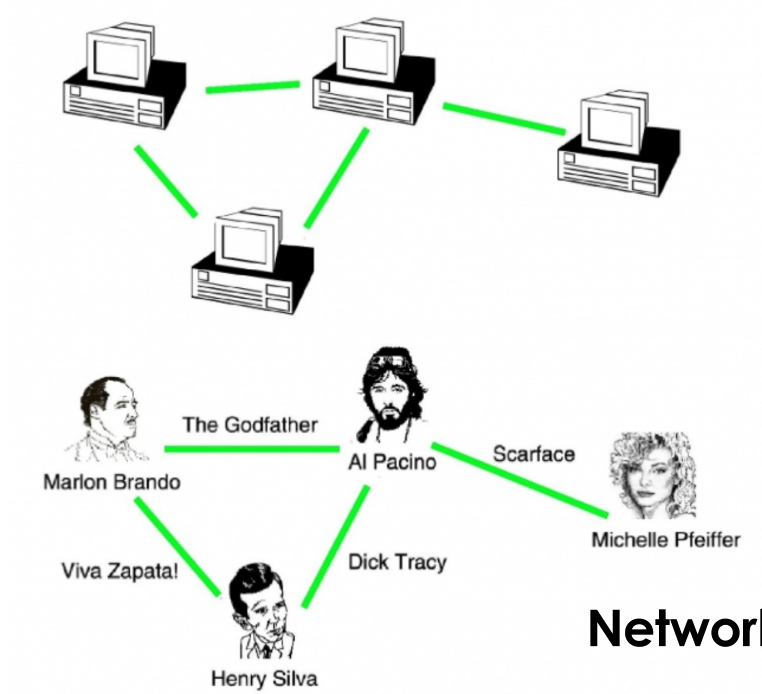
Network Analysis

AN INTRODUCTION FOR HUMANISTS

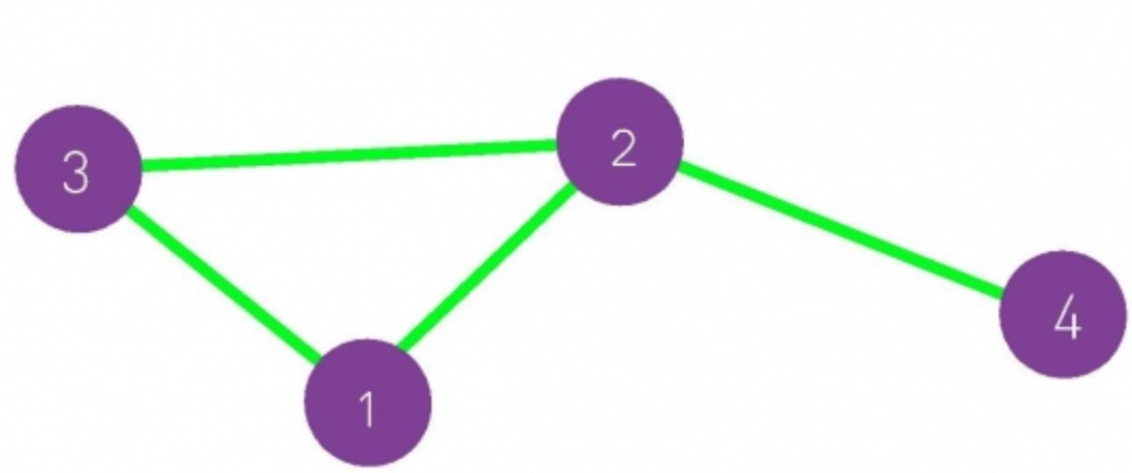
Dr Katarzyna Anna Kapitan

30 January 2025

Networks & Graphs



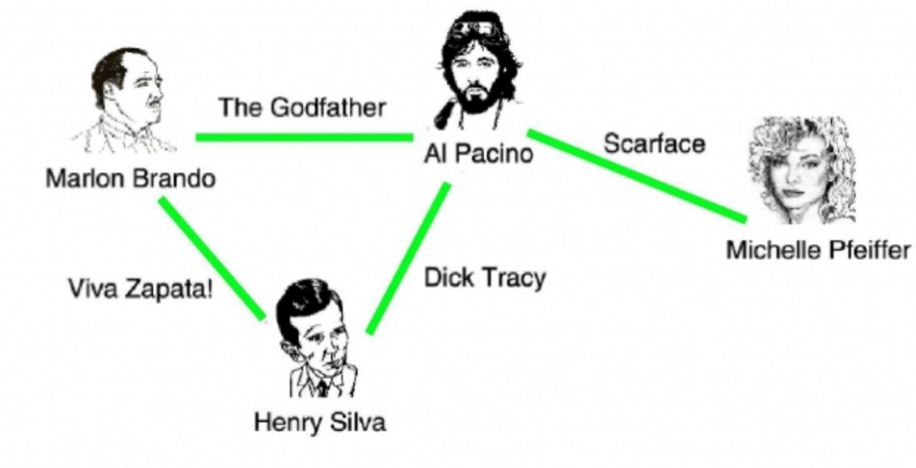
Network, node, link



Graph, vertex, edge

Source: Barabási, Network Science (https://networksciencebook.com)

Networks & Graphs



N, represents the number of components in the system (number of **nodes**).

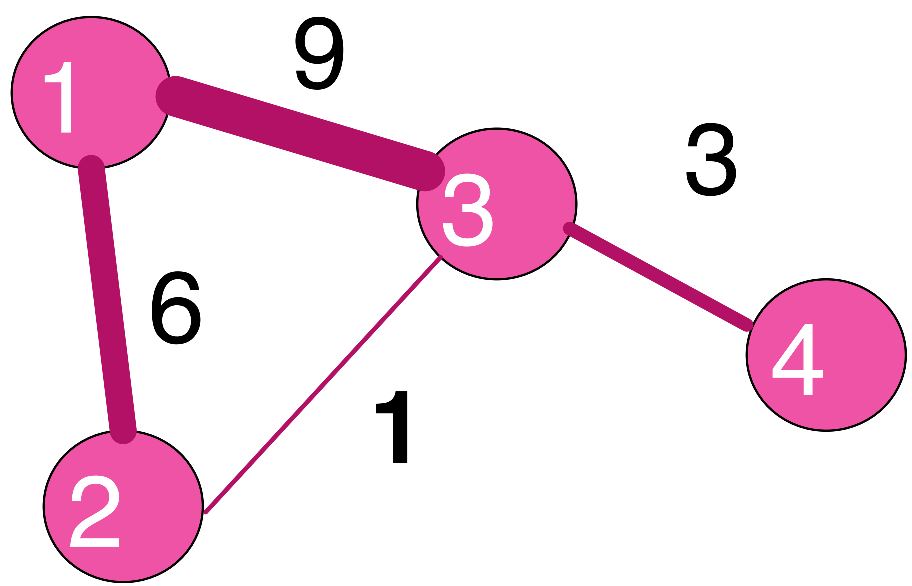
L, represents the total number of interactions between the nodes (number of **links**).

$N = 4$
$L = 4$

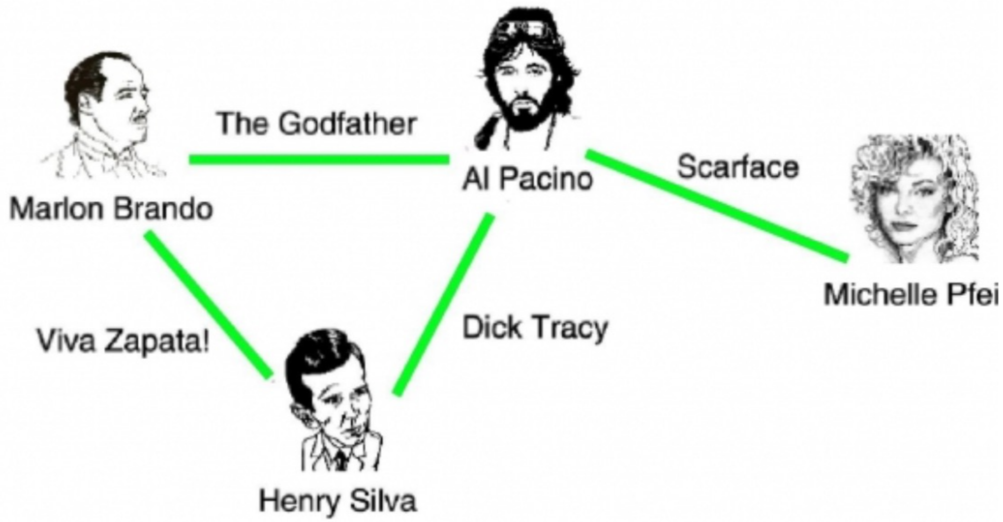
Source: Barabási, Network Science (https://networksciencebook.com)

Links (Weighted & Unweighted)

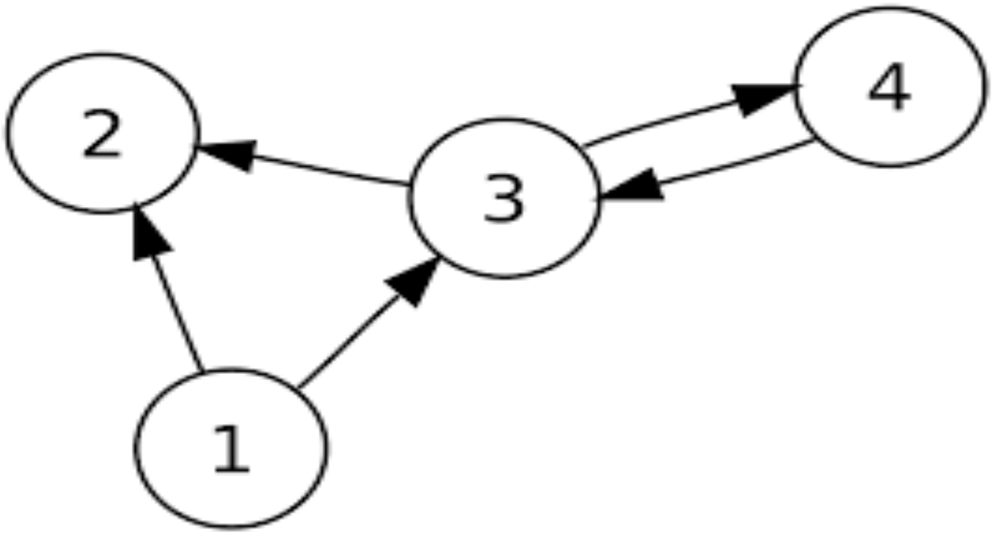
- A network can be **unweighted** or **weighted**.
- In a weighted network, links have associated **weights.** The **weighted link** *(i,j,w)* between nodes *i* and *j* has weight **w**.
- We can for example count the number of movies in which two actors played together and reflect this as a weight of the link between them.



Links (Directed & Undirected)



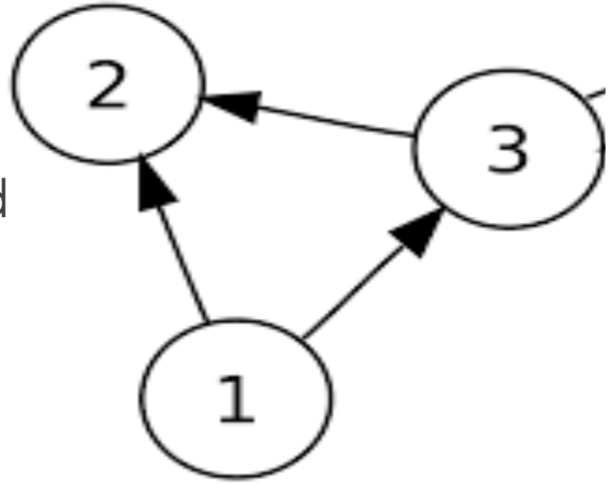
Undirected Links -> Undirected Network
Hollywood actor network; two actors are connected if they played in the same movie.
Source: Barabási, Network Science (https://networksciencebook.com)



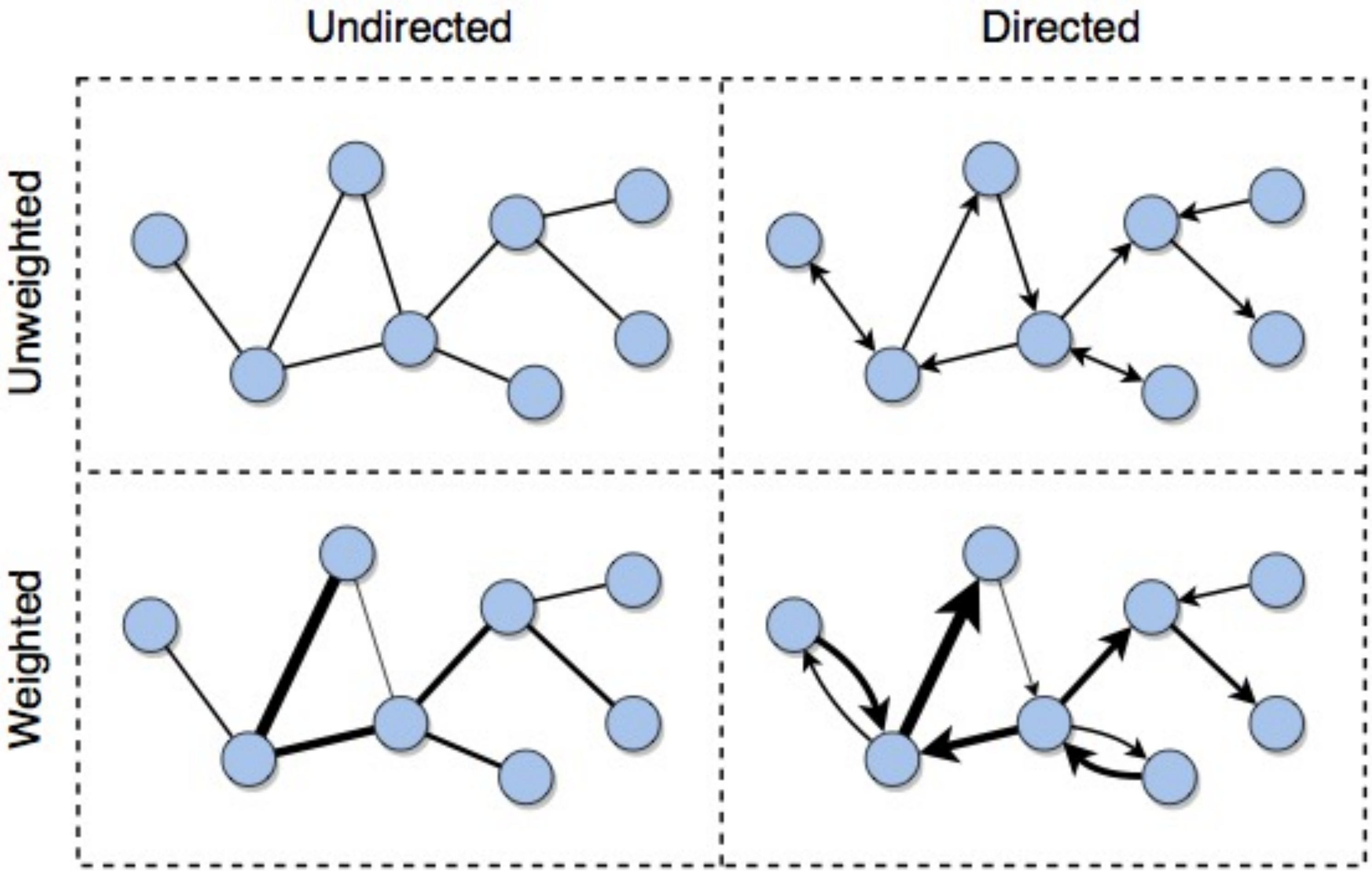
Directed Links -> Directed Network (Digraph)
For example, scholars' correspondence network; two scholars are connected if they sent or received a letter to/from each other; the direction of the link is denoted with the arrow, illustrating who sent a letter to whom.
Source: Wikipedia, Directed Graph.

Links (Directed & Undirected)

- **Link (*i, j*)** goes from the **source** node ***i*** to the **target** node ***j***.
- In **undirected networks**, all links are bi-directional and the order of the two nodes in a link does not matter; (1,2) is the same as (2,1), meaning there is a link between ***node 1*** and ***node 2.***
- In **directed networks** the order does matter(1,2) means that there is a link **from *node 1* to *node 2*** and (2,1) that there is a link **from *node 2* to *node 1***.



Which link is illustrated above, (3,1) or (1,3)?

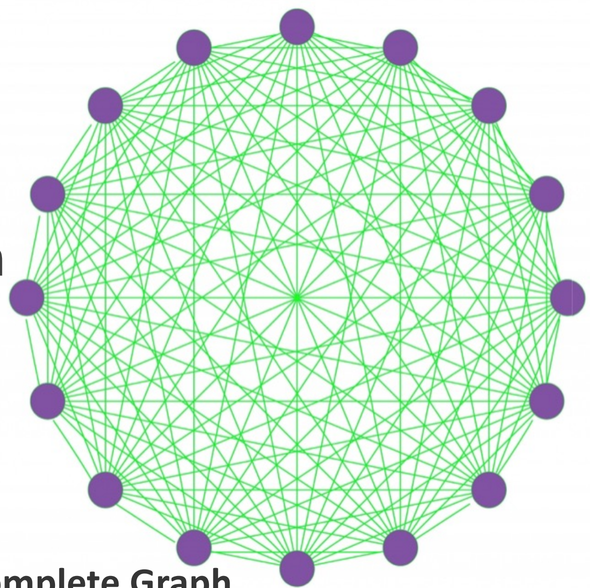


Source: Menczer, Fortunato, Davis, *A First Course in Network Science*, version 3 (2023).

Density and Sparsity

Density and Sparsity

- The density is the fraction of possible links in the network.
- To calculate the **density** of the network, we need to know the maximum number of links possible between its nodes.
- A network with the maximum number of links, in which all possible pairs of nodes are connected, is called a **complete network.**
- A complete network has maximal density which equals to 1.



Complete Graph

A complete graph with $N = 16$ nodes and $L_{max} = 120$ links.

Source: Barabási, Network Science (https://networksciencebook.com)

Density and Sparsity

d (Density)

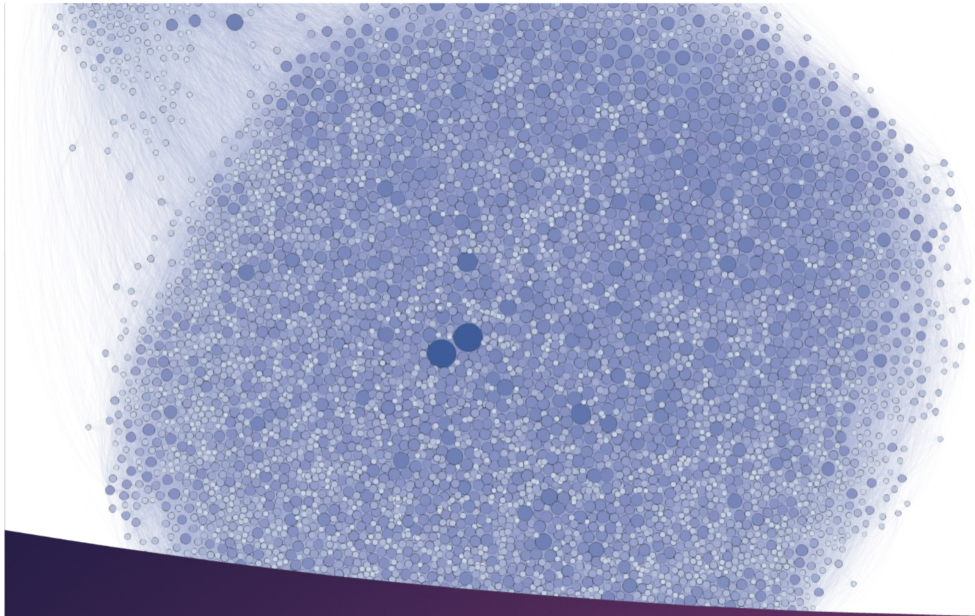
$\boldsymbol{d = L / L_{max}}$

The density ***d*** is the fraction of possible links in the network.

L = Actual number of links in the network

$\boldsymbol{L_{max}}$ = Maximum possible number of links in a network of this size.

.



Network	Type	Nodes (*N*)	Links (*L*)	Density (*d*)
Facebook Northwestern Univ.		10,567	488,337	0.009
IMDB movies and stars		563,443	921,160	0.000006
IMDB co-stars	W	252,999	1,015,187	0.00003
Twitter US politics	DW	18,470	48,365	0.0001
Enron email	DW	87,273	321,918	0.00004
Wikipedia math	D	15,220	194,103	0.0008
Internet routers		190,914	607,610	0.00003
US air transportation		546	2,781	0.02
World air transportation		3,179	18,617	0.004
Yeast protein interactions		1,870	2,277	0.001
C. elegans brain	DW	297	2,345	0.03
Everglades ecological food web	DW	69	[illegible]	[illegible]

Density and Sparsity

- **The network is sparse if *d* << 1**
- ***Most real systems are sparse.***

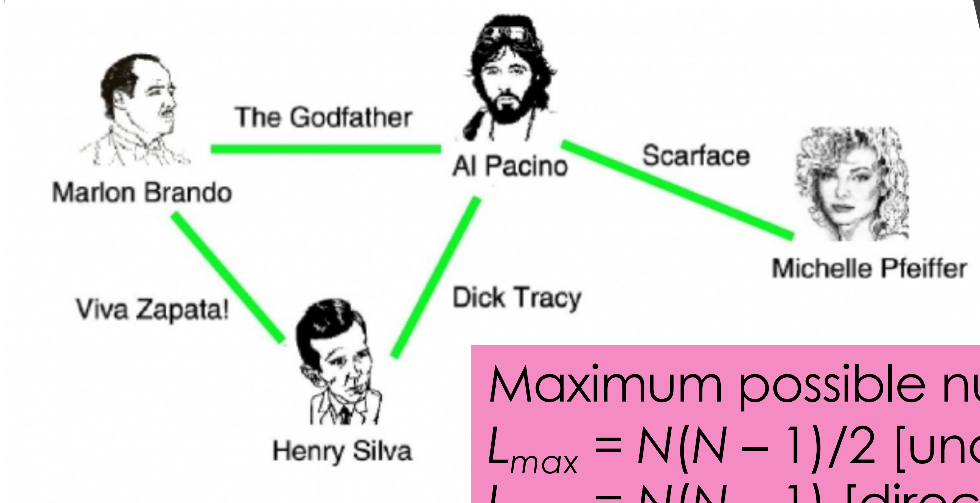
Source: Menczer, Fortunato, Davis, *A First Course in Network Science*, version 3 (2023)

Maximum Number of Links

The maximum number of links in a network is bounded by the possible number of distinct connections among the nodes.
In undirected networks only one connection can exist between two nodes, but in directed networks two.

- ***N*** = Number of nodes
- ***L*** = Number of links
- L_{max} = Maximum possible number of links
  - In undirected network: $L_{max} = N(N-1)/2$
  - In directed network $L_{max} = N(N-1)$

Density and Sparsity



Maximum possible number of links
L_{max} = $N(N-1)/2$ [undirected]
L_{max} = $N(N-1)$ [directed]
Density
$d = L / L_{max}$

What is the density of our actor network?

Is the network sparse?

Number of nodes N = ?

Number of links L = ?

L_{max} = ?

d = ?

Density and Sparsity

What is the density of this network?

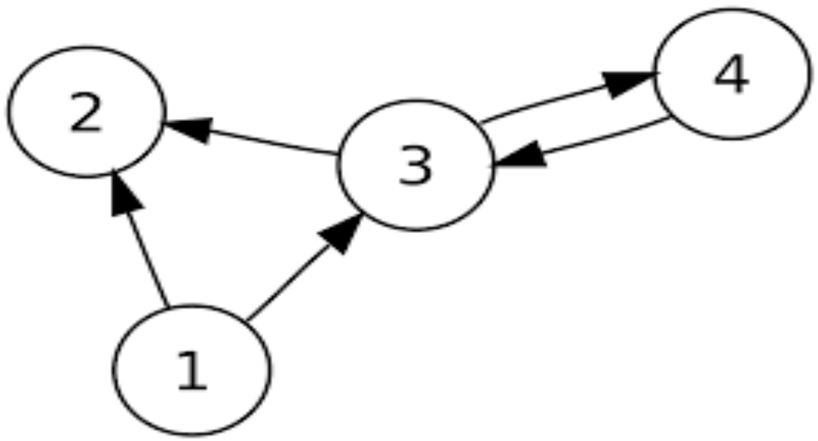
Is the network sparse?

Number of nodes N = ?

Number of links L = ?

L_{max} = ?

d = ?



$L_{max} = N(N-1)/2$ [undirected]
$L_{max} = N(N-1)$ [directed]
$d = L/ L_{max}$

Why should I care about density?

Degree

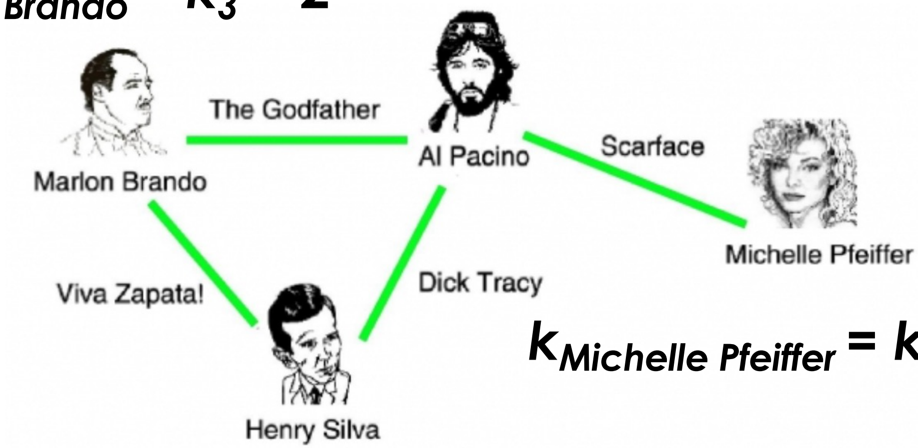
Katarzyna Anna Kapitan, Network Analysis for Humanists,
Paris 2025

Nodes & Degree in Undirected Networks

- A key property of each node is its **degree**
- **Degree** represents the number of links a node has to other nodes.
- We denote with $\boldsymbol{k_i}$ the degree of the $\boldsymbol{i^{th}}$ node in the network: $k_1=2$, $k_2=3$, $k_3=2$, $k_4=1$.

$\mathbf{k_{Marion\ Brando} = k_3 = 2}$

$\mathbf{k_{Al\ Pacino} = k_2 = 3}$



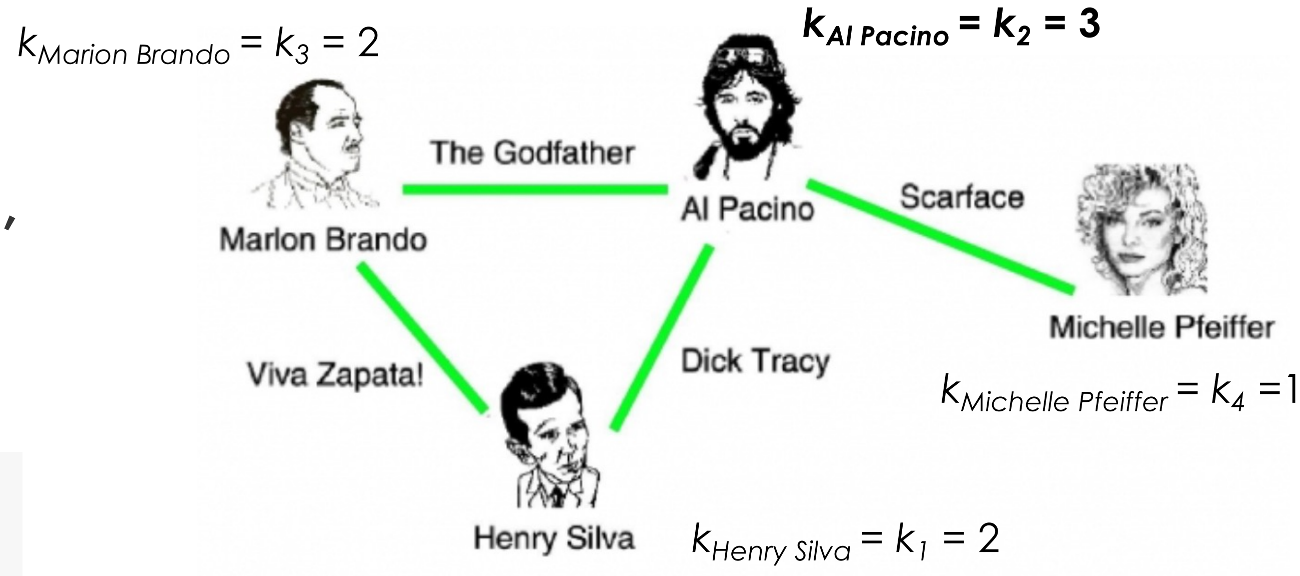
$\mathbf{k_{Michelle\ Pfeiffer} = k_4 = 1}$

$\mathbf{k_{Henry\ Silva} = k_1 = 2}$

In an **undirected network** **the *total number of links* (*L*)**, can be expressed as the **sum of the node degrees**.

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

** Here the 1/2 factor corrects for the fact that in the sum each link is counted twice, which we don't want in undirected networks.*



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

$L = (k_1 + k_2 + k_3 + k_4) / 2$

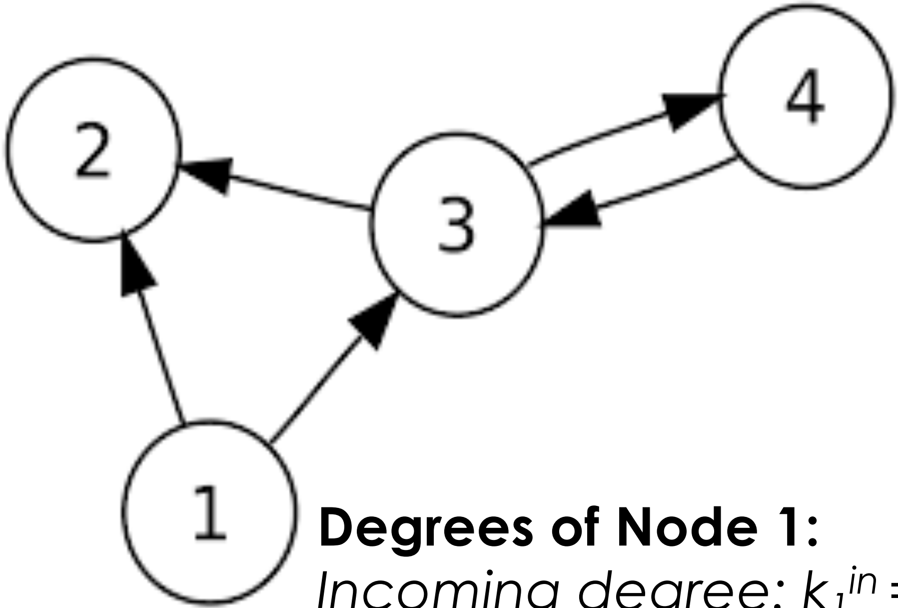
$L = (2 + 3 + 2 + 1) / 2$

$L = 8 / 2$

$L = 4$

Source: Barabási, Network Science (https://networksciencebook.com)

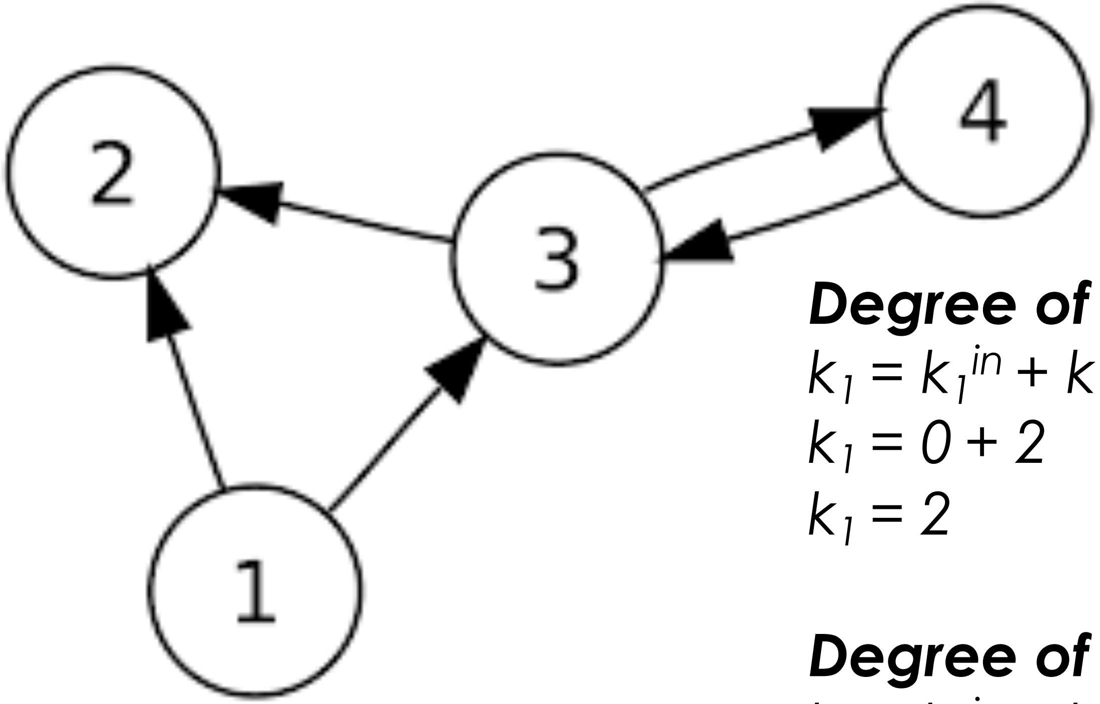
Nodes & Degree in Directed Networks



Degrees of Node 1:
Incoming degree: $k_1^{in} = 0$
Outgoing degree: $k_1^{out} = 2$

- In **directed networks** we distinguish between:
  - **incoming degree**, k_i^{in}, representing the number of links that point to node *i*,
  - **outgoing degree**, k_i^{out}, representing the number of links that point from node *i* to other nodes.

In **directed networks** a node's **total degree** (k_i) is a sum of its incoming and outgoing degrees: $\boldsymbol{k_i = k_i^{in} + k_i^{out}}$



Degree of node 1:

$k_1 = k_1^{in} + k_1^{out}$

$k_1 = 0 + 2$

$k_1 = 2$

Degree of node 2:

$k_2 = k_2^{in} + k_2^{out}$

$k_2 = 2 + 0$

$k_2 = 2$

What is the degree of node 3:

$k_3 = k_3^{in} + k_3^{out}$

$k_3 = ? + ?$

$k_3 = ?$

What is the degree of node 4:

$k_4 = k_4^{in} + k_4^{out}$

$k_4 = ? + ?$

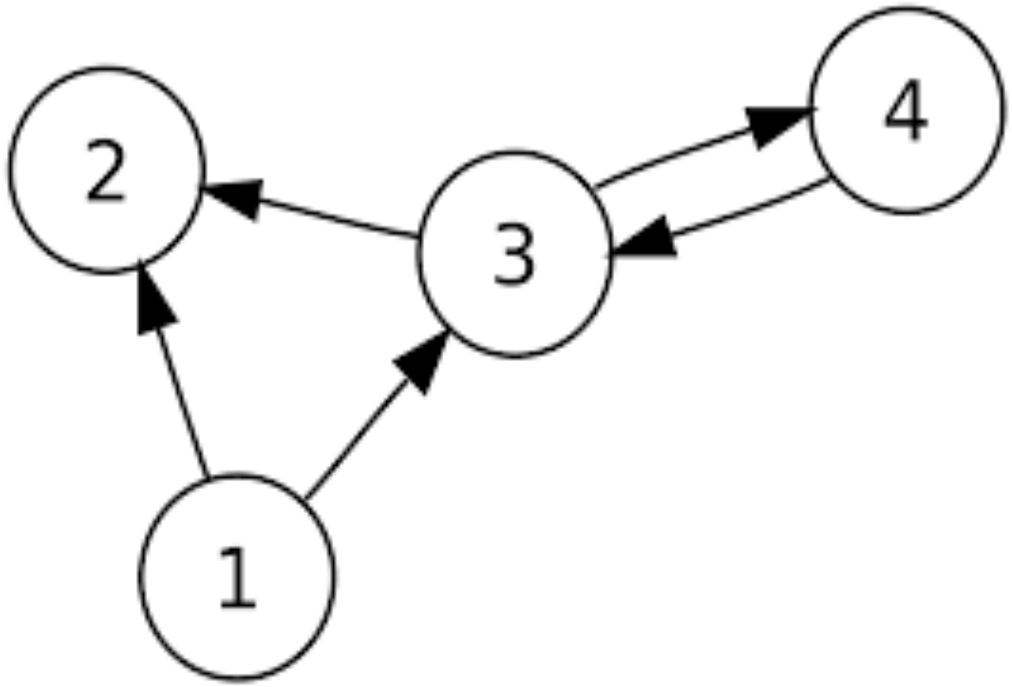
$k_4 = ?$

The **total number of links (*L*)** in a **directed network** is expressed as the sum of the incoming degrees (which is equal the sum of the outgoing degrees).

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

$L = k_1^{in} + k_2^{in} + k_3^{in} + k_4^{in}$
$L = 0 + 2 + 2 + 1$
$L = 5$

$L = k_1^{out} + k_2^{out} + k_3^{out} + k_4^{out}$
$L = 2 + 0 + 2 + 1$
$L = 5$



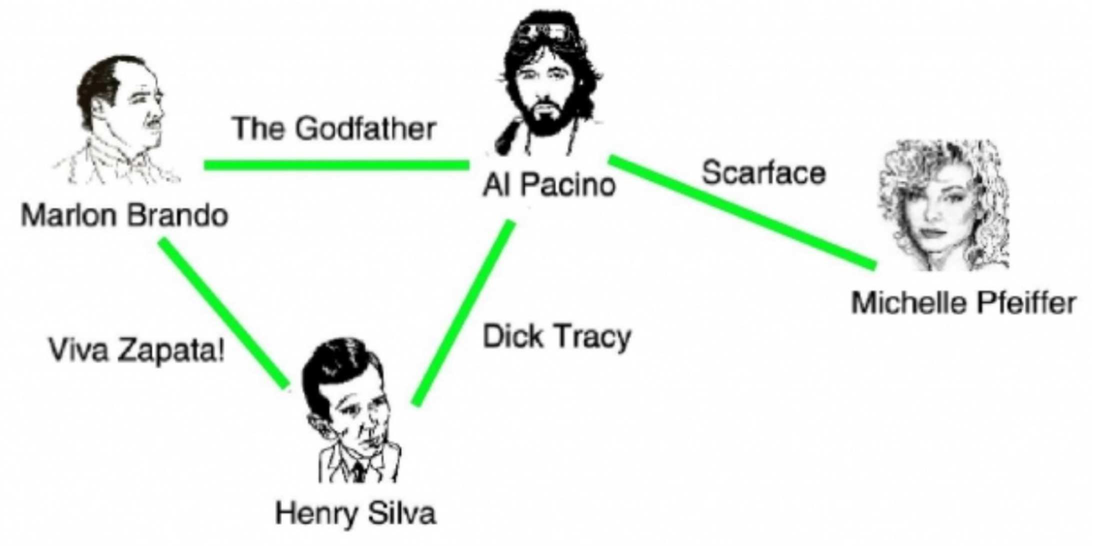
Source: Barabási, Network Science (https://networksciencebook.com)

Average Degree

⟨k⟩ - Average Degree

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

$$\langle k \rangle = (k_1 + k_2 + \ldots + k_N) / N$$

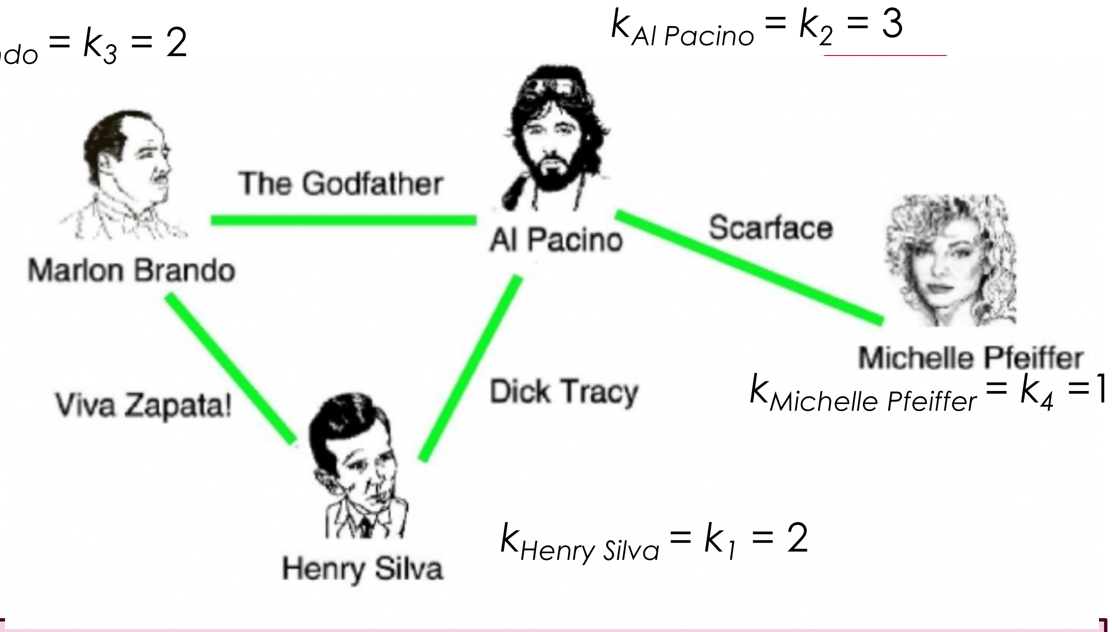


Source: Barabási, Network Science (https://networksciencebook.com)

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

$\langle k \rangle = (k_1 + k_2 + \ldots + k_N) / N$
$N = 4$
$k_1=2, k_2=3, k_3=2, k_4=1$

$\langle k \rangle = (2 + 3 + 2 + 1) / 4$
$\langle k \rangle = 8/4$
$\langle k \rangle = 2$



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

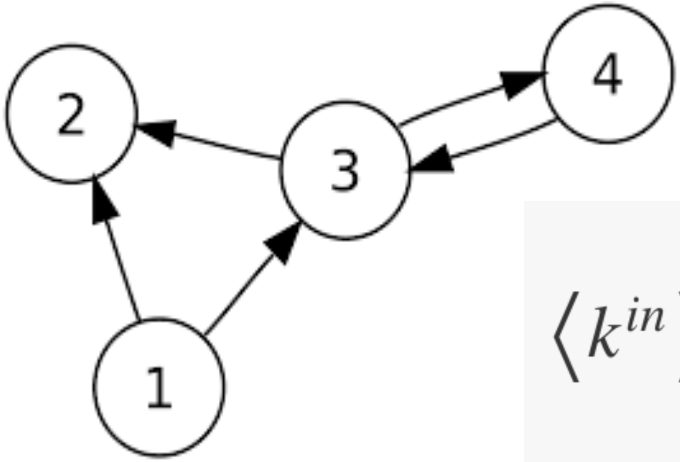
$N = 4$
$L = 4$
$\langle k \rangle = 2L/N$

$\langle k \rangle = 2*4/ 4$
$\langle k \rangle = 8/4$
$\langle k \rangle = 2$

Average Degrees (Directed Network)

$$\langle k^{in} \rangle = \langle k^{out} \rangle$$

Average Incoming Degree = Average Outgoing Degree

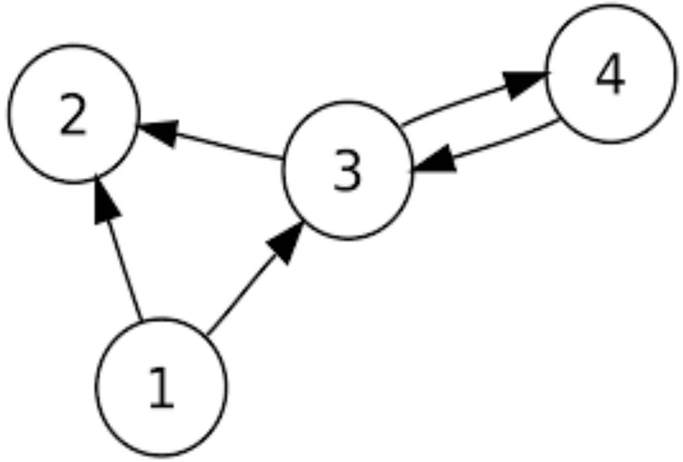


$$\left\langle k^{in} \right\rangle = \frac{1}{N} \sum_{i=1}^{N} k_i^{in} = \left\langle k^{out} \right\rangle = \frac{1}{N} \sum_{i=1}^{N} k_i^{out} = \frac{L}{N}$$

$\langle k^{in} \rangle = L / N$

$L = 5$

$N = 4$



$\langle k^{in} \rangle = L/N$

$\langle k^{in} \rangle = 5 / 4$

$\langle k^{in} \rangle = 1{,}25$

$\langle k^{in} \rangle = (k_1^{in} + k_2^{in} + \ldots + k_N^{in})/N$

$k_1^{in} = 0;\ k_2^{in} = 2;\ k_3^{in} = 2;\ k_4^{in} = 1$

$N = 4$

$\langle k^{in} \rangle = (0 + 2 + 2 + 1\ ) / 4$

$\langle k^{in} \rangle = 5 / 4$

$\langle k^{in} \rangle = 1{,}25$

Why should I care about degree?

Strength

Strength (or weighted degree)

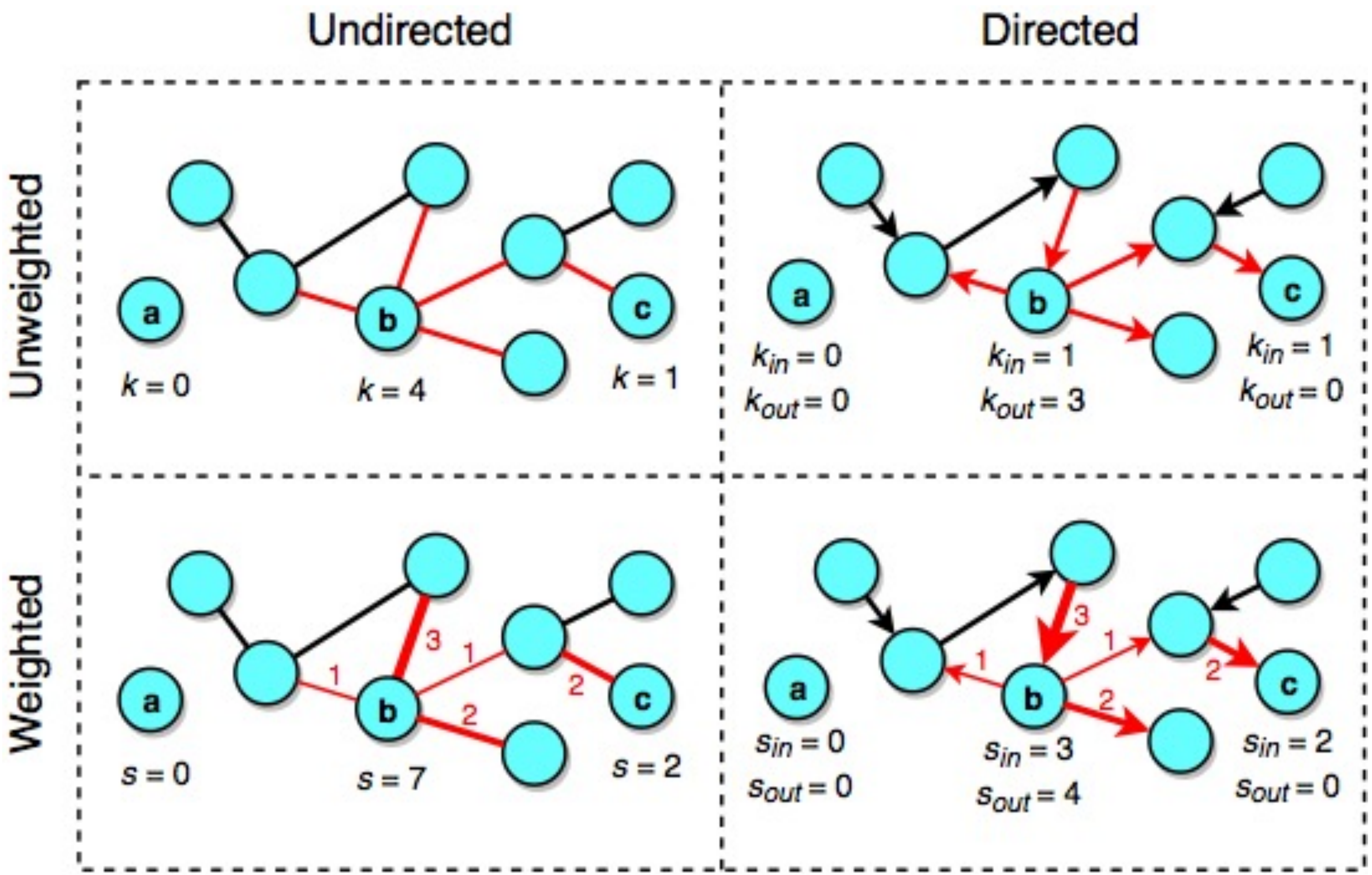
- In a **weighted network**, each edge has an associated weight
- The **strength of a node *i* (s_i)** is the sum of the weights of all edges connected to that node:

$$s_i = \sum_{j \in N(i)} w_{ij}$$

- **w_{ij}** is the weight of the edge between node *i* and *j*
- **N(i)** is the set of neighbours of node i

Strength (or weighted degree)

- In a **directed weighted network**, edges have both a **direction** and a **weight**, so we distinguish between **in-strength** and **out-strength** of a node.
- The **in-strength** of a node *i* is the sum of the weights of all edges **pointing to** the node
- The **out-strength** of a node *i* is the sum of the weights of all edges **leaving** the node
- The total **node strength** (sum of in-strength and out-strength) reflects a node's overall importance in a weighted network.



Source: Menczer, Fortunato, Davis, *A First Course in Network Science*, version 3 (2023).

Network Representations: Adjacency matrix

- **Adjacency matrix**: N x N matrix where each element $a_{ij} = 1$ if i and j are adjacent, $a_{ij} = 0$ otherwise.
- In undirected networks, the matrix is symmetric:
  $a_{ij} = a_{ji}$
- In directed networks, the adjacency matrix is **not** symmetric

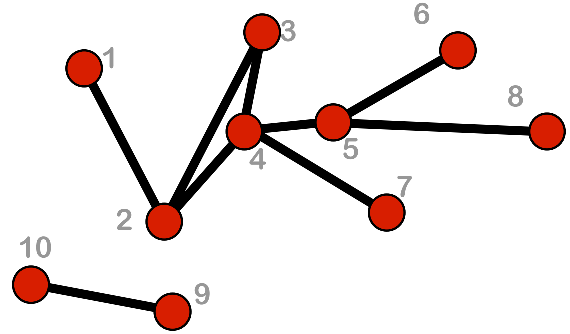
	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0
4	0	1	1	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0

Source: Menczer, Fortunato, Davis, *A First Course in Network Science*, version 3 (2023).

Edge List

- List of node pairs that are connected
- In directed networks the order of source and target matter!
- In weighted networks, each pair is replaced by a triple (*i*, *j*, *w*)

1	2
2	3
2	4
3	4
4	5
4	7
5	6
5	8
9	10



Source: Menczer, Fortunato, Davis, *A First Course in Network Science*, version 3 (2023).