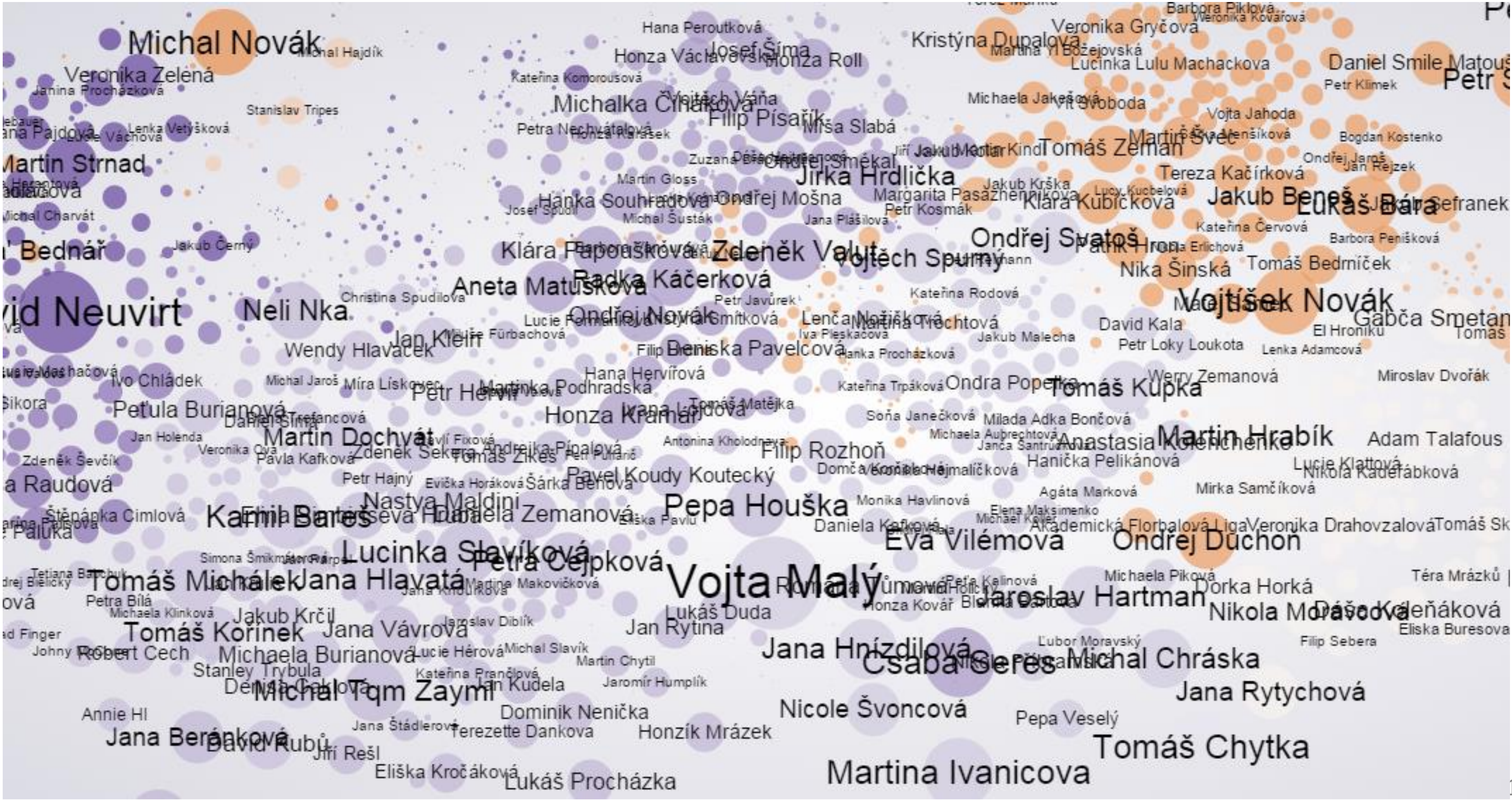
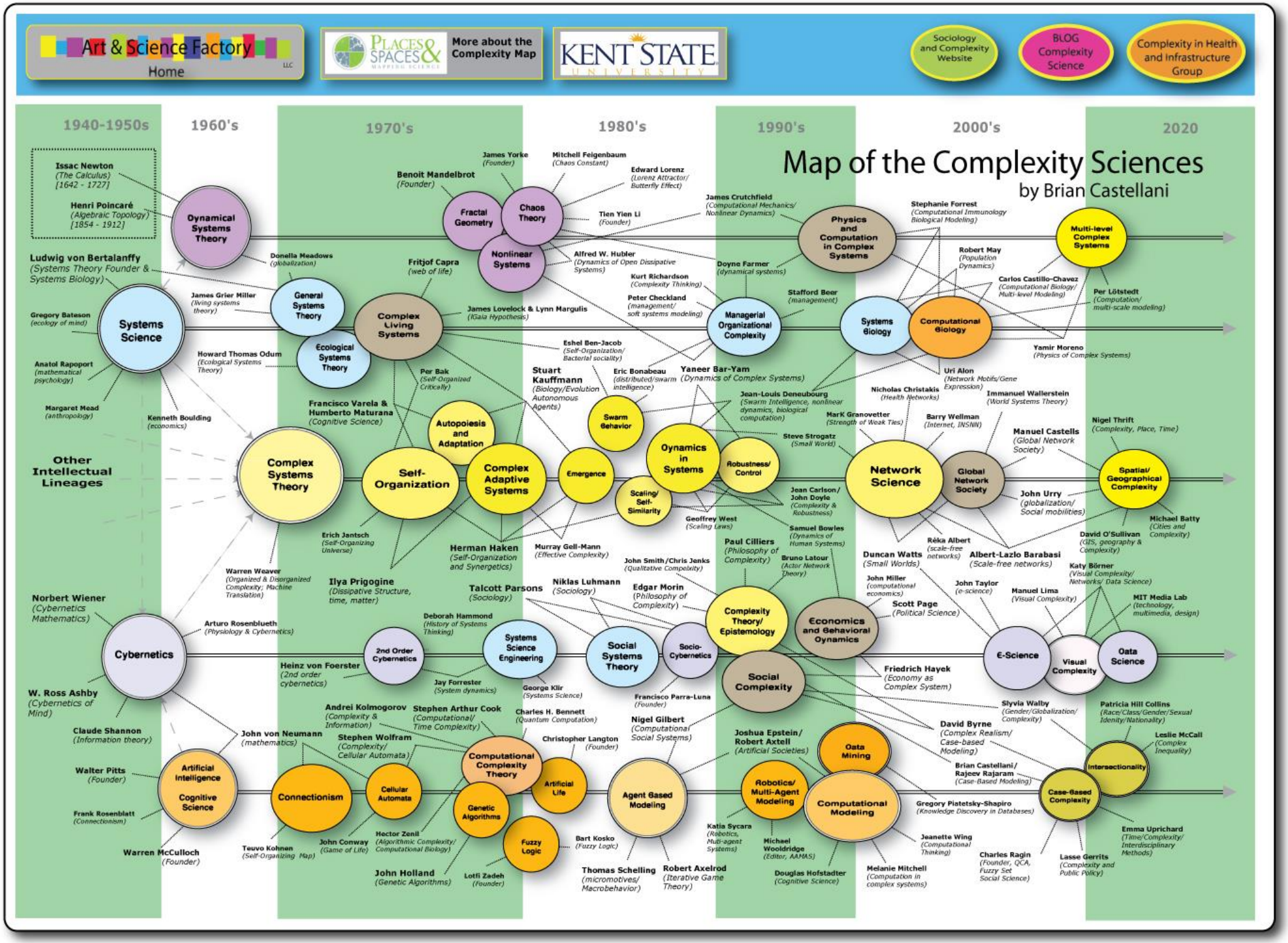
fm.vse.cz

Complexity Visualisation

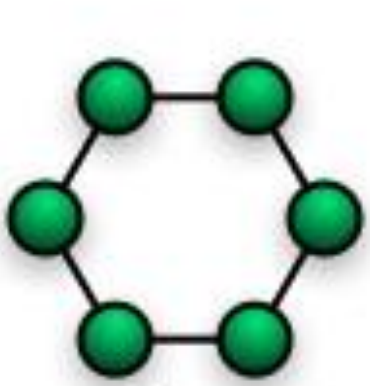


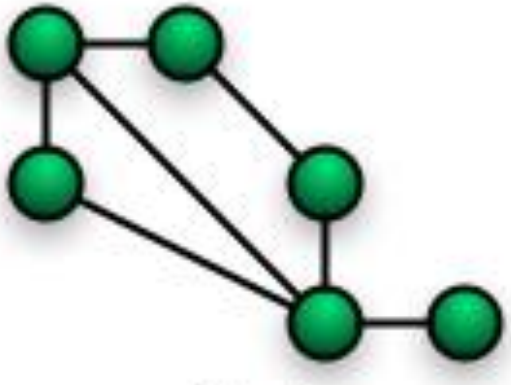

TED
Watch
Read
Attend
Participate
About
Log in
Eric Berlow:
Simplifying complexity
View interactive transcript
Watch later
Favorite
Download
951,349
Ecologist Eric Berlow doesn't feel overwhelmed when faced with complex systems. He knows that more information can lead to a better, simpler solution. Illustrating the tips and
Related playlists
Browse all



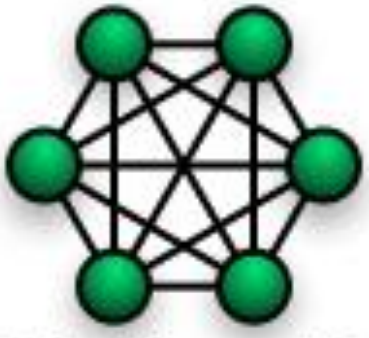
http://www.art-sciencefactory.com/complexity-map.html

Network Typology

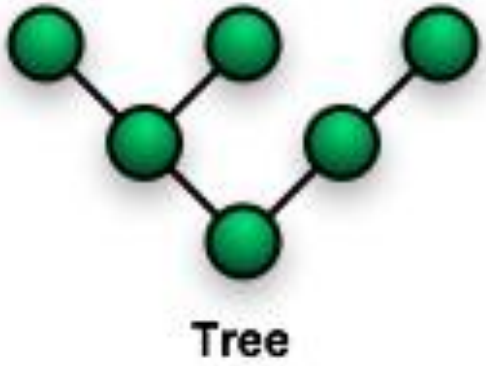


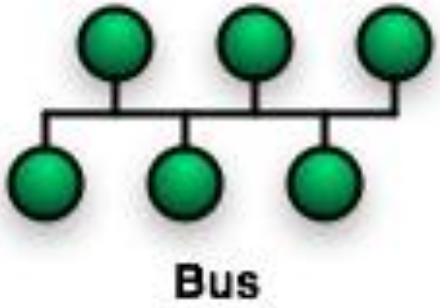






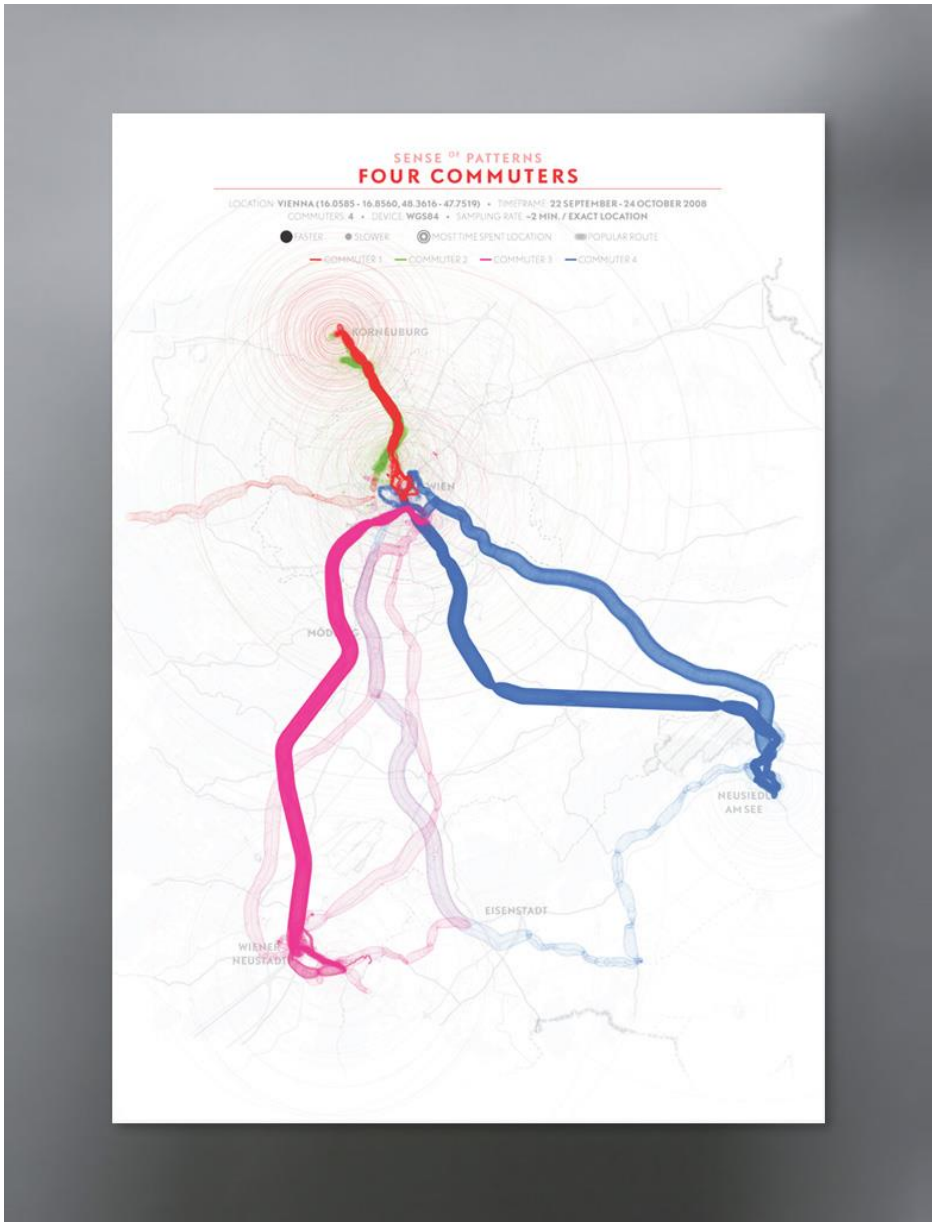






http://en.wikipedia.org/wiki/Network_topology

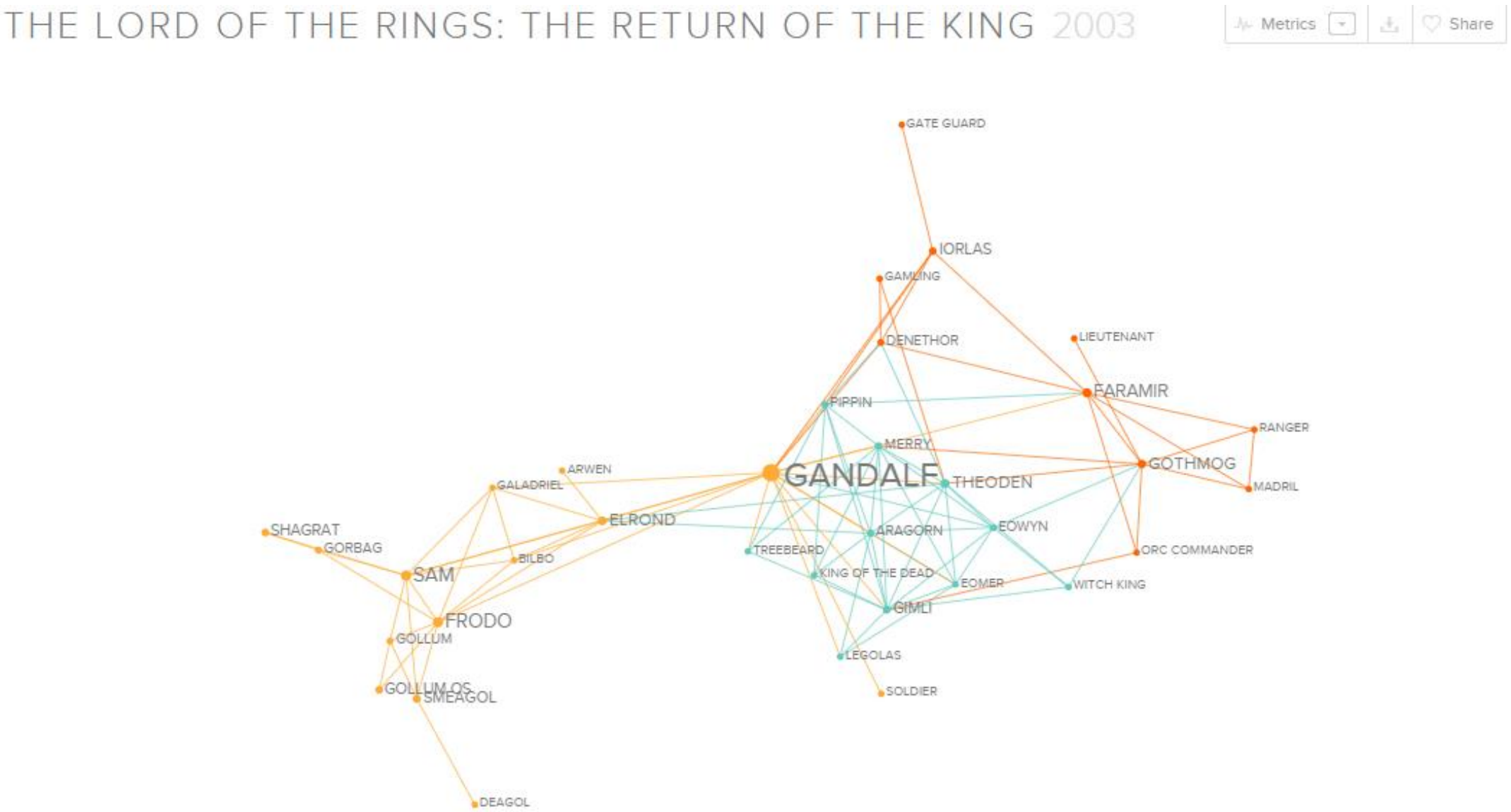
Possible Domains for Visualisation

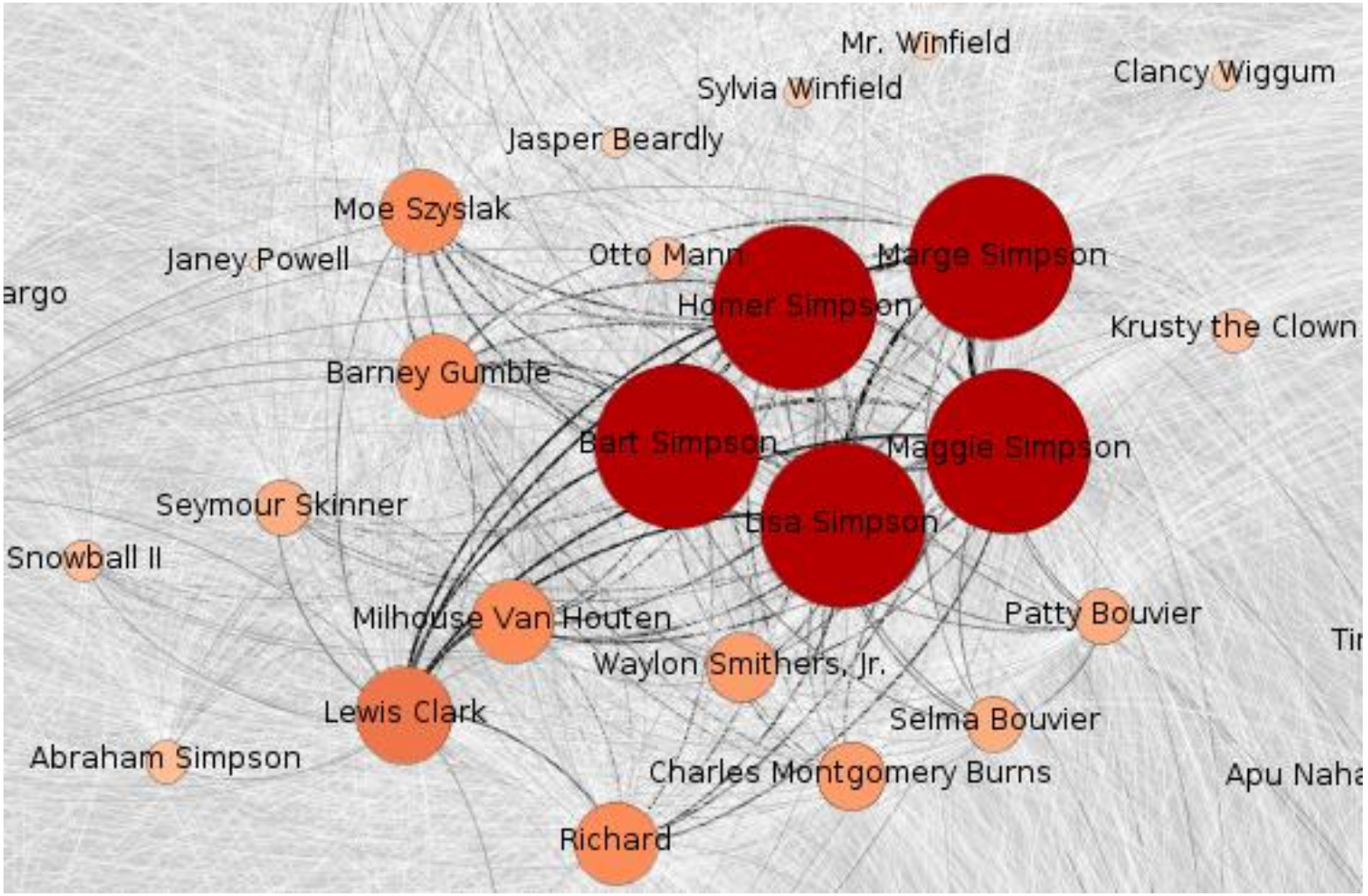


http://casualdata.com/senseofpatterns/

Possible Domains for Visualisation

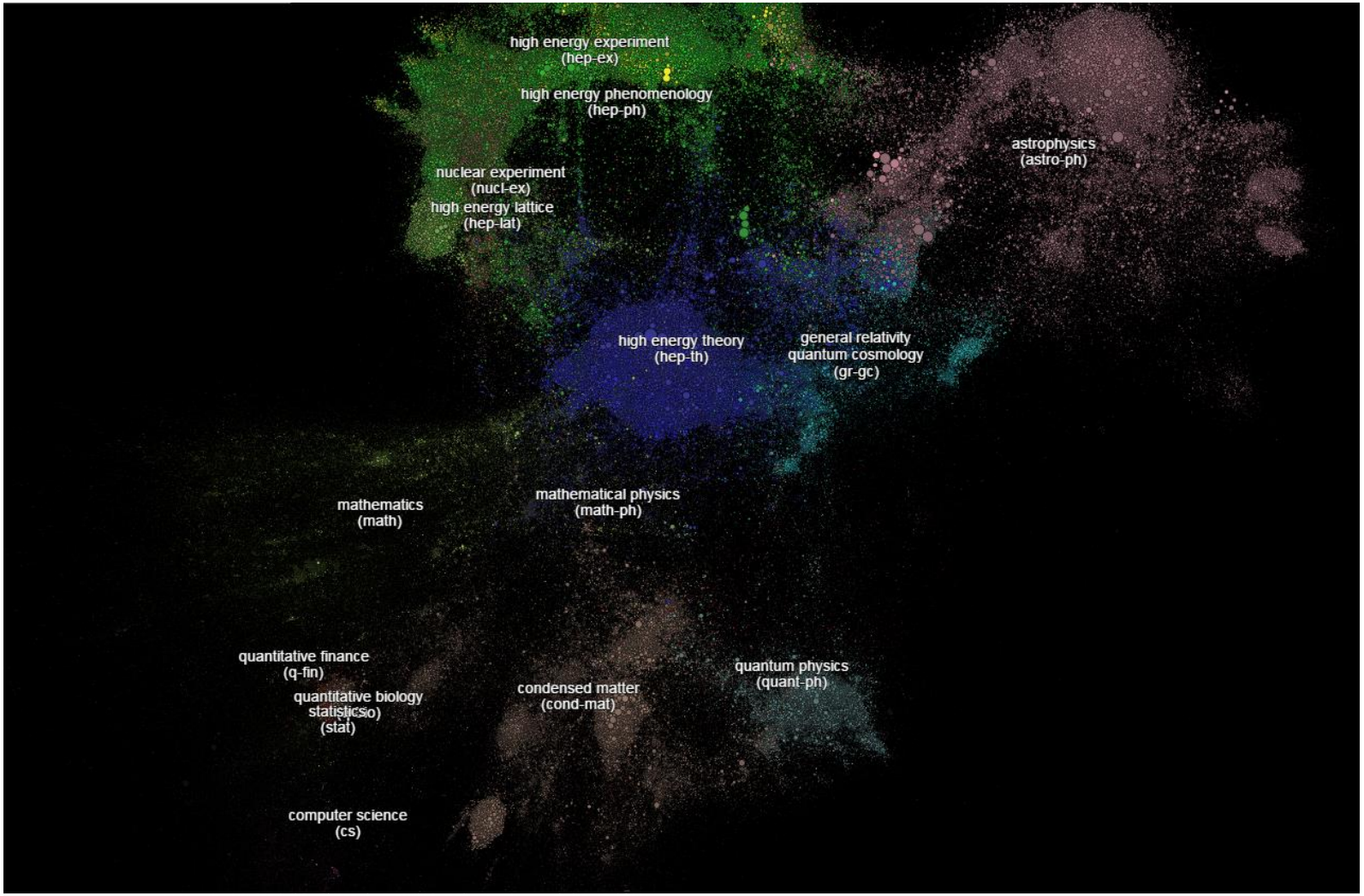
http://www.moviegalaxies.com/movies/513-The-Lord-of-the-Rings:-The-Return-of-the-King





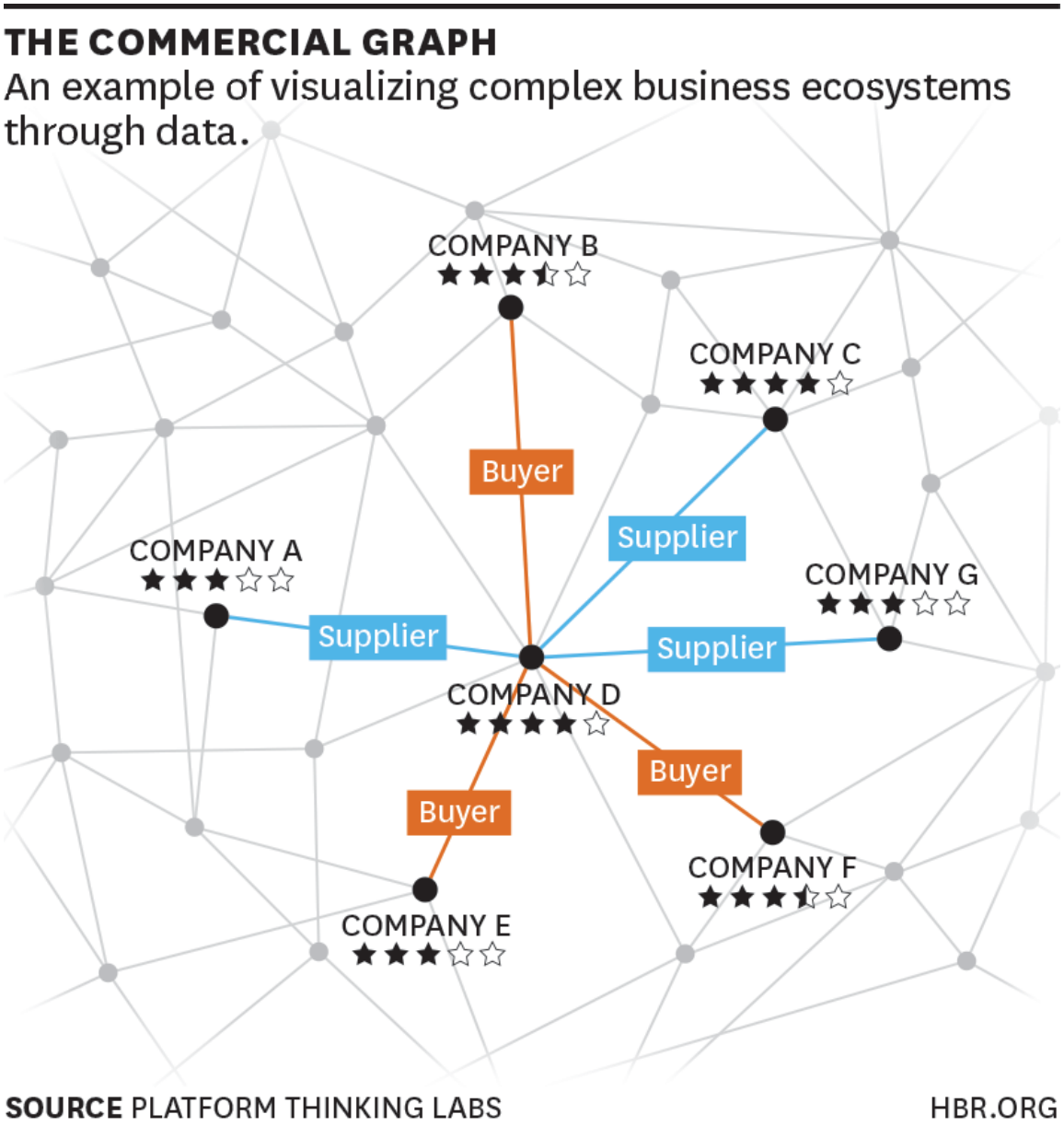
http://www.vislives.com/2014/09/the-simpons-social-netwok-season-1.html

Scientometrics



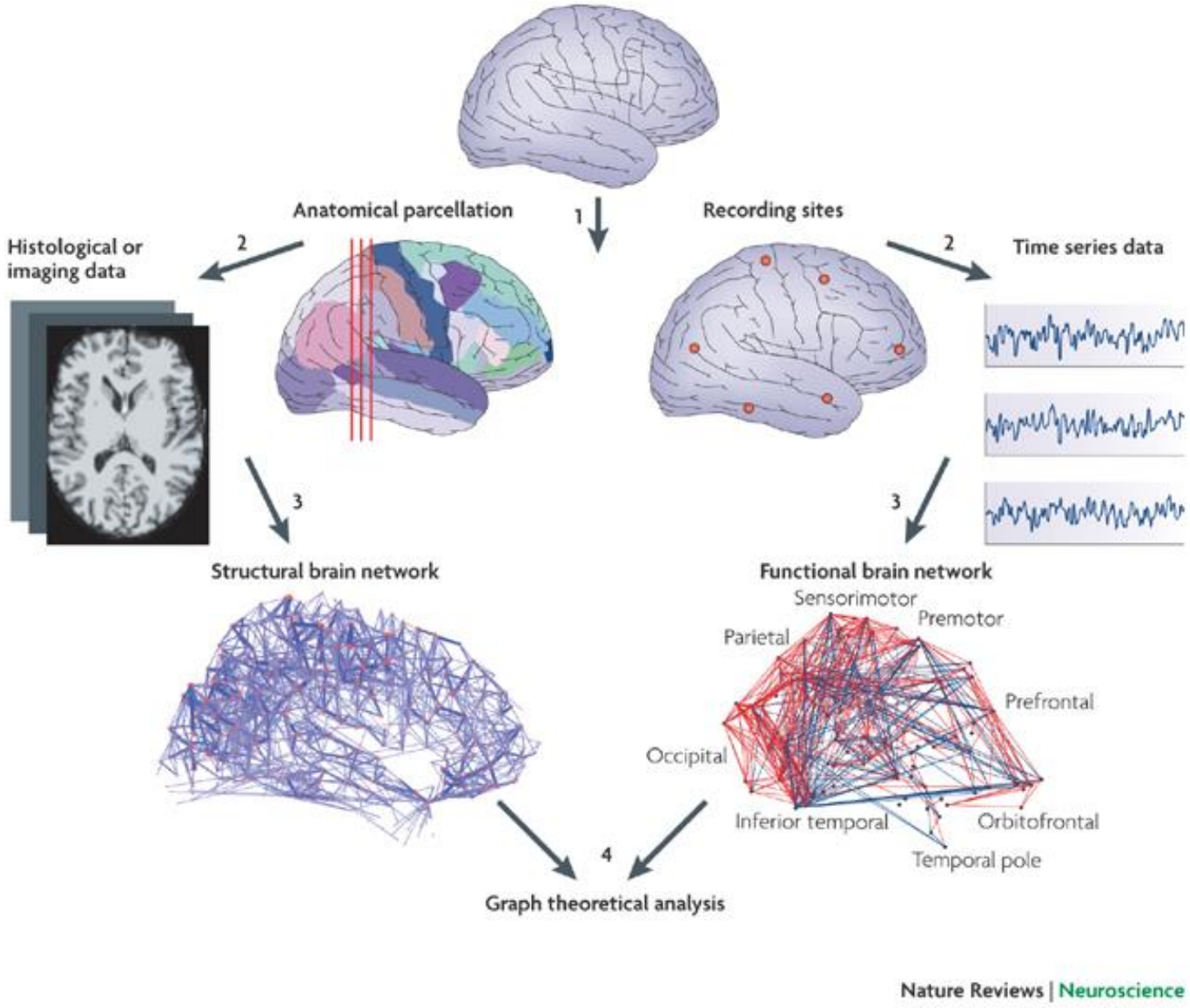
http://paperscape.org/

Business Ecosystems



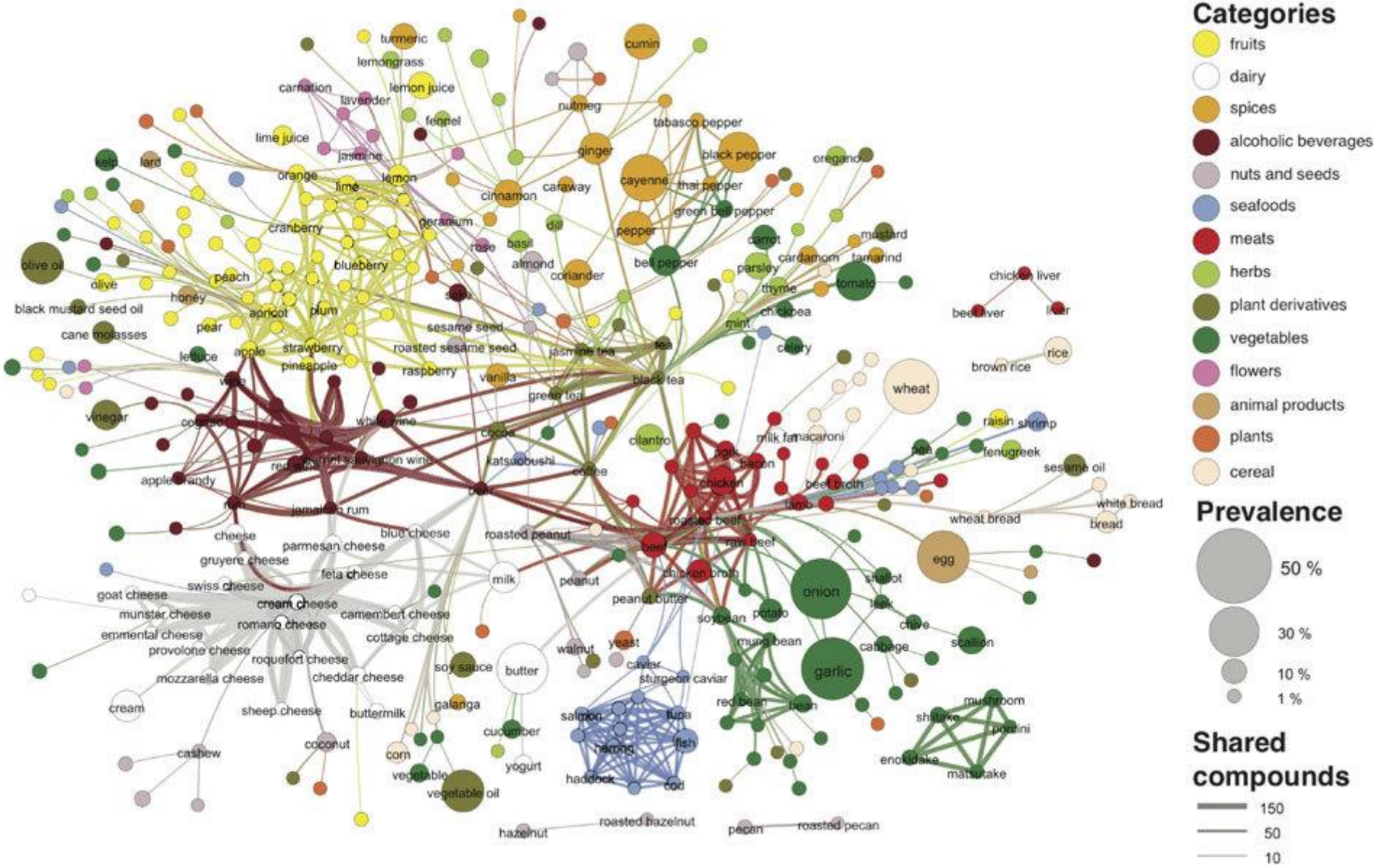
https://hbr.org/2015/02/the-rise-of-social-graphs-for-businesses

Biology



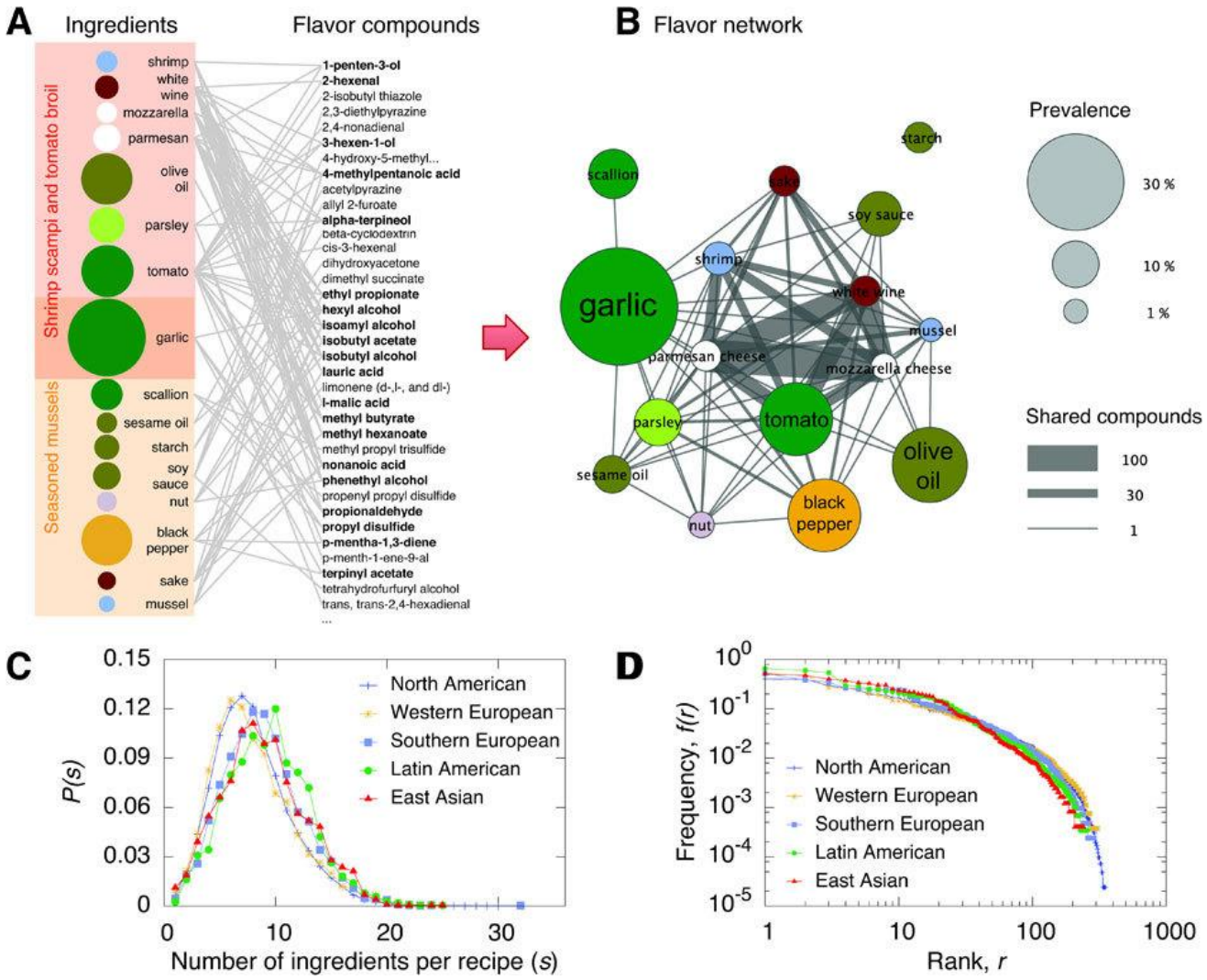
http://www.nature.com/nrn/journal/v10/n3/full/nrn2575.html

Gastronomy (1/3)



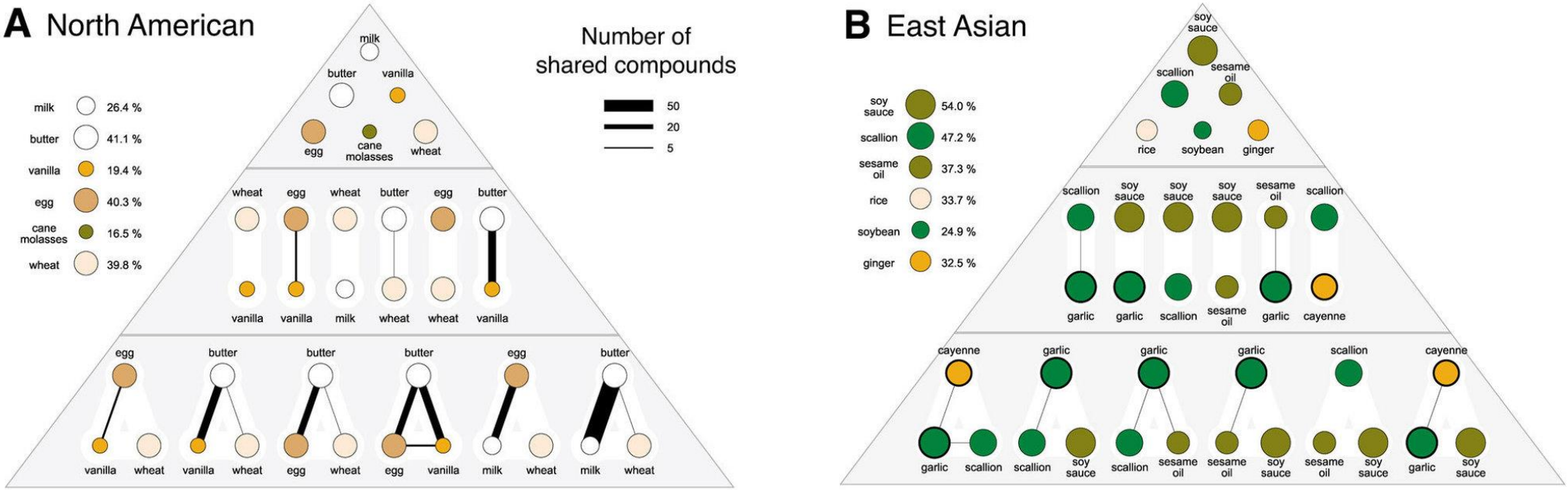
http://www.nature.com/srep/2011/111215/srep00196/full/srep00196.html

Gastronomy (2/3)

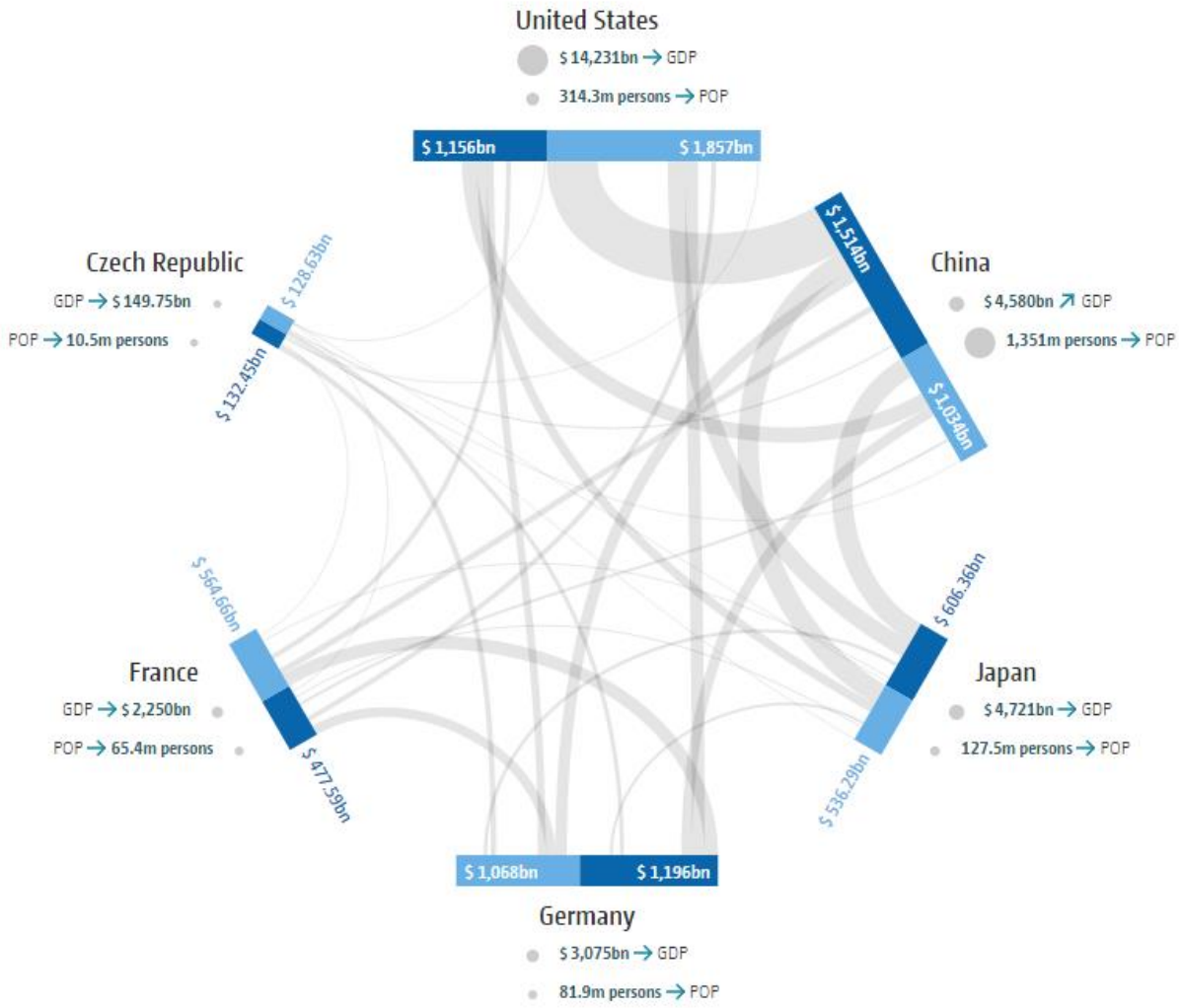


http://www.nature.com/srep/2011/111215/srep00196/full/srep00196.html

Gastronomy (3/3)

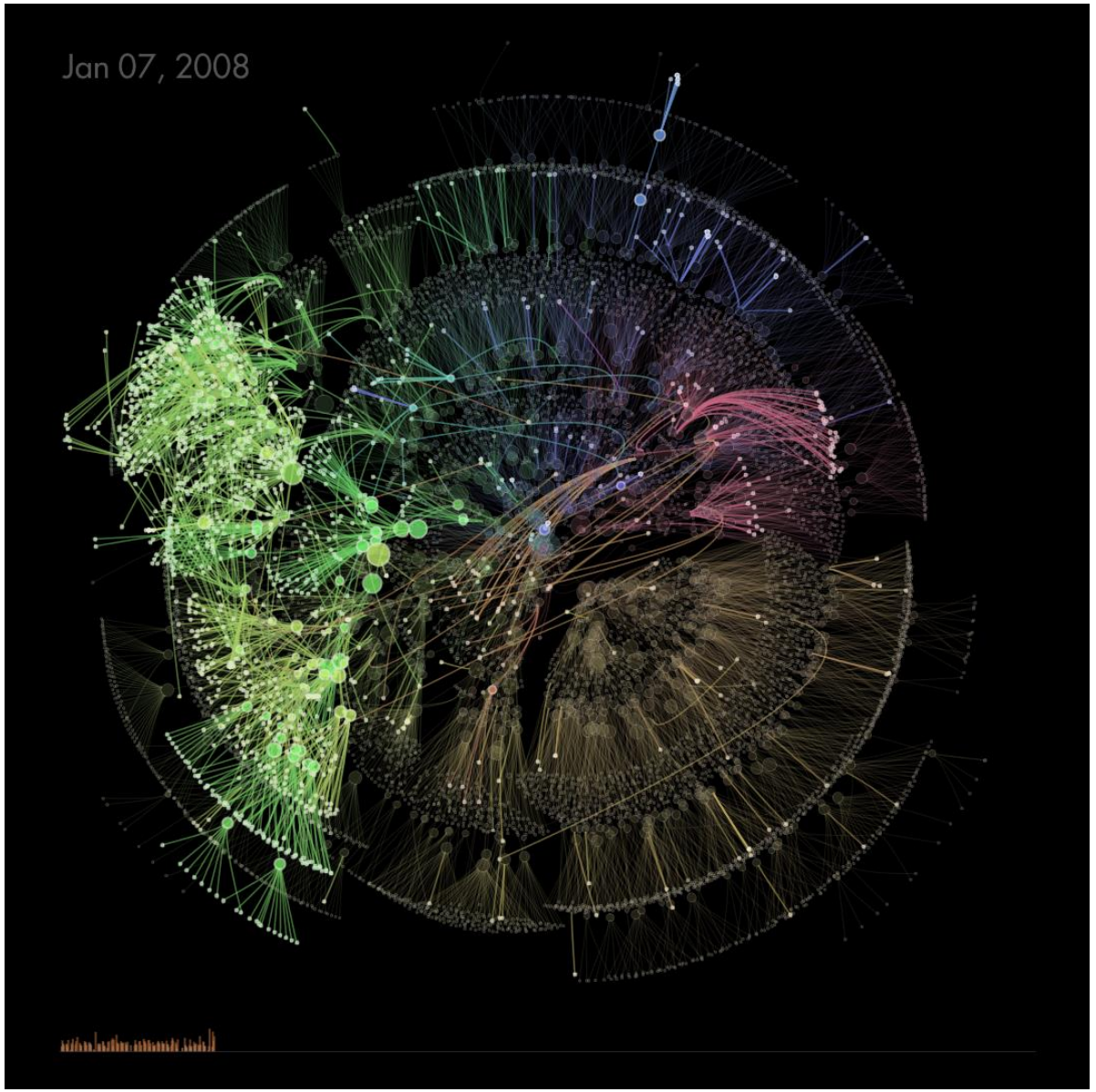


Economics



http://viz.ged-project.de/?lang=en

Dynamics of Organisational Changes



http://www.autodeskresearch.com/projects/orgorgchart

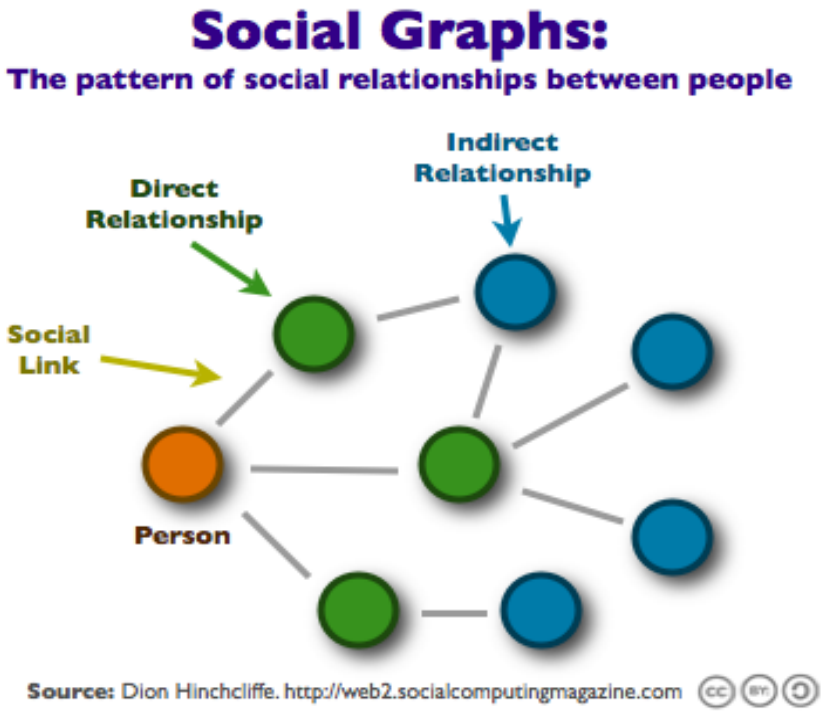
How social networks predict epidemics (Nicholas Christakis)

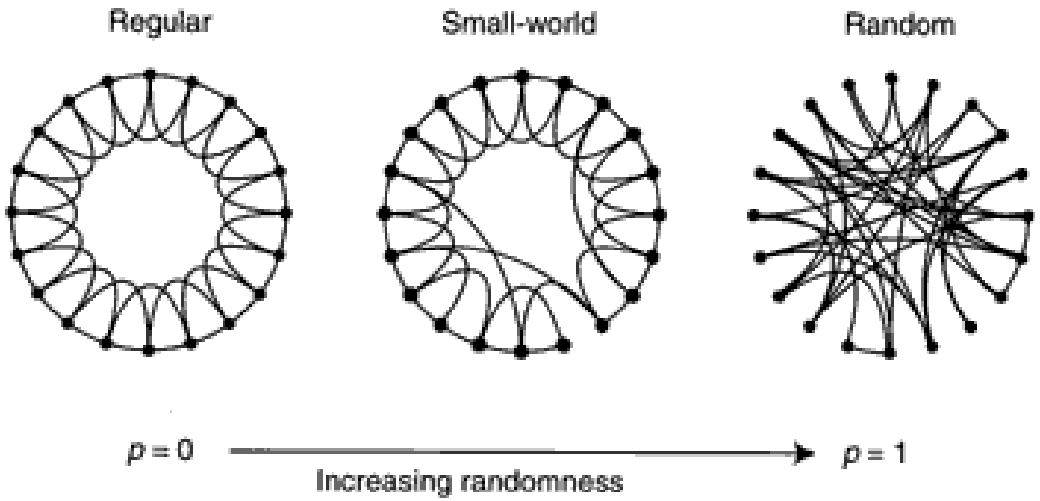


http://www.ted.com/talks/lang/cs/nicholas_christakis_how_social_networks_predict_epidemics.html

Social network (graph)

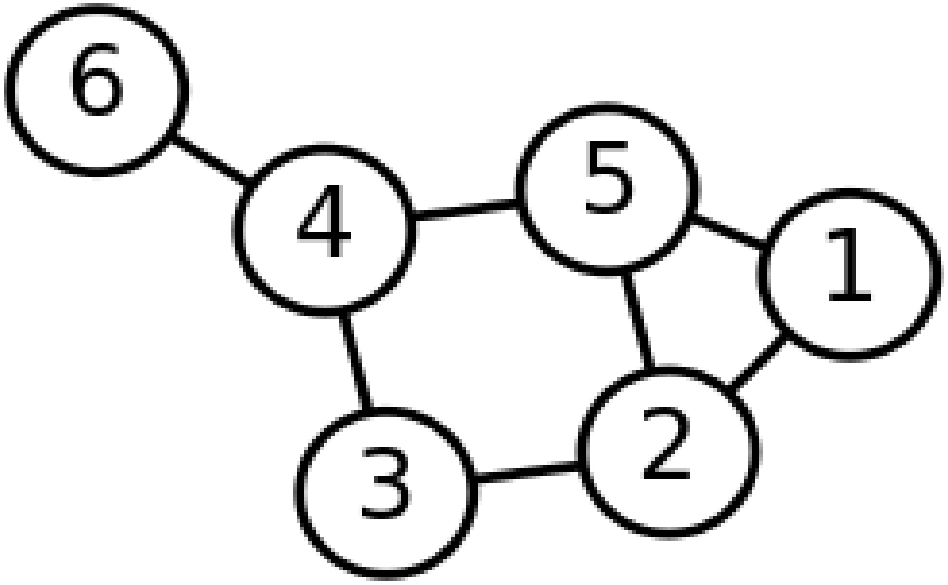
- Set of entities
- Relationship between entities
- Non-randomness
  - random | **complex network**

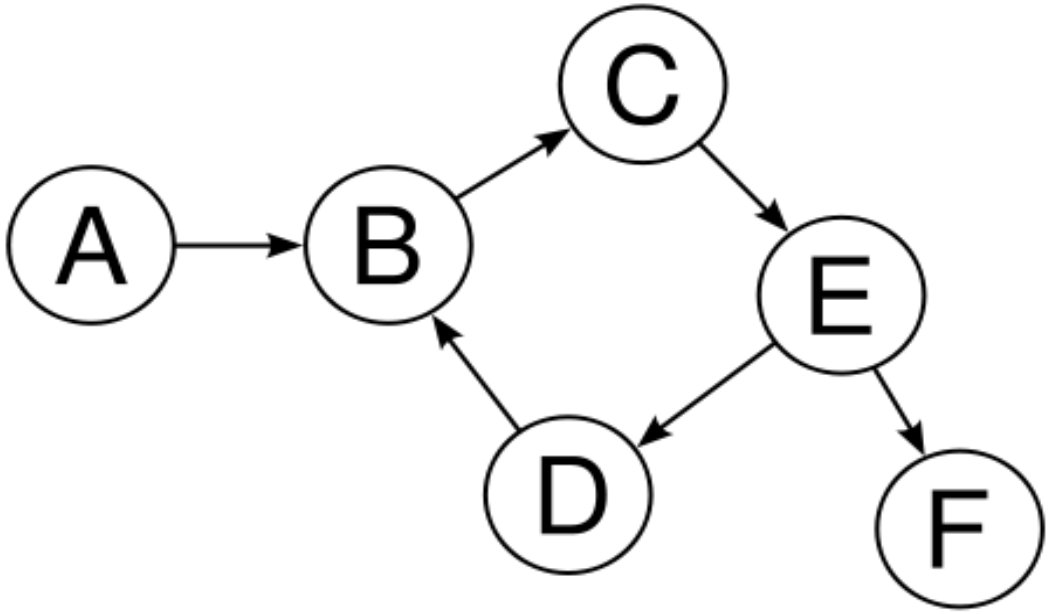




- Graphs
  - undirected × directed
  - weighted × unweighted

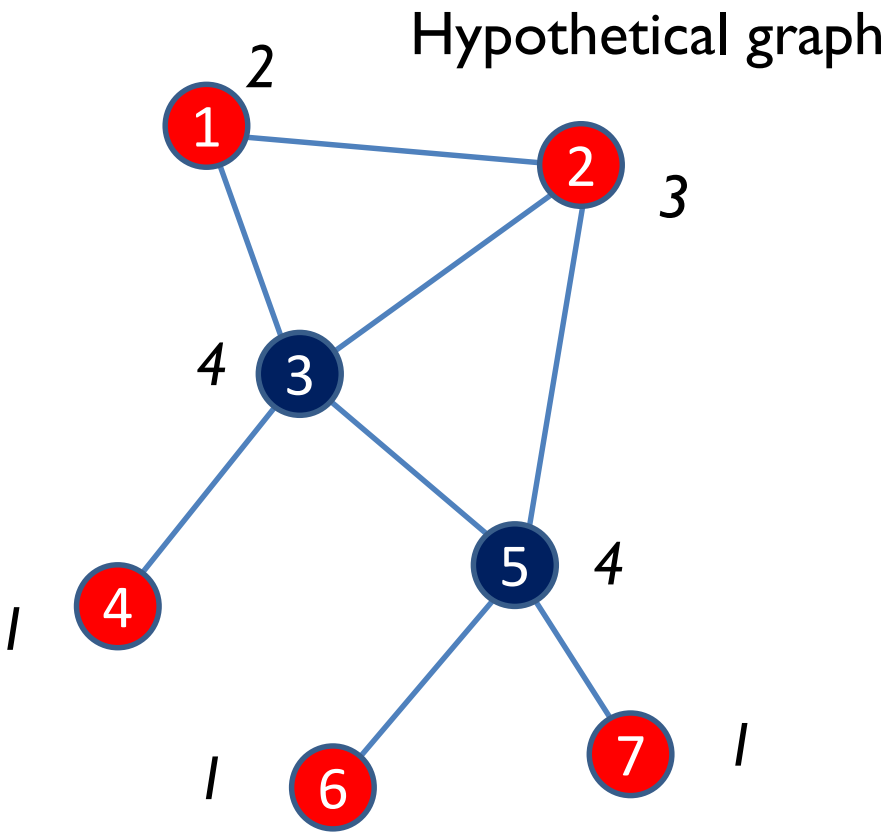
- Nodes (vertices, sites, actors)
- Edges (links, arcs, bonds, ties, relations)





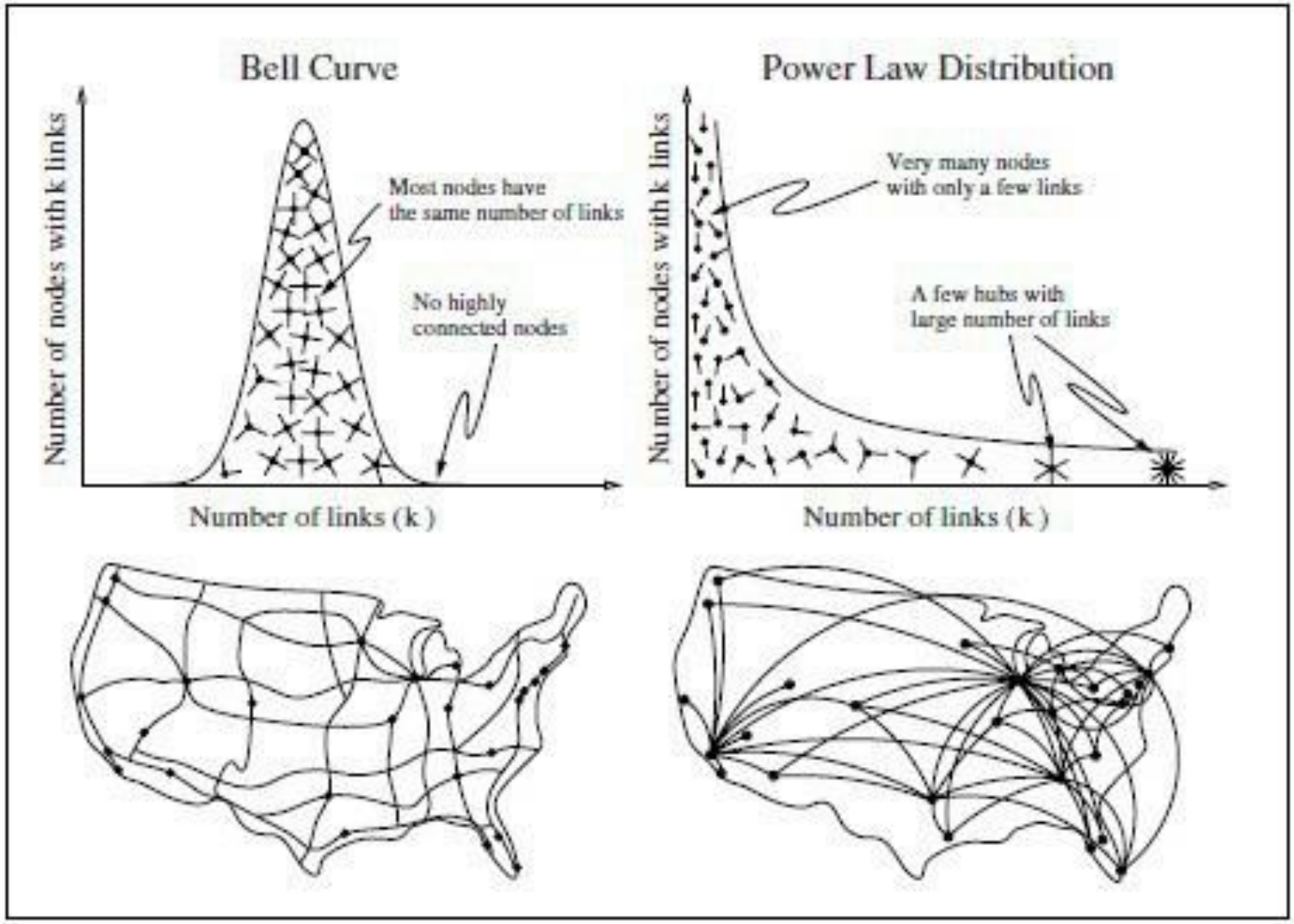
Degree centrality

- A node's (in-) or (out-)degree is the number of links that lead into or out of the node
- **In an undirected graph they are of course identical**
- Often used as measure of a node's degree of connectedness and hence also influence and/or popularity
- Useful in assessing which nodes are central with respect to spreading information and influencing others in their immediate 'neighborhood'



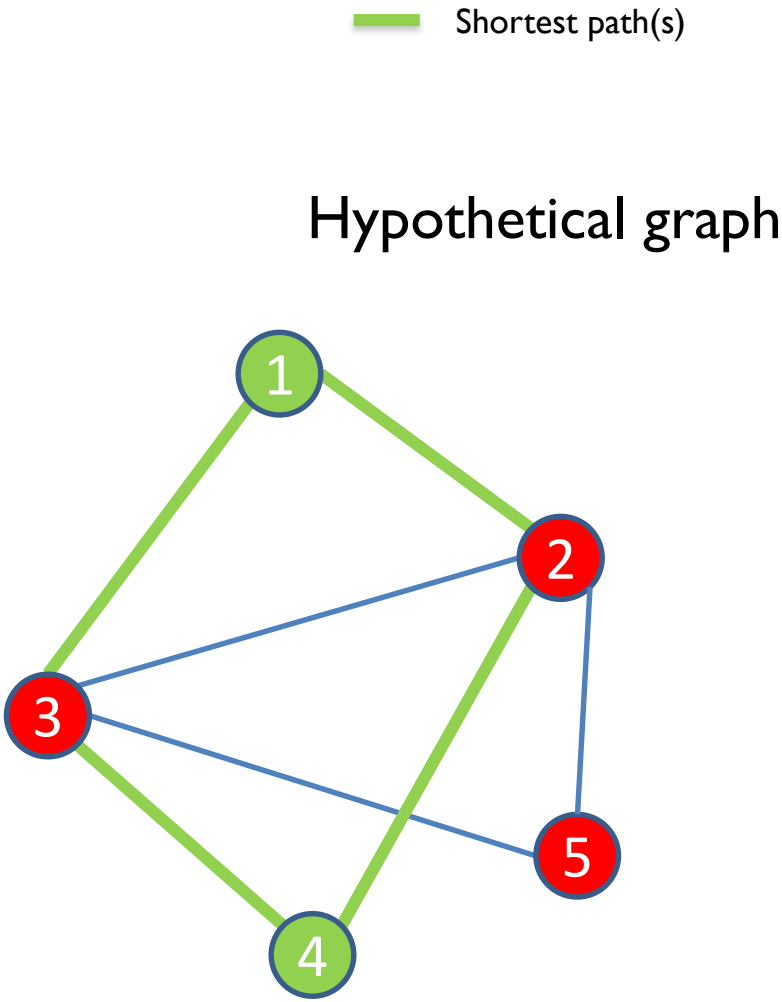
Nodes 3 and 5 have the highest degree (4)

Degree distribution



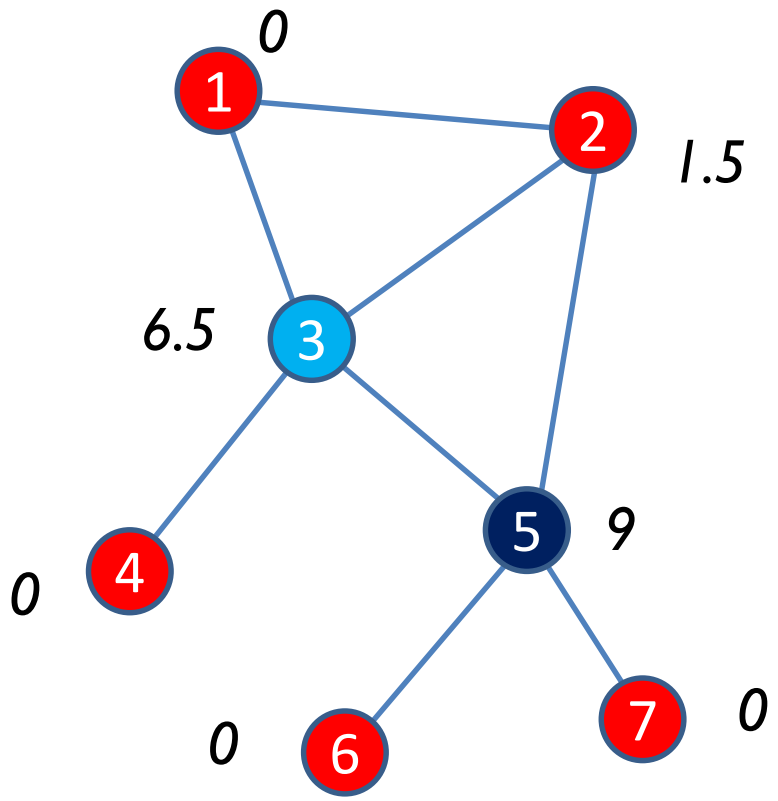
Paths and shortest paths

- A *path* between two nodes is any **sequence of non-repeating nodes** that connects the two nodes
- The *shortest path* between two nodes is the path that connects the two nodes with the shortest number of edges (also called the *distance* between the nodes)
- In the example to the right, between nodes 1 and 4 there are two shortest paths of length 2: {1,2,4} and {1,3,4}
- Other, longer paths between the two nodes are {1,2,3,4}, {1,3,2,4}, {1,2,5,3,4} and {1,3,5,2,4} (the longest paths)
- Shorter paths are desirable when speed of communication or exchange is desired (often the case in many studies, but sometimes not, e.g. in networks that spread disease)



Betweenness centrality

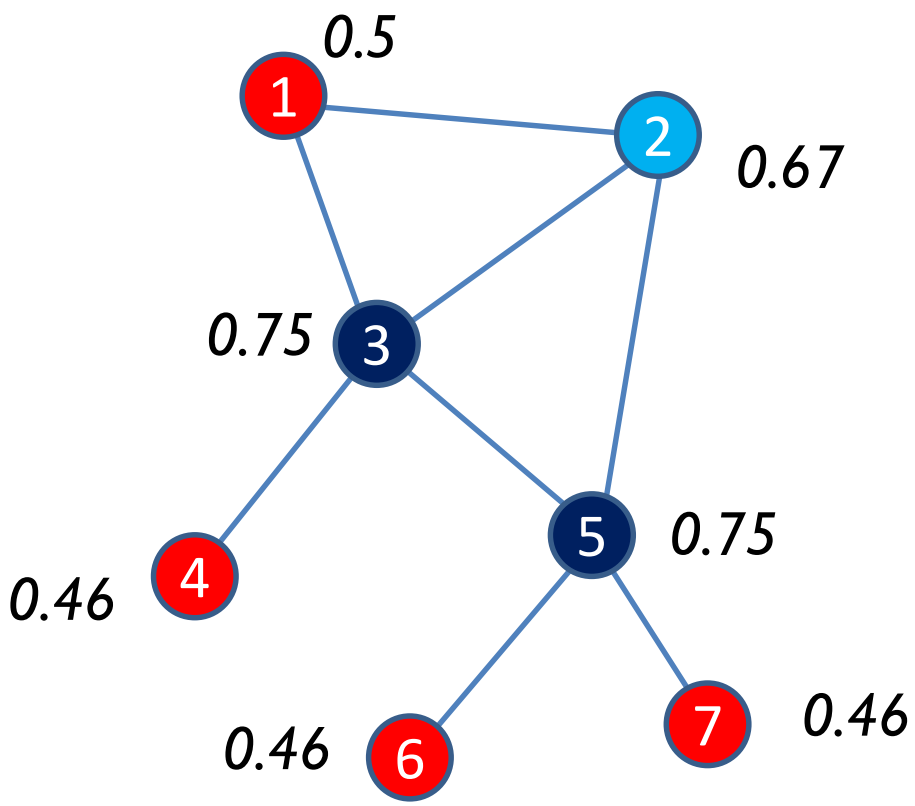
- For a given node *v*, calculate the number of shortest paths between nodes *i* and *j* that pass through v, and divide by all shortest paths between nodes *i* and *j*
- Sum the above values for all node pairs *i,j*
- Sometimes normalized such that the highest value is 1 or that the sum of all betweenness centralities in the network is 1
- Shows which nodes are more likely to be in communication paths between other nodes
- Also useful in determining points where the network would break apart (think who would be cut off if nodes 3 or 5 would disappear)



Node 5 has higher betweenness centrality than 3

Closeness centrality

- Small World Experiment (Stanley Milgram)
- Calculate the mean length of all shortest paths from a node to all other nodes in the network (i.e. how many hops on average it takes to reach every other node)
- Take the reciprocal of the above value so that higher values are 'better' (indicate higher closeness) like in other measures of centrality
- It is a measure of *reach*, i.e. the speed with which information can reach other nodes from a given starting node

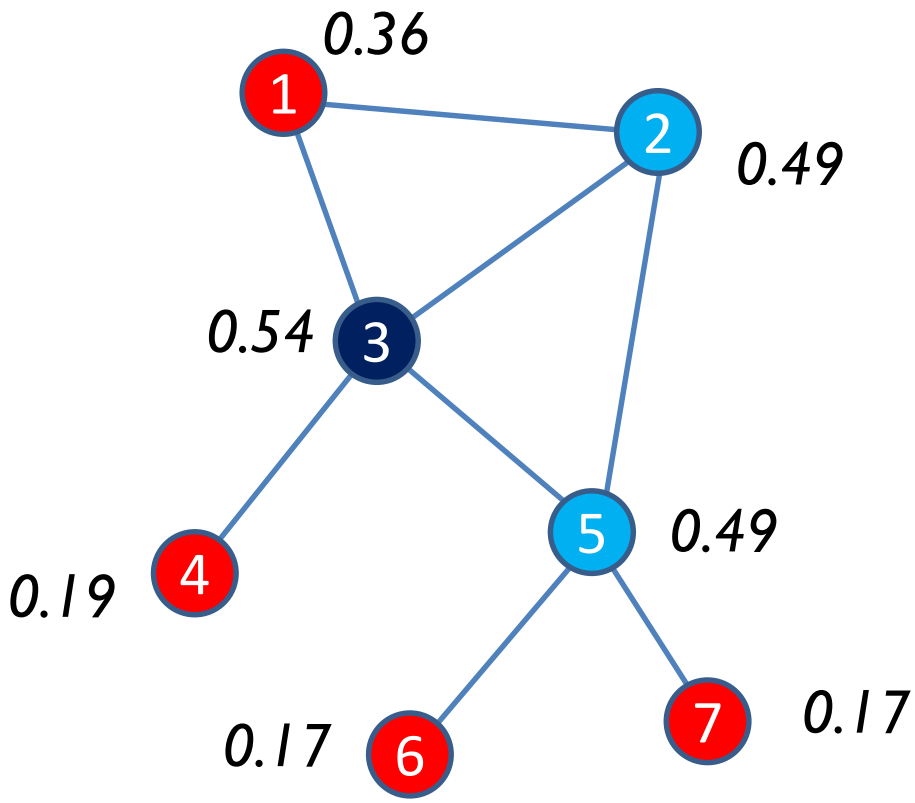


Nodes 3 and 5 have the highest (i.e. best) closeness, while node 2 fares almost as well

Note: Sometimes closeness is calculated without taking the reciprocal of the mean shortest path length. Then lower values are 'better'.

Eigenvector centrality

- A node's eigenvector centrality is proportional to the sum of the eigenvector centralities of all nodes directly connected to it
- In other words, a node with a high eigenvector centrality is connected to other nodes with high eigenvector centrality
- This is similar to how Google ranks web pages: links from highly linked-to pages count more
- Useful in determining who is connected to the most connected nodes



Node 3 has the highest eigenvector centrality, closely followed by 2 and 5

Interpretation of measures

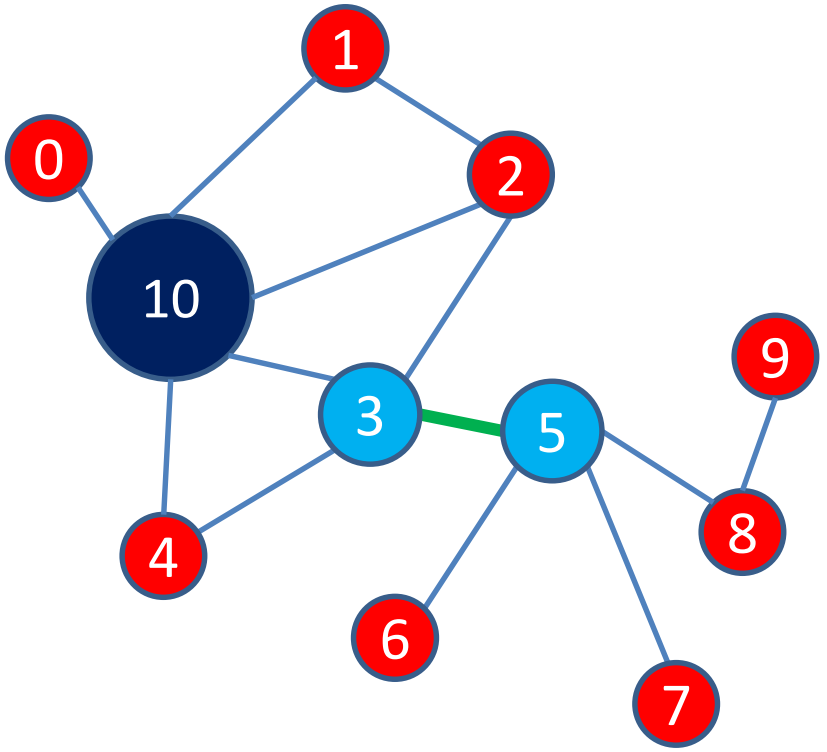
Centrality measure	Interpretation in social networks
Degree	How many people can this person reach directly?
Betweenness	How likely is this person to be the most direct route between two people in the network?
Closeness	How fast can this person reach everyone in the network?
Eigenvector	How well is this person connected to other well-connected people?

Interpretation application of measures

Centrality measure	Other possible interpretations…
Degree	In network of music collaborations: how many people has this person collaborated with?
Betweenness	In network of spies: who is the spy though whom most of the confidential information is likely to flow?
Closeness	In network of sexual relations: how fast will an STD spread from this person to the rest of the network?
Eigenvector	In network of paper citations: who is the author that is most cited by other well-cited authors?

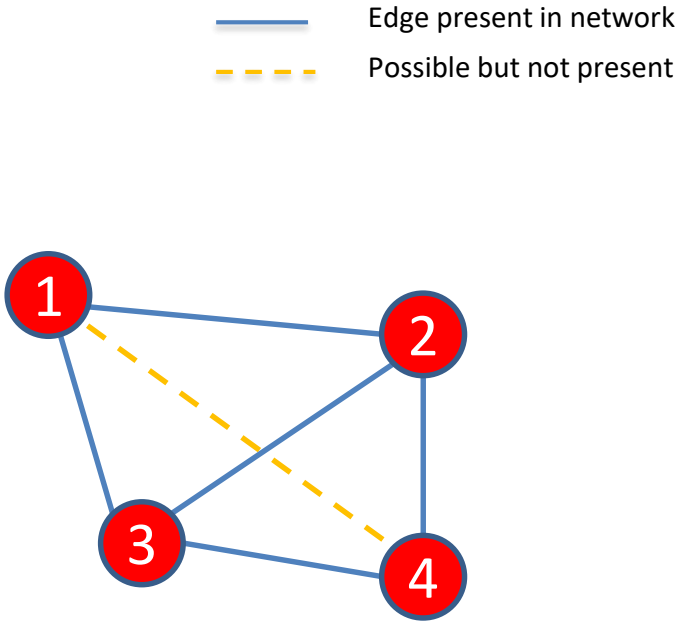
Identifying sets of key players

- In the network to the right, node 10 is the most central according to degree centrality
- But nodes 3 and 5 together will reach more nodes
- Moreover the tie between them is critical; if severed, the network will break into two isolated sub-networks
- It follows that other things being equal, players 3 and 5 together are more ‘key’ to this network than 10
- Thinking about sets of key players is helpful!

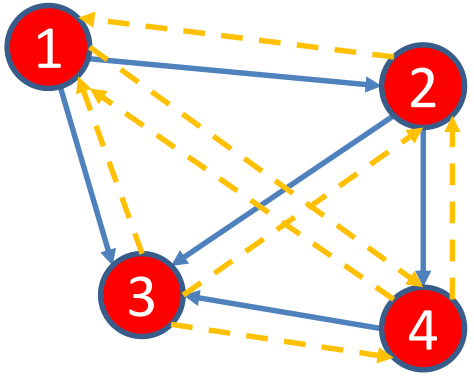


Density

- A network's *density* is the ratio of the number of edges in the network over the total number of possible edges between all pairs of nodes (which is *n(n-1)/2*, where *n* is the number of vertices, for an undirected graph)
- In the example network to the right density=5/6=0.83 (i.e. it is a fairly *dense* network; opposite would be a *sparse* network)
- It is a common measure of how well connected a network is (in other words, how closely knit it is) – a perfectly connected network is called a *clique* and has density=1
- A directed graph will have half the density of its undirected equivalent, because there are twice as many possible edges, i.e. *n(n-1)*
- Density is useful in comparing networks against each other, or in doing the same for different regions within a single network



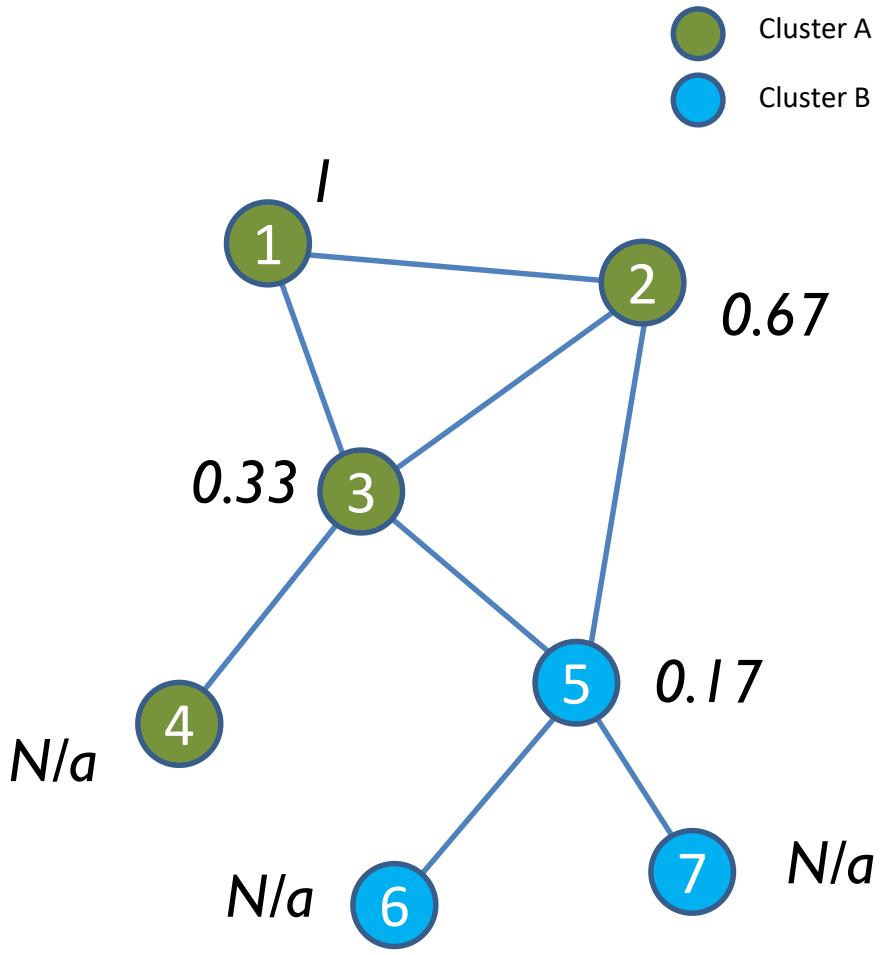
density = 5/6 = 0.83



density = 5/12 = 0.42

Clustering

- A node's *clustering coefficient* is the number of closed triplets in the node's neighborhood over the total number of triplets in the neighborhood. It is also known as *transitivity*.
- E.g., node 1 to the right has a value of 1 because it is only connected to 2 and 3, and these nodes are also connected to one another (i.e. the only triplet in the neighborhood of 1 is closed). We say that nodes 1,2, and 3 form a *clique*.
- Clustering algorithms identify clusters or 'communities' within networks based on network structure and specific clustering criteria (example shown to the right with two clusters is based on *edge betweenness*, an equivalent for edges of the betweenness centrality presented earlier for nodes)



Network clustering coefficient = 0.375
(3 nodes in each triangle x 2 triangles = 6 closed triplets divided by 16 total)

Complex Networks Layouting

- „Hairball“ problem



- „force-directed“ layouting (http://en.wikipedia.org/wiki/Force-directed_graph_drawing)
  - Attraction (drawing nodes closer together based on their similarity or relatedness)
  - Repulsion (unrelated or distantly related nodes further apart from one another)
  - Gravity (how nodes are drawn relative to the center of the graph)

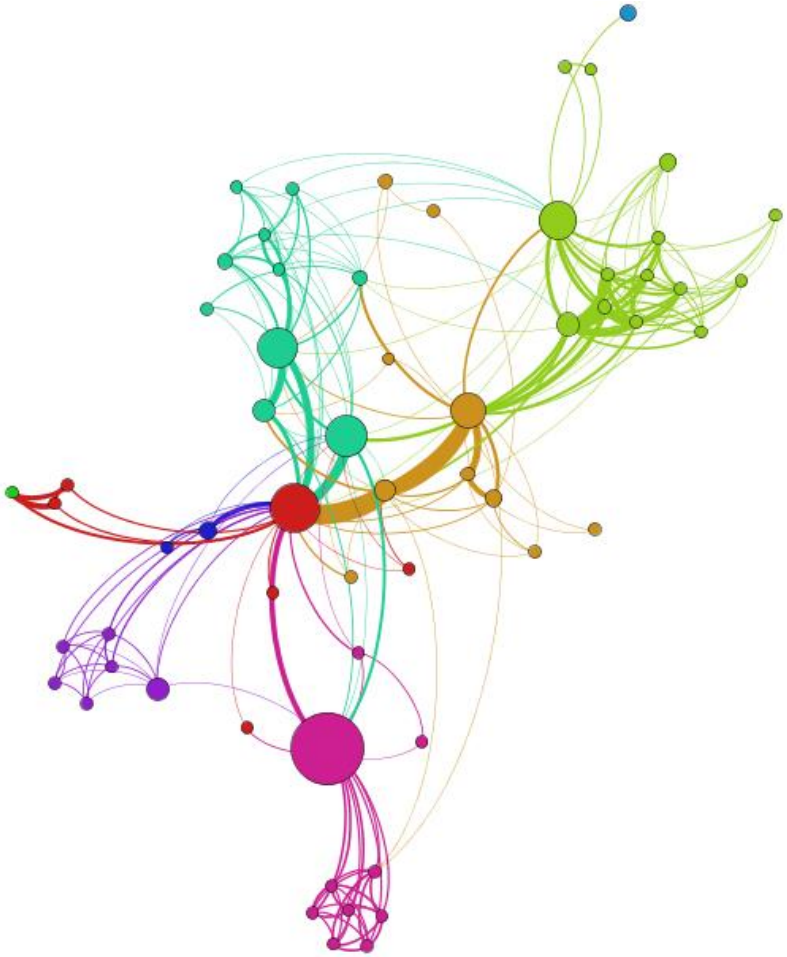
  - ARF, Force Atlas, Force Atlas 2, Fruchterman-Reingold, OpenOrd, Yifan Hu, …

- tree, circular, radial, geographic, layered, … layouting

- Gephi tutorial
  - http://www.slideshare.net/gephi/gephi-tutorial-layouts

Social Network Analysis Software

- CFinder
- **Gephi**
- NetMiner
- NetworkX
- NodeXL
- R
- ScioMetrica
- Tulip
- …

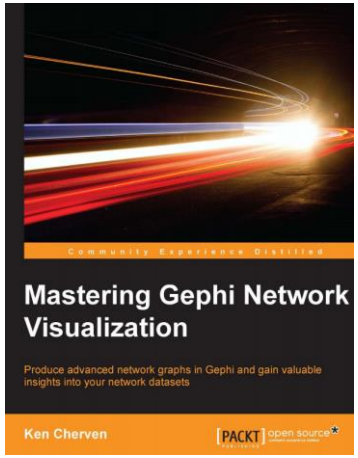


- Data repositories:
  - https://github.com/gephi/gephi/wiki/datasets
  - http://www-personal.umich.edu/~mejn/netdata/
  - http://snap.stanford.edu/data/
  - http://networkrepository.com/

Want to study more?

- (MOOC – Massive Open Online Course) Coursera.org
  - https://www.coursera.org/course/sna

- Ken Cherven – Mastering Gephi Network Visualization
  - http://www.amazon.com/Mastering-Gephi-Network-Visualization-Cherven/dp/1783987340



- Introduction to social network methods
  - http://faculty.ucr.edu/~hanneman/nettext/
- Social Network Analysis Instructional Web Site
  - http://www.analytictech.com/networks/
- Networks, Crowds, and Markets: Reasoning About a Highly Connected World
  - http://www.cs.cornell.edu/home/kleinber/networks-book/

- **Journal of Social Structure**
  - http://www.cmu.edu/joss/content/articles/volindex.html