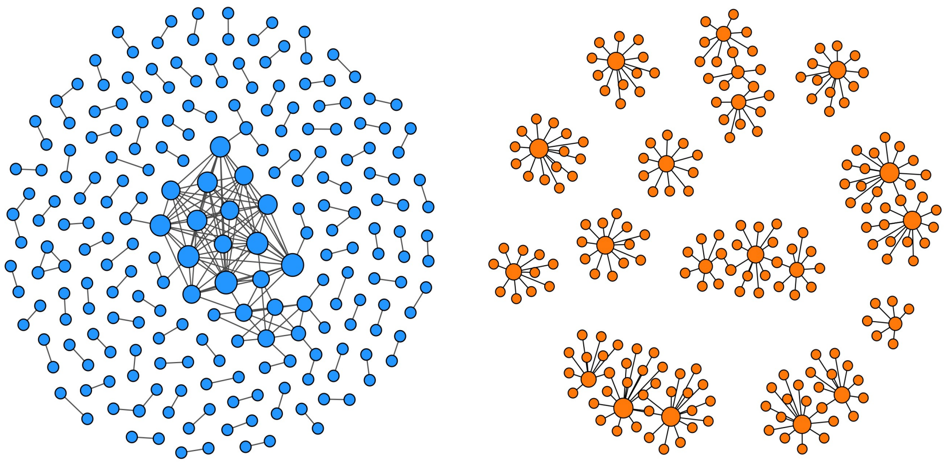
Network Analysis

AN INTRODUCTION FOR HUMANISTS

Dr Katarzyna Anna Kapitan
20 February 2025

Recap from last week

- Path, Distance, Diameter
- Average Shortest-Path Length & Small Worlds
- Connectedness
- Clustering coefficient
- Assortativity

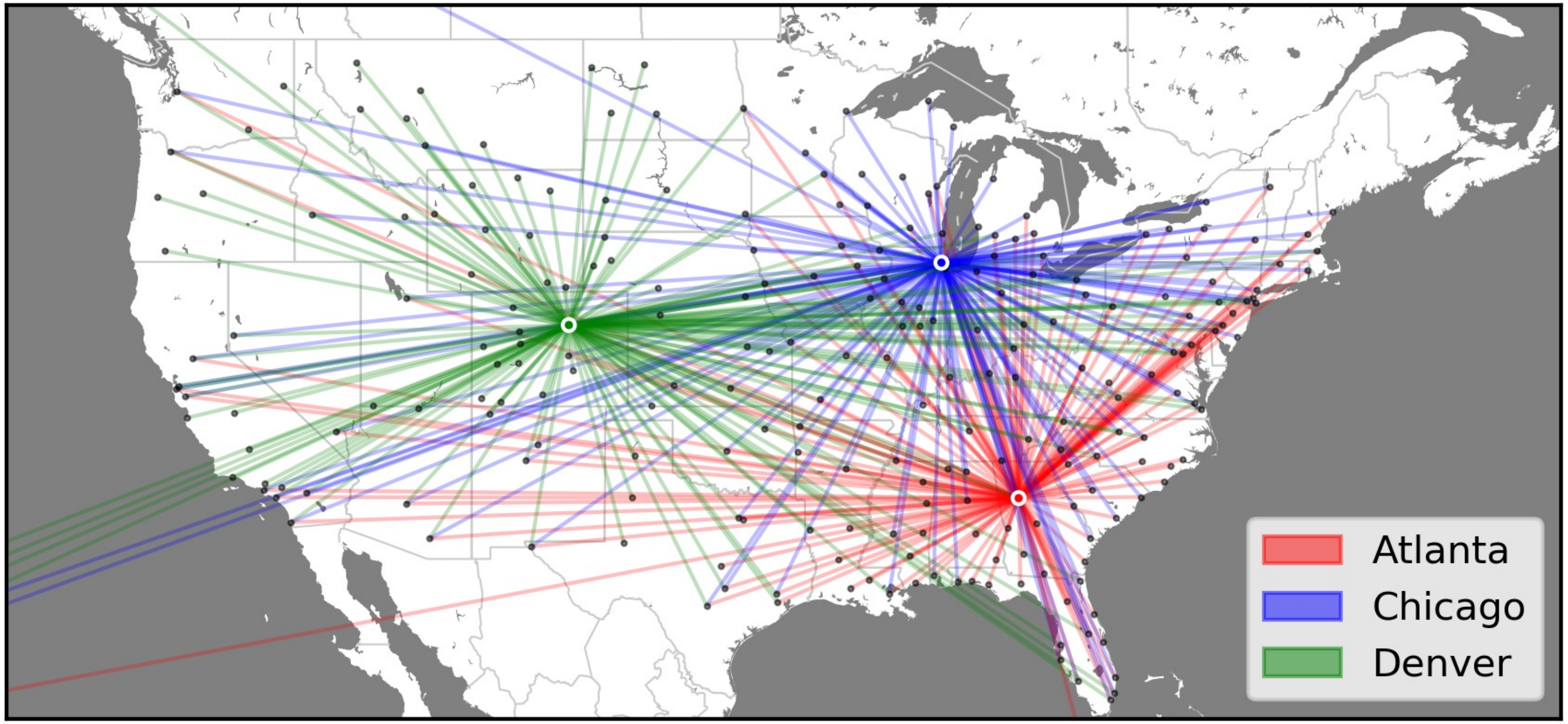


Source: Menczer, Fortunato, Davis, *A First Course in Network Science* (2023).

Hubs

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

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



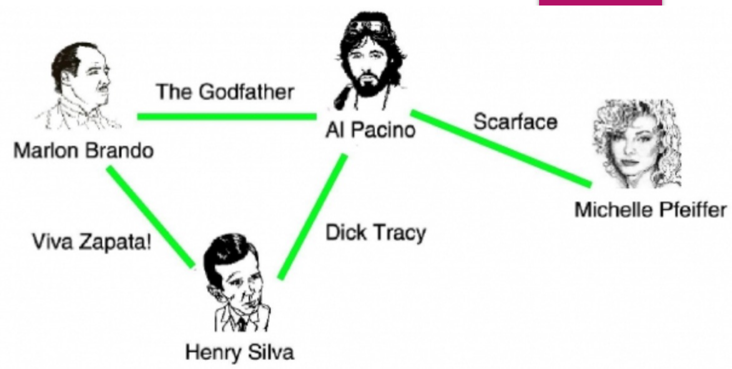
Source: Menczer, Fortunato, Davis, *A First Course in Network Science* (2023).

Hubs

- In general, high-degree nodes are called **hubs**
- Degree of a node is the **number of neighbours** of the node
- The definition of a "high-degree" is a bit less clear, but in principle any node with degree higher than network's average degree can be considered a **hub**.

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

Hubs



- What is the average degree of the actor network?
- Who 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)

Centrality Measures

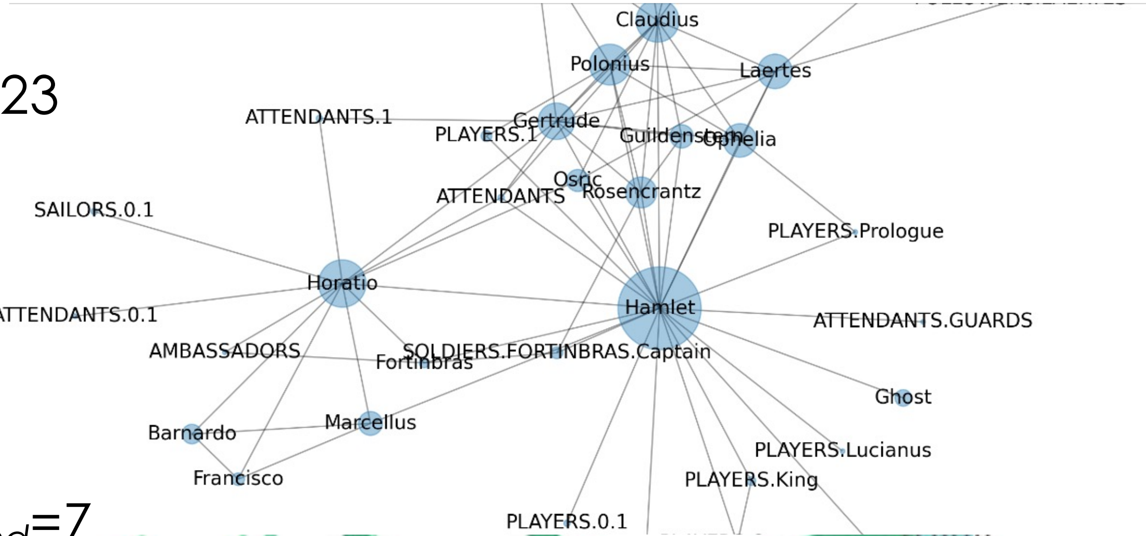
Centrality Measures

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

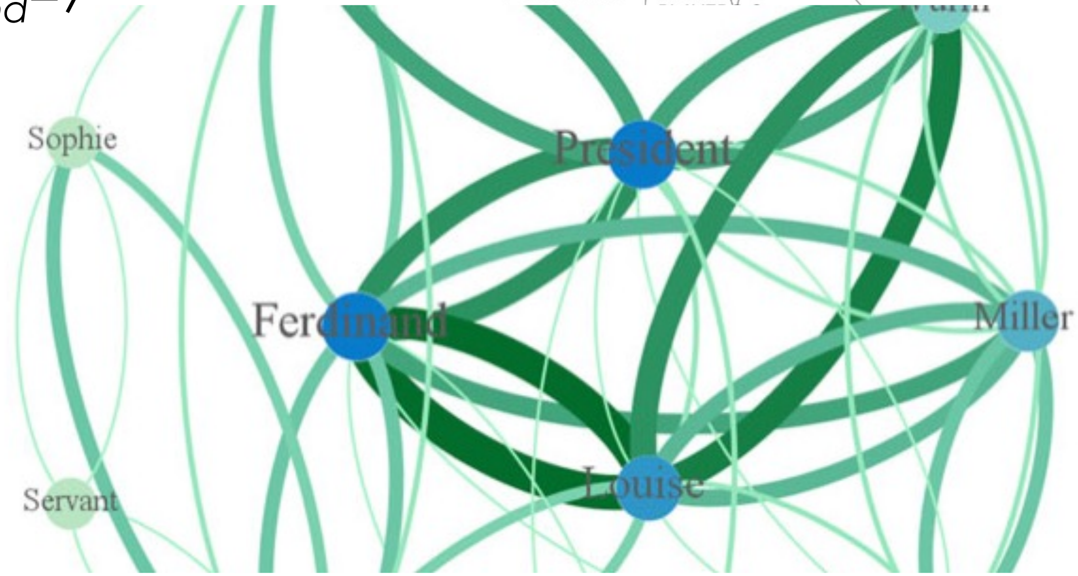
Degree Centrality and Degree Normalisation

- **Idea**: a node is the more important the *higher* the number of nodes to which it is directly connected
- When working with individual small networks degree is a useful measure in itself, but what if you want to compare different networks?
- What do should we keep in mind?

K_{Hamlet}=23



$K_{Ferdinand}$=7

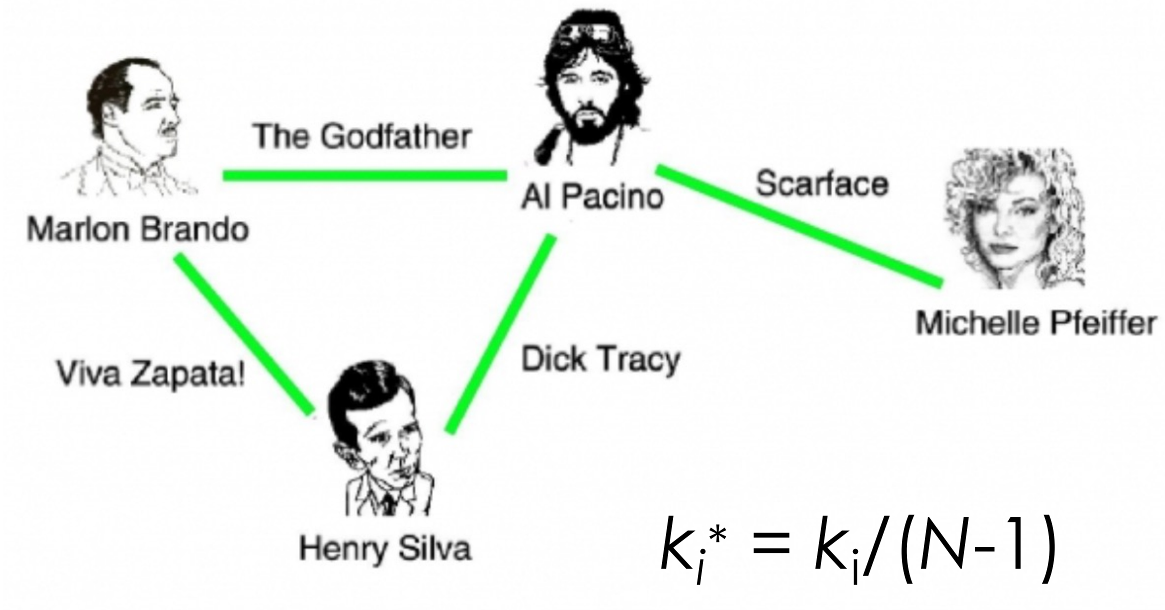


Source: Network of *Hamlet* by Karsdorp et al., (top), Network of *Love and Intrigue* by Pan et al. (bottom)

Degree Centrality and Degree Normalisation

- To compare the importance of hubs in different networks we have to **normalise** the degree scores.
- To **normalise** here means to divide the degree score by the maximum possible number of edges (or simply *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



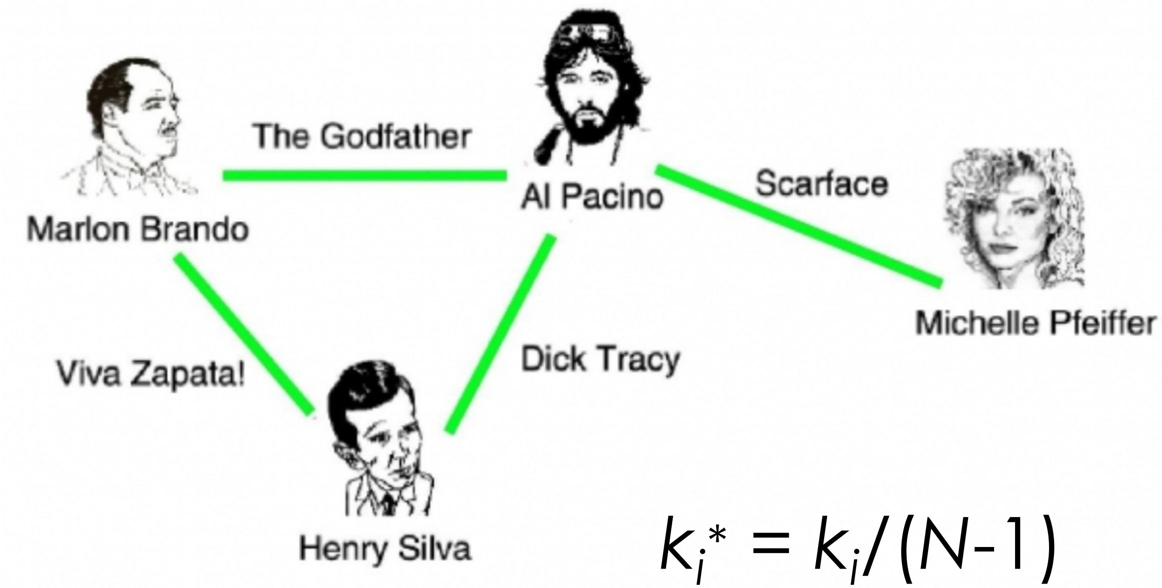
$$k_i^* = k_i/(N-1)$$

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

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

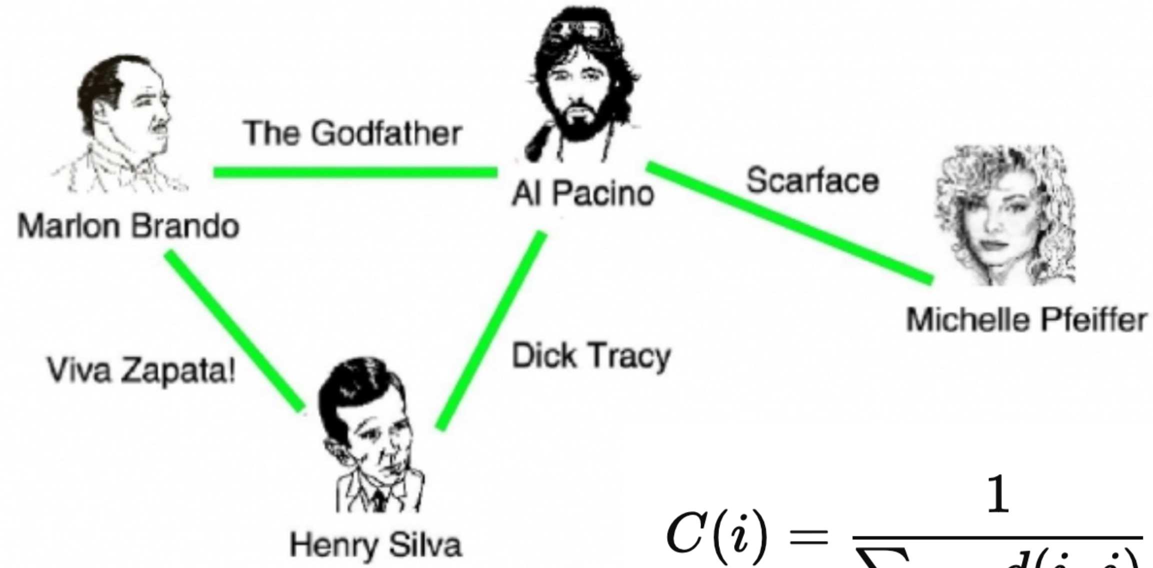
Closeness Centrality

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

$$C(i) = \frac{1}{\sum_{j \in V} d(i, j)}$$

- C(*i*) is the closeness centrality of node *i*,
- *d*(*i,j*) is the shortest path distance between node *i* and node *j*,
- *V* is the set of all nodes in the graph.

Closeness Centrality



$$C(i) = \frac{1}{\sum_{j \in V} d(i,j)}$$

What is Al Pacino's closeness centrality?

What is Michelle Pfeiffer's closeness centrality?

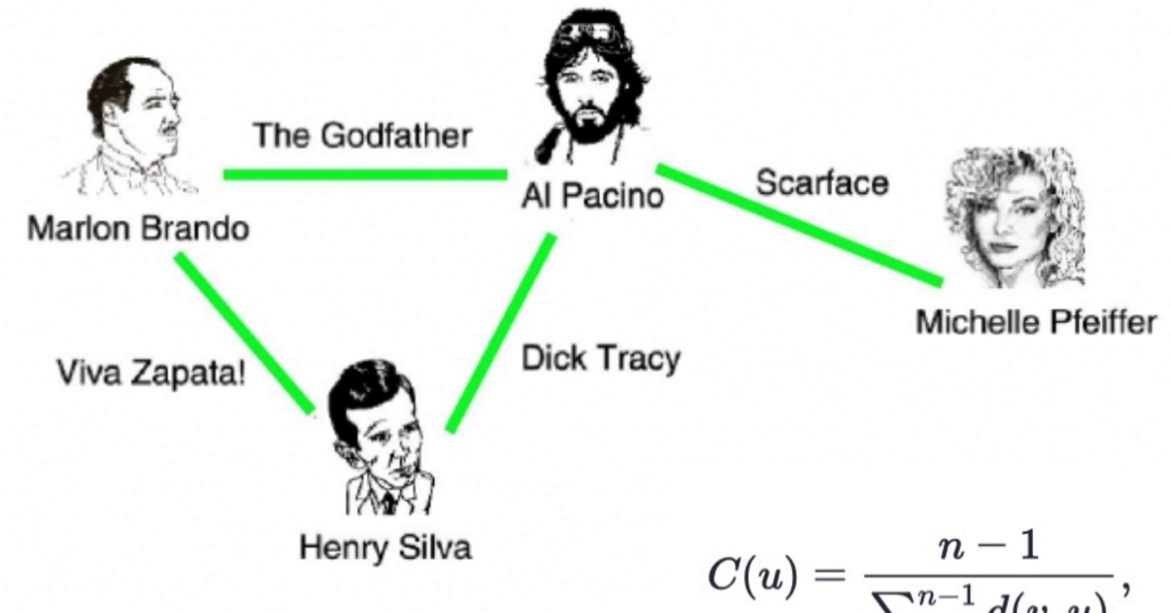
Normalised Closeness Centrality

- Again, we are not necessarily interested in absolute values, but rather in normalised measure which allows us to compare features across networks.

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)},$$

- $C(u)$ is the closeness centrality of node u,
- $d(v,u)$ is the shortest path distance between node v and node u,
- *n – number of nodes in the graph*

Normalised Closeness Centrality



$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)},$$

What is Al Pacino's normalised closeness centrality?

What is Michelle Pfeiffer's normalised closeness centrality?

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.

‘Closeness centrality of a node *u* is the reciprocal of the average shortest path distance to *u* over all n-1 reachable nodes. […] The closeness centrality is normalized to (n-1)/(|G|-1) where *n* is the number of nodes in the connected part of graph containing the node’. (Source: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.closeness_centrality.html)

Betweenness Centrality

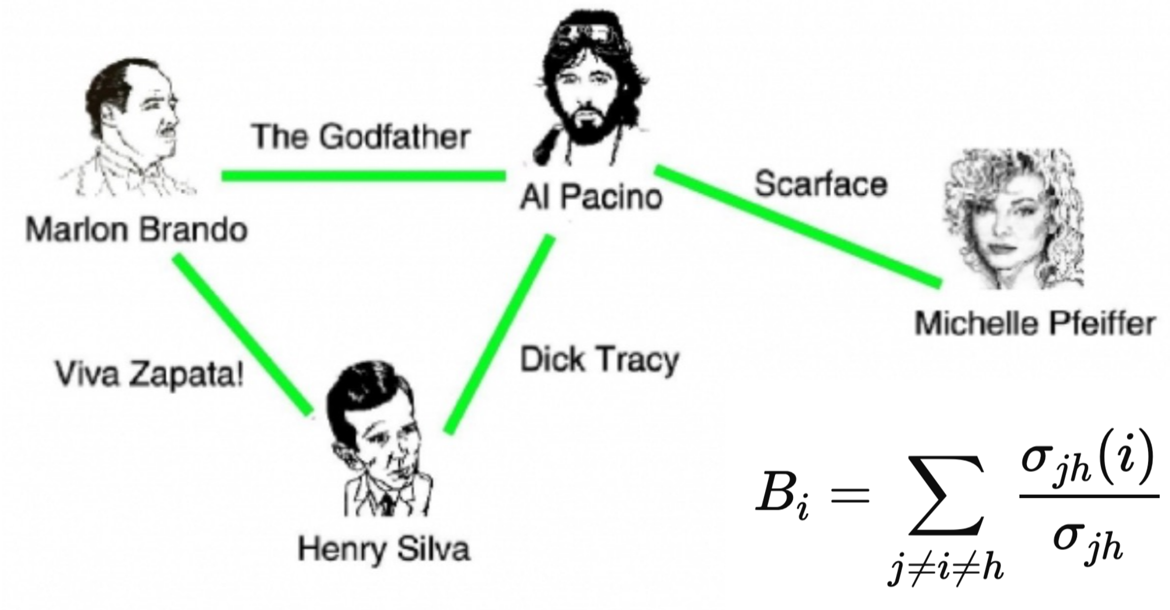
Betweenness Centrality

- **Idea**: a node is the more central the *more often it is crossed by paths*
- *Bi is betweenness centrality of node i*

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

- 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

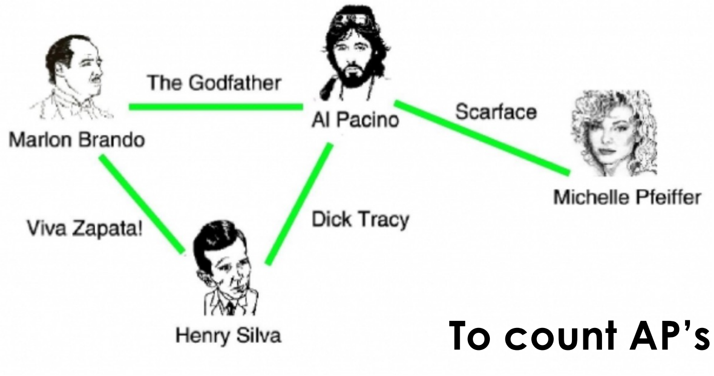
Betweenness Centrality



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

What is Al Pacino's betweenness centrality?

What is Michelle Pfeiffer's betweenness centrality?



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

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

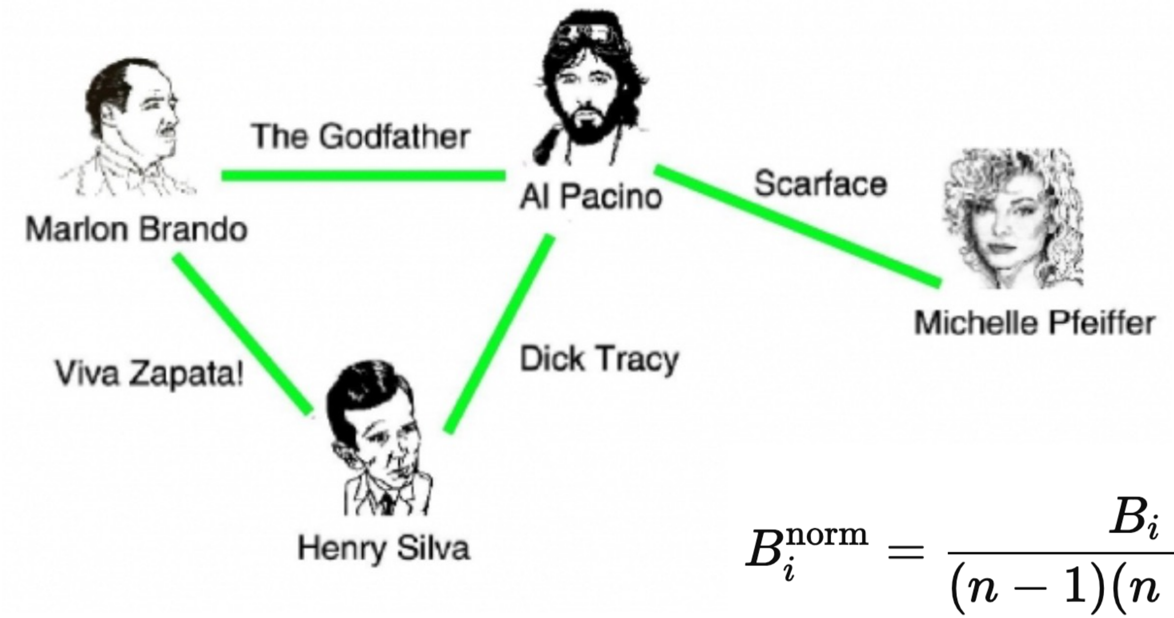
Normalised Betweenness Centrality

Again, we are not necessarily interested in absolute values, but rather in normalised measure which allows us to compare features across networks.

Therefore, in undirected network we divide the sum by 2/((n-1)(n-2)) and in directed networks by 1/((n-1)(n-2)) where n is the number of nodes in the graph.

$$B_i^{\mathrm{norm}} = \frac{B_i}{(n-1)(n-2)/2}$$

Normalised Betweenness Centrality



What is Al Pacino's normalised betweenness centrality?

What is Michelle Pfeiffer's normalised betweenness centrality?

$$B_i^{\text{norm}} = \frac{B_i}{(n-1)(n-2)/2}$$

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.

betweenness_centrality(*G*, *normalized*=*True*, *weight*=*None*, *endpoints*=*False*)

Normalized *bool, optional:* If True the betweenness values are normalized by 2/((n-1)(n-2)) for graphs, and 1/((n-1)(n-2)) for directed graphs where n is the number of nodes in G (**Source:** https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.betweenness_centrality.html)

Robustness

Robustness

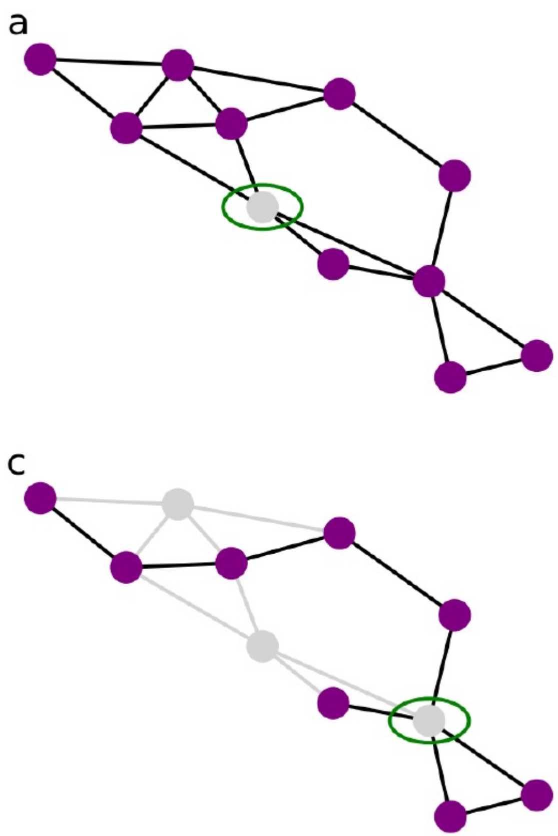
- A system is **robust** if the failure of some of its components does not affect its function
- **Question:** how can we define the robustness of a network?
- **Answer:** we remove nodes and/or links and see what happens to its structure
- **Key point:** *connectedness*

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

Robustness Test

- Checking how the connectedness of the network is affected as more and more nodes are removed
- Plot the relative size *S* of the largest connected component as a function of the fraction of removed nodes
- We suppose that the network is initially connected: there is only one component and $S = 1$
- As more and more nodes (and their links) are removed, the network is progressively broken up into components and *S* goes down

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



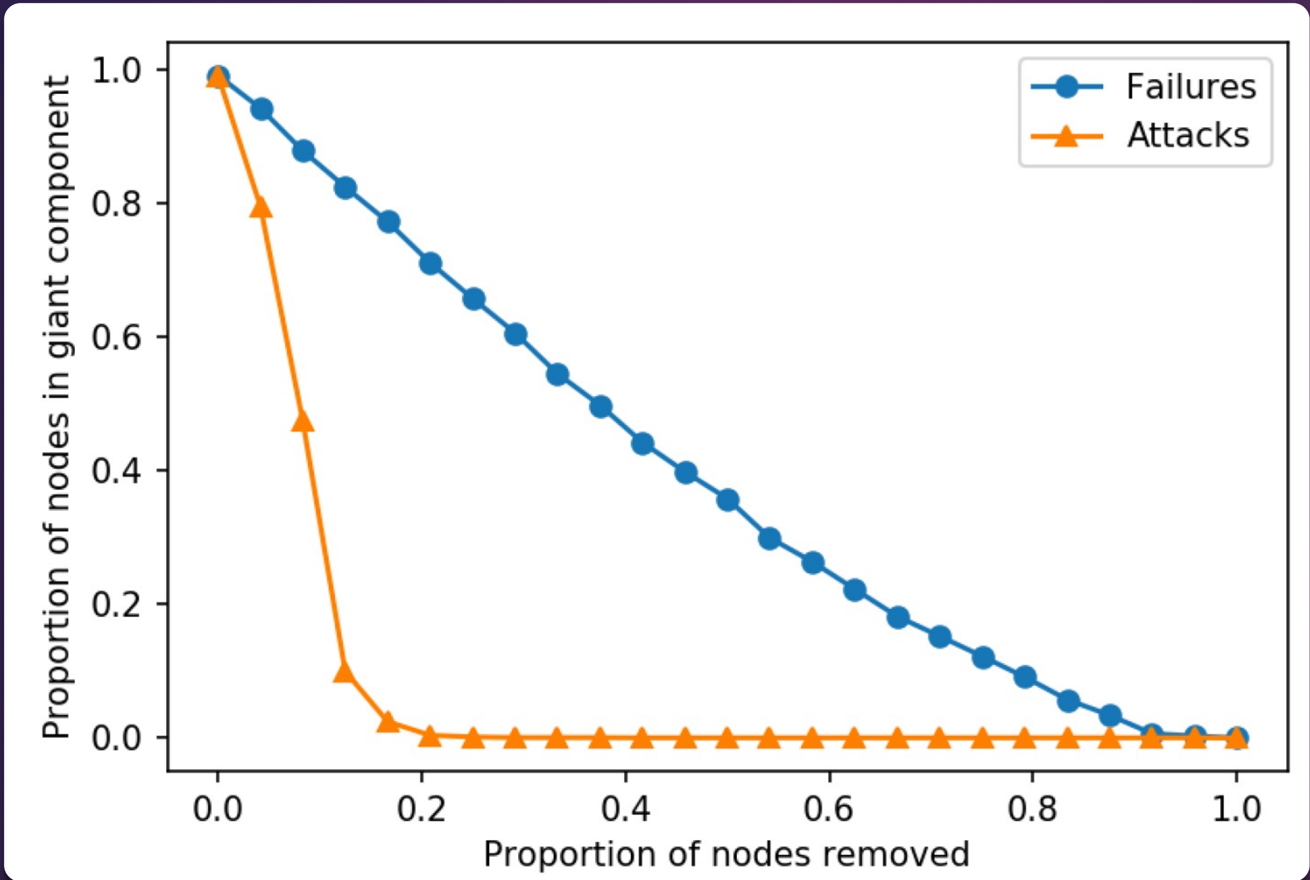
Source: Menczer, Fortunato, Davis, *A First Course in Network Science* (2023).

Robustness

Robustness

1. **Random failures:** nodes break down randomly, so they are all chosen with the **same probability**
2. **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).



Robustness

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