GET1030

Computers and the humanities

Lecture 7

Network analysis

Dr Miguel Escobar Varela

Objectives

- Define networks and review the history of network analysis
- Calculate basic network measurements by hand
- Review research that applies network analysis to the study of cultural objects

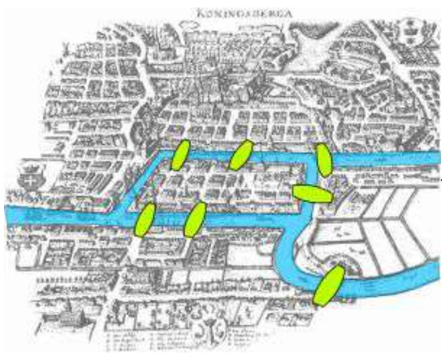
*Next lecture we will learn to calculate these measurements in Gephi and visualize them in Python.

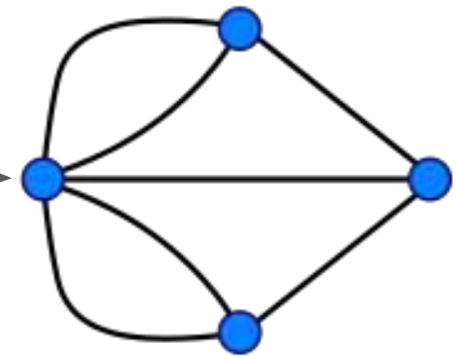
Lecture 6

Network analysis

Part 1: Network definitions and histories

The 7 bridges of Königsberg





https://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

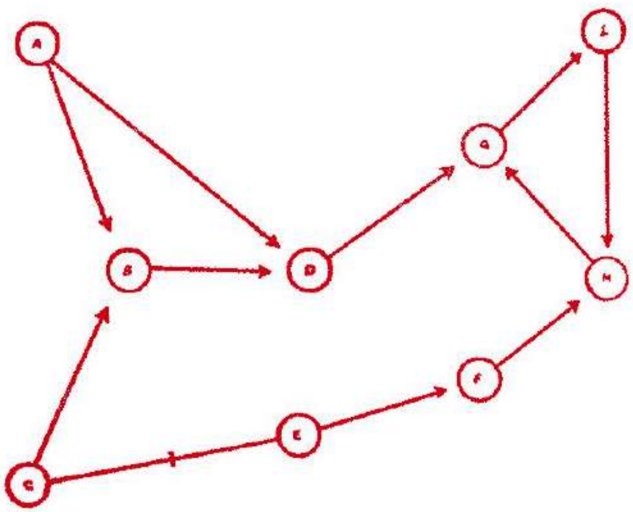
Can you cross each bridge once and only once?

Leonhard Euler (1736) proved mathematically that there is no solution to the problem.

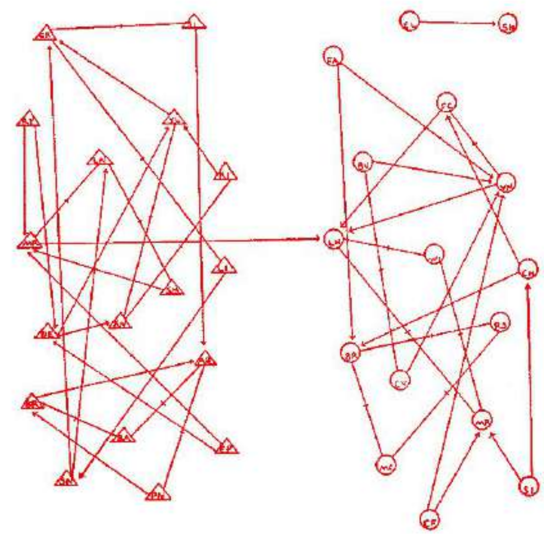
Leonhard Euler (1707-1783)

Social networks

- Jacob Levy Moreno



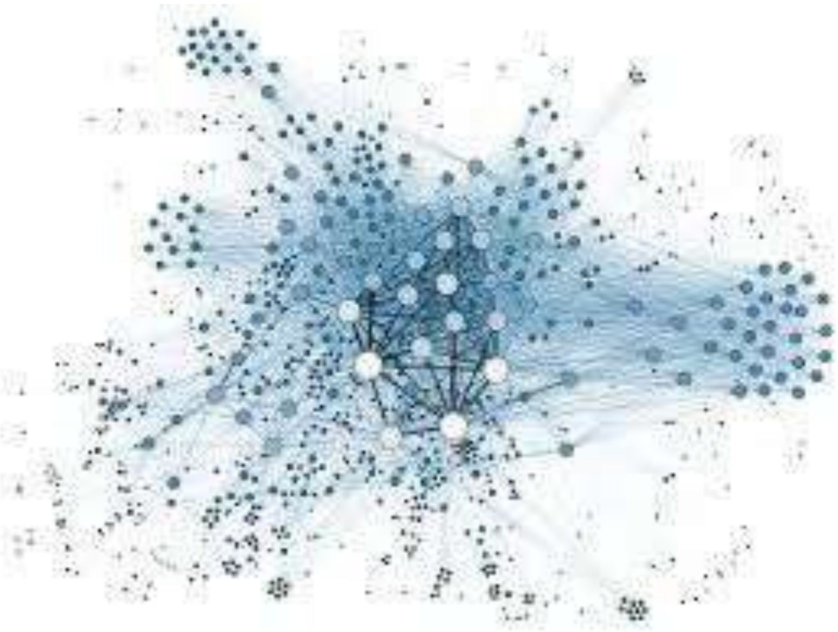
Moreno's Image of Who Recognized Whom Among a Collection of Babies (Moreno, 1934, p. 32).



Friendship Choices Among Fourth Graders (from Moreno, 1934, p. 38).

Current applications

- How things (ideas, pathogens) spread.
- What are the structural characteristics of a system or community.
- How systems or communities evolve over time.
- Applications in almost every field: sociology, biology, finance, history.

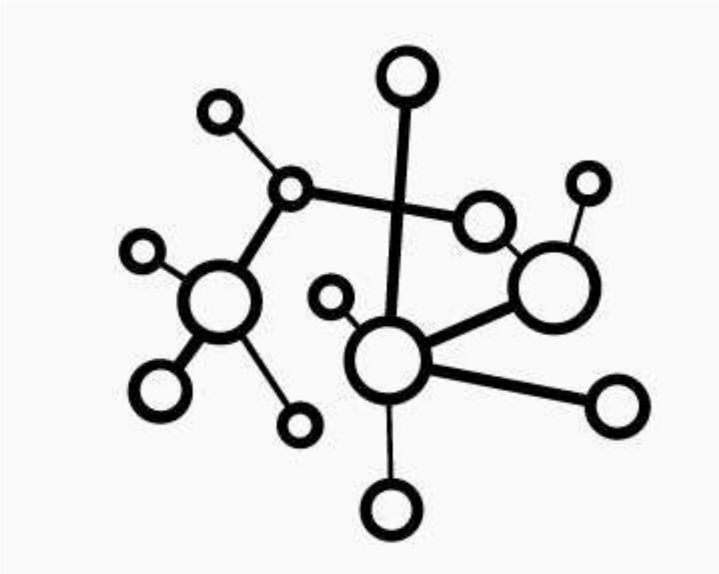


Definitions

A network consists of:

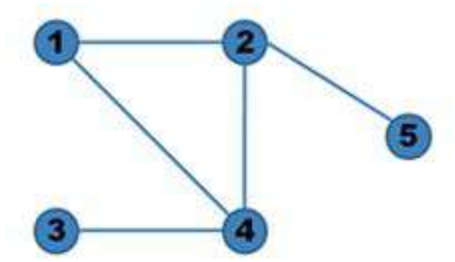
Nodes (things that are connected), often represented as circles.

Edges (connections between those things), often represented as lines joining the circles.

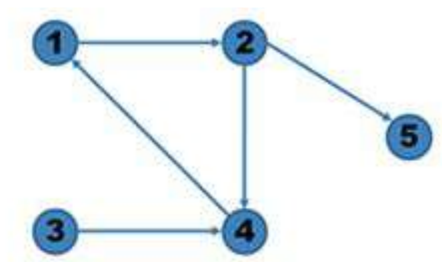


Directed or undirected

They can be directed or undirected.



In **undirected** networks, the edges are reciprocal. E.g Facebook friends.

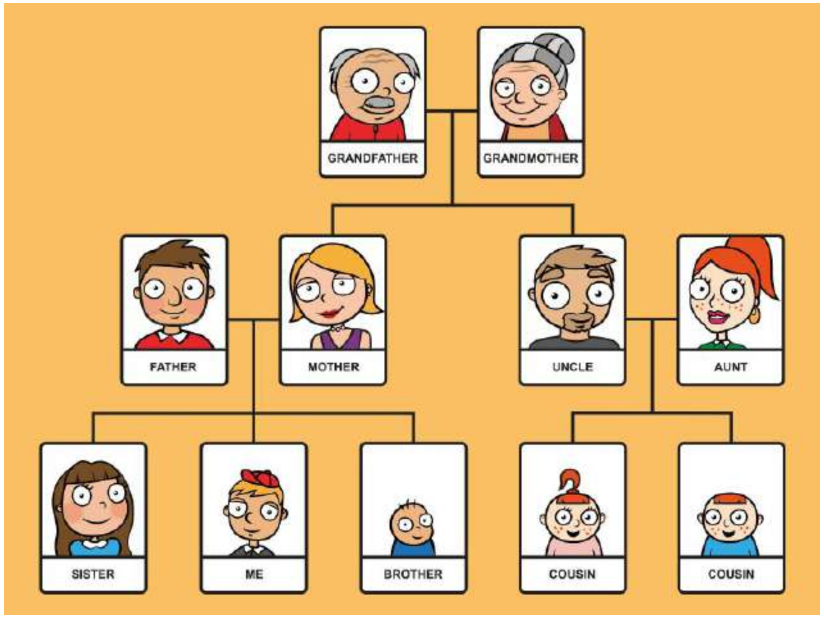


In **directed** networks, edges flow in one direction. E.g Twitter followers.

Edges

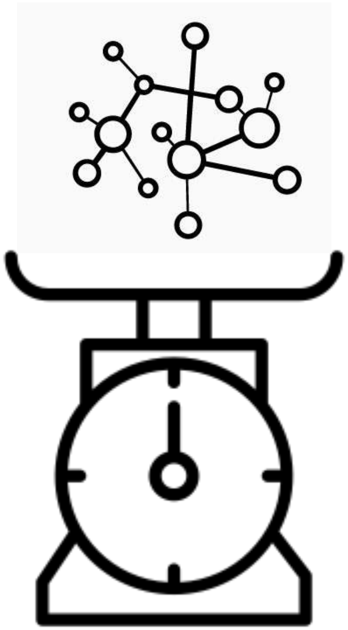
For the networks we will study, edges must be explicit, uniform connections between things.

That is, all edges must have the same meaning.



A family tree like this one is not a network, since some lines indicate marriage, and others indicate parentage.

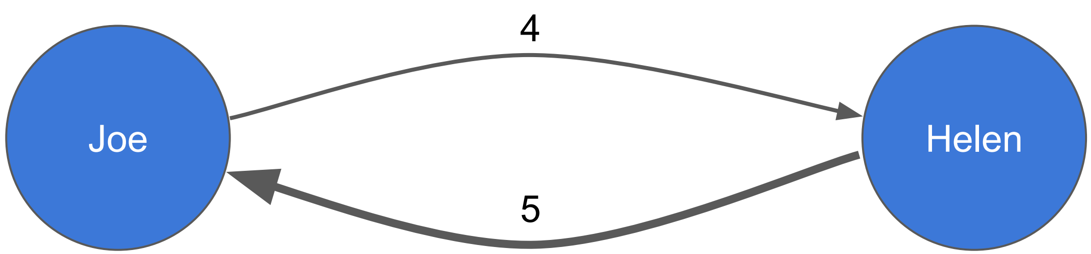
Weighted edges



Edges can be weighted. We can assign a numerical value to the connection between two nodes. This applies to both directed and undirected networks.

Weighted Directed Edges

Example: A network of direct messages on FaceBook.
Nodes: users
Edges: messages

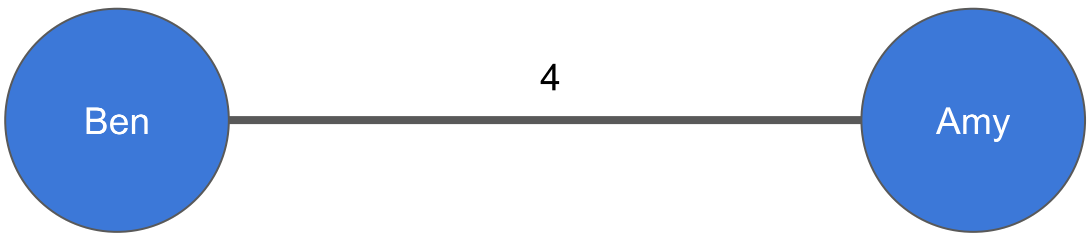


An edge from Joe to Helen means that he has messaged her at least once. Every time he messages her again, we can increase the weight of the edge, rather than add a new edge. A weight of 4 means that Joe has sent Helen 4 messages.

We can also model an edge in the other direction, between the same nodes. An edge with weight 5 from Helen to Joe, means she messaged him 5 times.

Weighted Undirected Edges

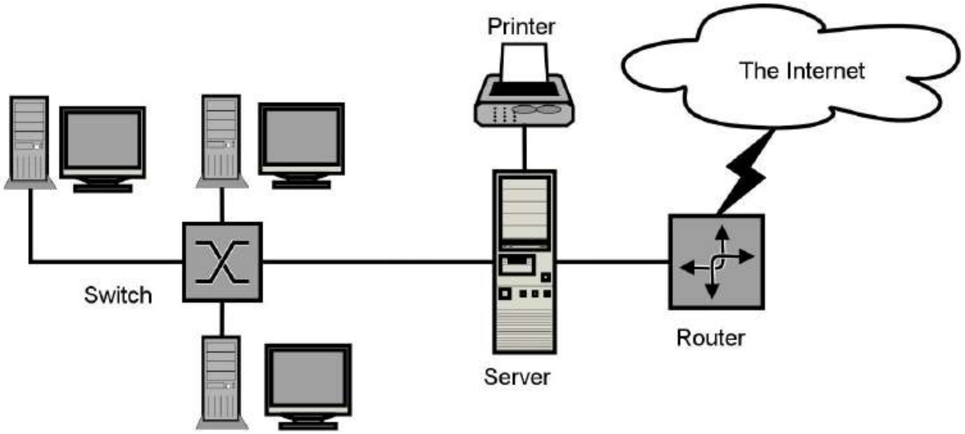
Example: a network of collaborations between artists.
Nodes: the artists
Edges: collaborations



An edge between Ben and Amy means they have collaborated at least once. With each new collaboration, the weight increases by 1. So a weight of 4 between Ben and Amy means that they have collaborated 4 times.

Nodes

For the networks we will study, nodes must all be the same kind of thing.



A computer “network” like this one falls outside our definition of networks.

Here you have different types of things: printers, servers, routers.

Nodes

We can add classifications to the nodes.

For example, we can classify a network of collaborators by nationality. But all nodes must be people and all edges must indicate collaboration.

This classification might help you see how different measurements vary across different types of nodes.

(We will see the usefulness of this next session).

Lecture 6

Network analysis

Part 2: Basic network measurements

Two kinds of questions

About the network as a whole:
How connected is it? ⇢ *density*
How far away is every node from other nodes? ⇢ *average path length*

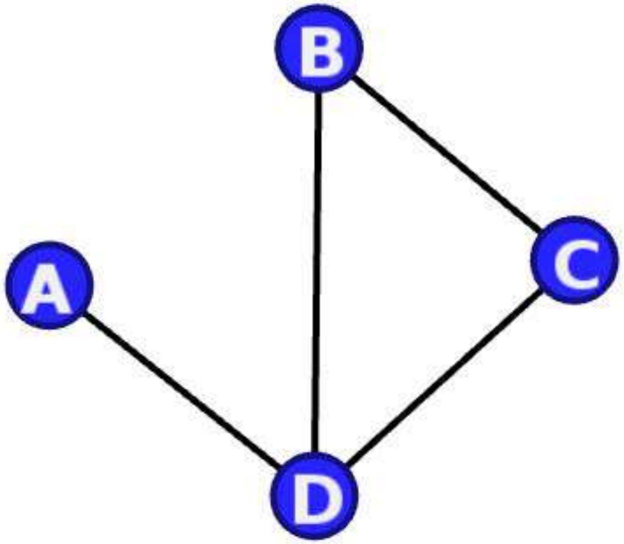
About a node:
How far is a node from the farthest node? ⇢ *eccentricity*
How many connections does it have? ⇢ *degree*
How important is a node to the connections in the network? ⇢ *betweenness*
How close is a network to other nodes on average? ⇢ *closeness*

NOTE: For the next examples we are only going to consider networks that are unweighted, undirected and complete.

GET1030

Density

Density. The fraction of potential connections present in a network. Let *T* be the total number of nodes.



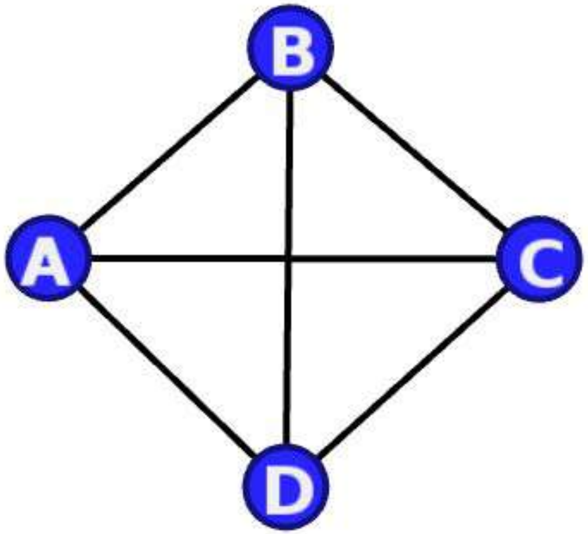
$$potential\ edges = \frac{T \cdot (T-1)}{2}$$

$$density = \frac{actual\ edges}{potential\ edges}$$

Potential edges = 4 * 3 / 2 = 6
Actual edges = 4
Density = 4/6 = 0.66

Density

Density. The fraction of potential connections present in a network. Let *T* be the total number of nodes.



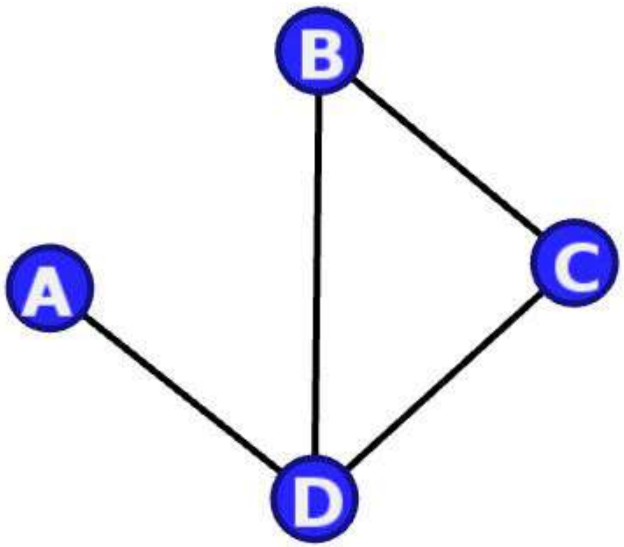
$$potential\ edges = \frac{T \cdot (T-1)}{2}$$

$$density = \frac{actual\ edges}{potential\ edges}$$

Potential edges = 4 * 3 / 2 = 6
Actual edges = 6
Density = 6/6 = 1

Average Path Length

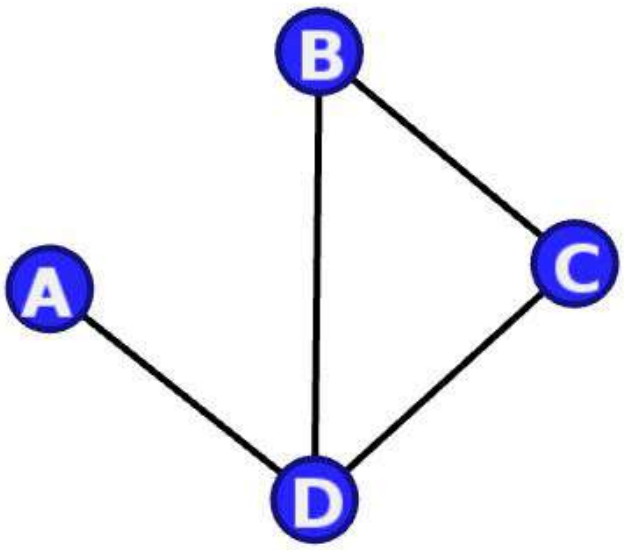
Average Path Length. Average number of steps along the shortest paths for all possible pairs of network nodes.



Node pairs	Shortest path length
A, B	2
A, C	1
A, D	1
B, C	2
B, D	1
C, D	1
Average	**1.33**

Diameter and radius

The largest *shortest path* is the diameter, and the smallest *shortest path* is the radius.



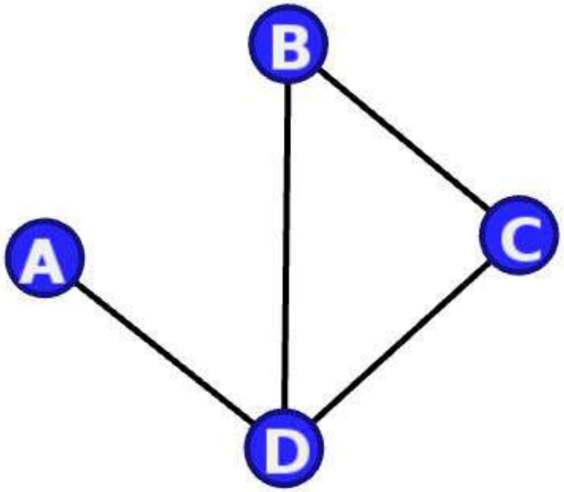
Node pairs	Shortest path length
A, B	2
A, C	1
A, D	1
B, C	2
B, D	1
C, D	1
Average	**1.33**

The diameter of this graph is 2 and the radius is 1.

Eccentricity

The eccentricity of a node *n* in a connected network is the maximum distance between *n* and any other node.

Consider the example for node A



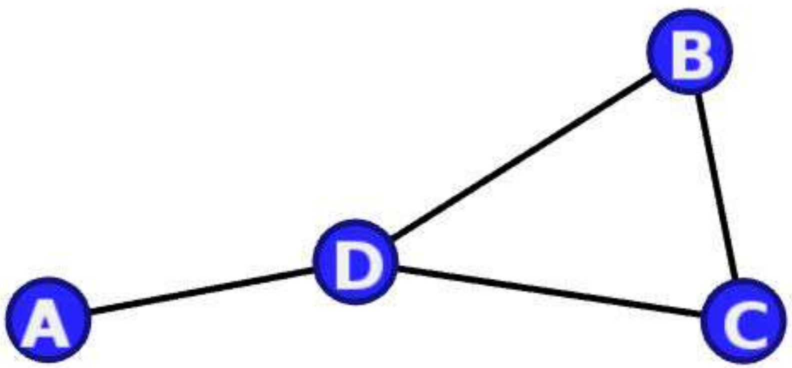
To node:	Distance
B	2
C	1
D	1

The maximum is 2, so the eccentricity of A is 2.

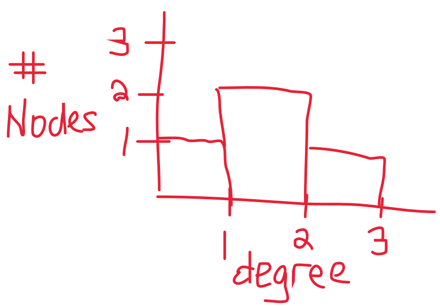
The maximum eccentricity is the graph diameter. The minimum graph eccentricity is called the graph radius.

Degree

Degree. Total number of edges a node has to other nodes.



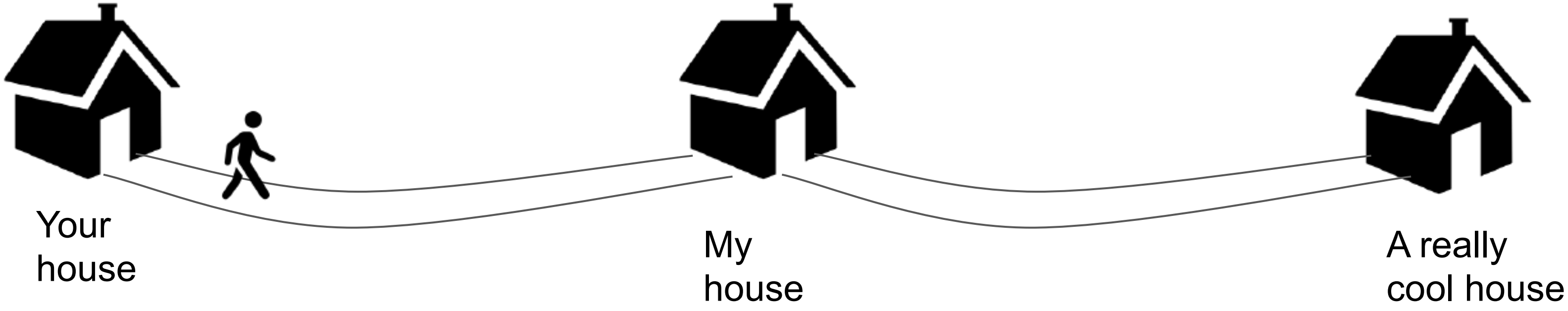
Node	Degree
A	1
B	2
C	2
D	3



Betweenness

How important is a node to the shortest paths in the network?

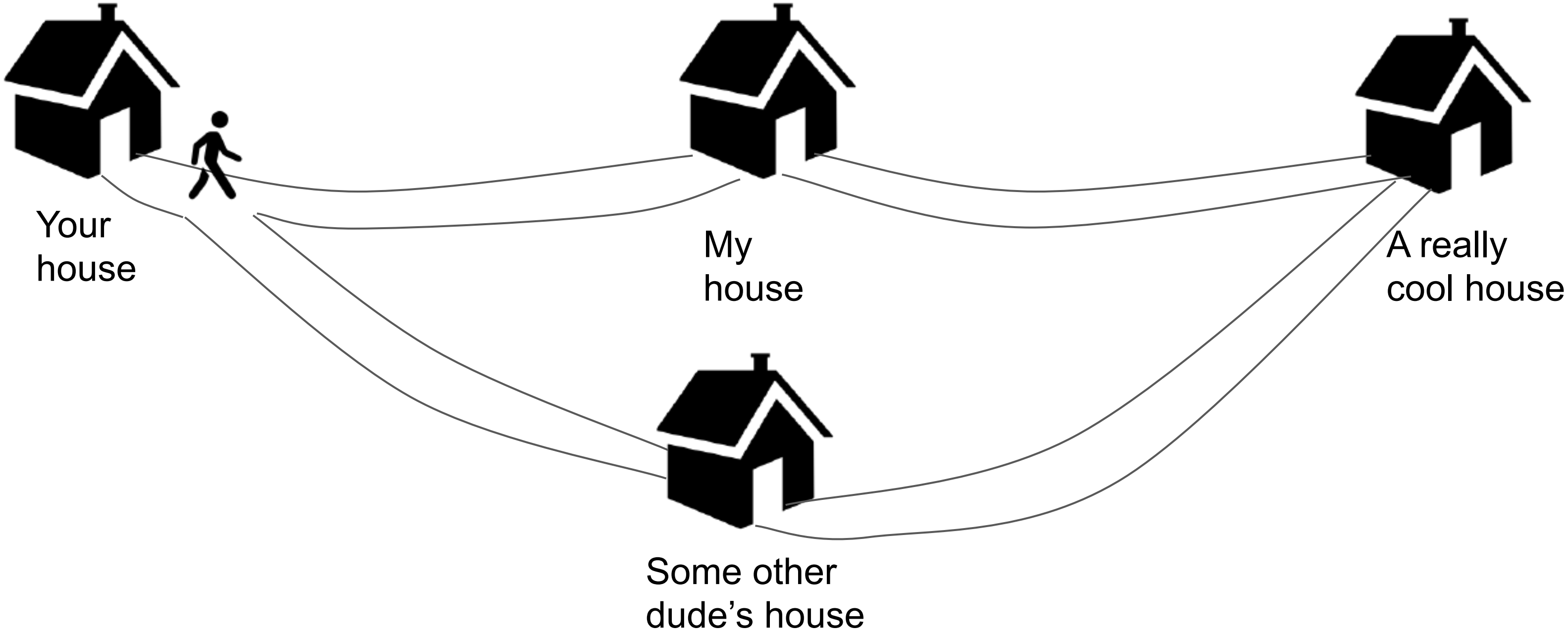
In the case below my house is very important to the shortest paths and should have a high betweenness.



Betweenness

How important is a note to the shortest paths in the network?

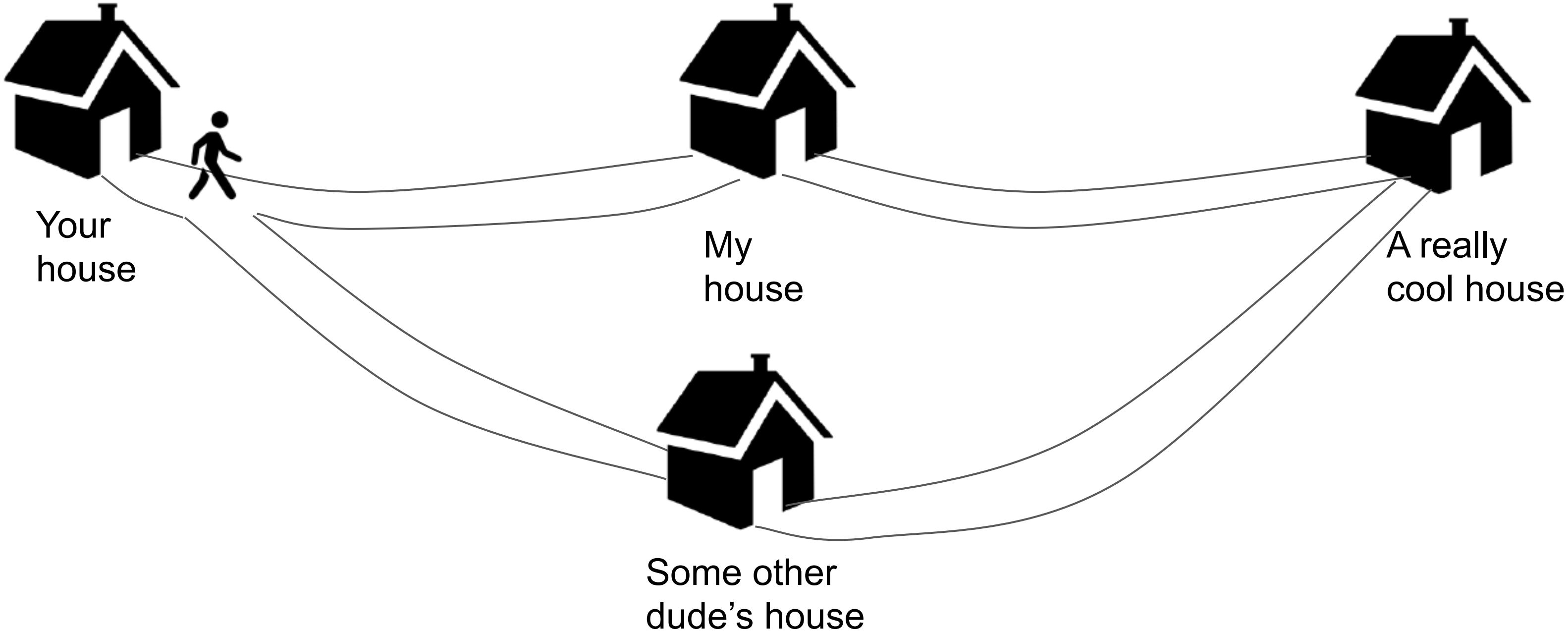
Now you've got two options, so my house is less important, it should have a lower betweenness.



Betweenness

And how important is my house to the other dude's house? Well, not at all. He has a direct route to the cool house and to your house.

The betweenness of my house is the fraction of all shortest paths pass through my house.

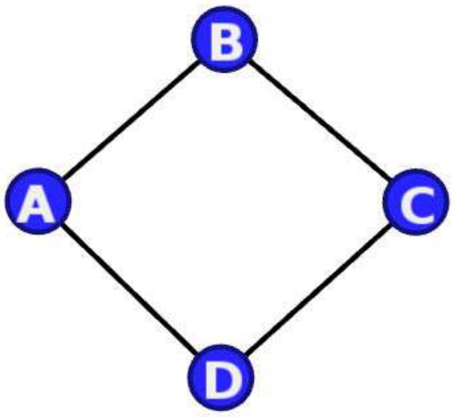


Betweenness

Betweenness: how important a node is to the shortest paths through the network.

The sum of the fractions of all shortest paths that pass through a node *n.*

To compute betweenness for a node *n*:
1. Select a pair of nodes *x,y*
2. Find all the shortest paths between x and y.
3. Compute the fraction of those paths that pass through *n*.
4. Repeat for every other pair of nodes *x,y*



For node A.

Pair B,D has 2 shortest paths

Shortest path	Passes through A?
B ↦ A ↦ D	Yes
B ↦ C ↦ D	No

Fraction that passes through A: 0.5

Pair B,C has 1 shortest path
Fraction that passes through A: 0

Pair C,D has 1 shortest path
Fraction that passes through A:0

The betweenness of node A is 0.5.

Normalized Betweenness

The problem of this is that bigger networks tend to get bigger numbers.

To *normalize* the value [that is, to get a number between 0 and 1], betweenness can be divided by the maximum number of shortest paths, which is

$$max\ shortest\ paths = \frac{(T-1)\cdot(T-2)}{2}$$

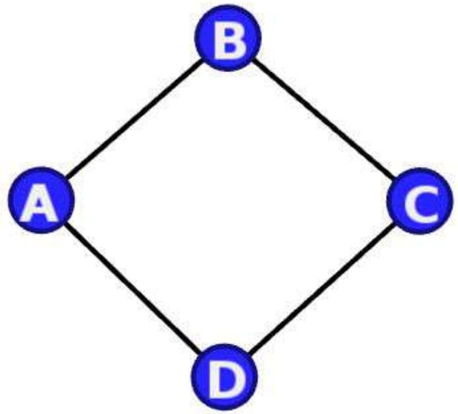
Where *T* is the total number of nodes in the network.

*This is the same as the *density* for the subgraph that excludes *n.*

The raw betweenness of node A is 0.5.

Max shortest paths = (3x2)/2 = 3

Normalized betweenness = 0.5 / 3 = 0.166



Closeness

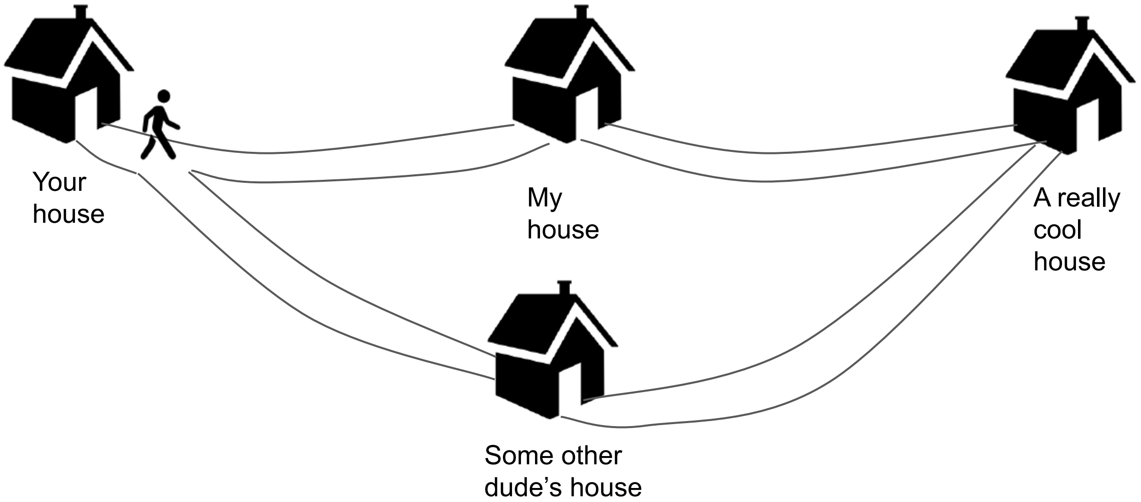
How close is your house to other houses? (consider only the shortest paths).

- To the cool house - 2 steps
- To my house - 1 steps
- To other dudes - 1 step

So on average, it takes you 1.33 steps to go anywhere. But this is an indication of *farness*. To know how close you are, you can take the inverse of the farness 1/1.33 = 0.75

This makes sense because the maximum theoretical closeness is when it would take you 1 step to go anywhere. So 1/1=1 is the maximum theoretical closeness.

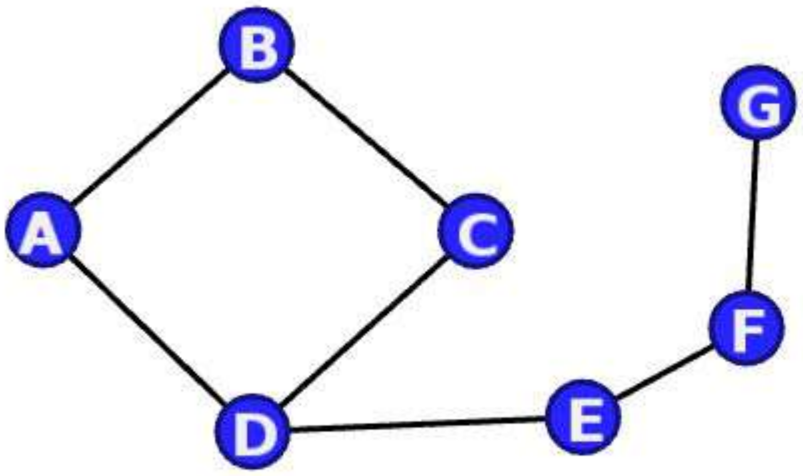
*Note that taking the inverse is the same as dividing the total number of houses (excluding yours) by the sum of all shortest paths, or 3/4= 0.75.



Closeness

Closeness centrality indicates how close a node *n* is to all other nodes in the network.

Let's consider the shortest paths from A.



To node:	Distance:
B	1
C	2
D	1
E	2
F	3
G	4
Sum	13
(Number destination nodes / sum)	6/13 = 0.46

The closeness centrality of A is 0.46

Mathematical notation

Density

$$d = \frac{2E}{T \cdot (T-1)}$$

Where *d* is density, *E* is the total number of edges, and *T* is the total number of nodes.

Average path length

$$l = \frac{2}{T \cdot (T-1)} \cdot \sum_{i \neq j} d(n_i, n_j)$$

Where *l* is average path length and *T* is the total number of nodes. Let $d(n_i, n_j)$ denote the shortest distance between n_i and , n_j

Betweenness centrality

$$b(n) = \sum_{i \neq n \neq j} \frac{\sigma_{ij}(n)}{\sigma_{ij}}$$

Where *b(n)* is the betweenness centrality of node *n*, $\boldsymbol{\sigma}_{ij}$ is the total number of shortest paths from node *i* to node *j* and $\boldsymbol{\sigma}_{ij}$ is the number of those paths that pass through *n.*

Normalized betweenness centrality

$$nb(n) = \frac{2}{(T-1) \cdot (T-2)} \cdot \sum_{i \neq n \neq j} \frac{\sigma_{ij}(n)}{\sigma_{ij}}$$

Where *nb(n)* is the normalized betweenness centrality of node n, *T* is the total number of nodes.

Closeness centrality

$$c(n) = \frac{N-1}{\sum_i d(i,n)}$$

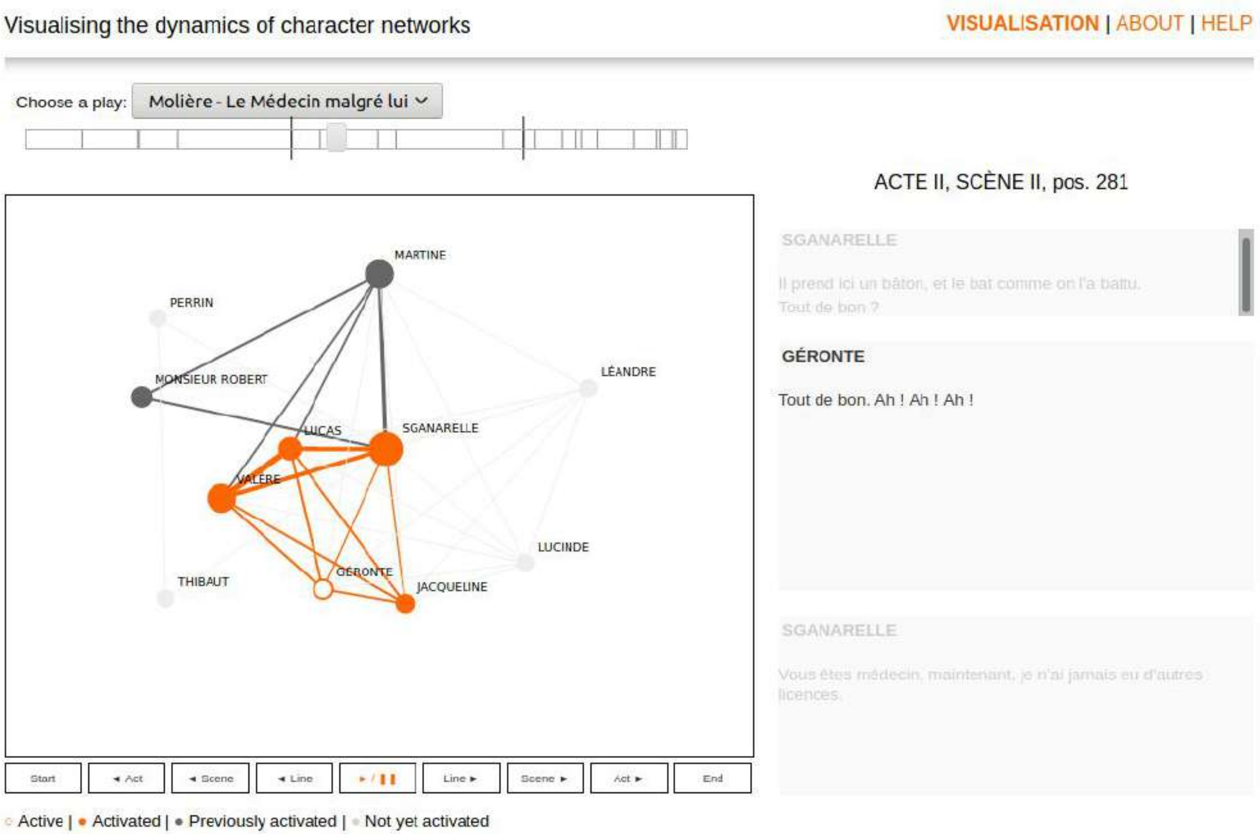
Where *c(n)* is the closeness centrality of node *n* and *d(i,n)* is the distance between nodes *i* and *n,* for every node *i* in the network.

Lecture 6

Network analysis

Part 3: Network analysis for art and literature

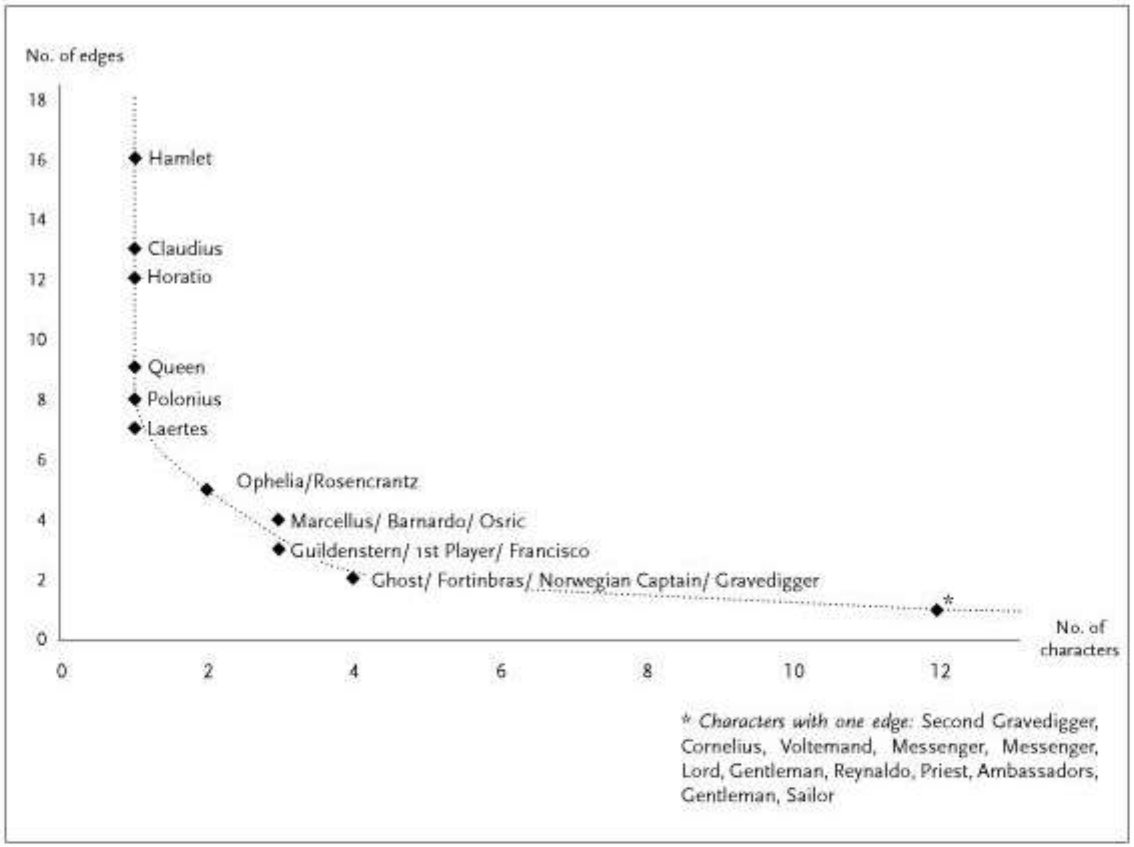
Molière’s plays as networks



https://maladesimaginaires.github.io/intnetviz/?author=moliere&play=le_medecin_malgre_lui

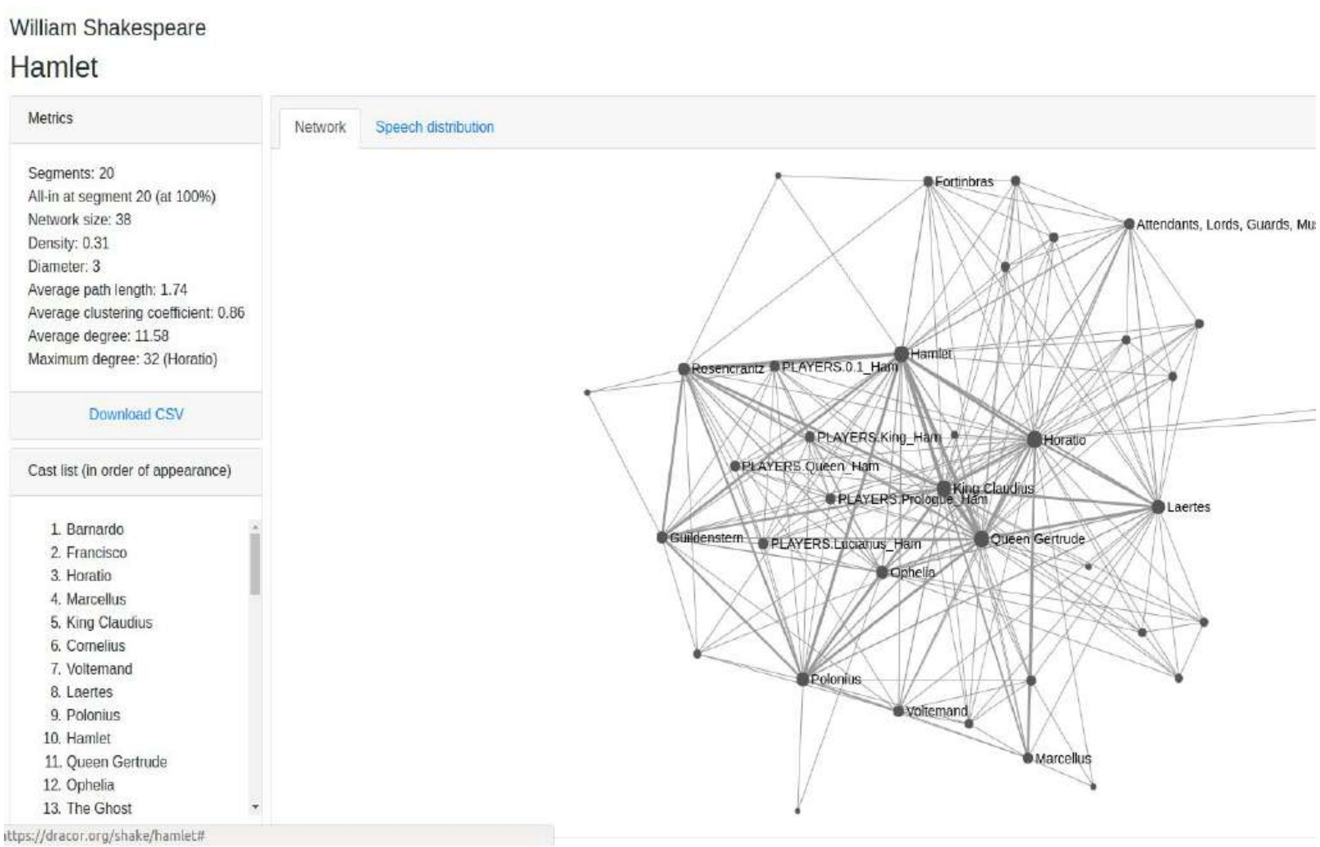
Moretti on Hamlet

Degree for characters in *Hamlet*



Moretti (2011)

Shakespeare networks



https://dracor.org/shake

German drama

Table 1: Network Measures, by Genre

	N=	Number of Characters (Median)	Max Degree (Median)	Average Degree (Average)	Density (Average)	Average Path Length (Average)
Corpus	465	16	13	9,01	0,59	1,46
Tragedy	101	19	16	9,57	0,52	1,56
Comedy	92	14	11	8,61	0,67	1,36
Libretto	56	16	13,5	9,09	0,64	1,39
Other	216	17	14	8,88	0,59	1,48

https://dlina.github.io/Network-Values-by-Genre/

German drama

Average path length (average)

Fischer et al (2015) https://dlina.github.io/200-Years-of-Literary-Network-Data/

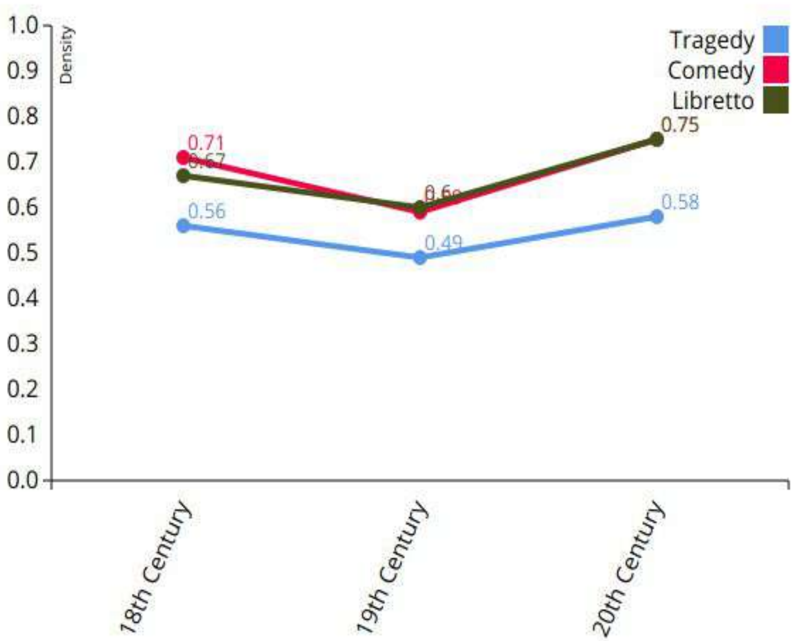
German drama

Max degree (median)

Fischer et al (2015) https://dlina.github.io/200-Years-of-Literary-Network-Data/

German drama

Fig. 5: Density (Mean), by Genre and Century



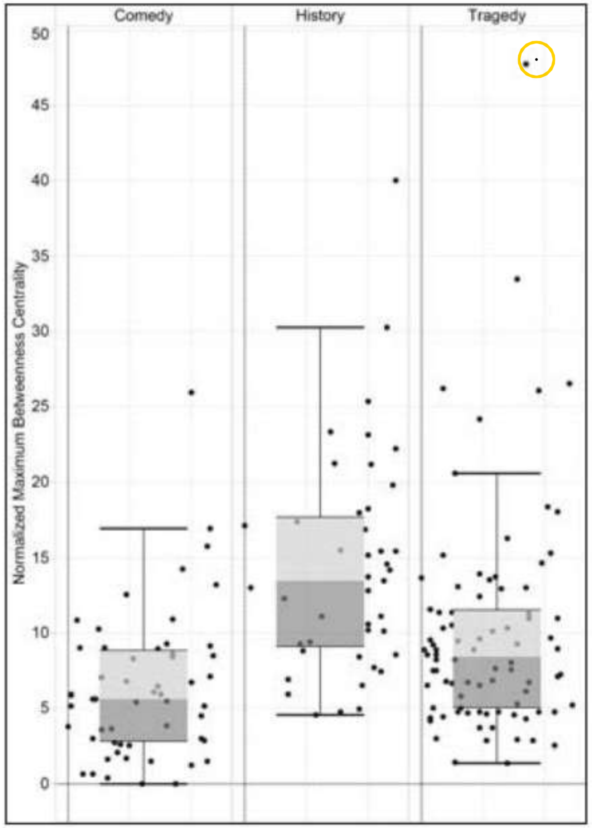
Fischer et al (2015), https://dlina.github.io/Network-Values-by-Genre/

Algee-Hewitt's work on dramatic networks

- Network properties of 3,568 English dramatic texts written 1550 and 1900 taken from ProQuest Literature Online drama corpus.
- Most probable recipient for each speech using a rule-based system (speech as directed edges).
- Betweenness centrality can be used to estimate plays where character are mediators between other characters (e.g., political intrigue).
- The graph shows the maximum *normalized betweenness centrality* for a range of plays, classified by type.

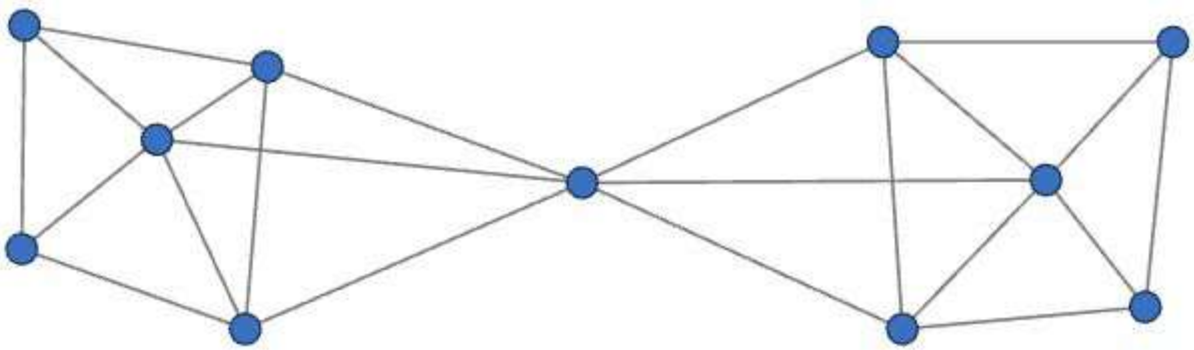
This identifies "go-betweens"
If the value is low: no go-betweens among groups.
If the value is high: at least one go-between among groups.

This is not the same as the total amount of interactions. If one character was in all interactions, then his/her eccentricity would be 1.

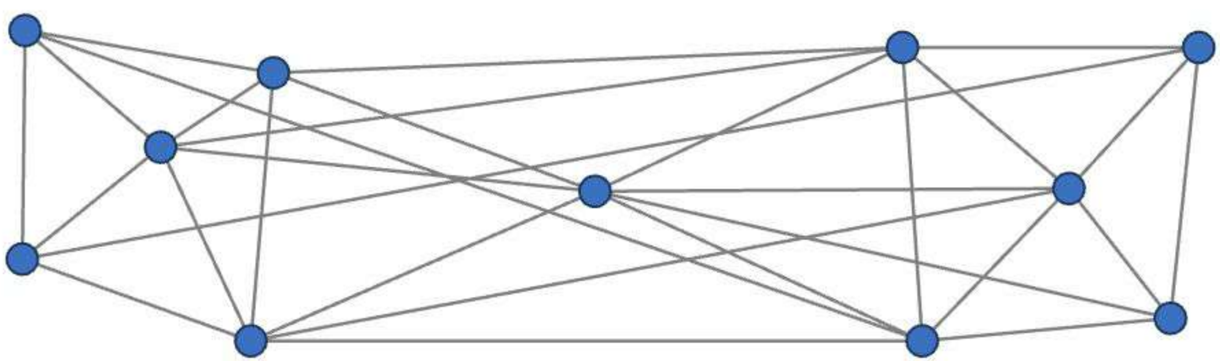


Which one would have maximum betweenness?

A)

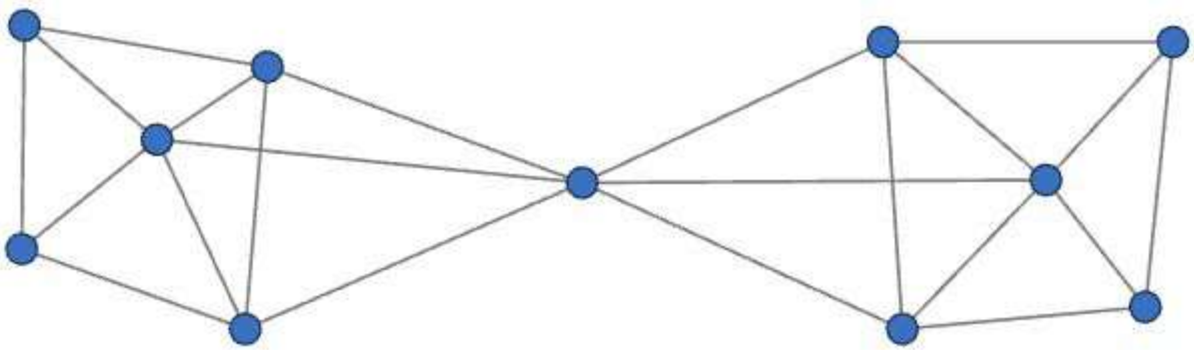


B)



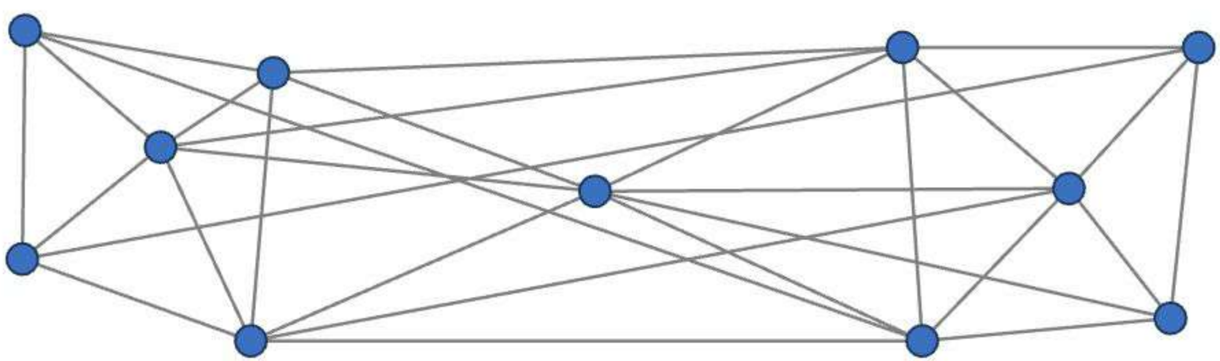
Which one would have maximum betweenness?

A)



Max ***b***	25
Max ***nb***	0.55

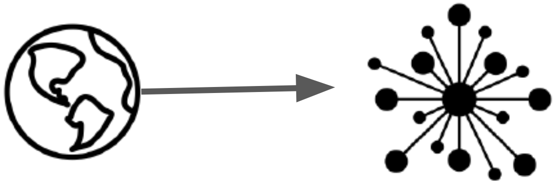
B)



Max ***b***	4.26
Max ***nb***	0.094

Critical perspective

What does it mean to model something as a network?



We saw many examples of theatre plays, where characters are modelled as nodes.

But what should the edges be?

- Co-presence (characters are in the same scene)
- Interactions (a character speaks to another one)

Both options make sense, but they are not neutral. They represent an opinion on what matters.
In you projects, pay attention to the assumptions and make sure you explicitly state them.

References

Algee-Hewitt, Mark. “Distributed Character: Quantitative Models of the English Stage, 1550–1900.” *New Literary History* 48, no. 4 (2017): 751–82. https://doi.org/10.1353/nlh.2017.0038.

Moretti, Franco. “Network Theory, Plot Analysis.” *New Left Review*, no. 68 (March 1, 2011): 80.

Trilcke, Peer, and Frank Fischer. “Dracor.Org.” Accessed September 5, 2018. https://dracor.org/.

Trilcke, Peer, Frank Fischer, Mathias Göbel, and Dario Kampkaspar. “Comedy vs. Tragedy: Network Values by Genre.” *DLINA: Digital Linterary Analysis* (blog), 2015. https://dlina.github.io/Network-Values-by-Genre/.

Trilcke, Peer, Frank Fischer, and Dario Kampkaspar. “Digital Network Analysis of Dramatic Texts.” In *Digital Humanities 2015 Conference Abstracts*. Sydney, Australia, 2015.

Xanthos, Aris, Isaac Pante, Yannick Rochat, and Martin Grandjean. “Visualising the Dynamics of Character Networks.” In *Digital Humanities 2016 Conference Abstracts*. Krakow, Poland, 2016.

Błoch, Agata, Demival Vasques Filho, and Michał Bojanowski. 2020. “Networks from Archives: Reconstructing Networks of Official Correspondence in the Early Modern Portuguese Empire.” *Social Networks*, September. https://doi.org/10.1016/j.socnet.2020.08.008.

Park, Doheum, Arram Bae, Maximilian Schich, and Juyong Park. 2015. “Topology and Evolution of the Network of Western Classical Music Composers.” *EPJ Data Science* 4 (1): 2. https://doi.org/10.1140/epjds/s13688-015-0039-z.