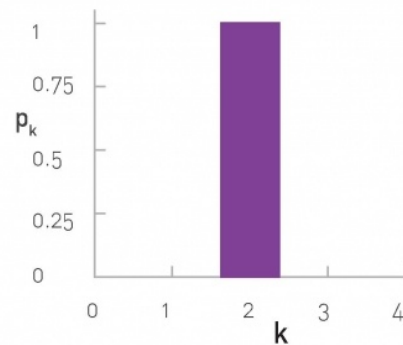
Exercise
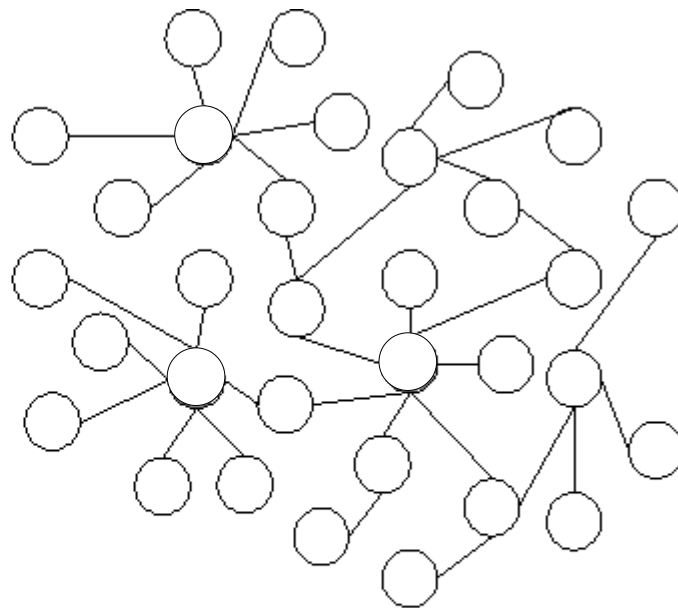
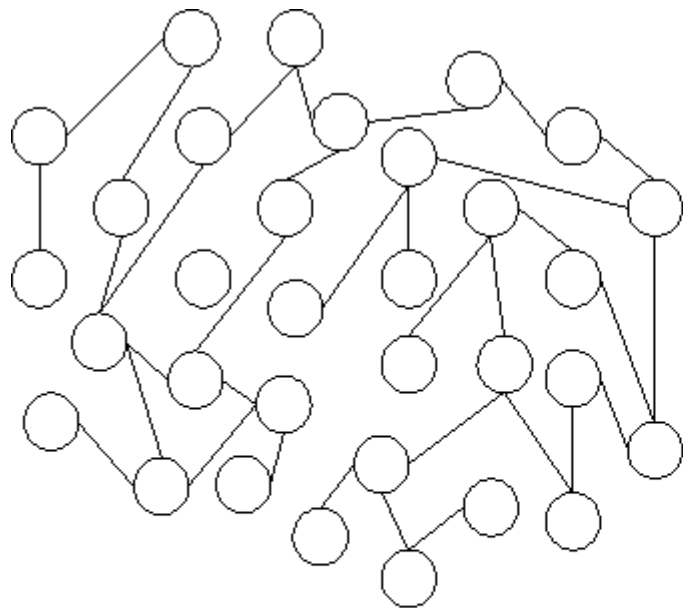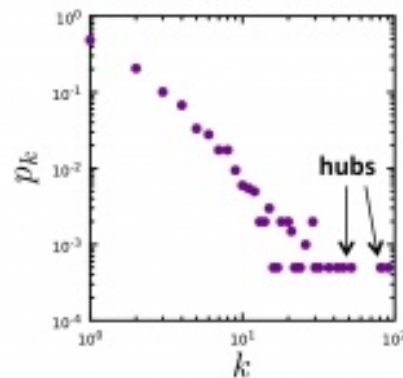Draw the degree distribution of these graphs

Degree distribution; real graph
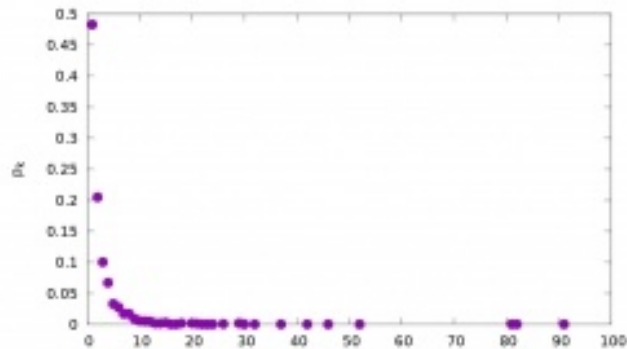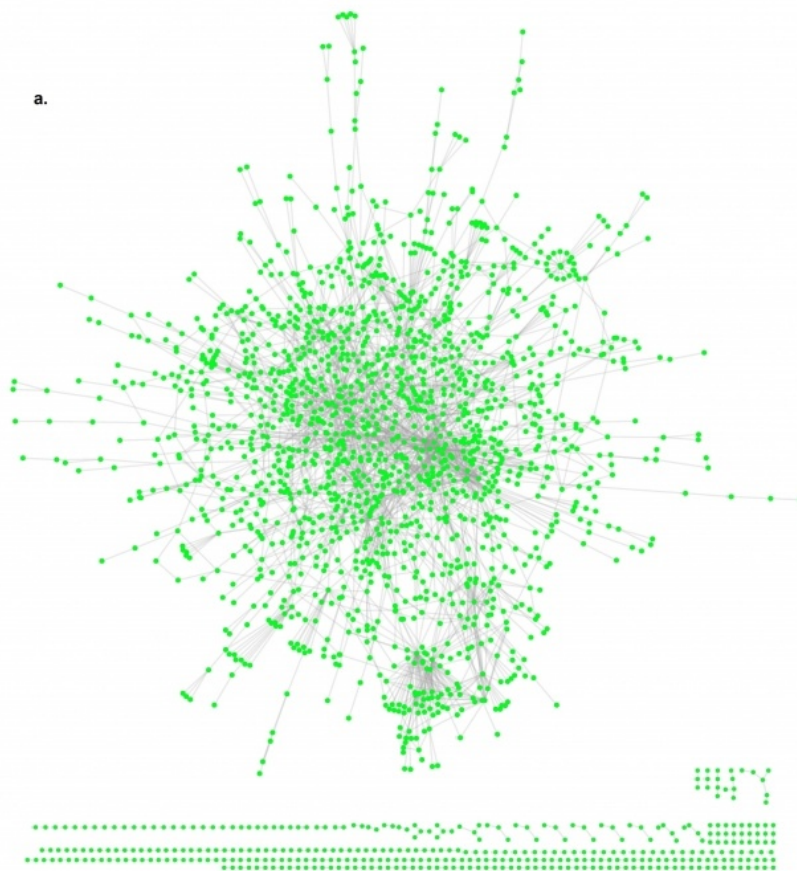
Linear scale
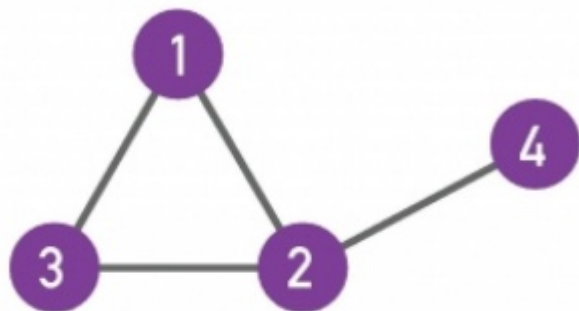
Log-log scale

Adjacency matrix
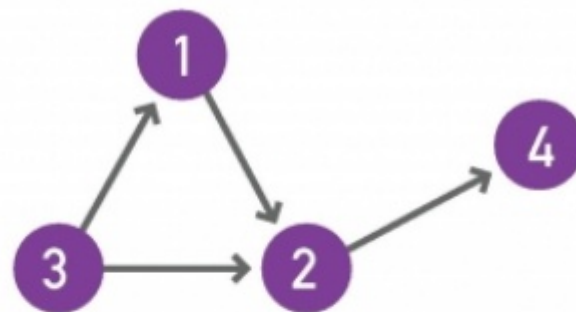
What is an adjacency matrix

- 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 $(i,j) \notin E$

- **A_{ij} always means row i, column j**

  – Sometimes Barabási's book has this wrong
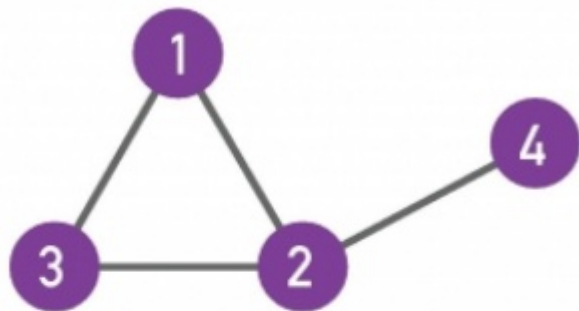
Examples

Undirected graph

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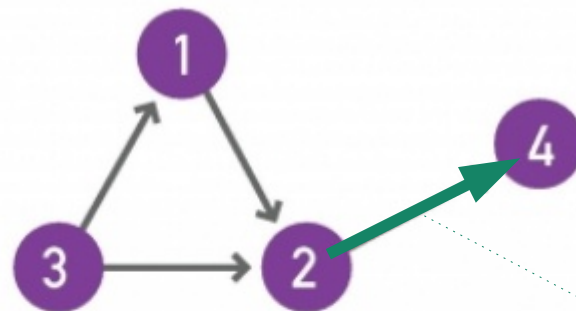
Directed graph

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

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

Row 2
Column 4

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

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 Question

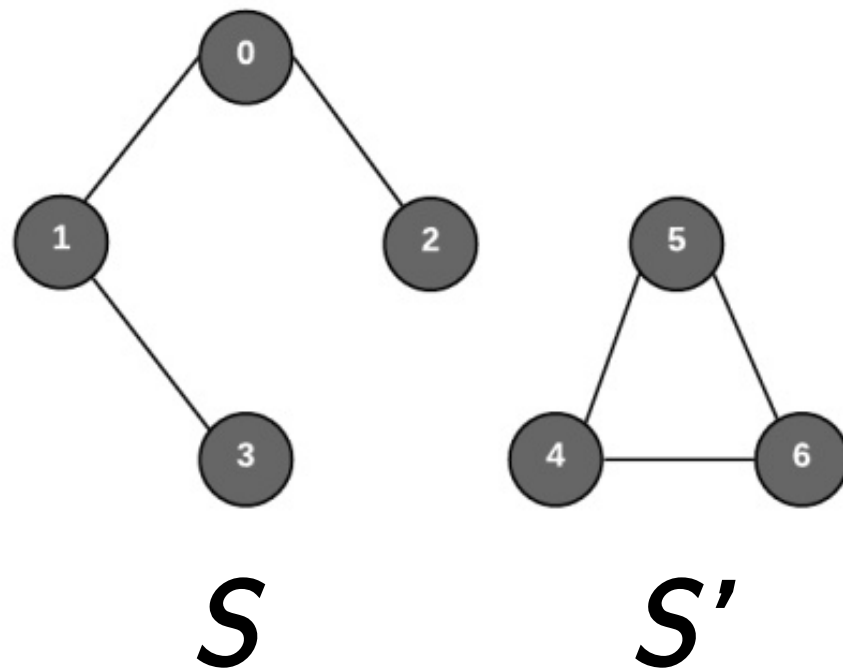- In terms of A, what is the expression for:

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

If a graph is disconnected

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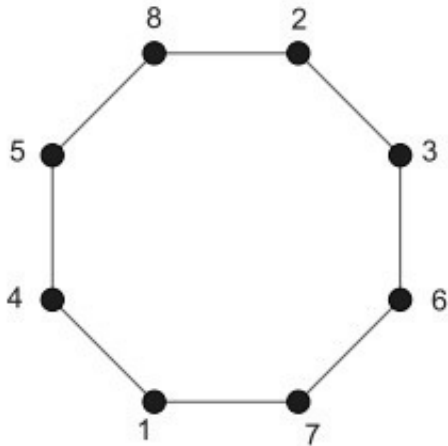have adjacency matrices
with **block structure**

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

S   *S'*

More concepts

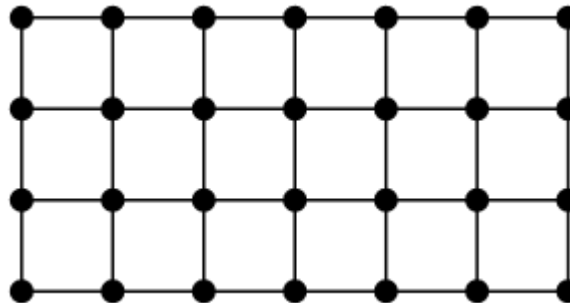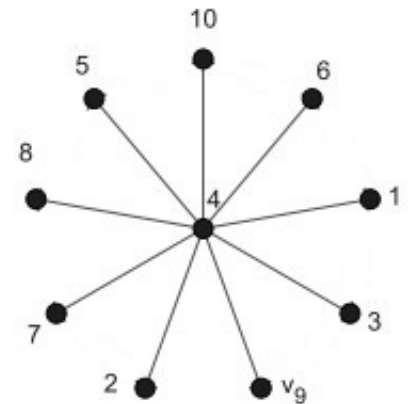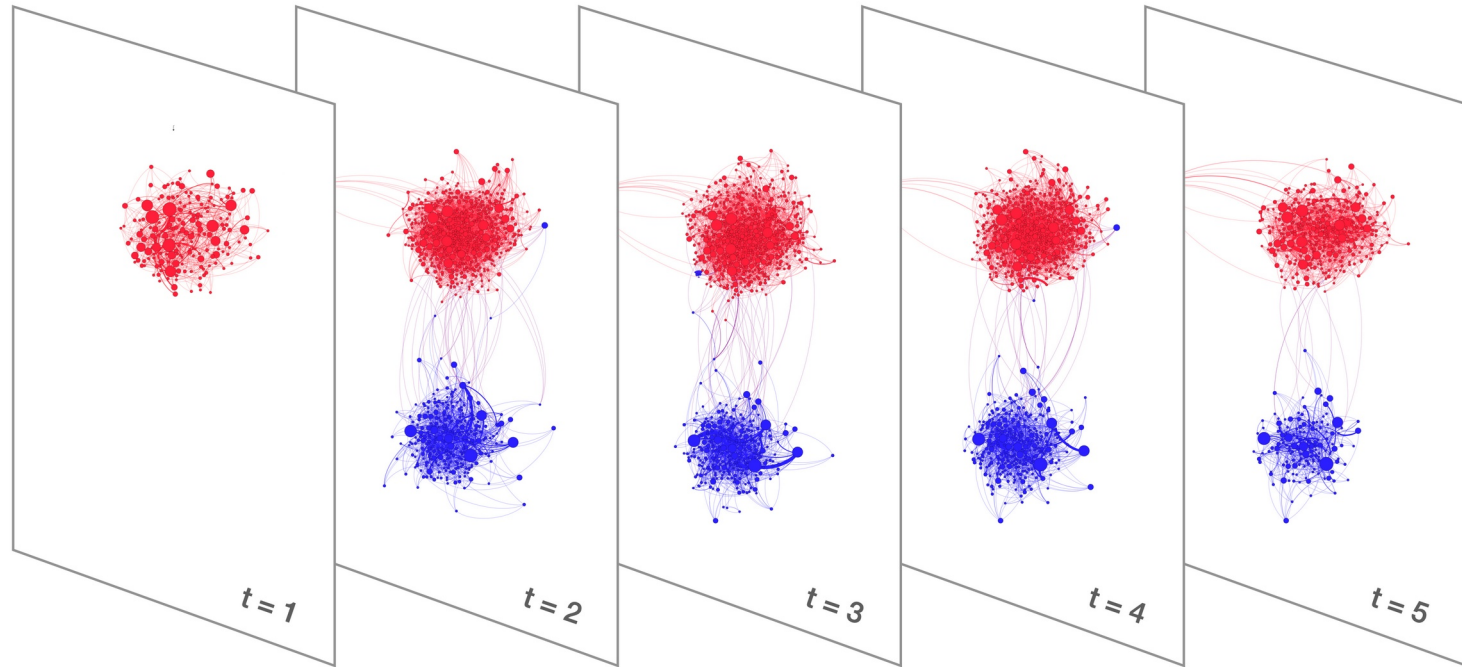# Some graphs have a name

Line

Cycle

Lattice

Star

Some graphs change over time

In a **temporal, or "time-evolving"** graph, at each timestep, we have a snapshot of the network
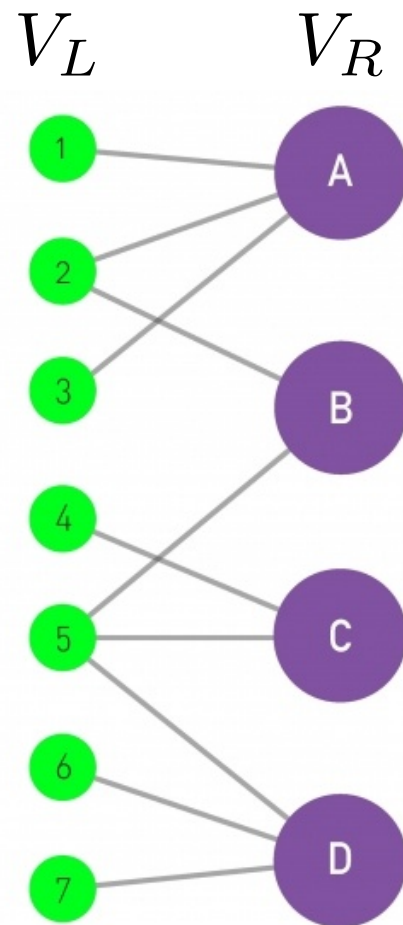

t = 1   t = 2   t = 3   t = 4   t = 5

Some graphs are "bi-partite"

V_L $\qquad$ V_R

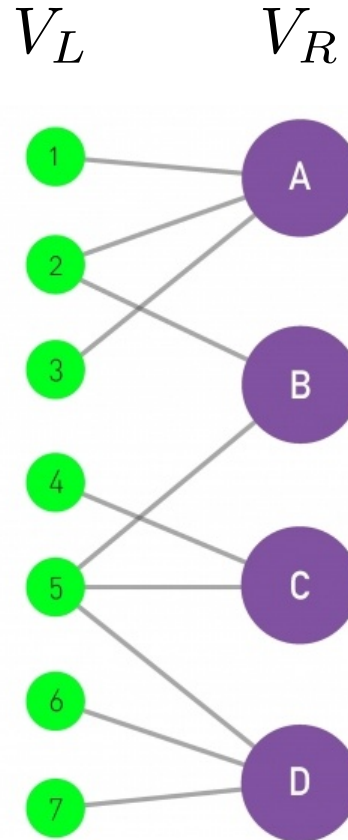- A bipartite graph is a graph

  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
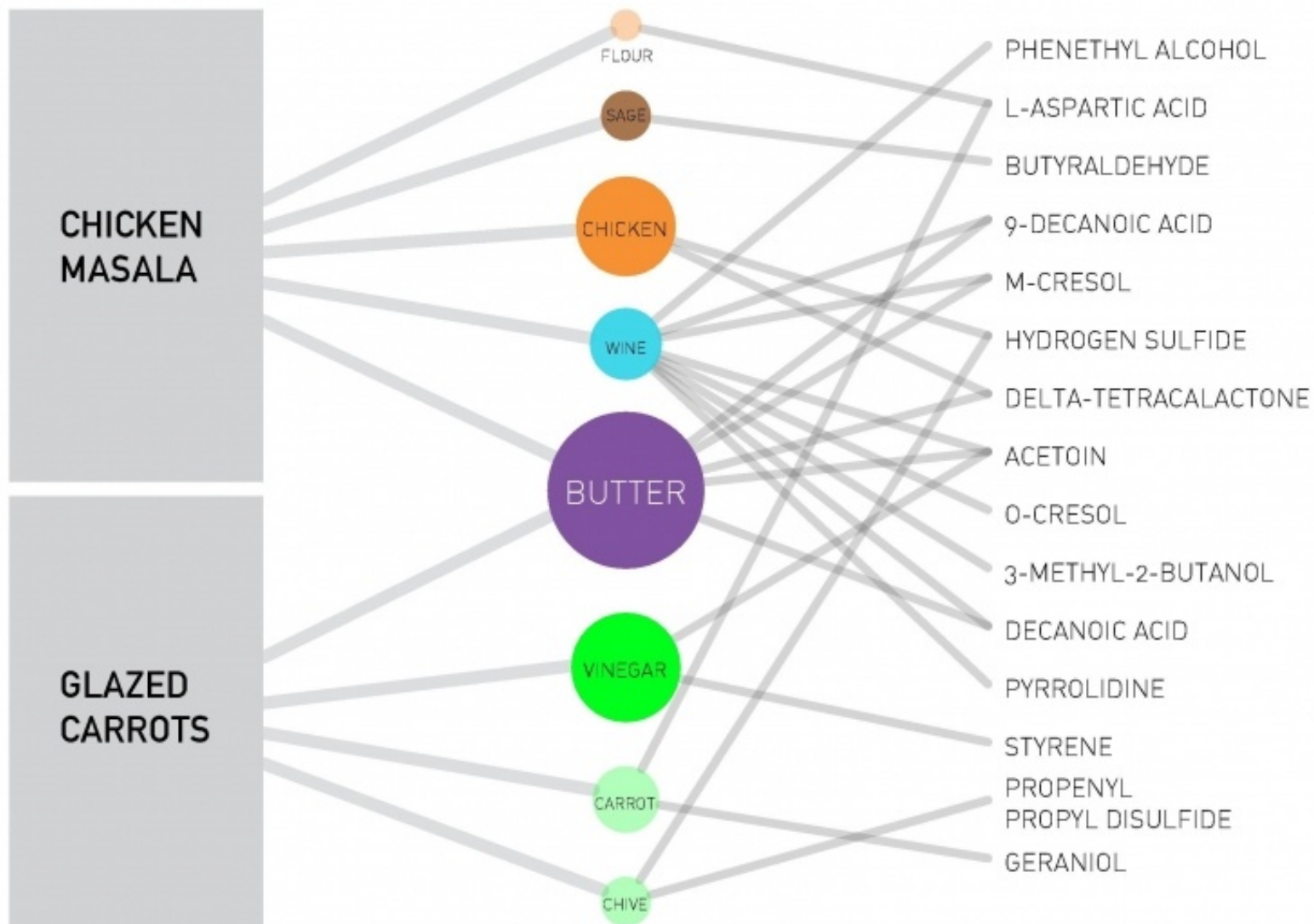
V_L V_R

?

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

Summary

Things to remember

- Definitions: degree, in-degree, out-degree, line graph, cycle graph, star graph, lattice, bi-partite graph

- Writing the adjacency matrix of a graph, and drawing a graph given its adjacency matrix

- Plotting the degree distribution of a graph

- Project a bi-partite graph

Practice on your own

Draw the indegree, outdegree, degre distribution

Write the adjacency matrix
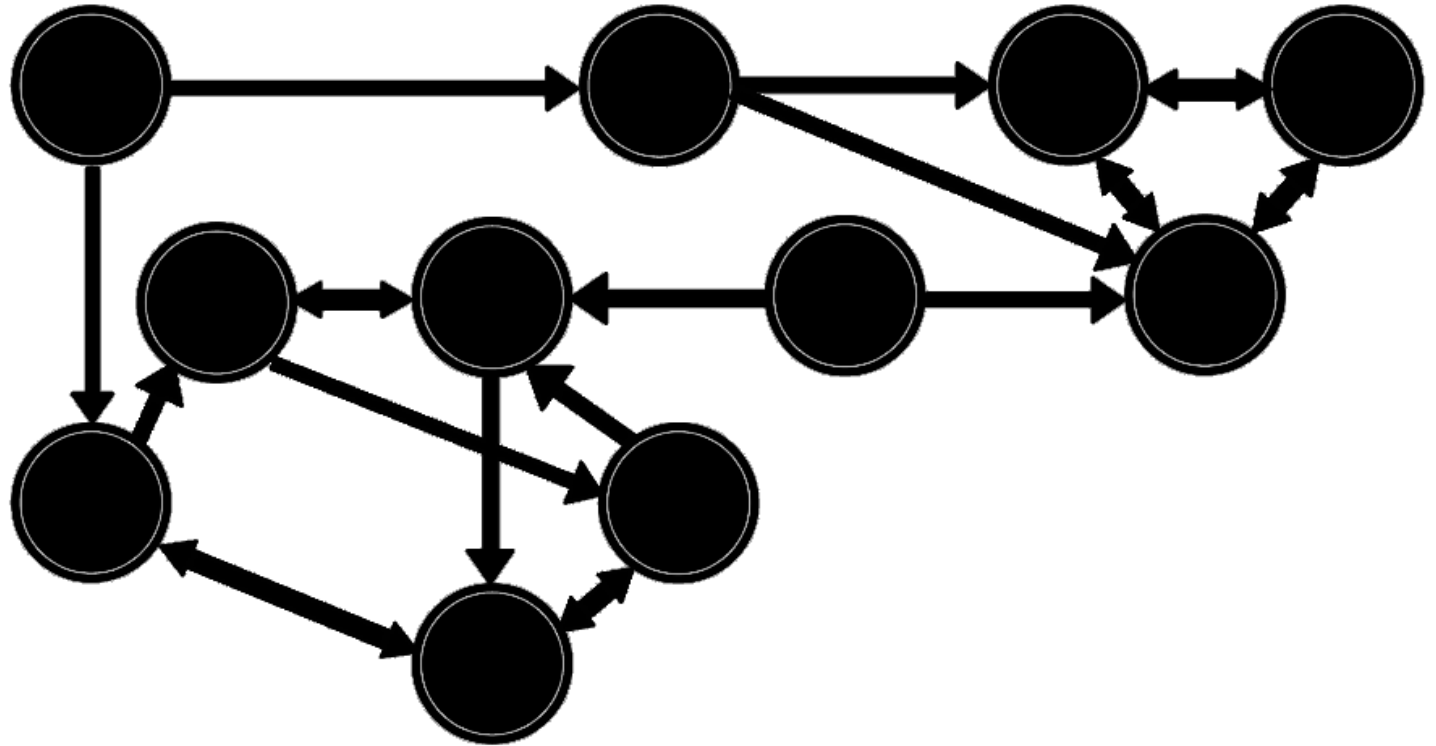
Practice on your own

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

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