Tutorial

AN INTRODUCTION TO NETWORK SCIENCE FOR THE DIGITAL HUMANITIES

Shakti N. Menon

The Institute of Mathematical Sciences, Chennai

Mar 27th 2024

Bits & Strings 2024:
Workshop on Computational Epigraphy

Social Media Network

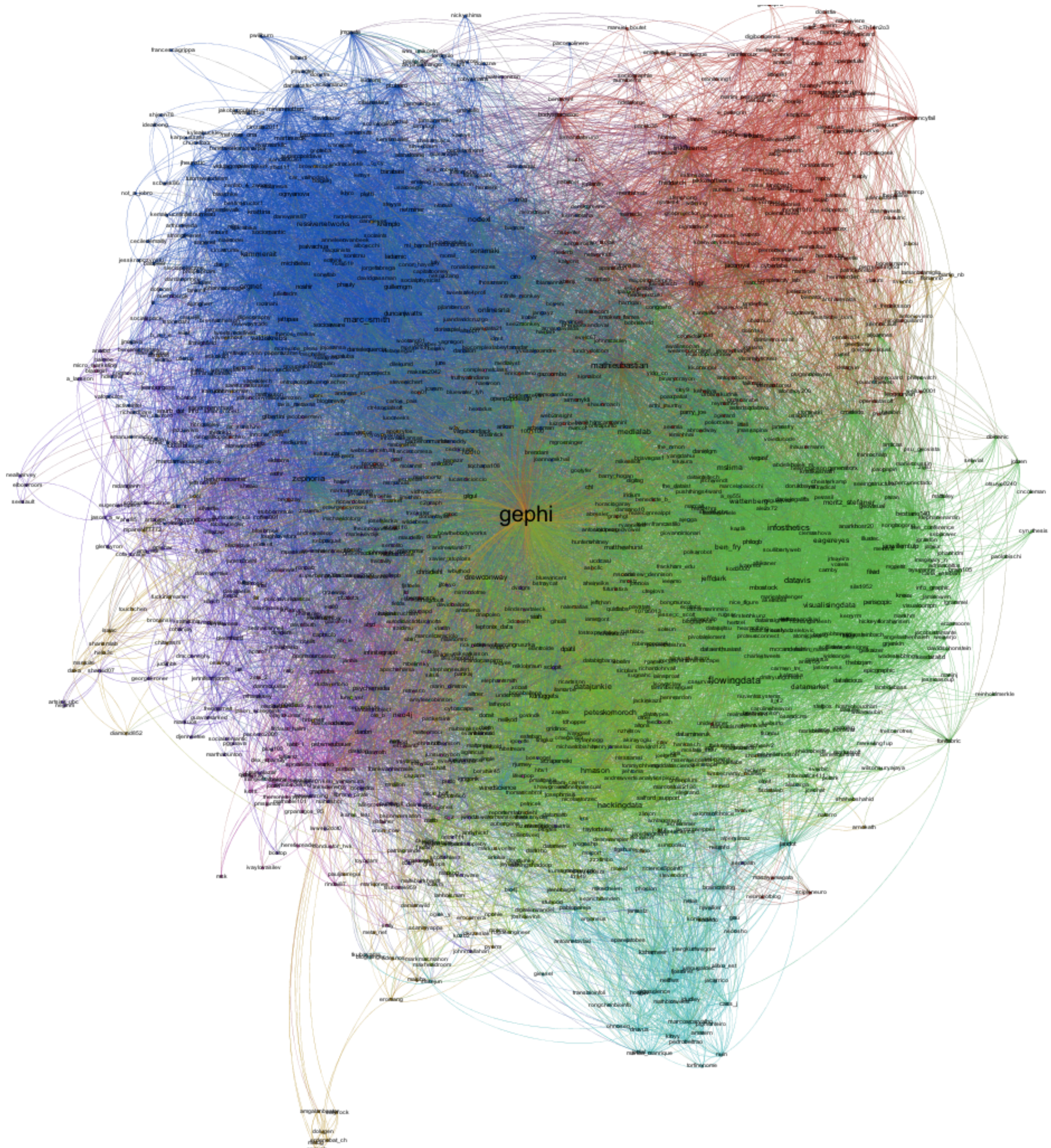


image: http://social-dynamics.org/a-gephi-visualization-of-gephi-on-twitter/

Similarity Network of Musicians

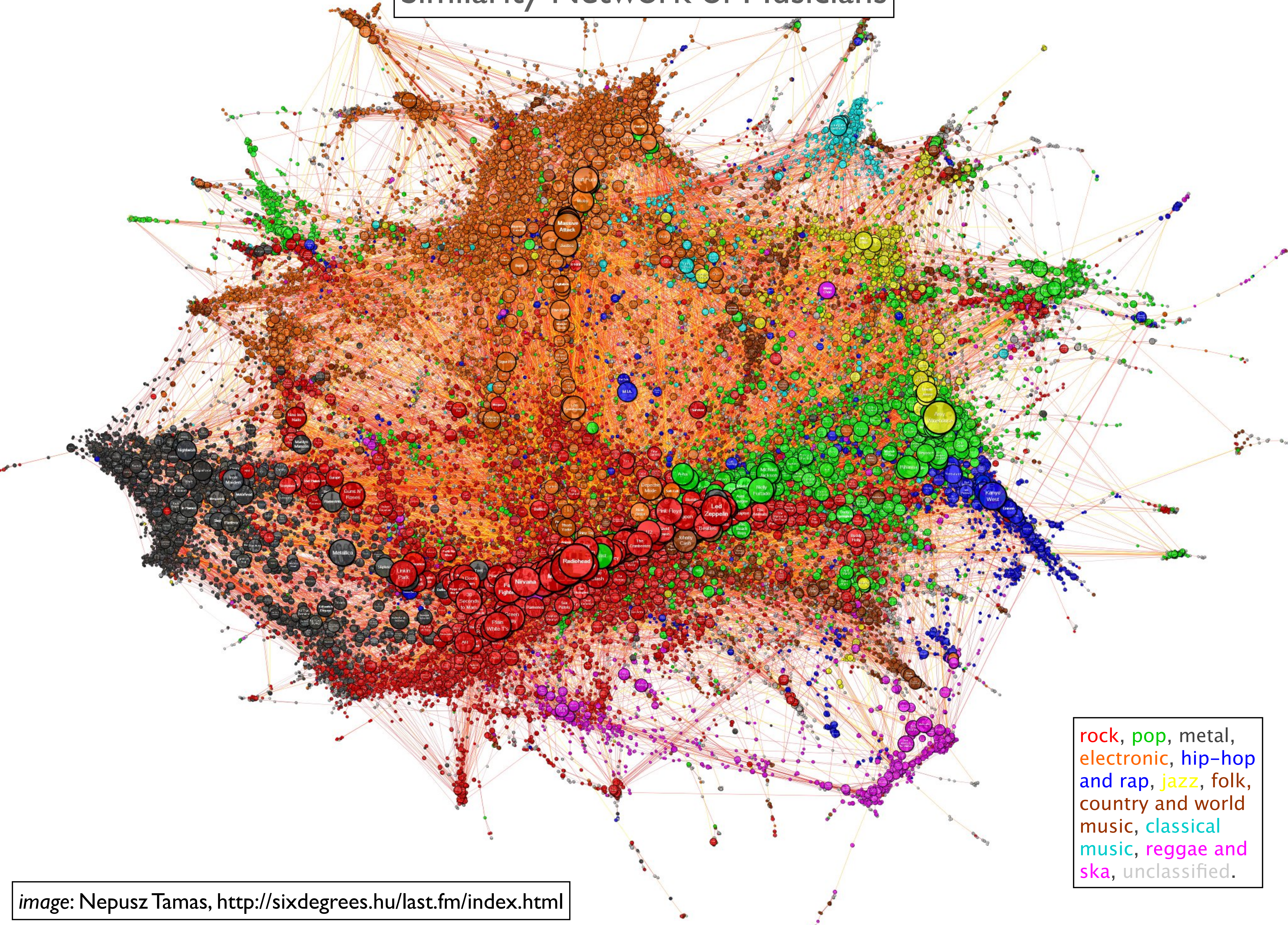


image: Nepusz Tamas, http://sixdegrees.hu/last.fm/index.html

The Political Blogosphere and the 2004 U.S. Election

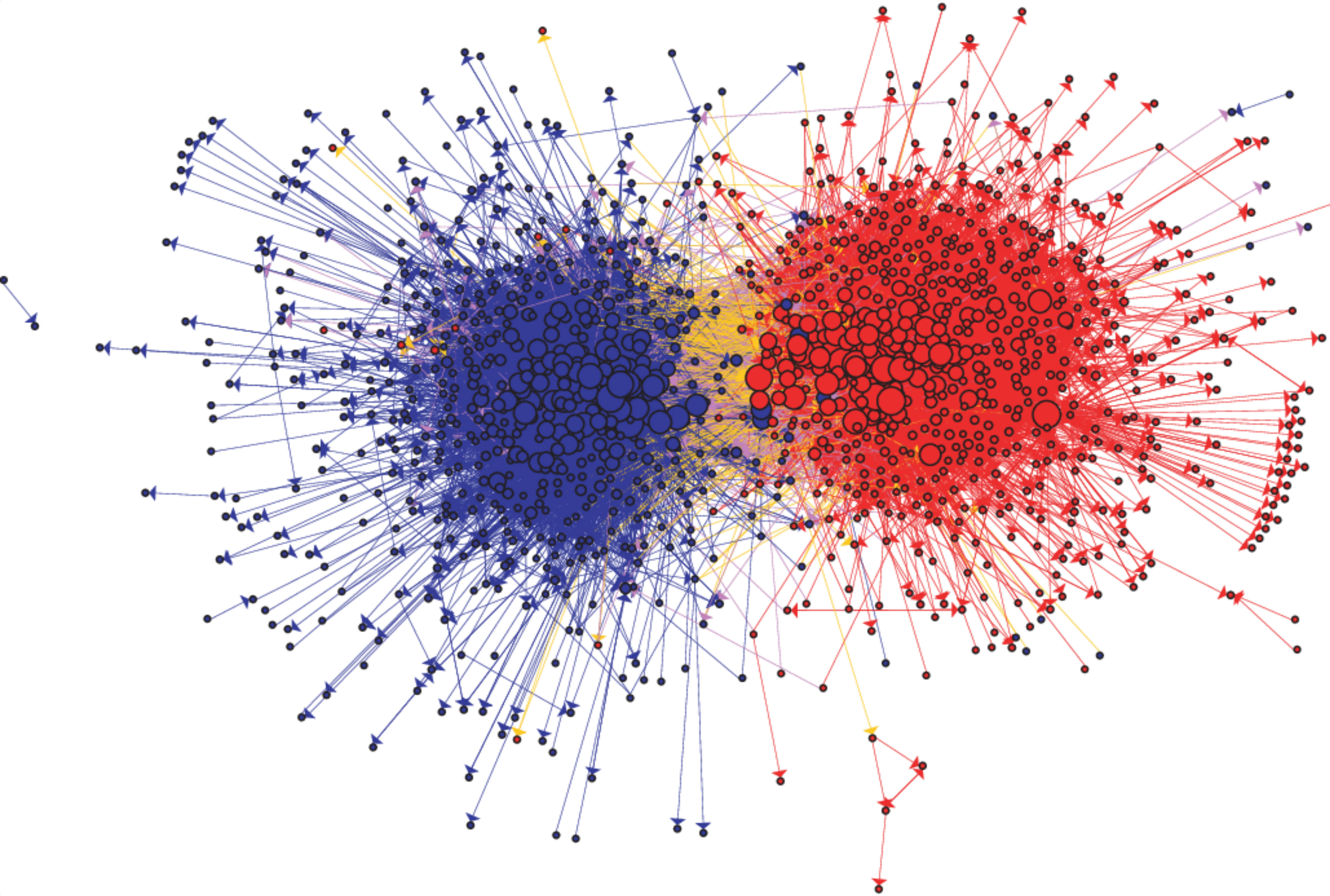


image: Lada Adamic & Natalie Glance, Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem (2005).

Social network in ancient Athens

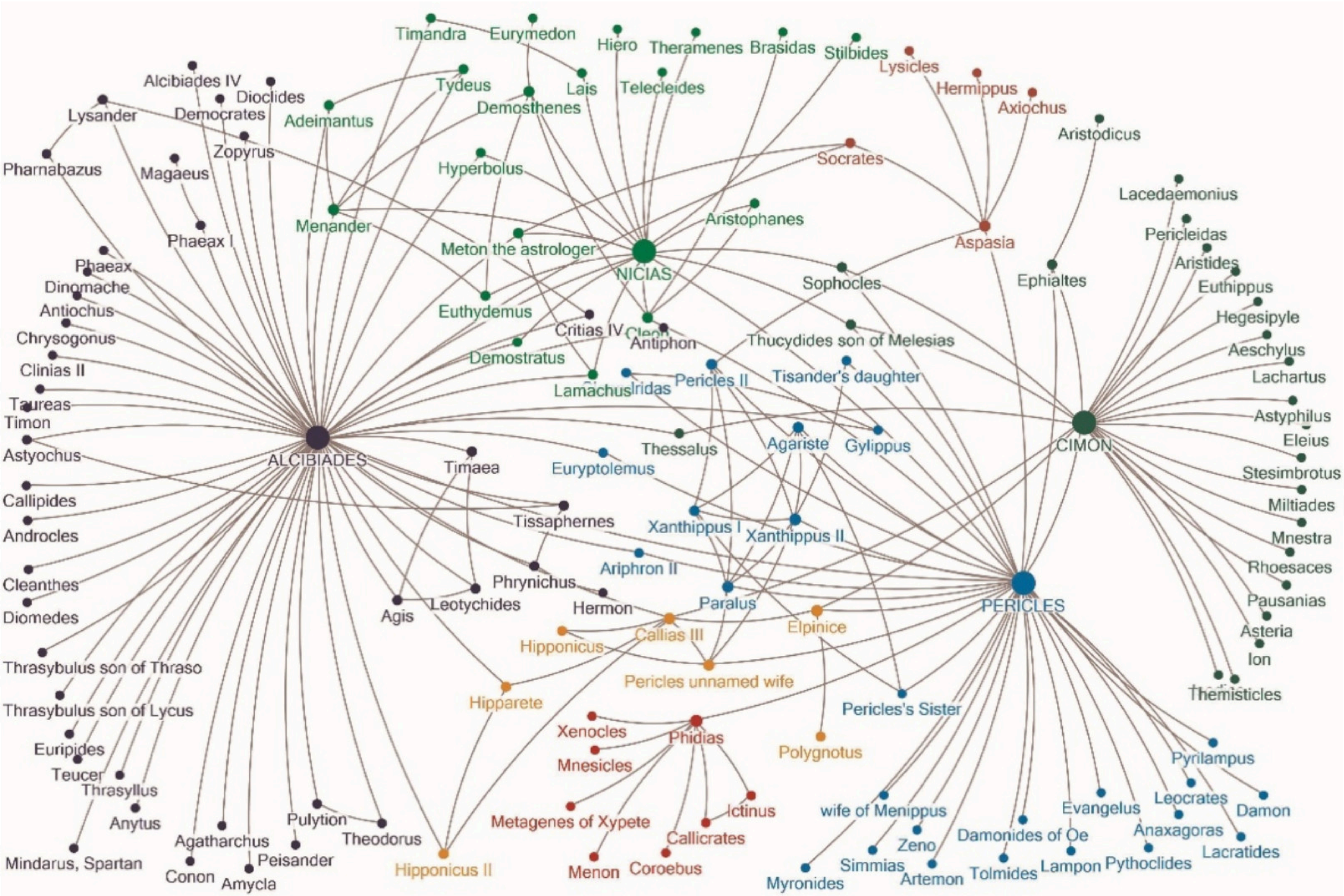


image: Cline, DH, "Athens as a Small World", Journal of Historical Network Research 4, 36-56 (2020).

Family tree of Louis III, Duke of Württemberg (1585)

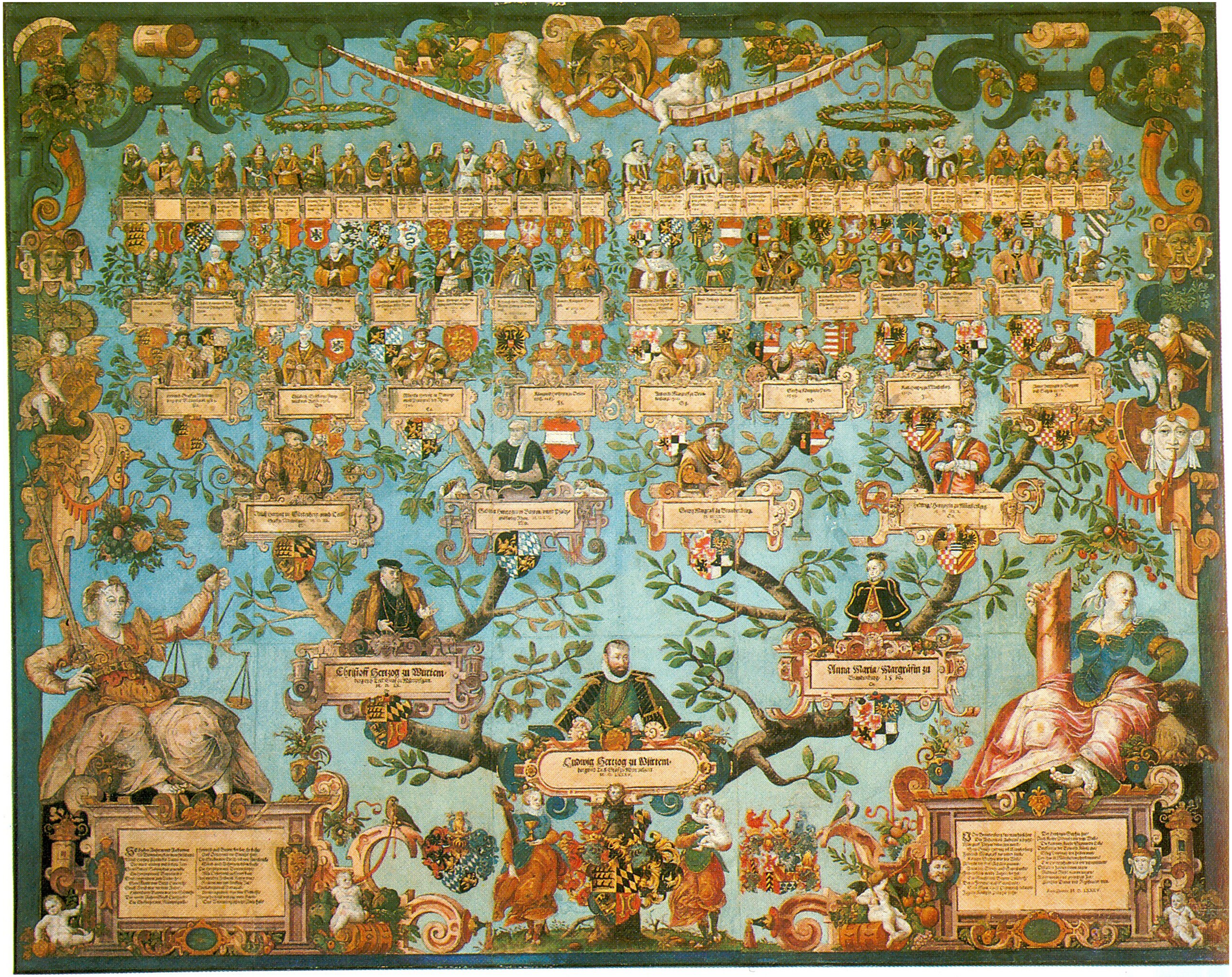
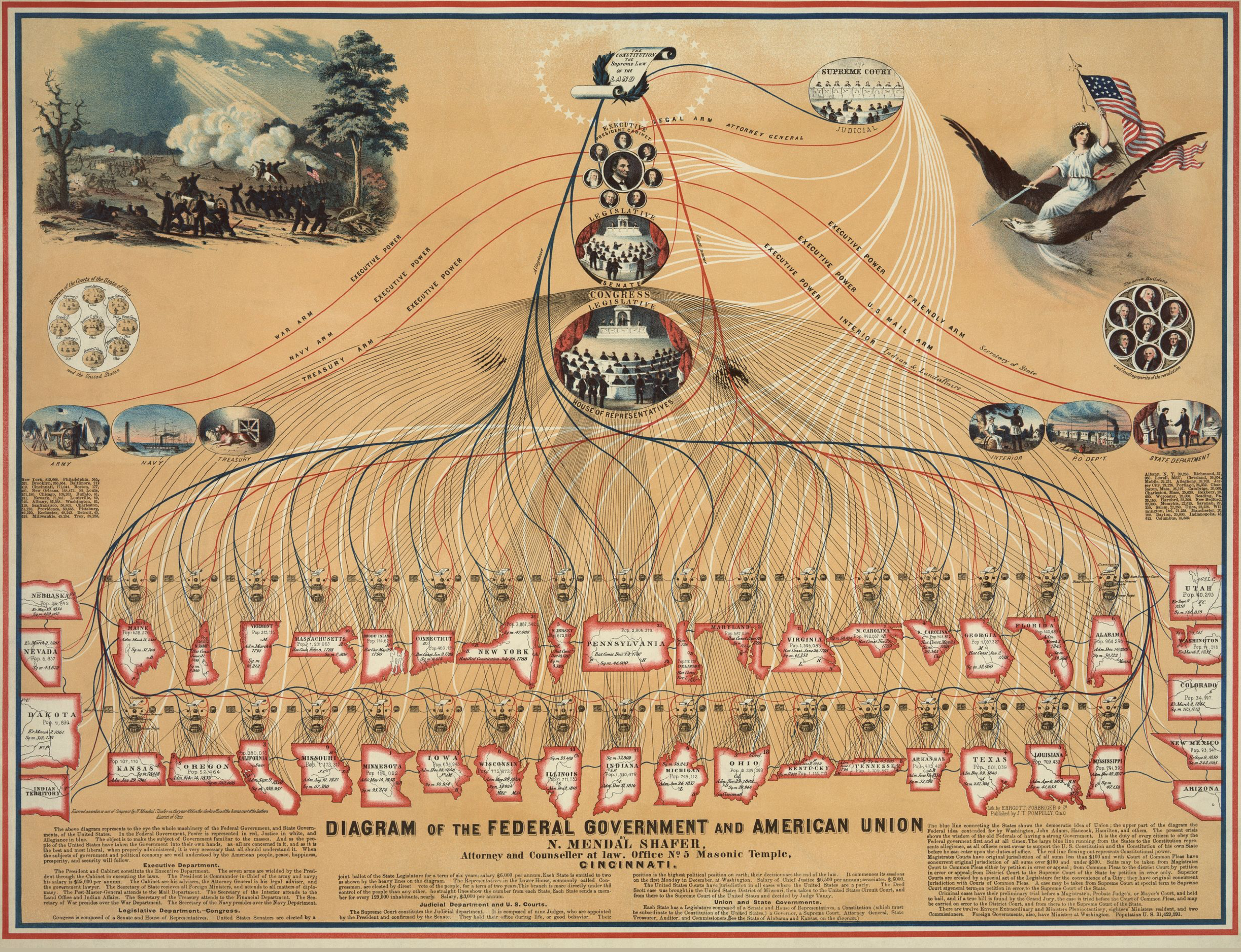


Diagram of the Federal Government and American Union (1862)



Shakespeare's tragedies

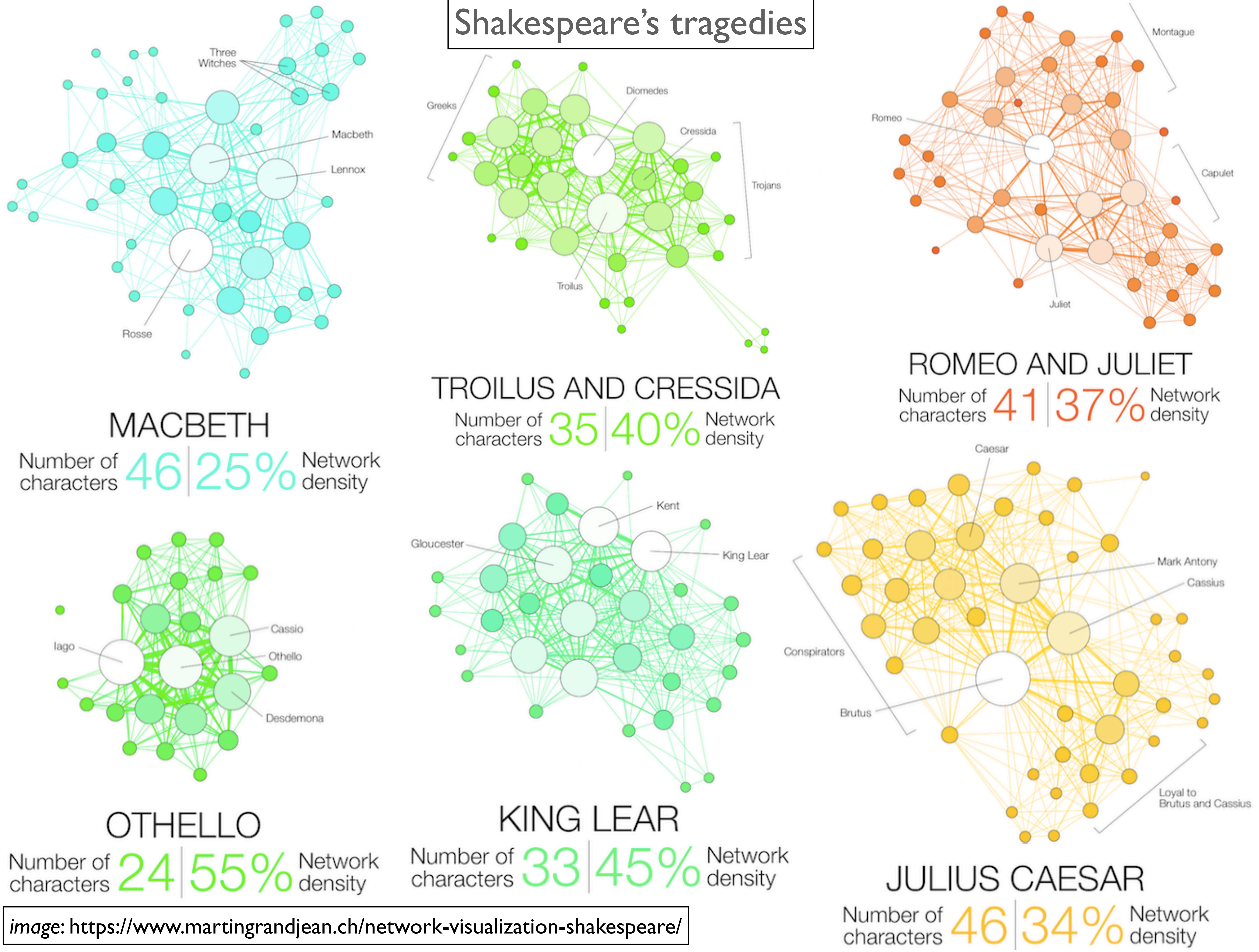
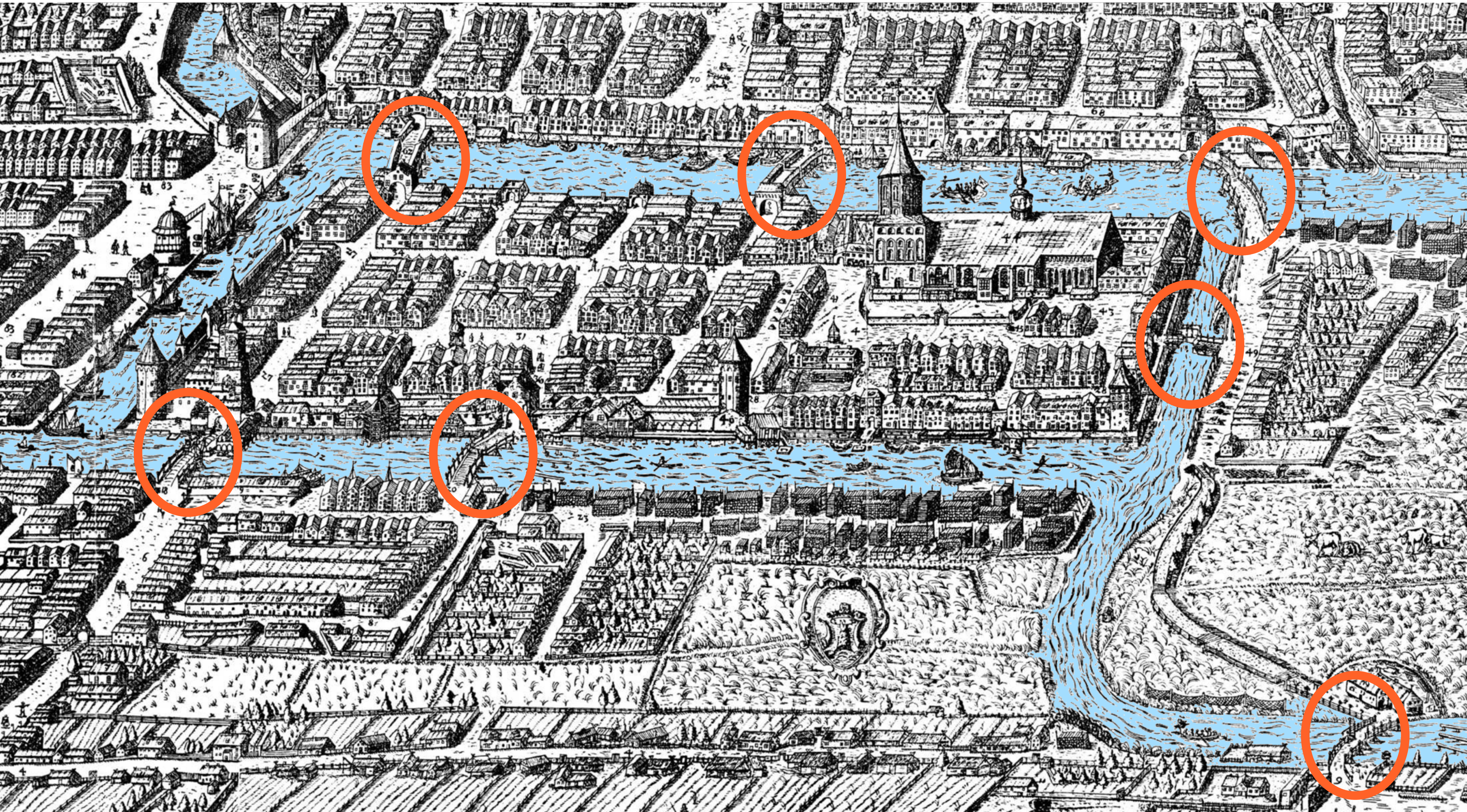
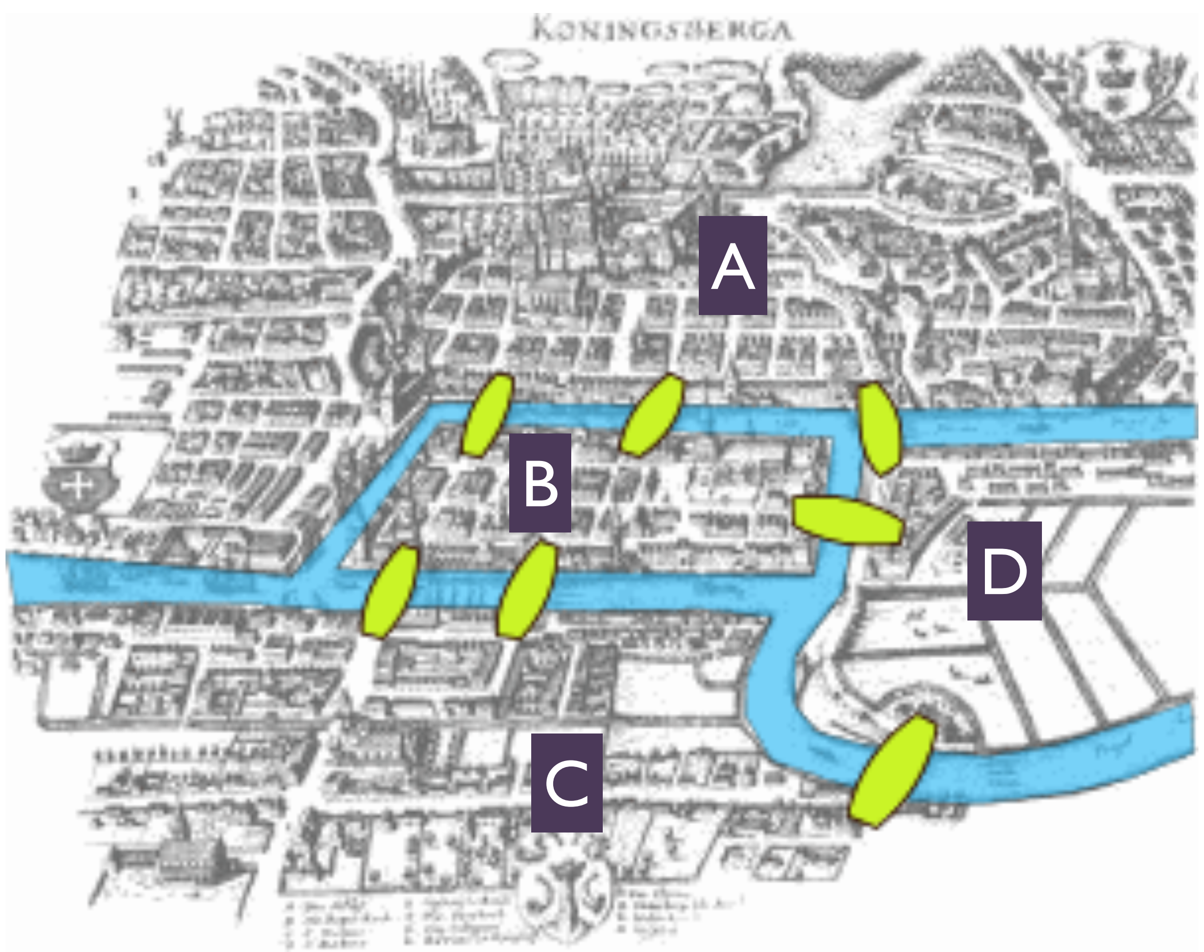


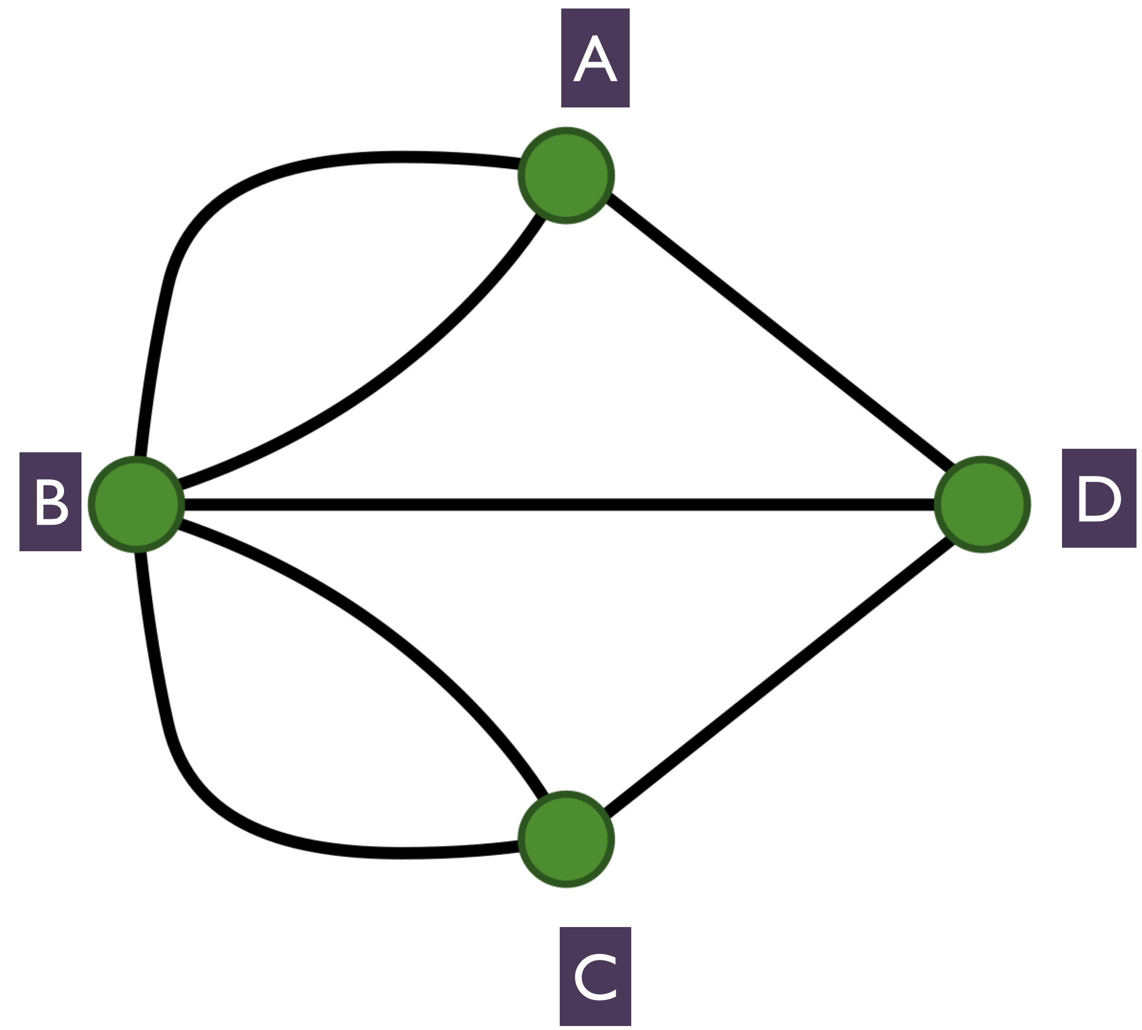
image: https://www.martingrandjean.ch/network-visualization-shakespeare/

THE SEVEN BRIDGES OF KÖNIGSBERG



EULER'S SOLUTION IN 1736



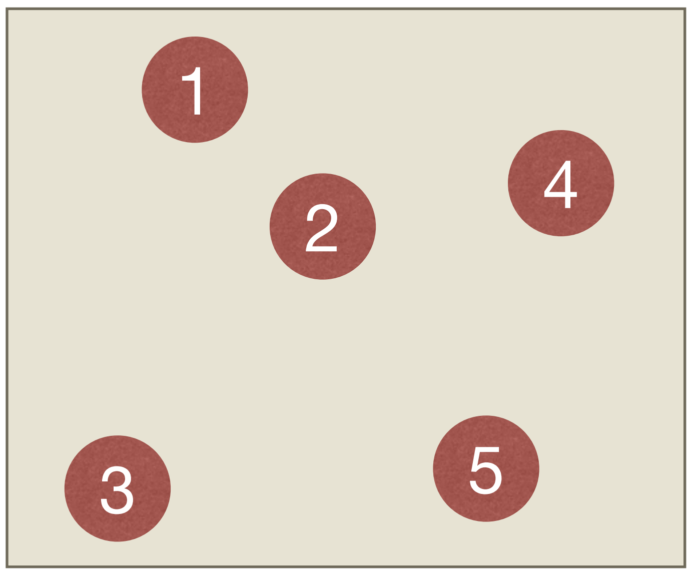


Each land mass can be viewed as a "vertex" and each bridge as a "link".

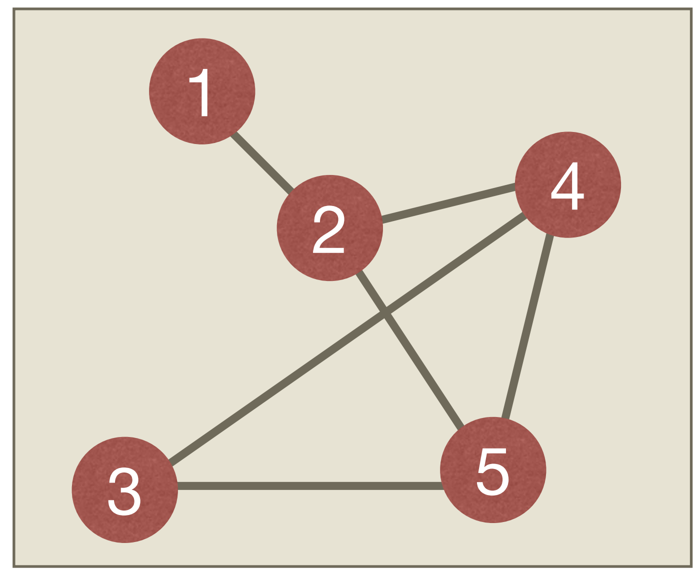
Only terminal vertices can have an odd number of links.

FUNDAMENTAL CONCEPTS

Nodes (or “Vertices”)

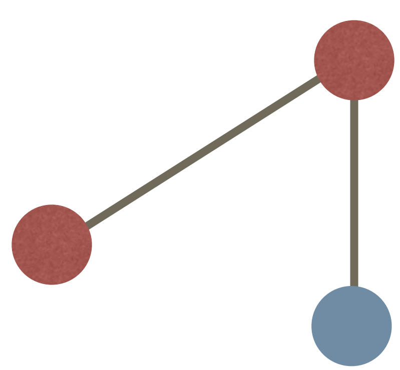


Links (or “Edges”)

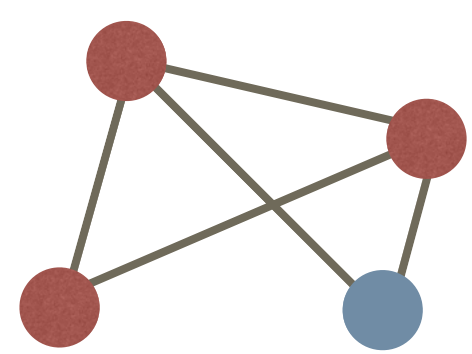


NODE	DEGREE
1	1
2	3
3	2
4	3
5	3

The total number of links associated with a node it its **degree** (k).

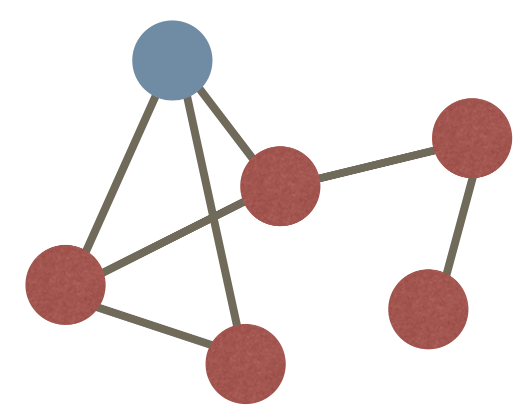


$k = 1$



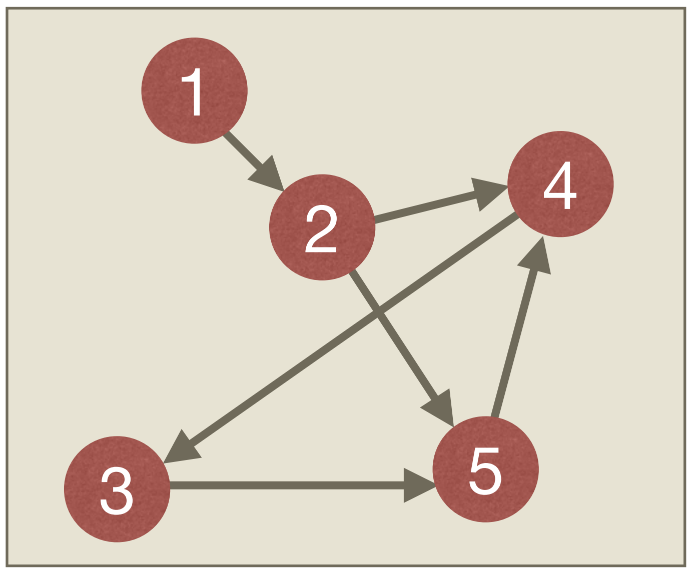
$k = 2$

$k = 3$

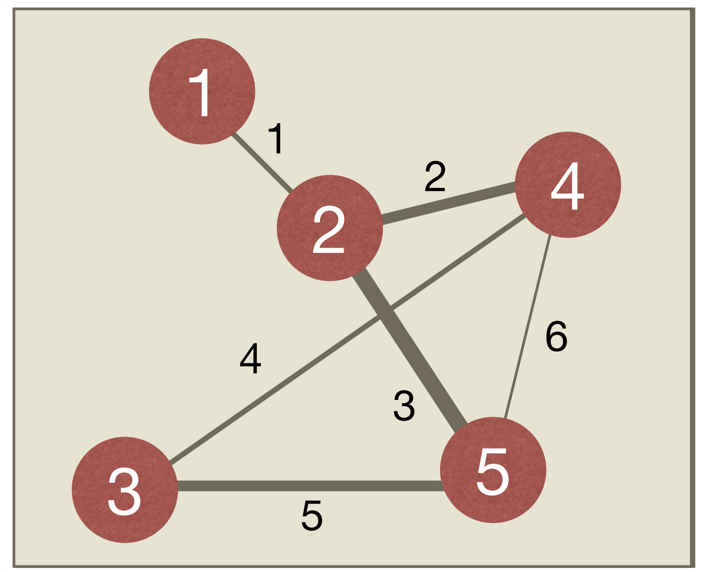


FUNDAMENTAL CONCEPTS

Directed network



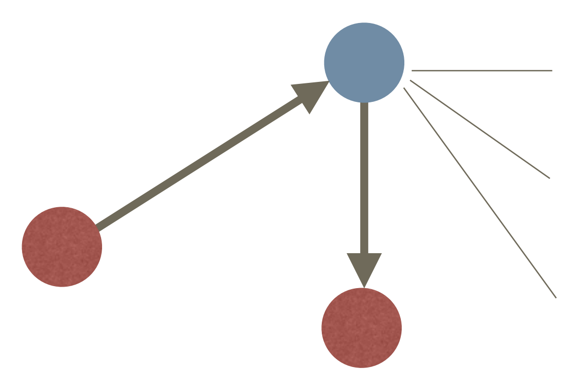
Weighted network



EDGE	WEIGHT
1	1
2	2
3	3
4	1
5	2
6	0.5

NODE	IN-DEGREE	OUT-DEGREE
1	0	1
2	1	2
3	1	1
4	2	1
5	2	1

In a directed network a node can have an in-degree different to its out-degree

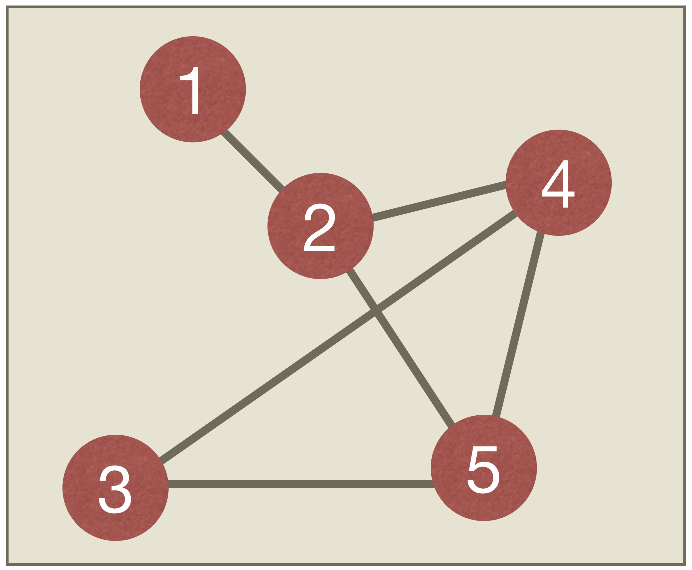


$k^{in} = 1$

$k^{out} = 1$

$k^{total} = 2$

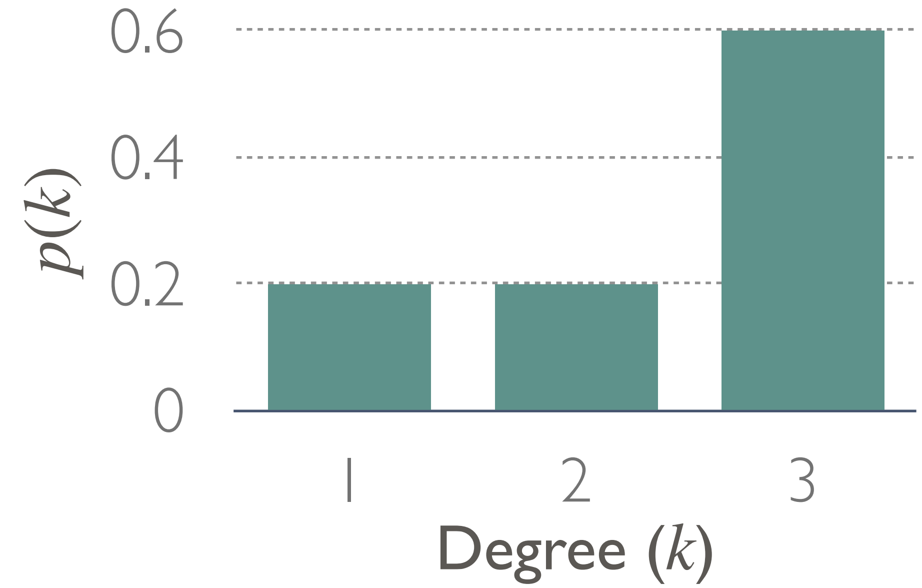
FUNDAMENTAL CONCEPTS

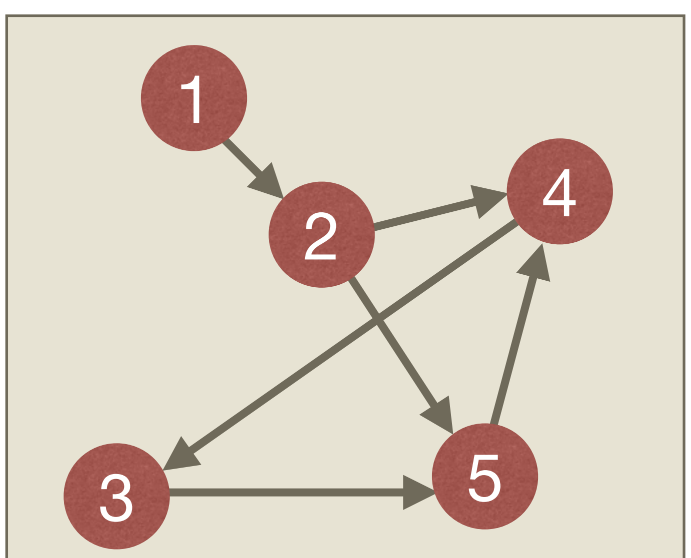


Adjacency matrix

	1	2	3	4	5
1	0	1	0	0	0
2	1	0	0	1	1
3	0	0	0	1	1
4	0	1	1	0	1
5	0	1	1	1	0

Degree distribution





target →

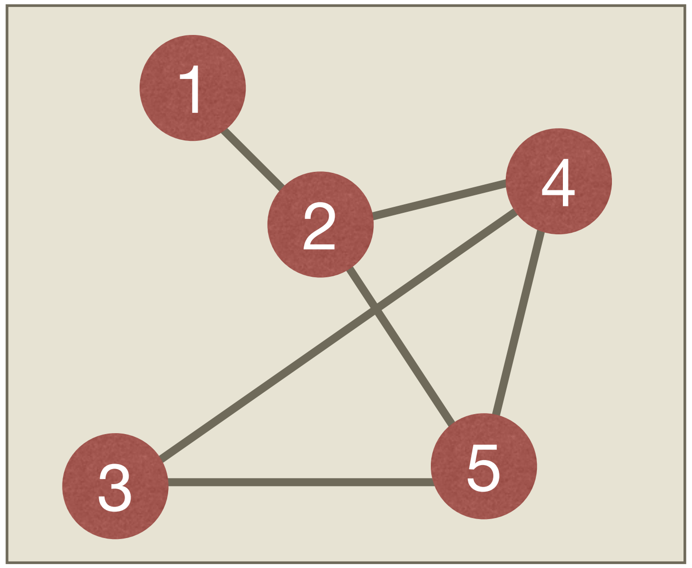
source ↓

	1	2	3	4	5
1	0	1	0	0	0
2	0	0	0	1	1
3	0	0	0	0	1
4	0	0	1	0	0
5	0	0	0	1	0

Average degree

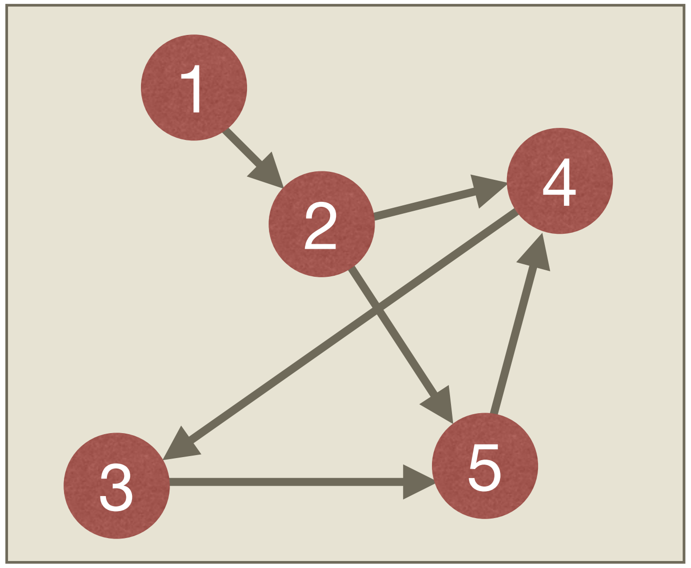
$$\langle k \rangle = \frac{1}{N} \sum_i k_i$$

FUNDAMENTAL CONCEPTS



i-j	1-2	1-3	1-4	1-5	2-3	2-4	2-5	3-4	3-5	4-5
d_{ij}	1	3	2	2	2	1	1	1	1	1

The shortest path length d_{ij} between two nodes i and j is the minimum number of links one has to cross to travel between them.

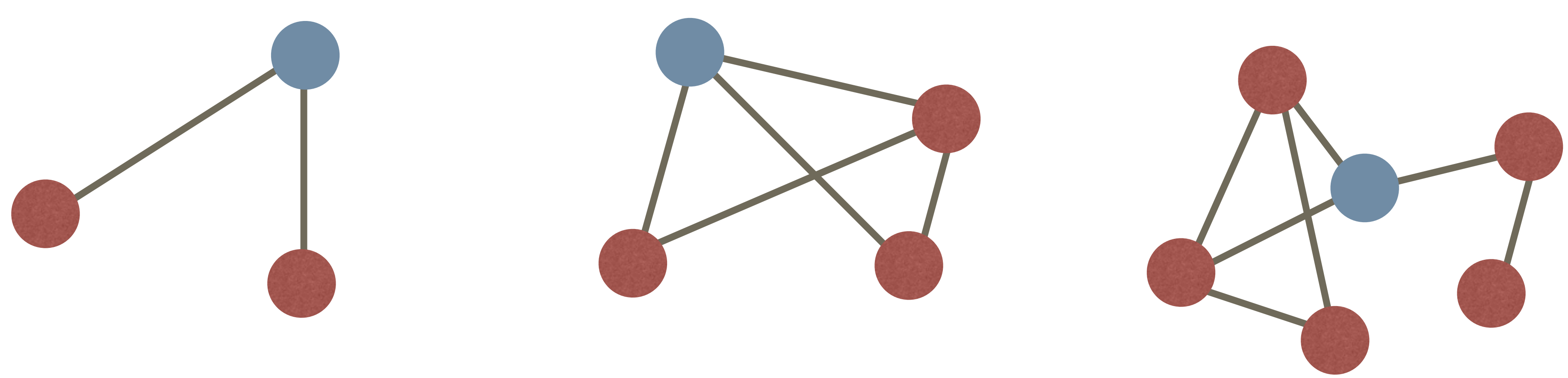


Can you work out what the shortest path length is between every pair of nodes of this directed network?

CLUSTERING COEFFICIENT

- The (local) clustering coefficient of a node measures the extent of connectivity of its local neighbourhood, i.e. how close they are to being a "clique" or a complete subgraph.
- If a node i in an undirected network has k_i neighbours, there can be a maximum of $k_i(k_i - 1)/2$ links between them.
- The clustering coefficient C_i of node i is the fraction of these links that exist.

What is the clustering coefficient of the blue nodes?



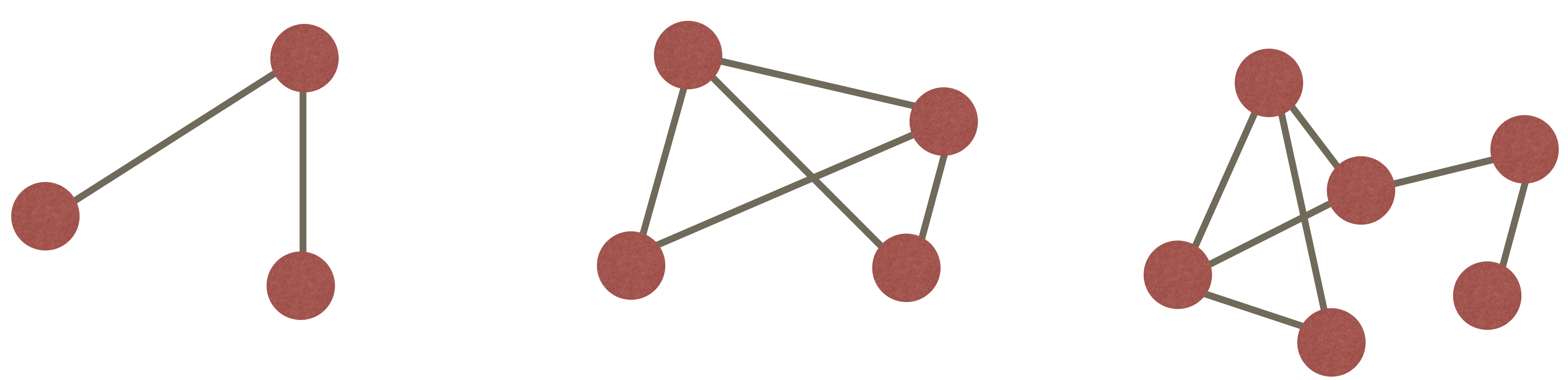
AVERAGE PATH LENGTH

- The average path length is the average of the shortest path lengths between every pair of nodes in the network.

- For a network comprising N nodes, if $d(i, j)$ is the shortest number of steps between nodes i and j, then the average path length is:

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

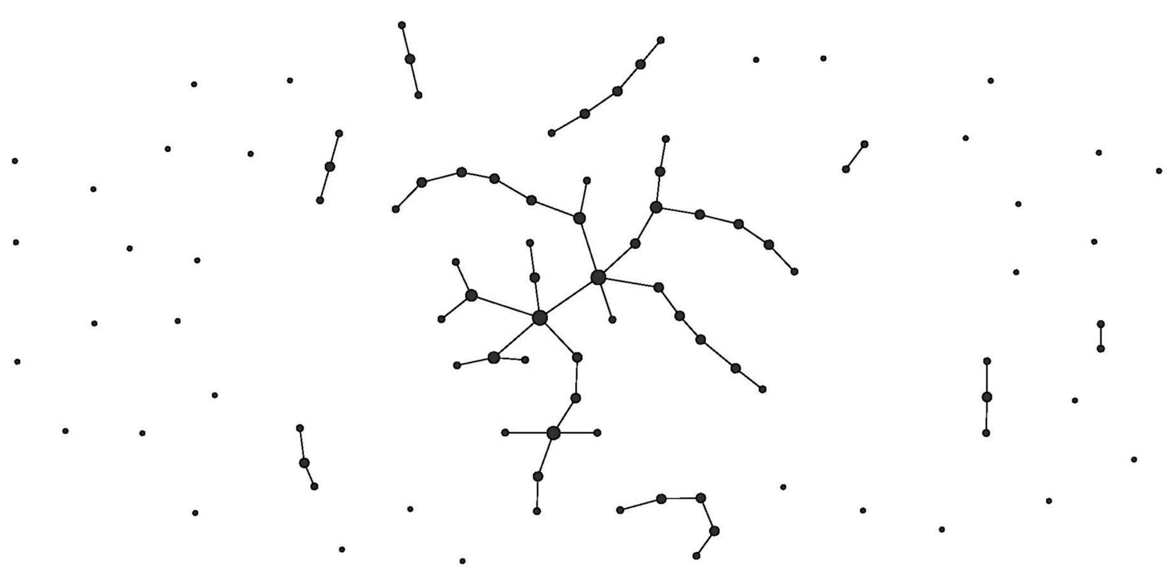
What is the average path length of these networks?



RANDOM NETWORKS

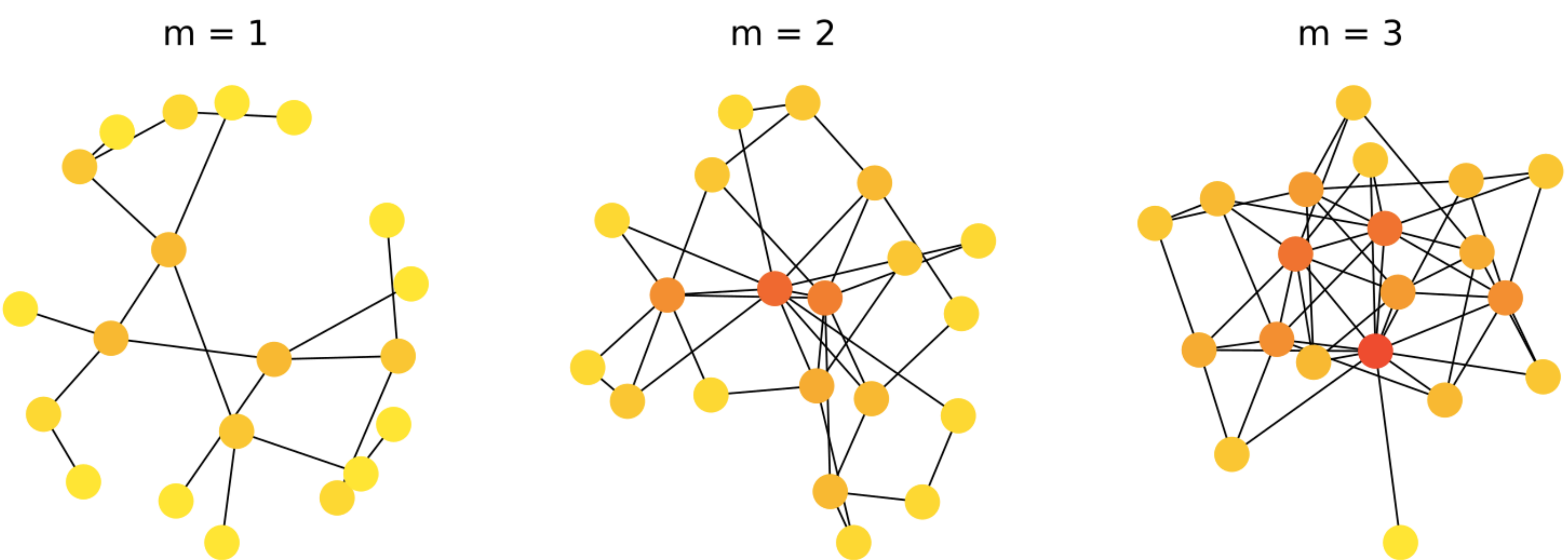
- In 1959 two related models for generating random networks were proposed.
- In the more commonly used version of the model, we specify the number of nodes N, and connect each pair of nodes with a probability p.

Paul Erdős Alfréd Rényi Edgar Gilbert



- These random networks are commonly referred to as Erdős-Rényi (ER) networks.
- Depending on the product Np, the resulting network may have multiple connected components.
- There exists no path between nodes of two different components.

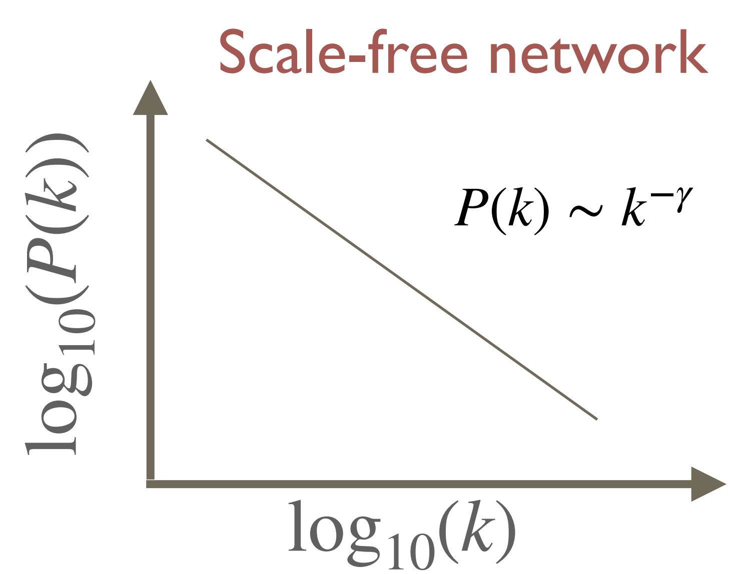
RANDOM NETWORKS



Albert-László Barabási

Réka Albert

- The Barabási-Albert (BA) model generates random "scale-free" networks using a preferential attachment mechanism.
- Nodes are sequentially added to the network and each connects to m random existing nodes.
- The probability that a new node connects to an existing node i is: $p_i = k_i / \Sigma_j k_j$.

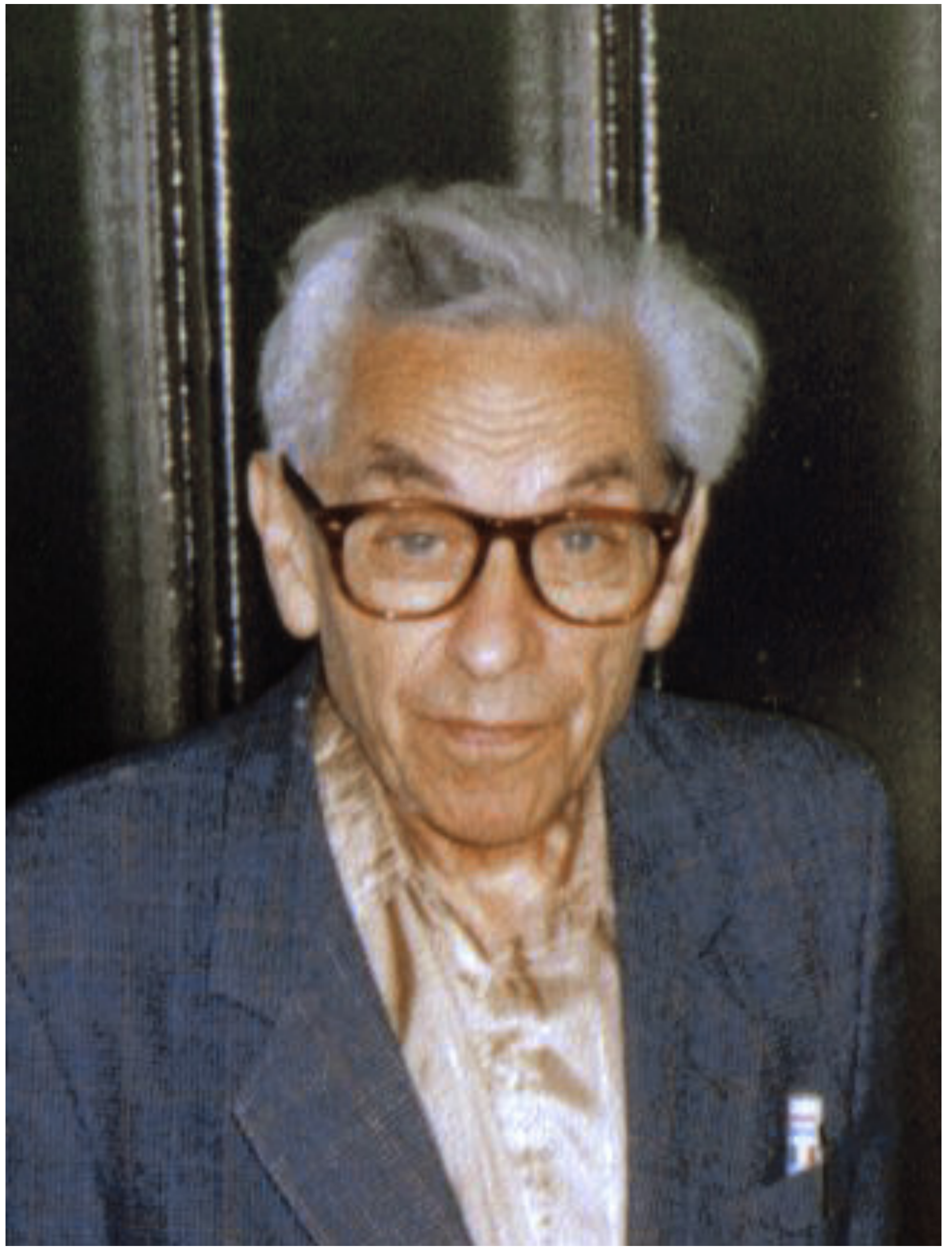


"The worker knows the manager in the shop, who knows Ford; Ford is on friendly terms with the general director of Hearst Publications, who last year became good friends with Árpád Pásztor, someone I not only know, but is to the best of my knowledge a good friend of mine - so I could easily ask him to send a telegram via the general director telling Ford that he should talk to the manager and have the worker in the shop quickly hammer together a car for me, as I happen to need one."

Frigyes Karinthy, "*Láncszemek (Chains)*" (1929).

WHAT DO THESE TWO HAVE IN COMMON?

Kevin Bacon



Paul Erdős

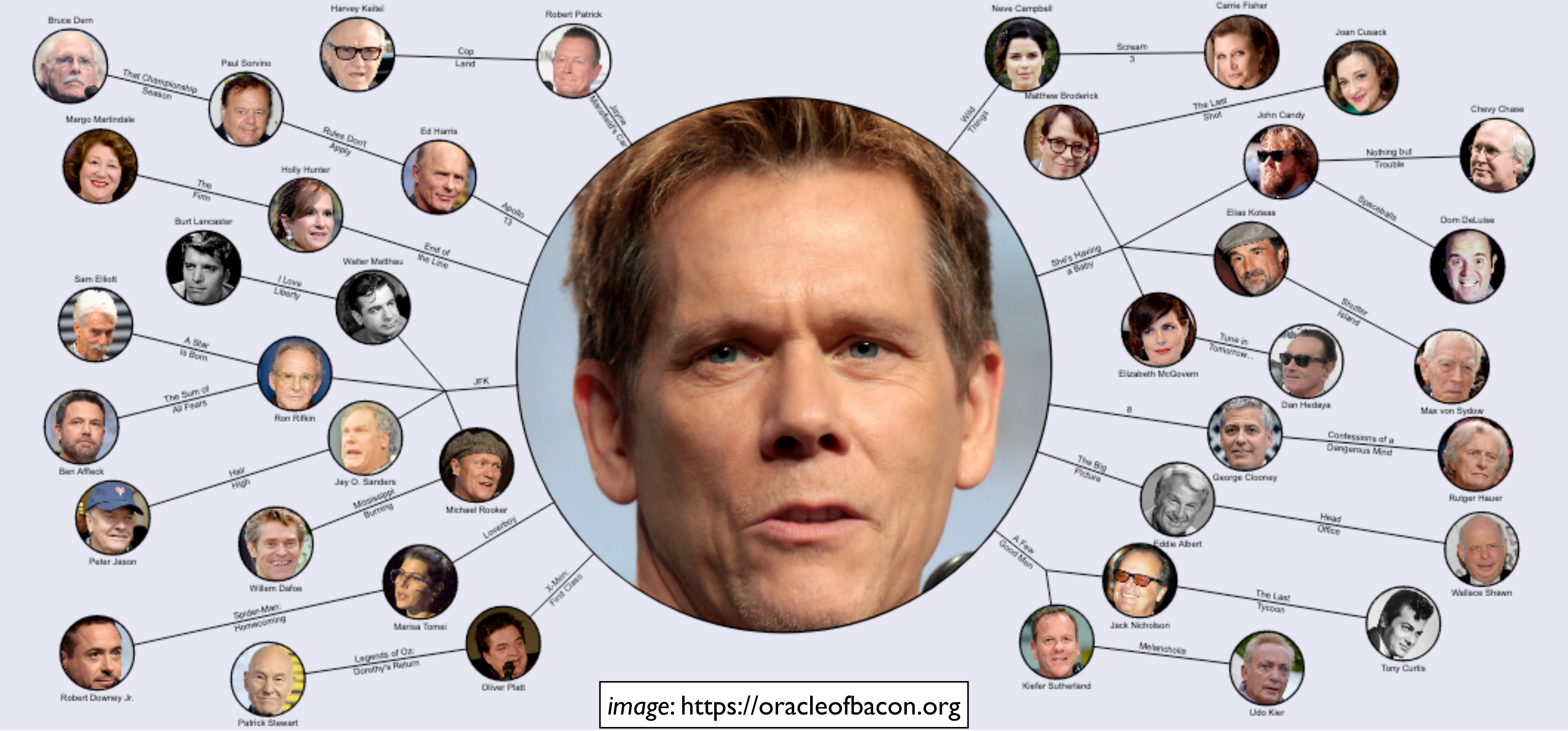
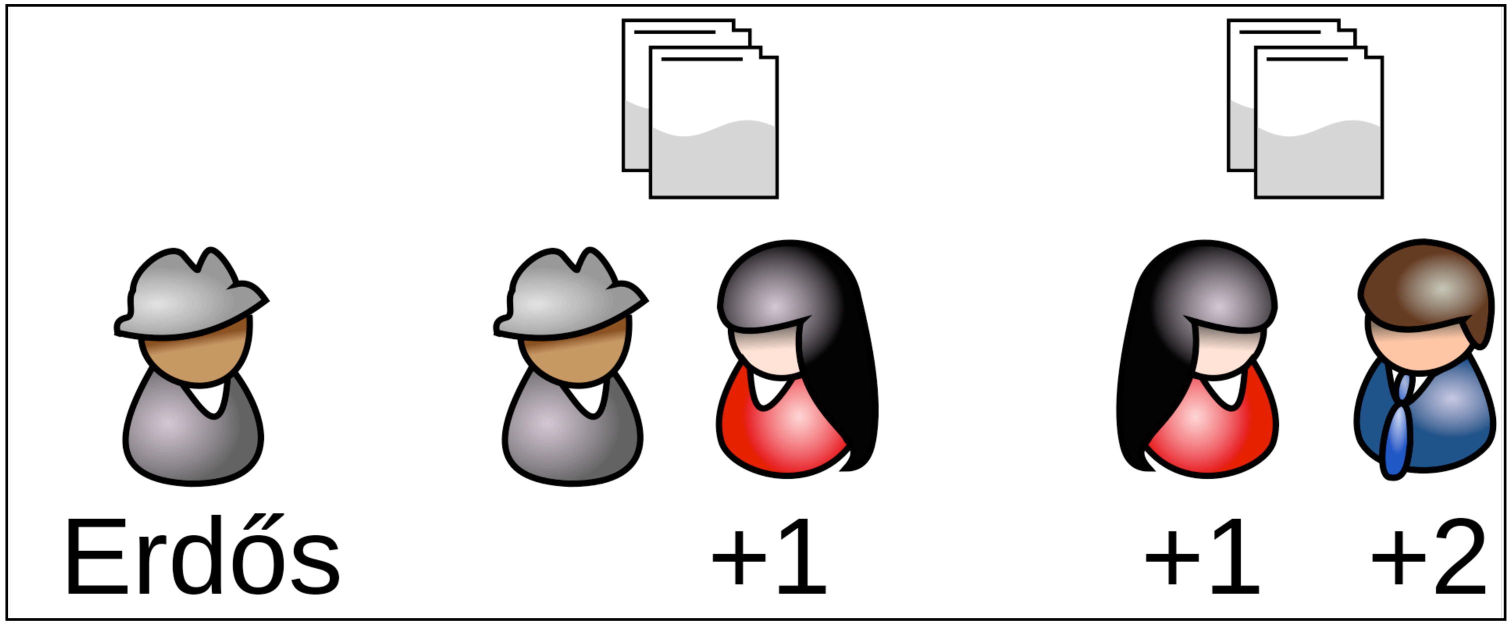


image: https://oracleofbacon.org

Both their names are popularly associated with the concept of “six degrees of separation”



MILGRAM'S LETTER EXPERIMENT

- Information packets were sent to 296 randomly selected individuals in Nebraska and Kansas.
- They were asked to forward the letter to someone who they think might know a specified person in Boston.
- Of the letters that reached the final target, the average path length was ~ 5.2.

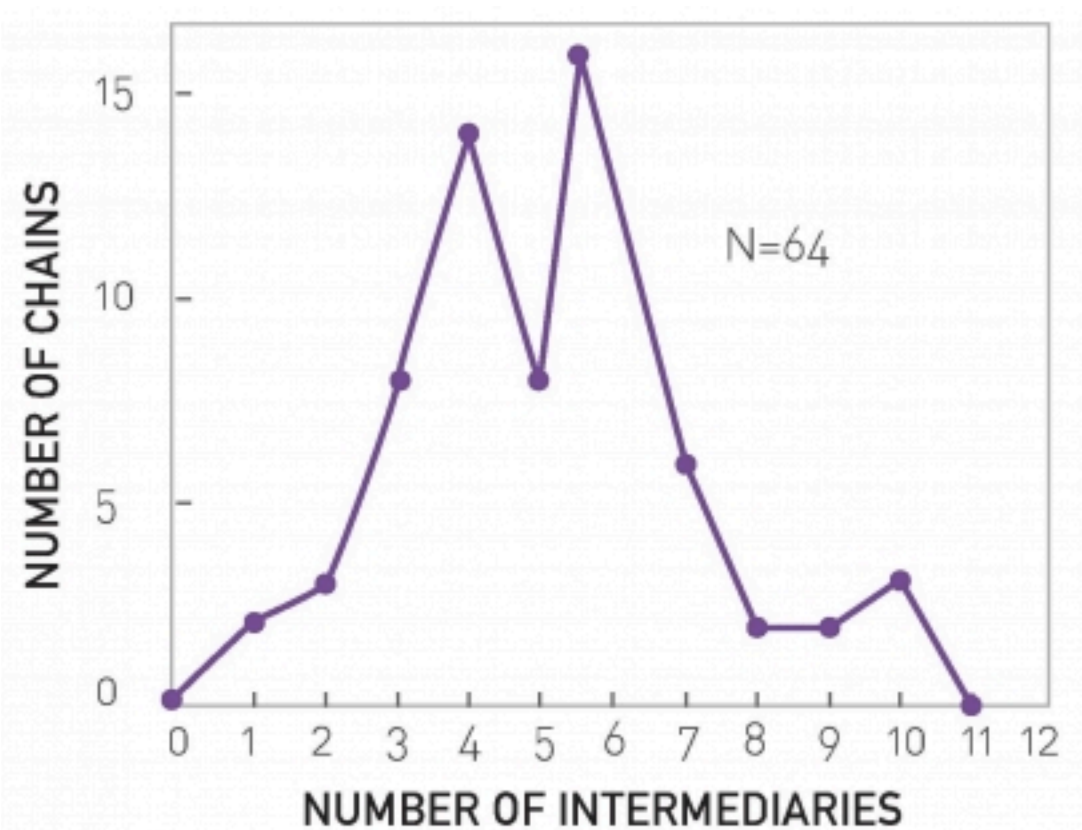
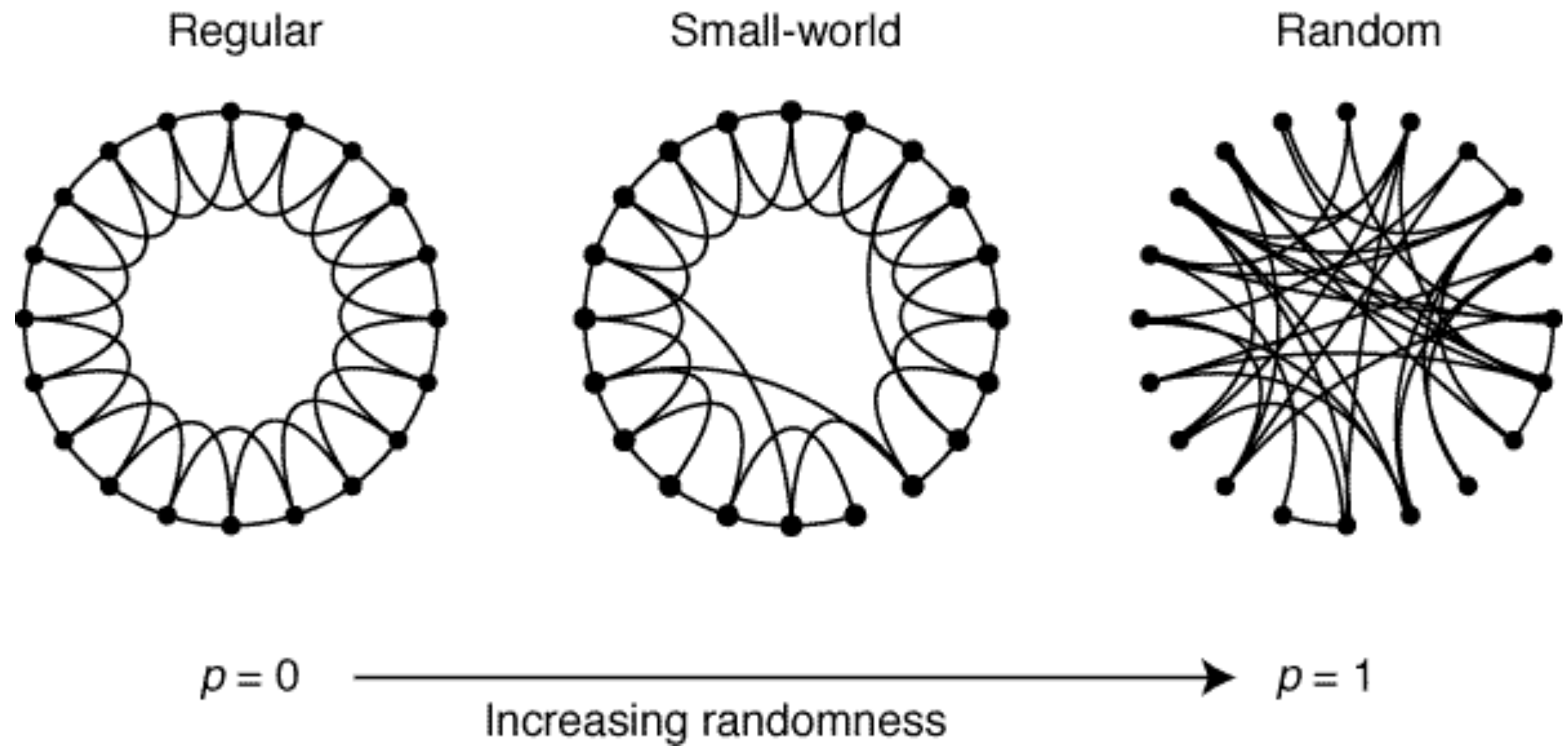
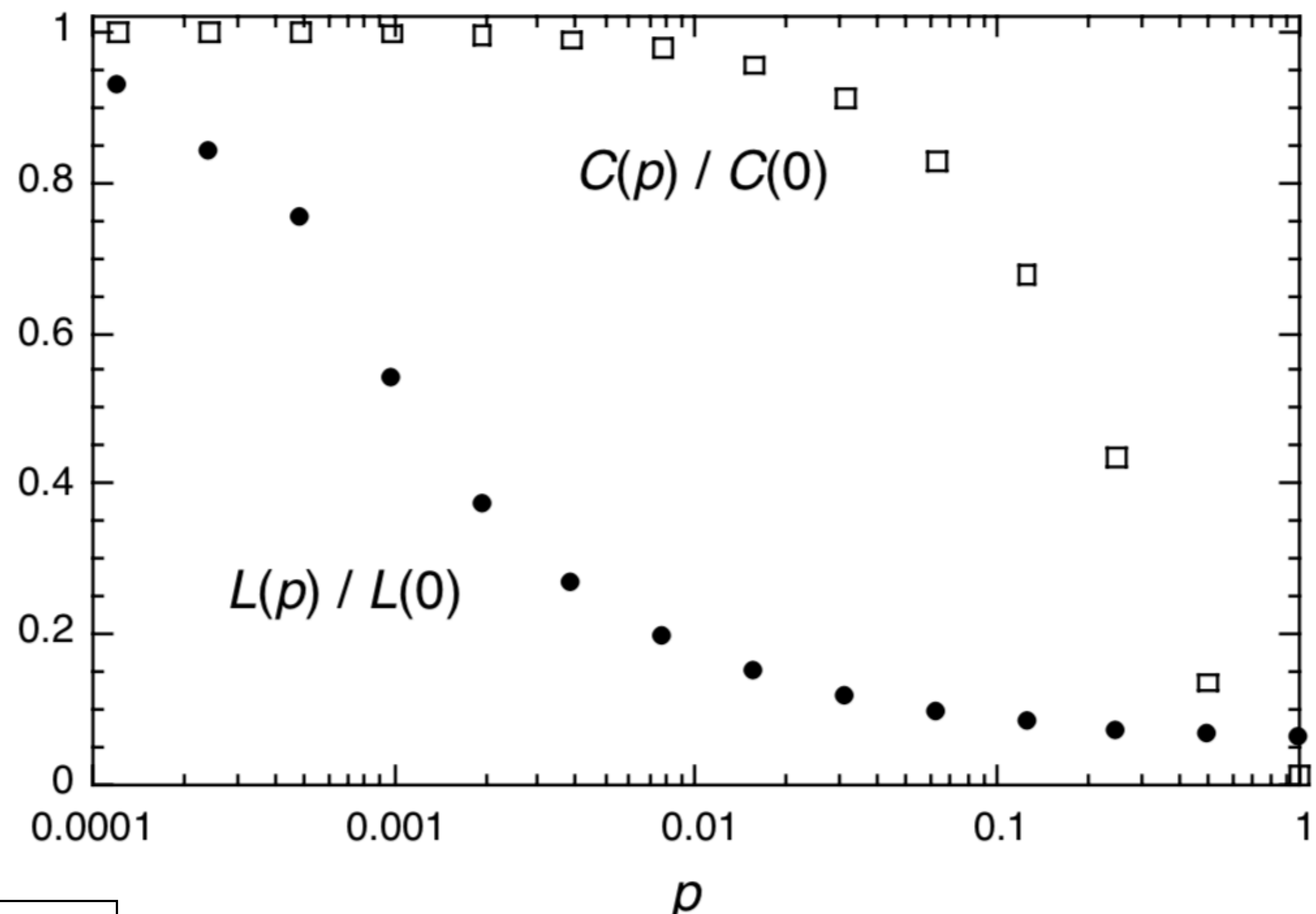


image [left]: Barabási, A.-L., "Network Science" (http://networksciencebook.com/)

SMALL WORLD NETWORKS



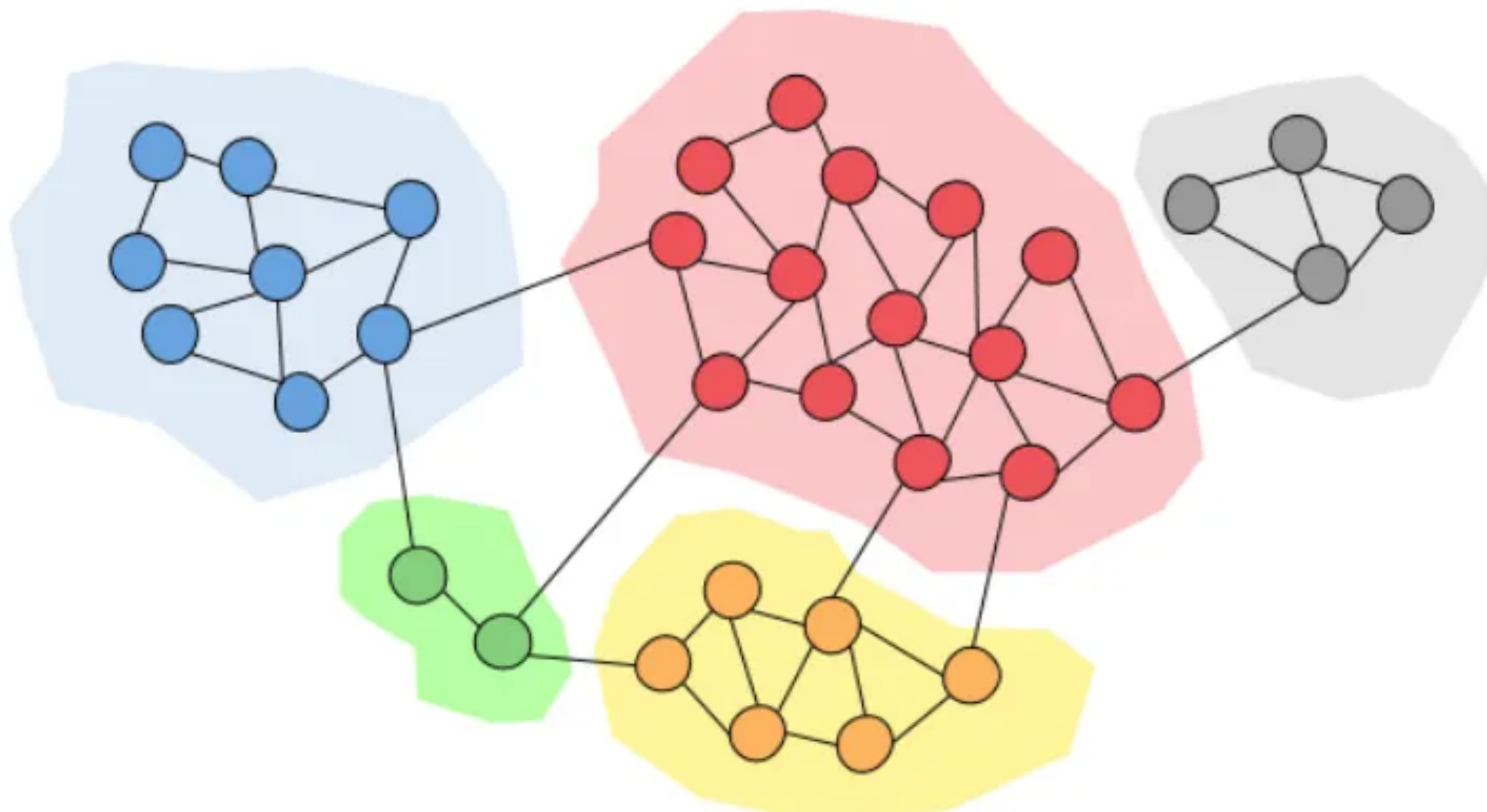
“Small world” networks are characterized by a low average path length L and a high clustering coefficient C.



images: Watts, D. & Strogatz, S., *Nature* **393**, 440-442 (1998).

MODULARITY

A network is said to have a modular structure if there exist groups (or “communities”) of nodes that have a higher density of connections than that between groups.



In practice, one has to first specify the modules/communities and then check if the density of intra-connections is more than that of the inter-connections.

image: https://towardsdatascience.com/community-detection-algorithms-9bd8951e7dae

Brain networks are modular

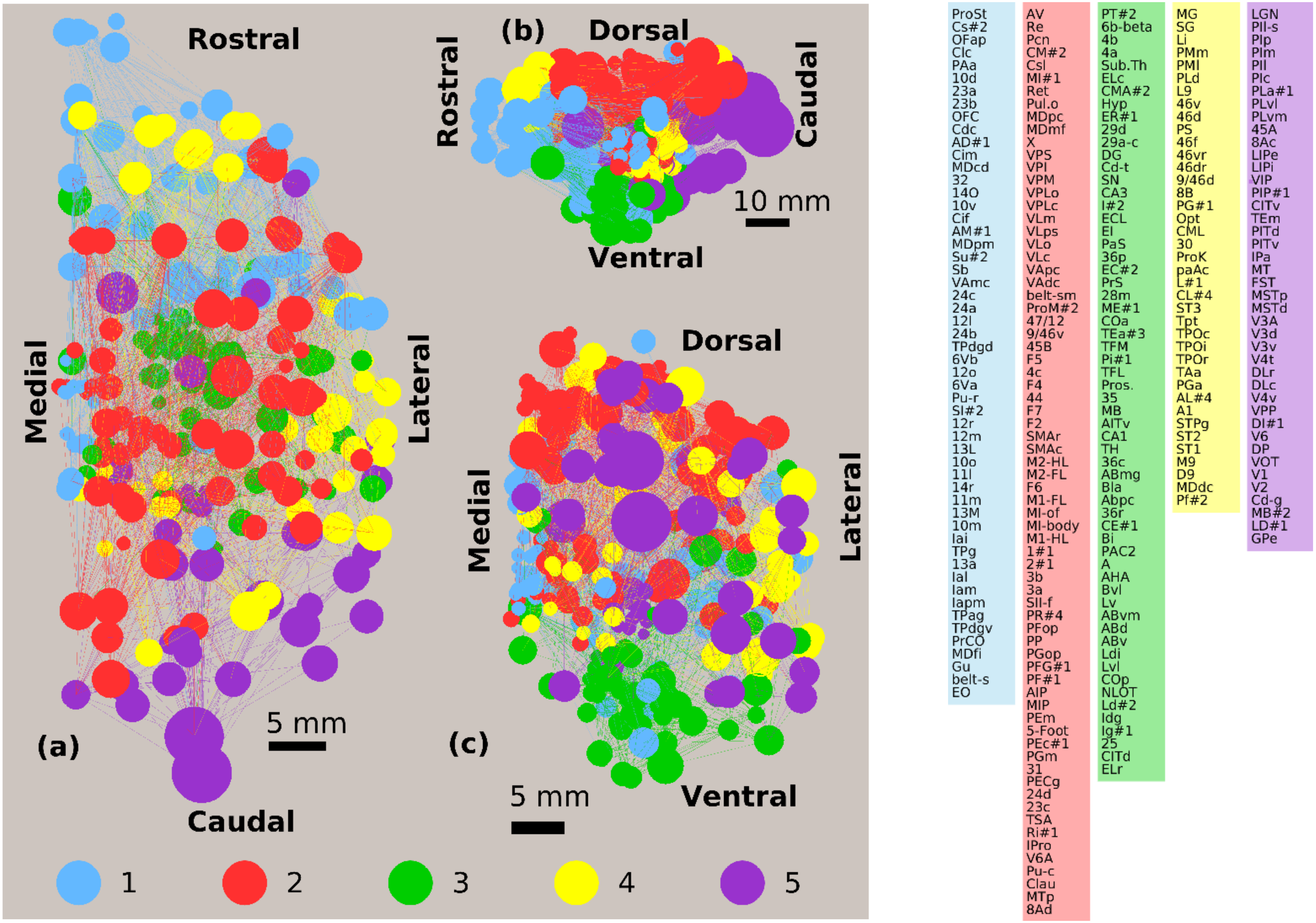
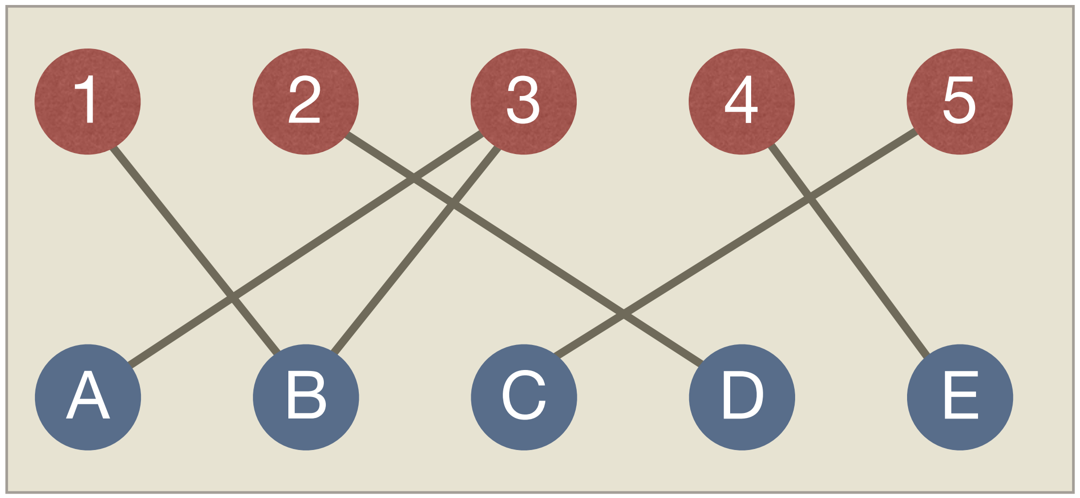


image: Pathak, A., Menon, S. N. and Sinha, S., *Phys. Rev. E* **106**, 054304 (2022).

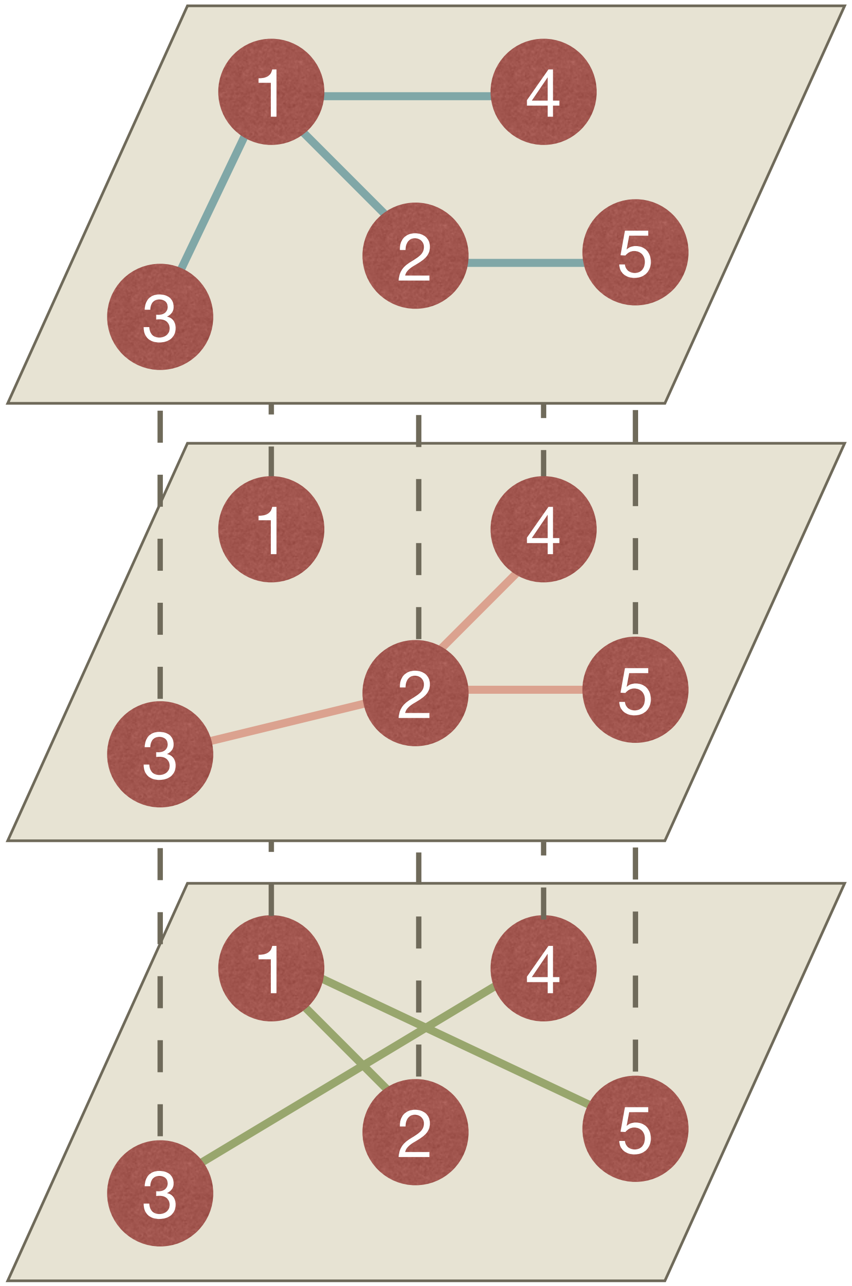
SOME OTHER TYPES OF NETWORKS

Networks that describe relations between two different classes of objects are known as Bipartite networks.

Networks in which there may be different types of links between nodes are known as Multiplex networks.



Bipartite network



Multiplex network

REFERENCES

ARTICLES:

- Erdős, P. & Rényi, A. *Publicationes Mathematicae* **6**, 290–297 (1959). [Random network model]
- Gilbert, E.N., *Annals of Mathematical Statistics* **30**: 1141–1144 (1959). [Random network model]
- Travers, J. & Milgram, S., *Sociometry* **32**, 425-443 (1969). [Milgram's experiment]
- Watts, D. & Strogatz. S., *Nature* **393**, 440-442 (1998). [Small-world construction]
- Barabási, A.-L. & Albert, R., *Science* **286**, 509–512 (1999). [Scale-free networks]

BOOKS:

- Barabási, A.-L., "*Network Science*" (*http://networksciencebook.com/*)
- Newman, M., *"Networks"* (OUP, 2018).