Network Analysis

AN INTRODUCTION FOR HUMANISTS

Dr Katarzyna Anna Kapitan
18 February 2026

Recap

- Path, Distance, Diameter
- Small Worlds
- Connectedness
- Clustering coefficient
- Assortativity

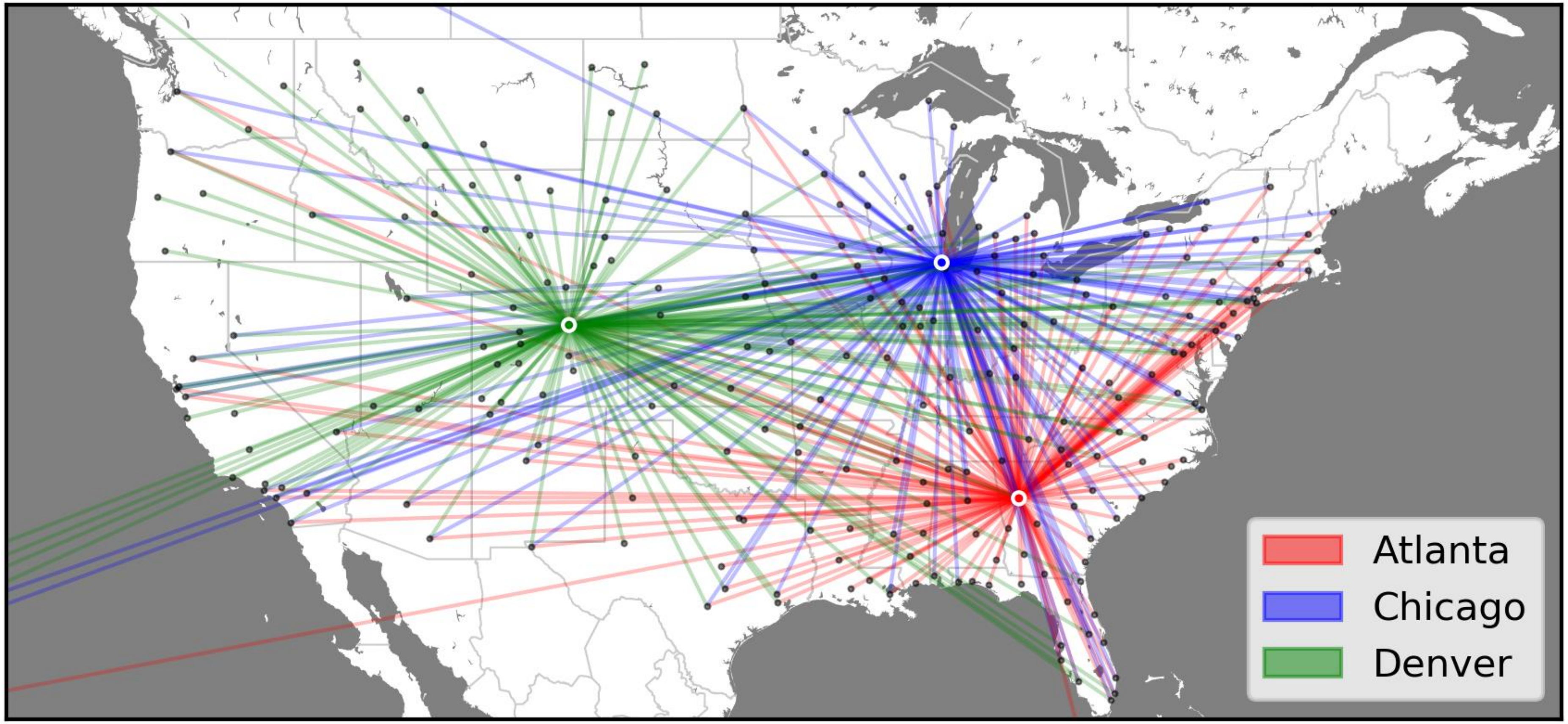
Today

- Hubs
- Centrality Measures
  - Degree
  - Closeness
  - Betweenness
- Robustness

Hubs

Hub – the central and most important part of a particular place or activity

(**Source**: Oxford Advanced Learner's Dictionary, acceseds 19/02/2024)



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

Hubs

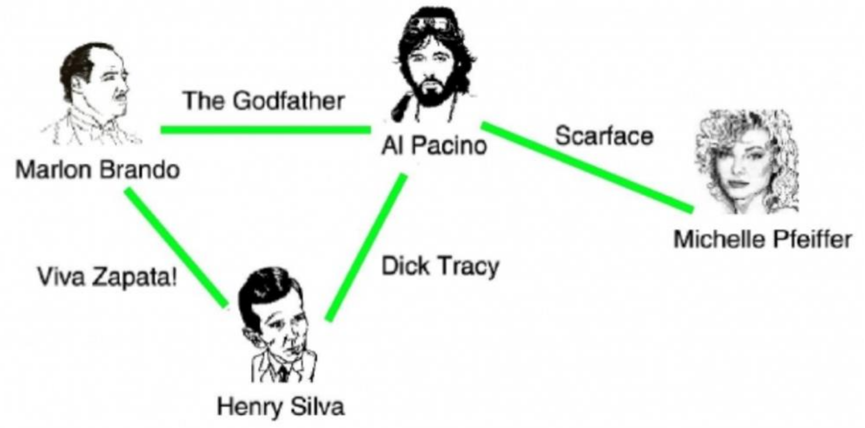
- Hubs ~ High-Degree Nodes (?)
  - nodes with degree ***much higher*** than **network's average degree** could be considered a **hub**

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

The average degree of the network (undirected network)

Hubs

- What is the average degree of the actor network?
- Who, in your opinion, are the hubs and why?
- What are their degrees?

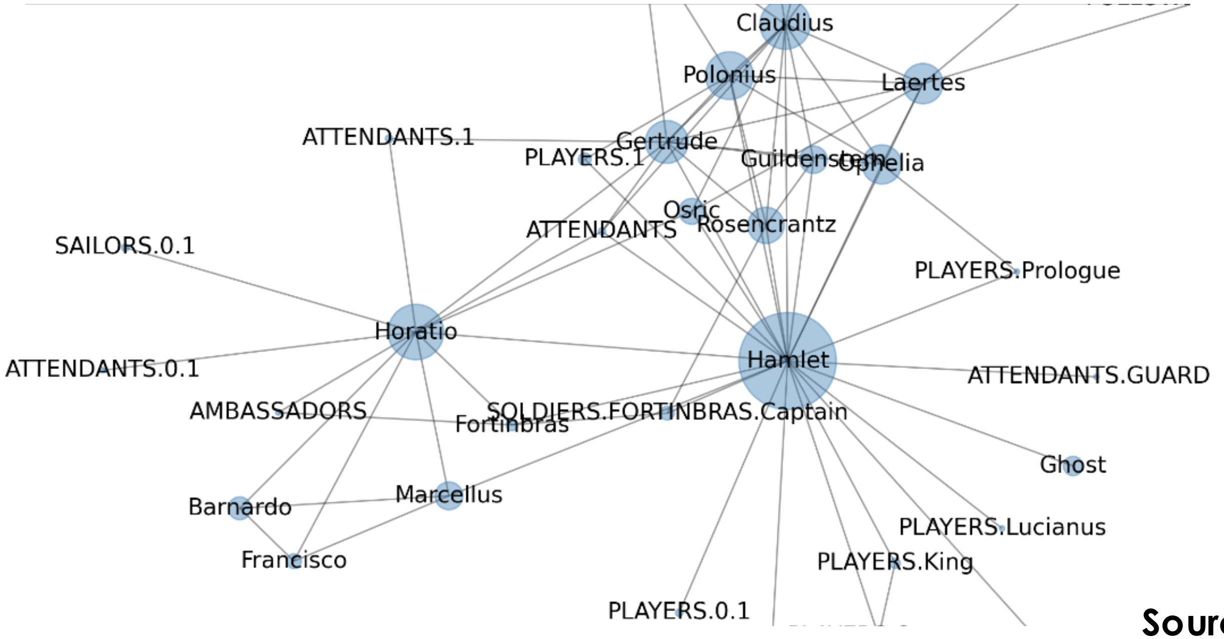


$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N}$$

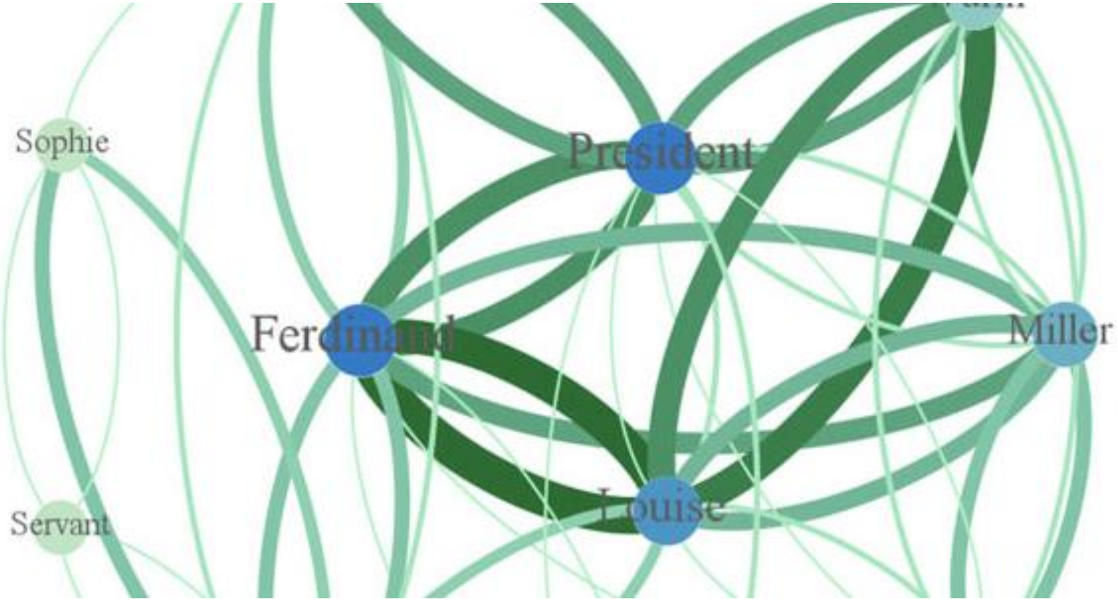
Source: Barabási, Network Science (https://networksciencebook.com)

Degree: Not the Whole Story

K_{Hamlet}=23



$K_{Ferdinand}$=7



Source: Network of *Hamlet* by Karsdorp et al., (left), Network of *Love and Intrigue* by Pan et al. (right)

Centrality Measures

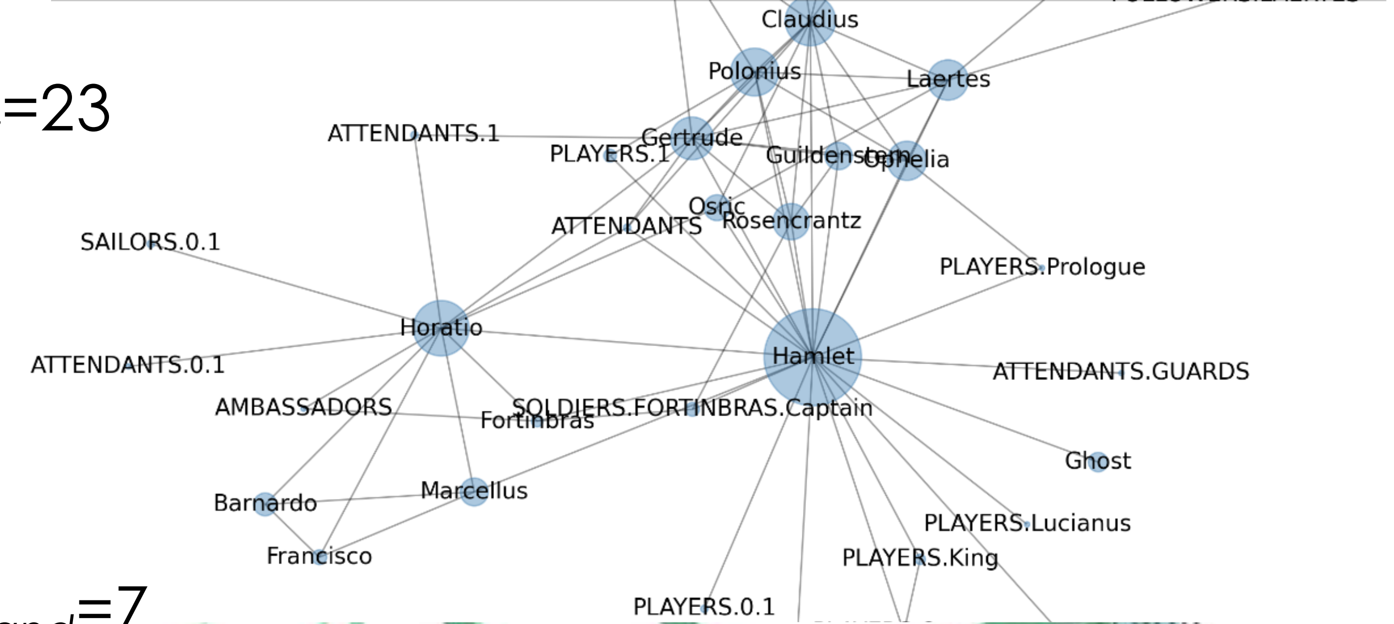
Centrality Measures

- **Centrality** is a measure of importance of a node in a system
  1. Degree Centrality
  2. Closeness Centrality
  3. Betweenness Centrality

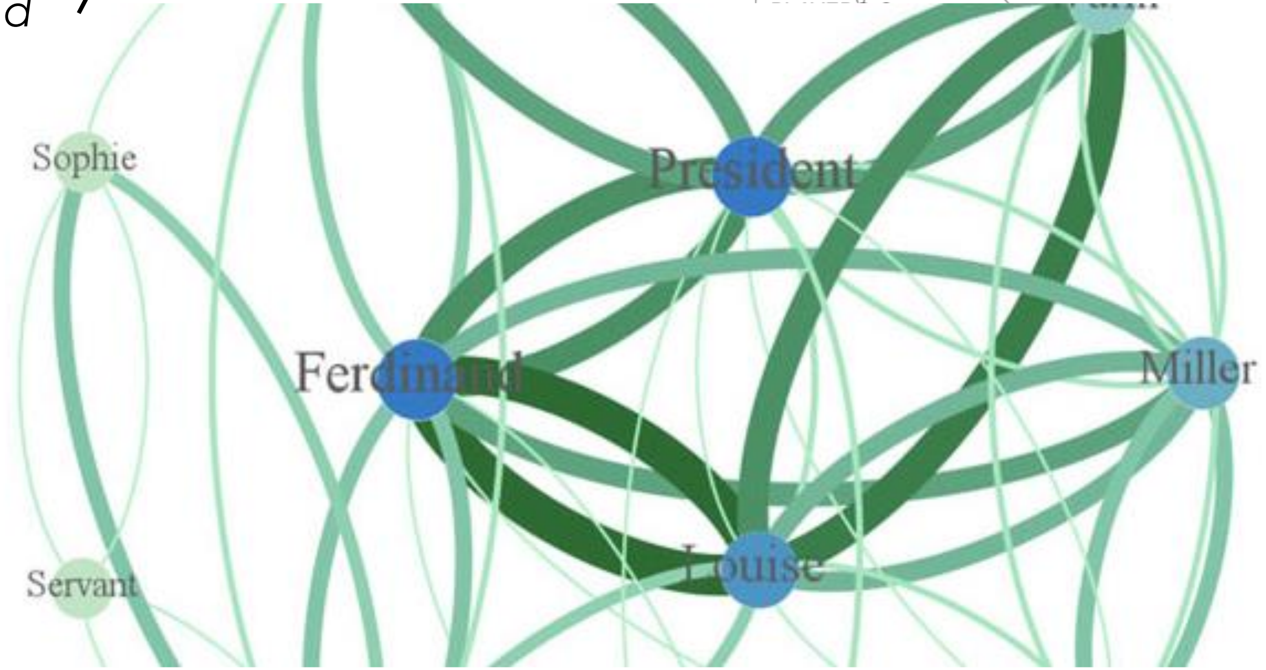
Degree Centrality and Degree Normalisation

A node is the more important the *higher* the number of nodes to which it is directly connected

K_{Hamlet}=23



$K_{Ferdinand}$=7

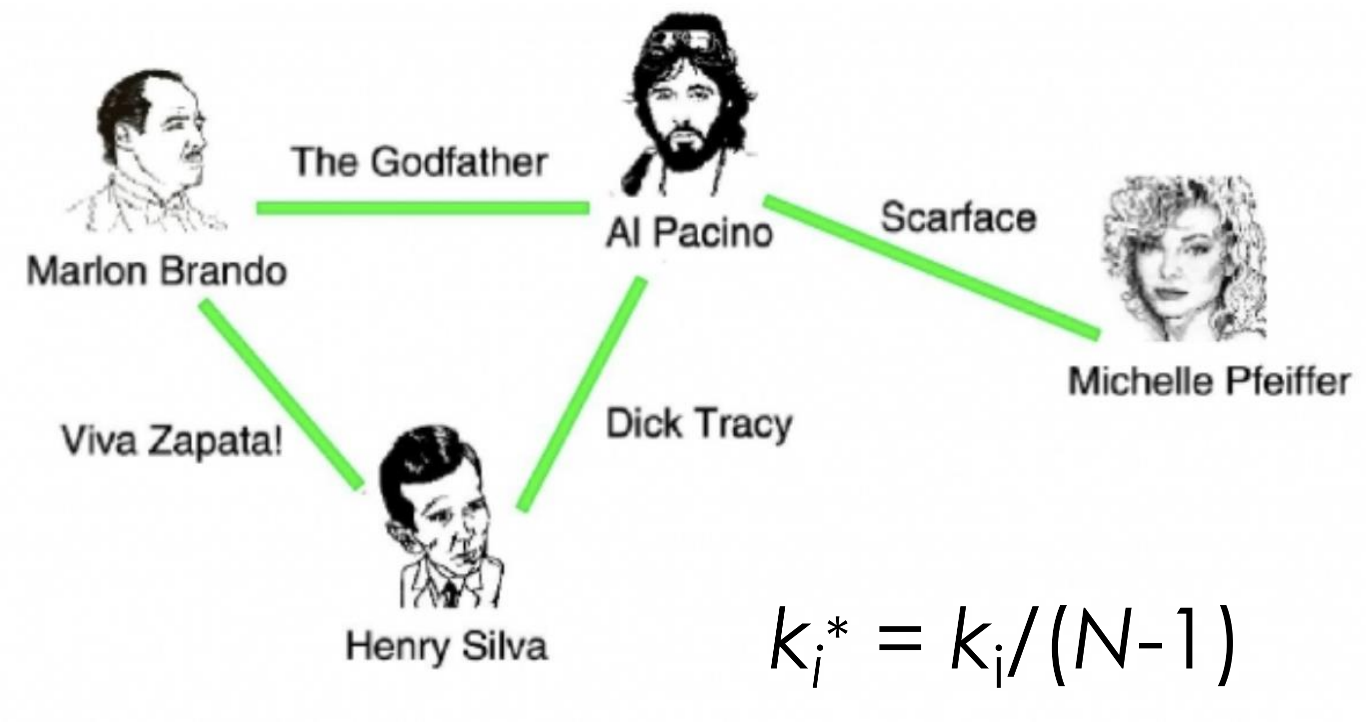


Source: Network of *Hamlet* by Karsdorp et al., (top), Network of *Love and Intrigue* by Pan et al. (bottom)

Degree Centrality and Degree Normalisation

- To compare the importance of hubs in different networks we have to **normalise** the degree scores.
- To **normalise** means to divide the degree score by the maximum possible degree of the node (*N*-1).
  - $k_i^* = k_i/(N-1)$
  - This places all scores in the range of 0 to 1; a node with a normalised degree of 1 is connected to all nodes in the network.

Degree Centrality and Degree Normalisation



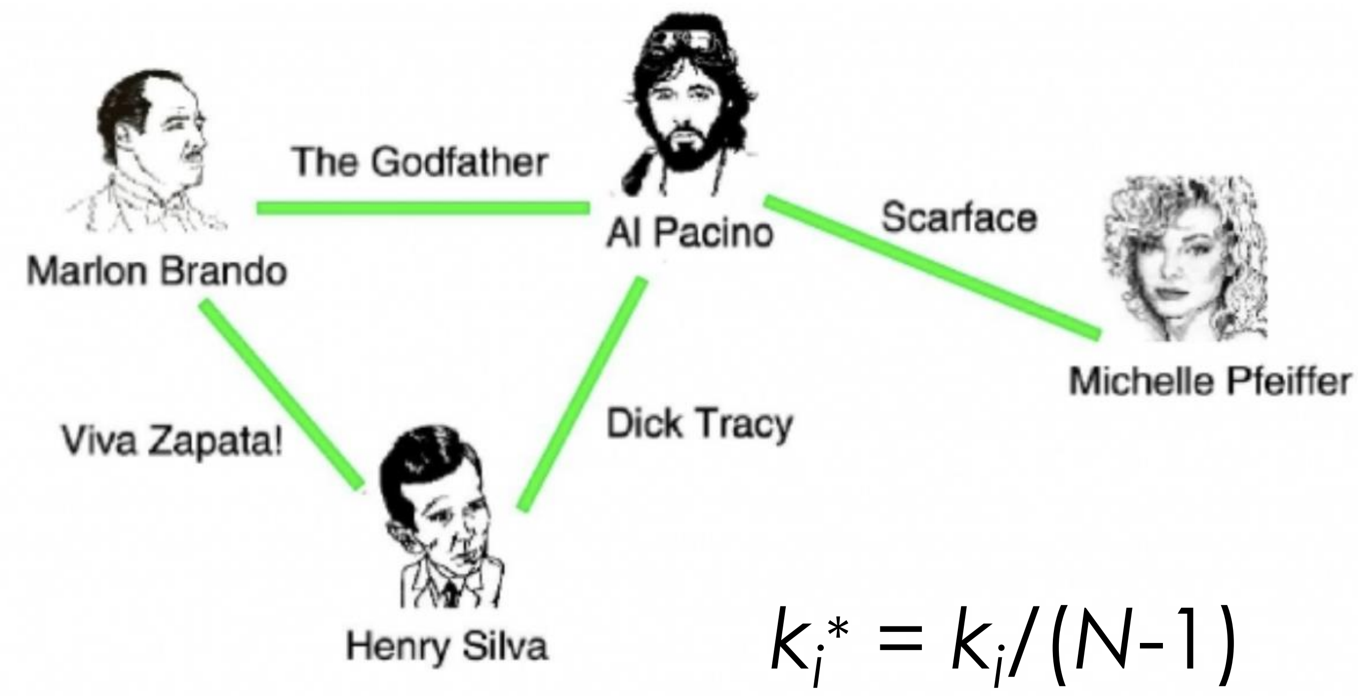
$$k_i^* = k_i/(N-1)$$

What is Al Pacino's normalised degree?

What is Marlon Brando's normalised degree?

What is Michelle Pfeiffer's normalised degree?

Degree Centrality and Degree Normalisation



$$k_i^* = k_i/(N-1)$$

What is Al Pacino's normalised degree?

$k_{Al\ Pacino}^* = 3/(4-1)$

$k_{Al\ Pacino}^* = 3/3$

$k_{Al\ Pacino}^* = 1$

What is Marlon Brando's normalised degree?

$k_{Marlon\ Brando}^* = 2/(4-1)$

$k_{Marlon\ Brando}^* = 2/3$

$k_{Marlon\ Brando}^* = 0.66$

Degree Centrality and Degree Normalisation

$$\widetilde{C_D}(i) = k_i$$

Unnormalised degree centrality = Degree

$$C_D(i) = \frac{k_i}{N-1} \qquad C_D(i) \in [0,1]$$

Normalised degree centrality (takes value between 0 and 1)

$$N: \text{ the total number of nodes in the network} k_i: \text{ the degr}$$

Degree & Degree Centrality

***Note**: in NetworkX you get different measures when you use nx.degree_centrality(G) and G.degree()*

- **nx.degree_centrality(G)** => The degree centrality values are normalized by dividing the actual degree of a node by the maximum possible degree in a simple graph *N*-1 where *N* is the number of nodes in G.
- **G.degree()** => The node degree is the number of edges adjacent to the node.

Closeness Centrality

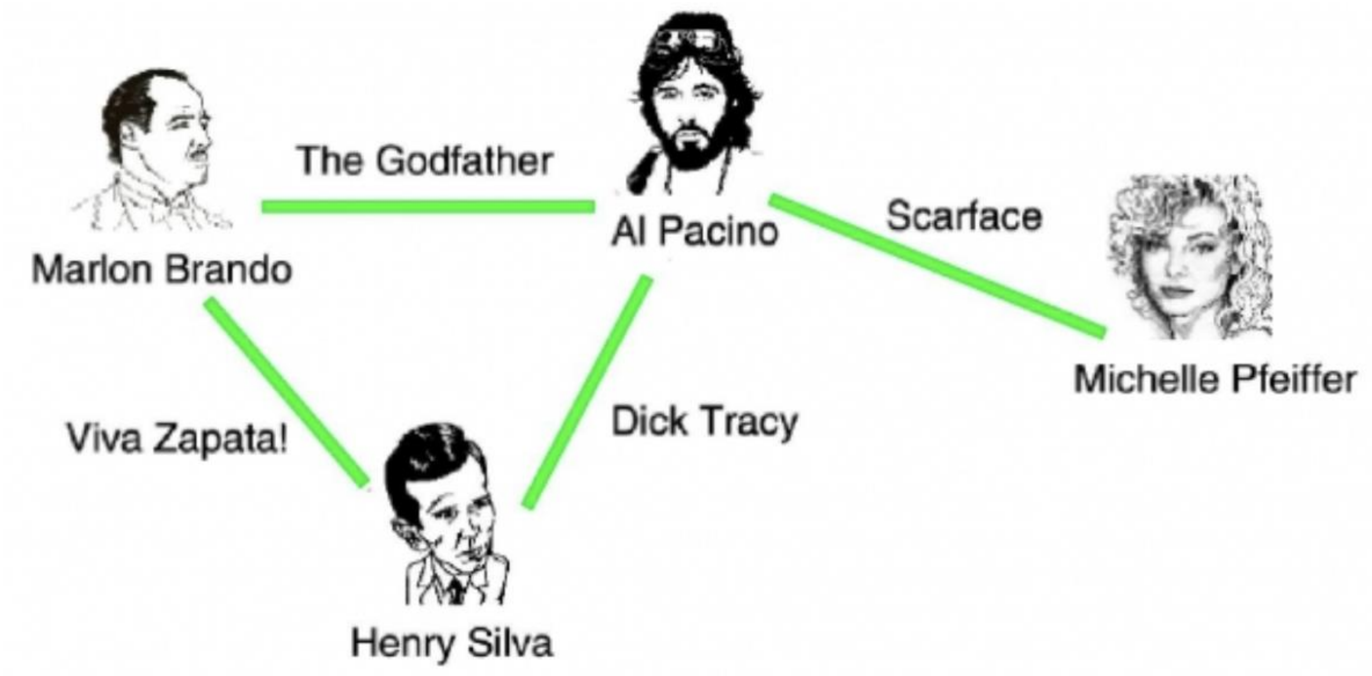
Closeness Centrality

- **Idea**: a node is the more central the *closer* it is to the other nodes, on average

$$\widetilde{C_C}(i) = \frac{1}{\sum_{j \neq i} d(i,j)}$$

- $\tilde{C}_C(i)$ is the unnormalized closeness centrality of node *i*,
- *d*(*i,j*) is the length of the shortest path (distance) between node *i* and node *j*,

Closeness Centrality



What is Al Pacino's closeness centrality?

What is Michelle Pfeiffer's closeness centrality?

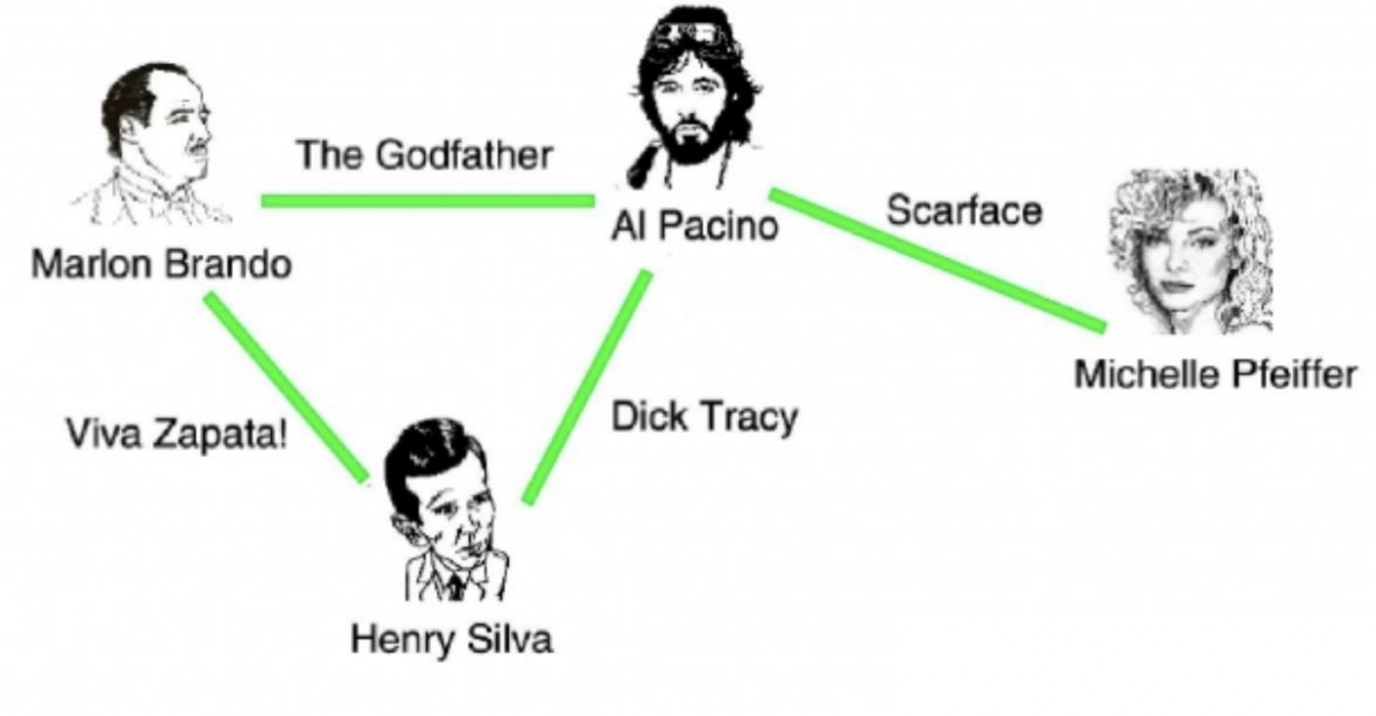
$$\widetilde{C_C}(i) = \frac{1}{\sum_{j \neq i} d(i,j)}$$

Normalised Closeness Centrality

- $C_C(i) = \frac{N-1}{\sum_{j \neq i} d(i,j)}$

- Cc(*i*) is the normalised closeness centrality of node *i*,
- *d*(*i,j*) is the length of the shortest path (distance) between node *i* and node *j*,
- *N – number of nodes in the graph*

Normalised Closeness Centrality



What is Al Pacino's normalised closeness centrality?

What is Michelle Pfeiffer's normalised closeness centrality?

$$C_C(i) = \frac{N-1}{\sum_{j \neq i} d(i,j)}$$

Normalised Closeness Centrality

***Note**: in NetworkX you get always the normalised value*

nx.closeness_centrality(G) => The closeness centrality values are normalized by adjusting it to the size of the network

(Source: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.closeness_centrality.html)

Betweenness Centrality

Betweenness Centrality

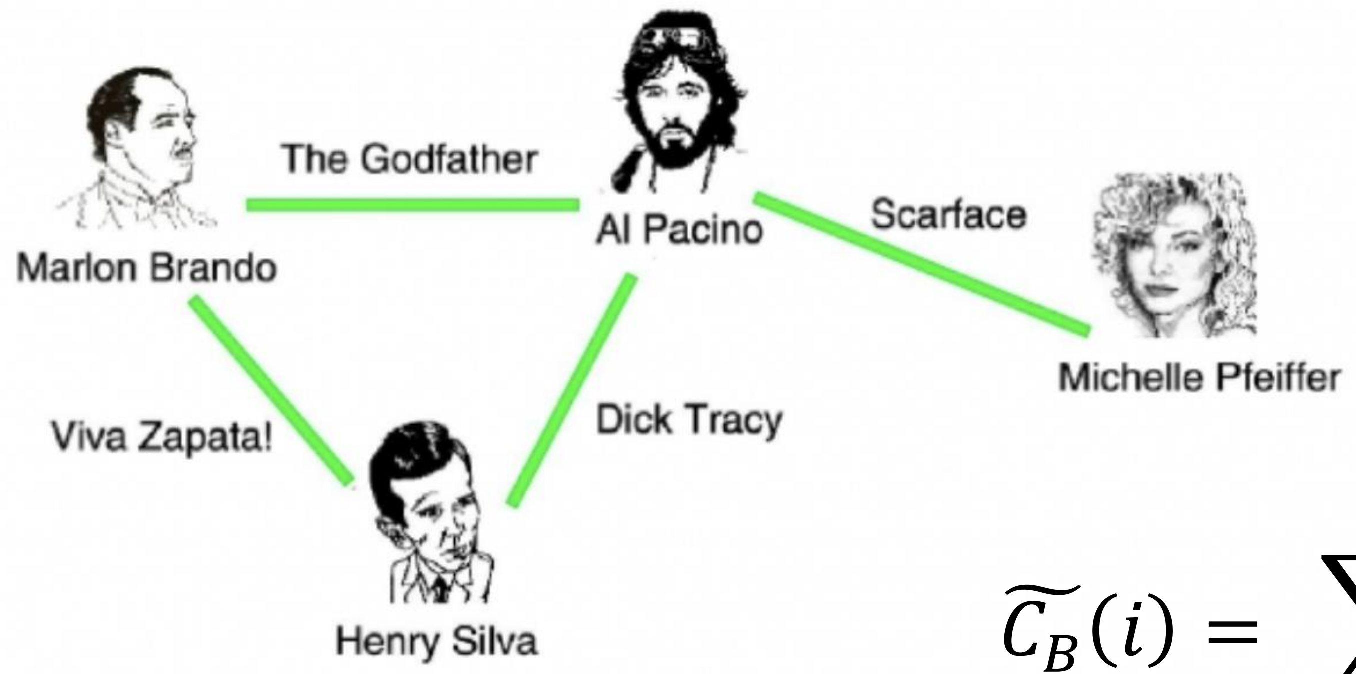
- **Idea**:
  - a node is the more central the *more often it is crossed by paths*

$$\widetilde{C_B}(i) = \sum_{j \neq i \neq h} \frac{\sigma_{jh}(i)}{\sigma_{jh}}$$

σ_{hj} : number of shortest paths from h to j

$\sigma_{hj}(i)$: number of those paths from h to j that run through i

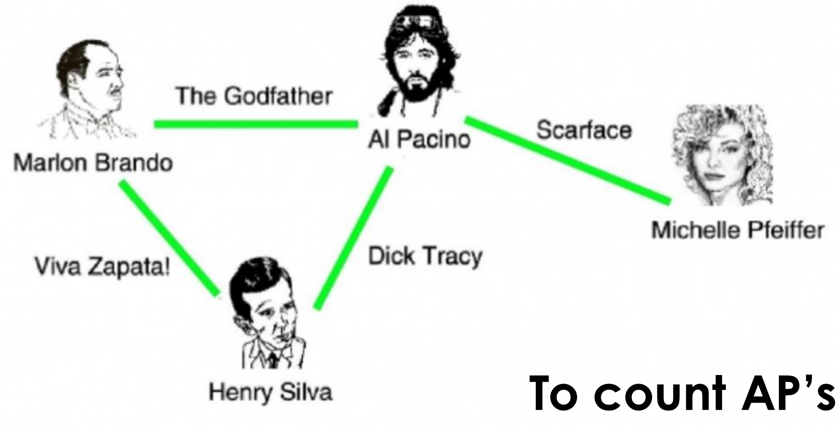
Betweenness Centrality



What is Al Pacino's betweenness centrality?

What is Michelle Pfeiffer's betweenness centrality?

$$\widetilde{C_B}(i) = \sum_{j \neq i \neq h} \frac{\sigma_{jh}(i)}{\sigma_{jh}}$$



$$\widetilde{C_B}(i) = \sum_{j \neq i \neq h} \frac{\sigma_{jh}(i)}{\sigma_{jh}}$$

To count AP's centrality

	No of shortest paths (x)	No of shortest paths going through AP (y)	y/x
MP - HS	1	1	1
MP - MB	1	1	1
MB - HS	1	0	0
		Sum of y/x	2

Note that *AP* is not an endpoint of any of these paths

Normalised Betweenness Centrality

In **undirected network** to normalise we divide the sum by the number of pairs of nodes excluding *i* where *n* is the number of nodes in the graph:

$$C_B(i) = \frac{\widetilde{C_B}(i)}{\frac{(N-1)(N-2)}{2}}$$

For a **directed graph**, the normalisation factor is different.
Can you guess what is it?

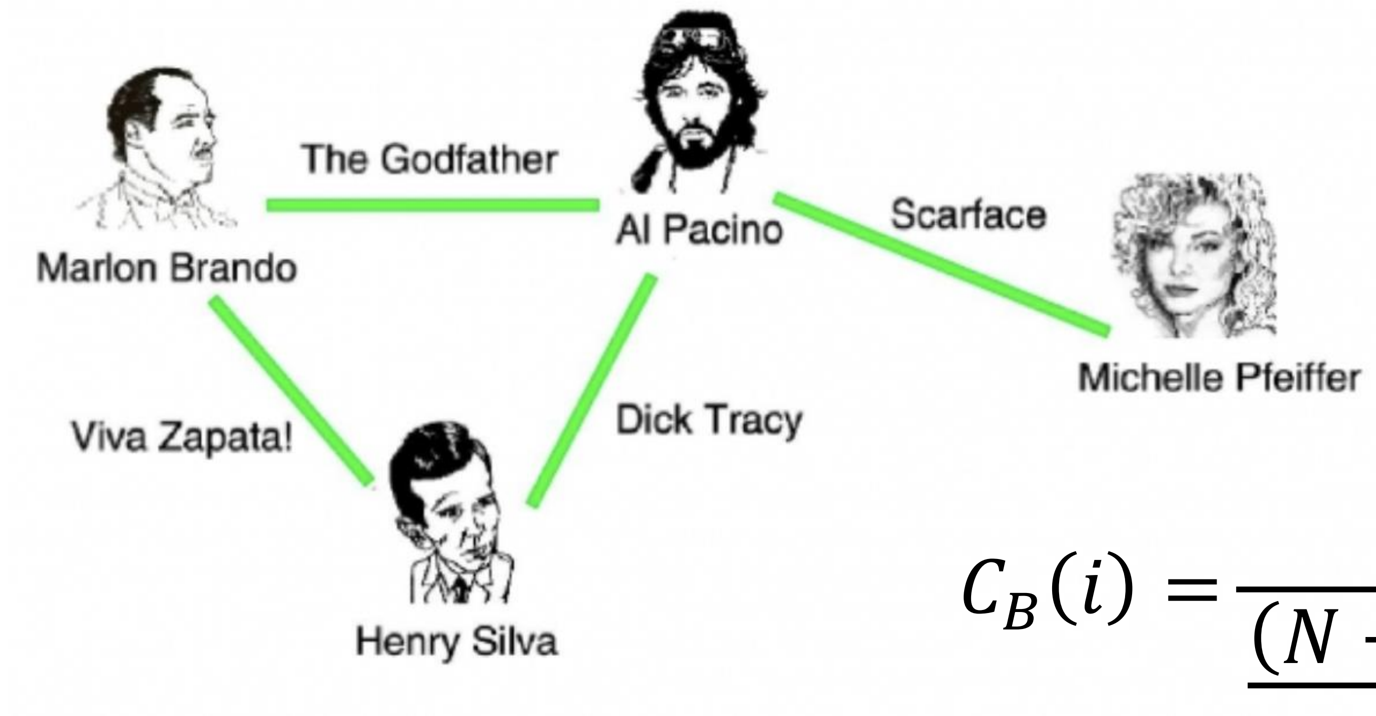
Normalised Betweenness Centrality

In **undirected network** to normalise we divide the sum by the number of pairs of nodes excluding *i* where *n* is the number of nodes in the graph:

$$C_B(i) = \frac{\widetilde{C_B}(i)}{\frac{(N-1)(N-2)}{2}}$$

For a **directed graph**, the normalisation factor is different.
Can you guess what is it?
It's (N-1)(N-2), so no division by 2.

Normalised Betweenness Centrality



What is Al Pacino's normalised betweenness centrality?

What is Michelle Pfeiffer's normalised betweenness centrality?

$$C_B(i) = \frac{\widetilde{C_B}(i)}{\frac{(N-1)(N-2)}{2}}$$

Normalised Betweenness Centrality

***Note**: in NetworkX by default you get the normalised value*

nx. betweenness_centrality(G) => Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that pass through v.

(**Source:** https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.betweenness_centrality.html)

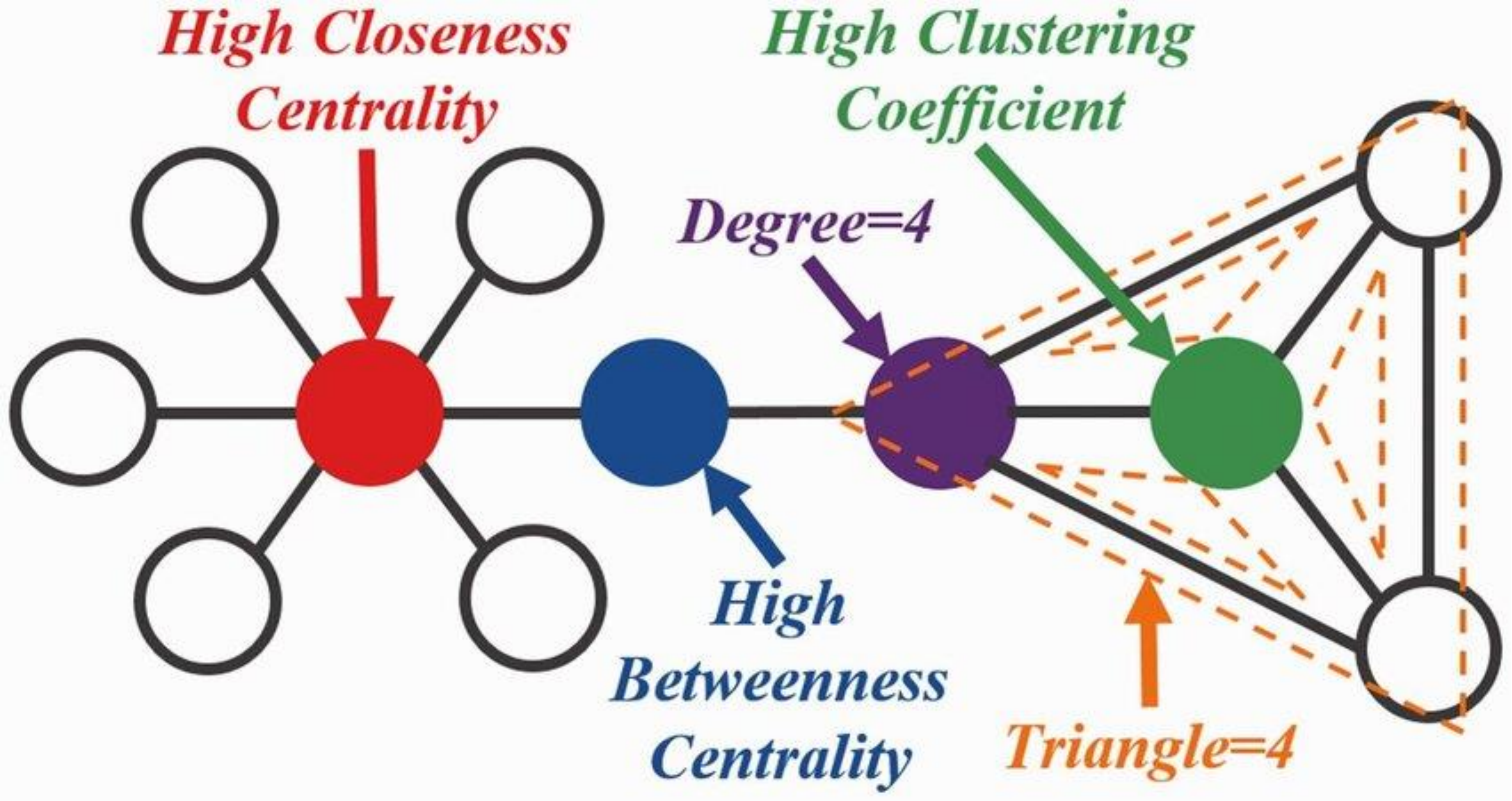
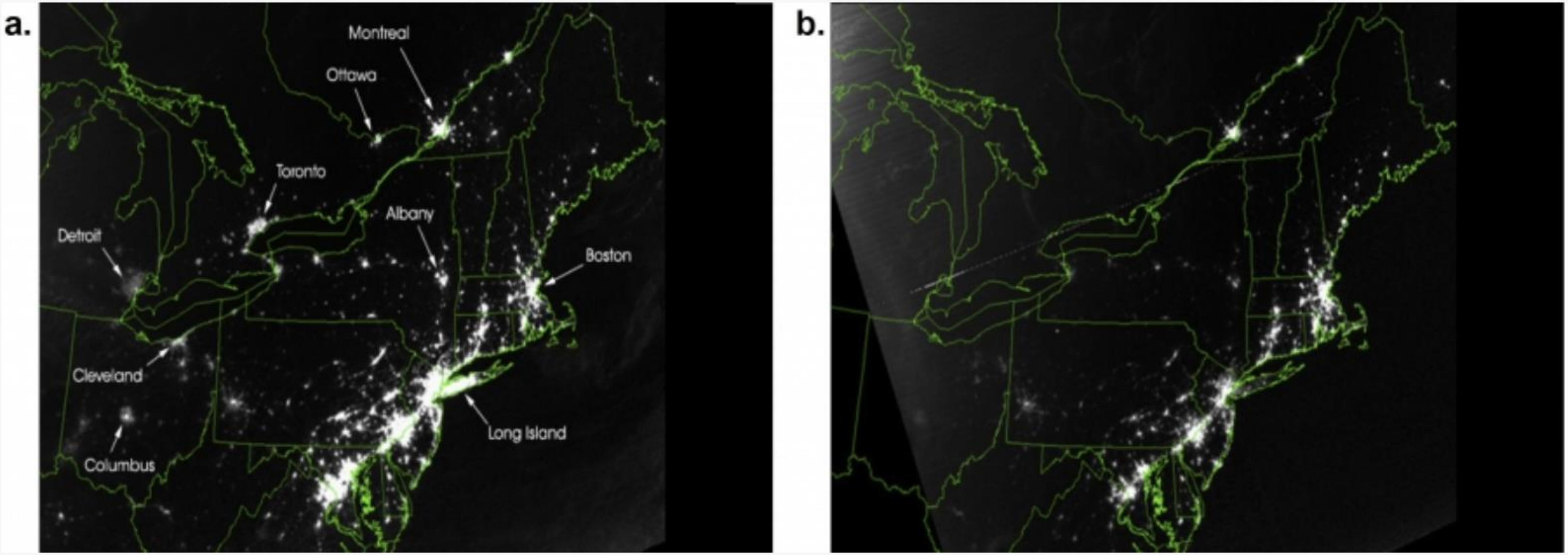


Image source: Zhang et al. (2016) A protein network descriptor server and its use in studying protein, disease, metabolic and drug targeted networks. DOI: 10.1093/bib/bbw071

Robustness



North American Blackout - 45 million people without power in USA and 10 million in Canada
a. Satellite Image of Northeast Unites Stated and Southeast Canada on 13 August 2003 at 9:29 pm
b. The same but 5 hour later.
Source: Barabási, Network Science (https://networksciencebook.com)

Robustness

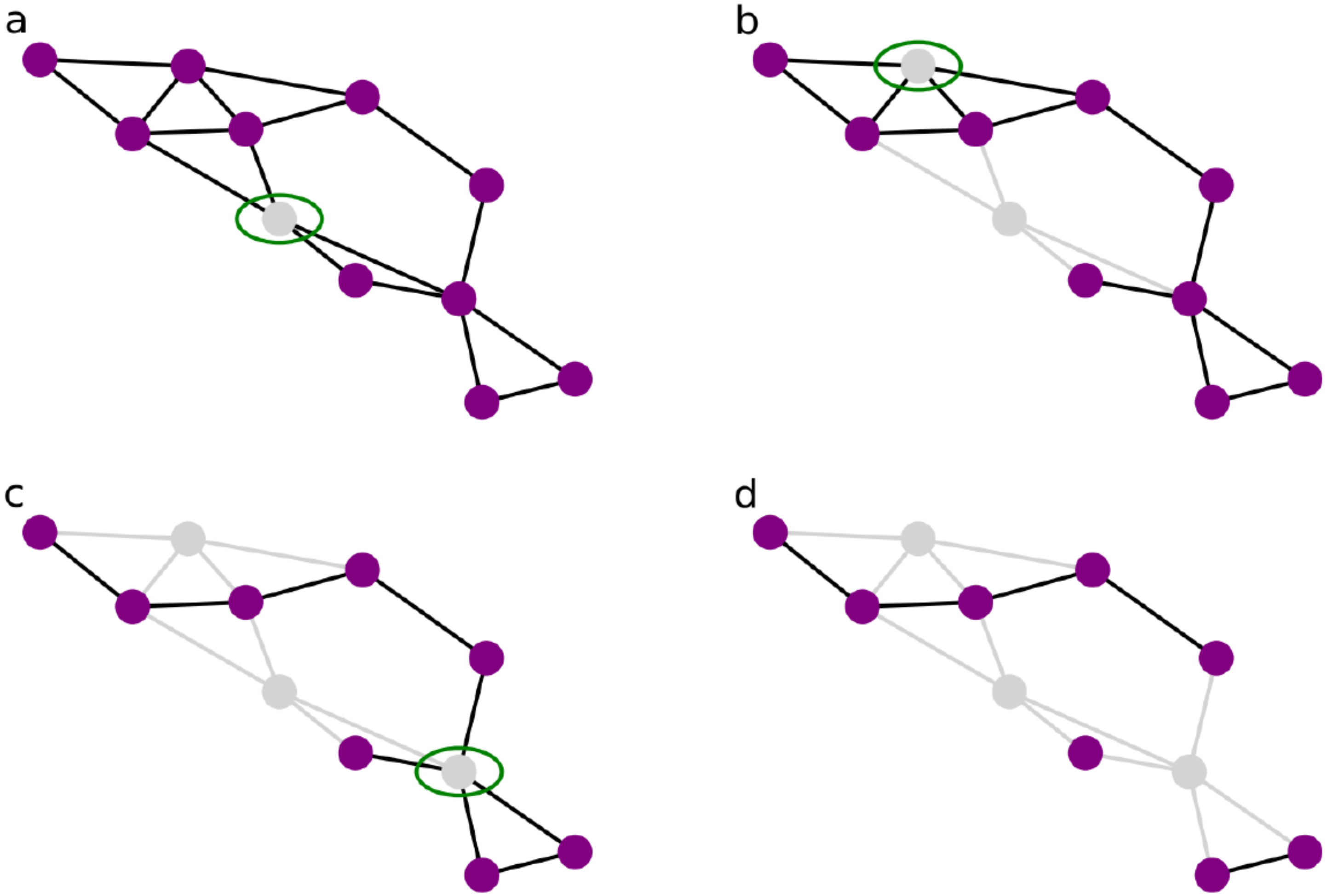
- A system is **robust** if the failure of some of its components does not affect its function
- **Question:** how can we define the robustness of a network?

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

Robustness

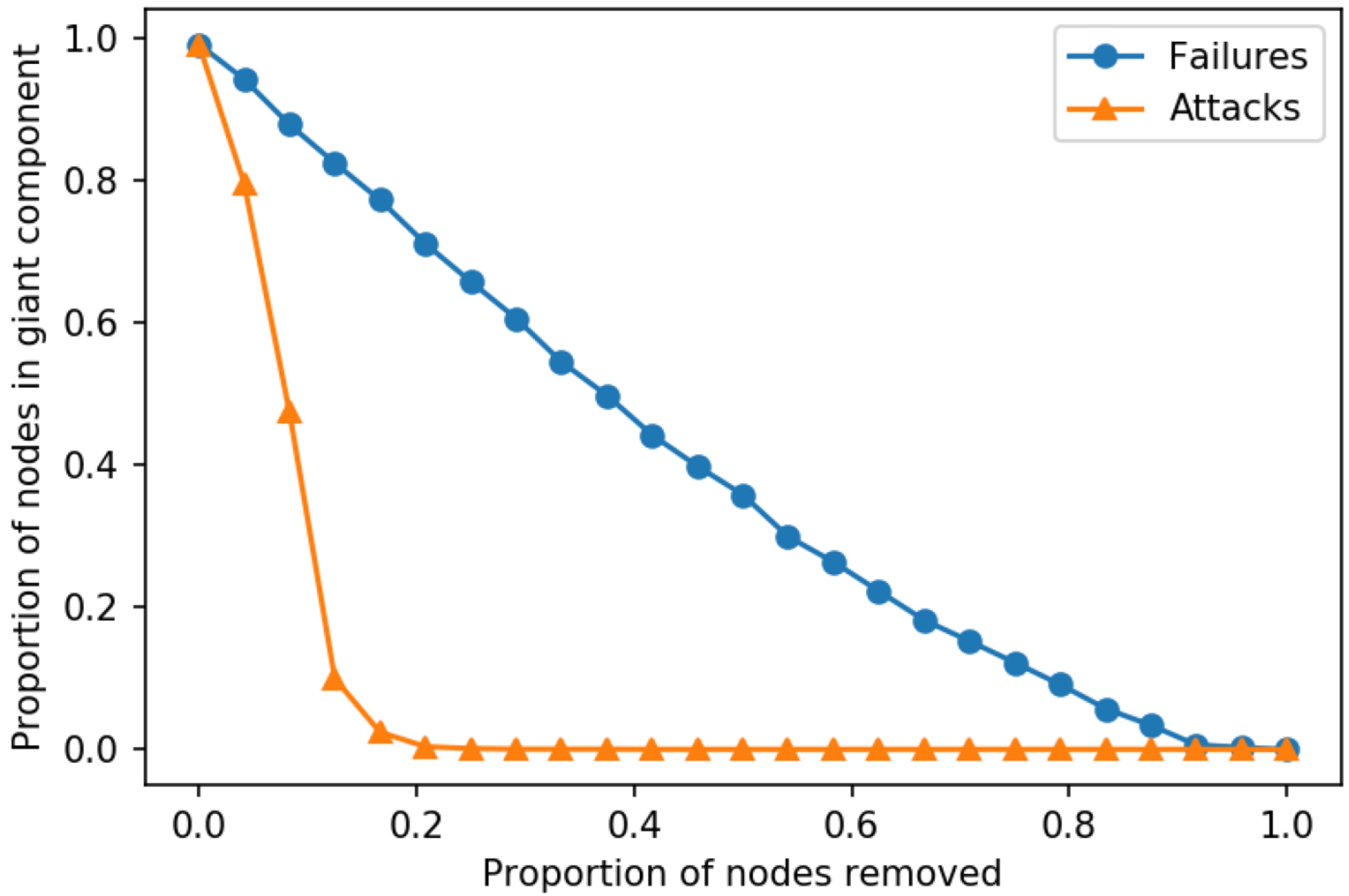
- A system is **robust** if the failure of some of its components does not affect its function
- **Question:** how can we define the robustness of a network?
- **Answer:** we remove nodes and/or links and see what happens to its structure
- **Key point:** connectedness *(A network is **connected** if there is a path between any two nodes – Week3)*

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



Robustness

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



Robustness

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

Robustness

- **Random failures:** nodes break down randomly, so they are all chosen with the **same probability**
- **Attacks:** hubs are deliberately targeted — the larger the **degree**, the higher the probability of removing the node

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

Questions ?