

Lecture 6: Introduction to Networks

Complex Systems 530

Outline

- ABM & networks
- Types of dynamic models using networks
- Basic terminology
- Network metrics
- Random networks

Networks

- Very flexible! Can capture many kinds of relationships, from concrete to abstract
- Network theory (graph theory) has a long history in math & computer science literature
- Many models can be written or thought of as a network & this perspective can often help understand the model (e.g. there is a whole huge theory just on networks of ODEs)

Networks

- Links between webpages, twitter followers, facebook friends, common use of hashtags/interests/etc.
- Family trees, friendship networks, contact networks, collaboration networks (mention Erdős & Bacon Numbers)
- Food webs among species, gene regulatory networks
- Diplomatic relationships, financial relationships
- Concept maps, causal diagrams, language/text, etc!

However...

- Just because you can think of a network representation of a system does not make it a meaningful representation of that system
- Need to consider what the network perspective gains you & how it can be useful

Networks & ABM

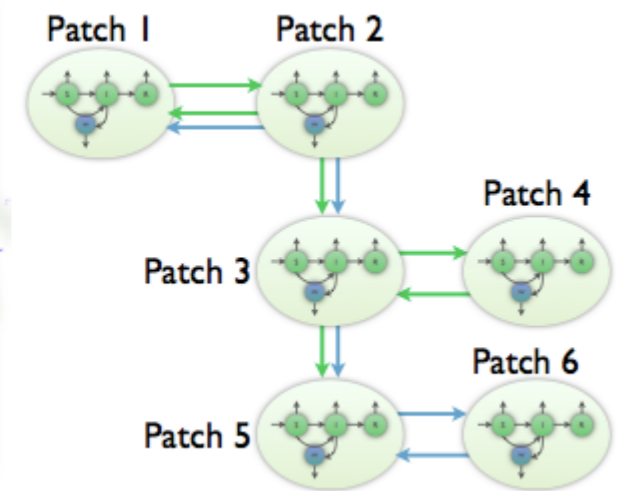
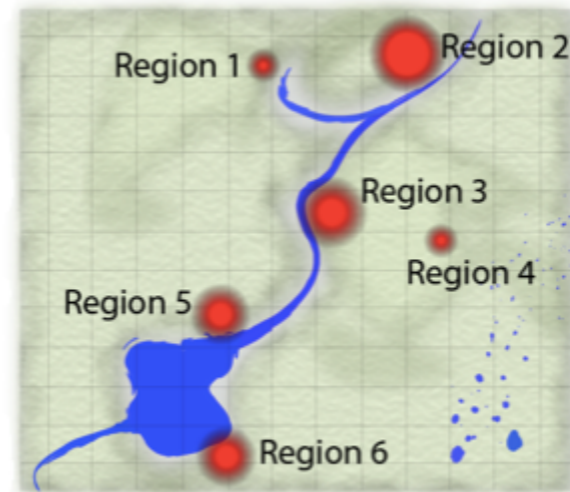
- Networks allow us to examine non-homogeneous interaction structures in emergent processes
 - Compare to grids or random/complete mixing (also types of networks!)
- Understanding of networks often proves to be important in computational model of complex systems
- Many network models are ABM, and many other kinds of models can be cast as network models too

Types of network dynamics

- **Dynamics on networks:** models where the processes of interest occur over a fixed network structure
- **Dynamics of networks:** models of the dynamic changes over time of the network topology itself
- **Adaptive networks:** models looking at the interplay of the two (both the processes on the network, and how the network changes)

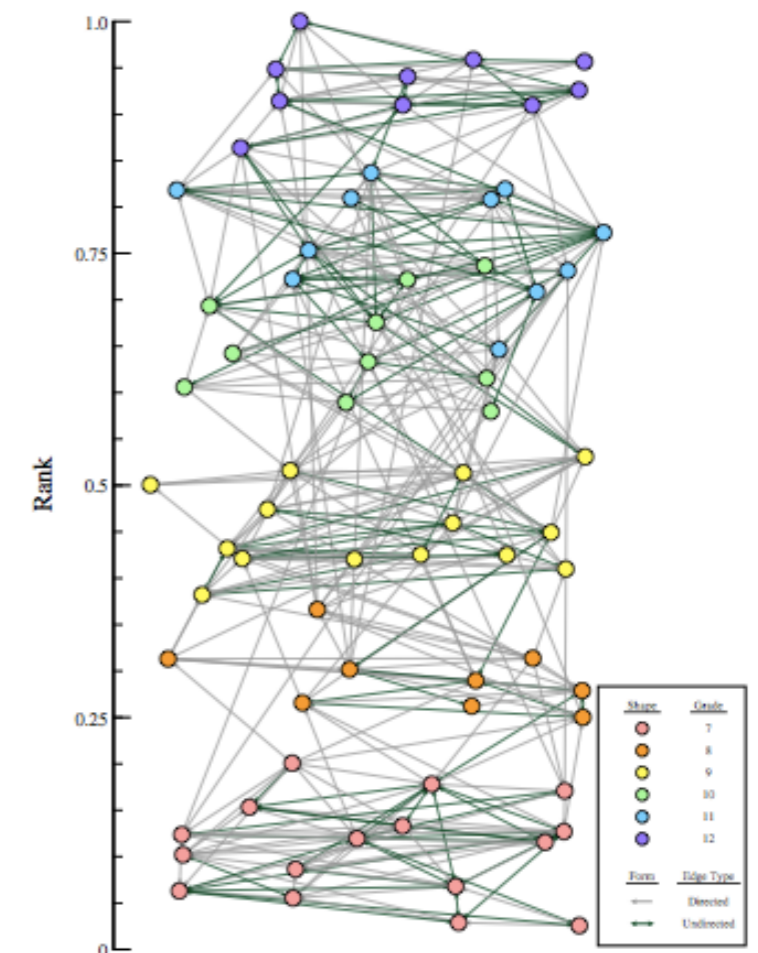
Networks

- Network (graph) = nodes & edges
- Node (vertex) - an object, can be people, communities, locations, water sources, signaling molecules, genes, etc.
- Edge - a connection between two nodes



Types of Networks

- **Directed graph** - edges have a direction associated with them (e.g. friendships that go one way)
- Edges sometimes called arcs
- E.g. friendship networks & social status (Newman & Ball)
- Disease Transmission

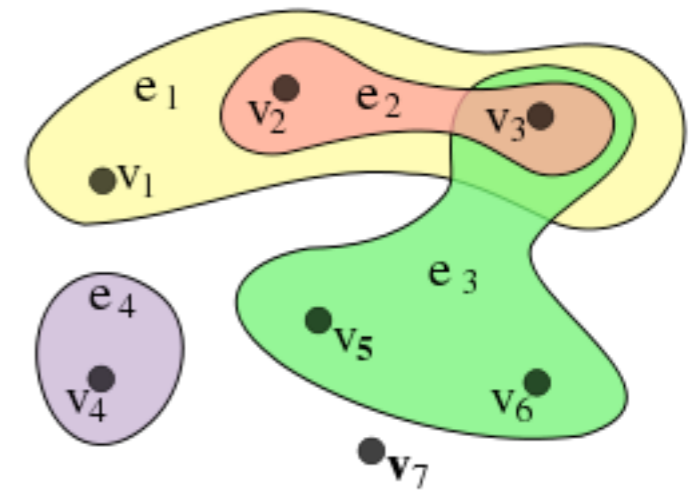


Types of Networks

- **Weighted graph** - assigns a number (weight) to each edge/node
 - E.g. association strength, parameter value, disease status
 - Weighting can also be thought of as a type or state instead of number (e.g. S, I, R, or cancer stage, etc.)
 - One of the most common for modeling
 - Can have weighted edges, nodes, or both

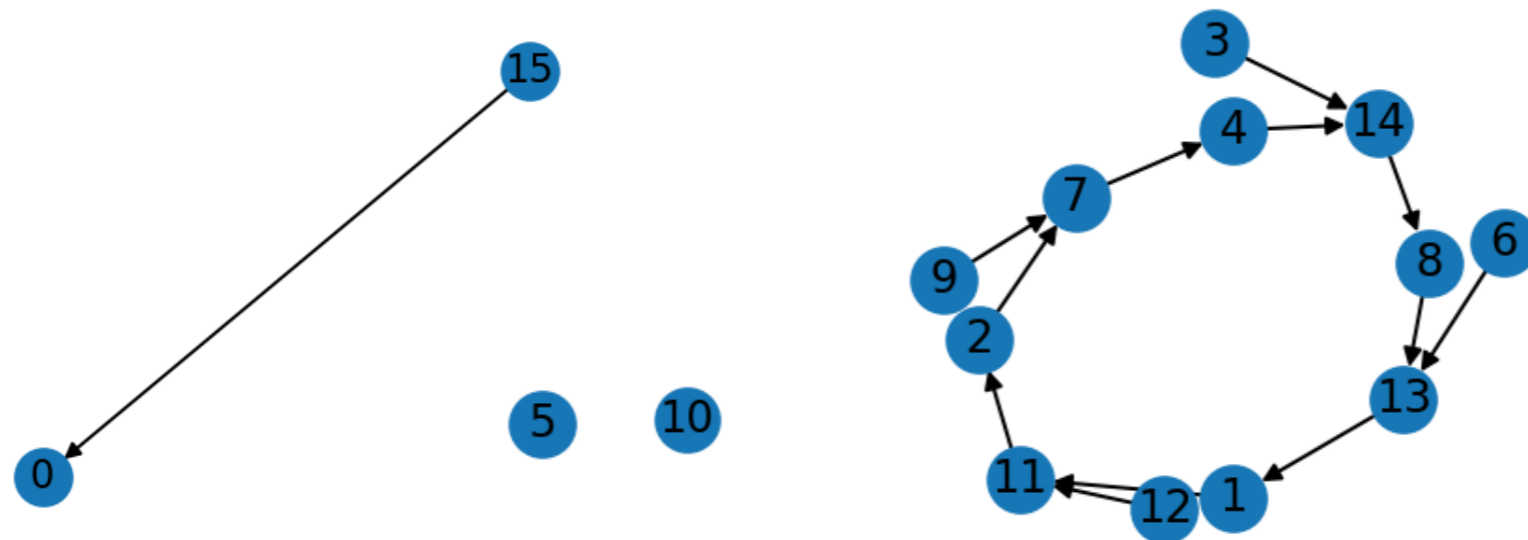
Types of Networks

- **Multigraph** - multiple edges allowed between nodes
- **Hypergraph** - edges can have more than two vertices attached



Key definitions/vocab

- Graphs can be connected or disconnected
- **Connected graph** - Graph in which every node is “reachable” from every other
- **Connected component** - Subgraph that is connected w/in itself but not the rest of the graph



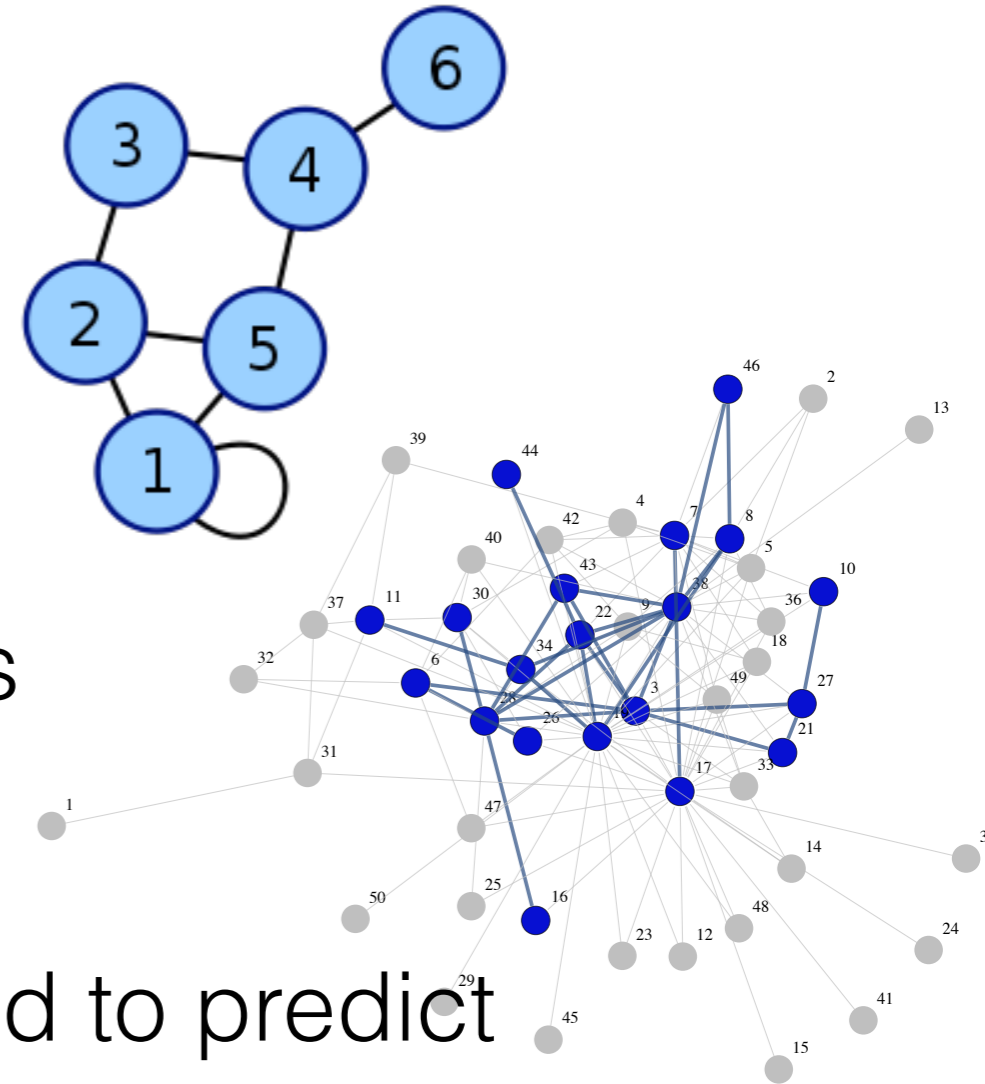
Key definitions/vocab

- **Loop** - edge connected to same vertex at both ends

- **Subgraph** - a subset of a graph

- **Neighborhood** of node x - nodes that are adjacent to x

- Often want to use neighborhood to predict effects on individual, e.g. infectious disease, behavioral influence

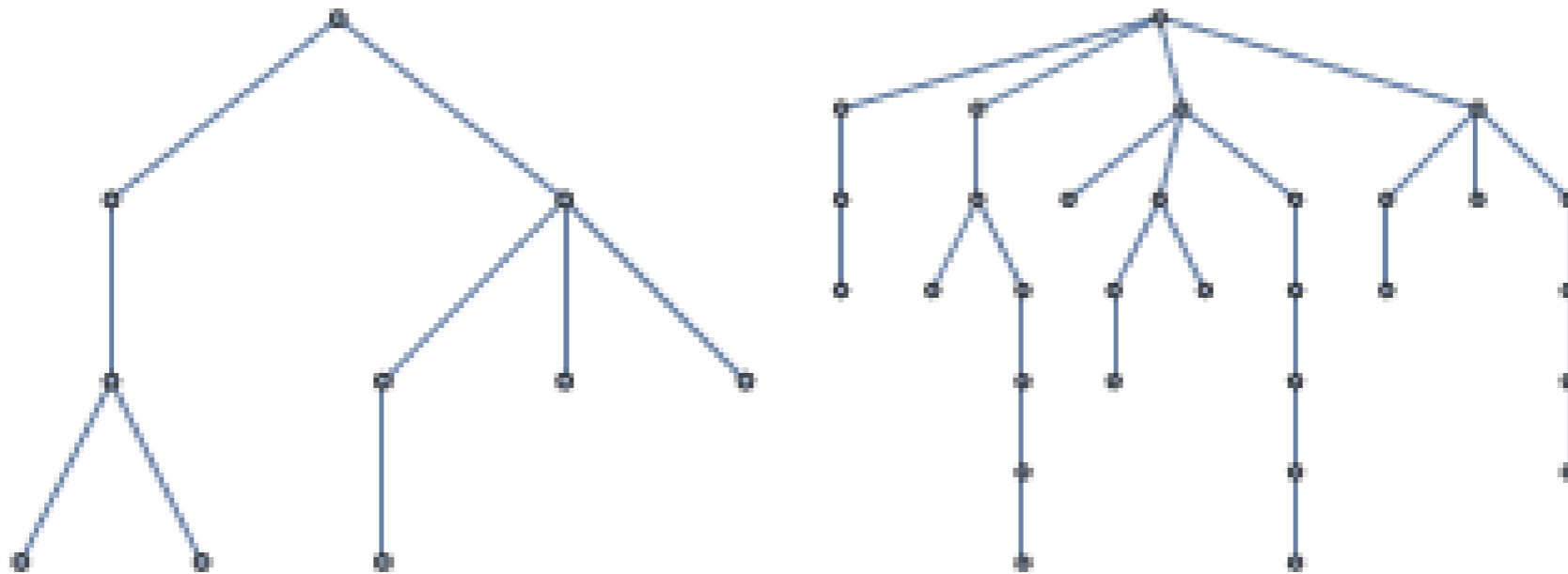


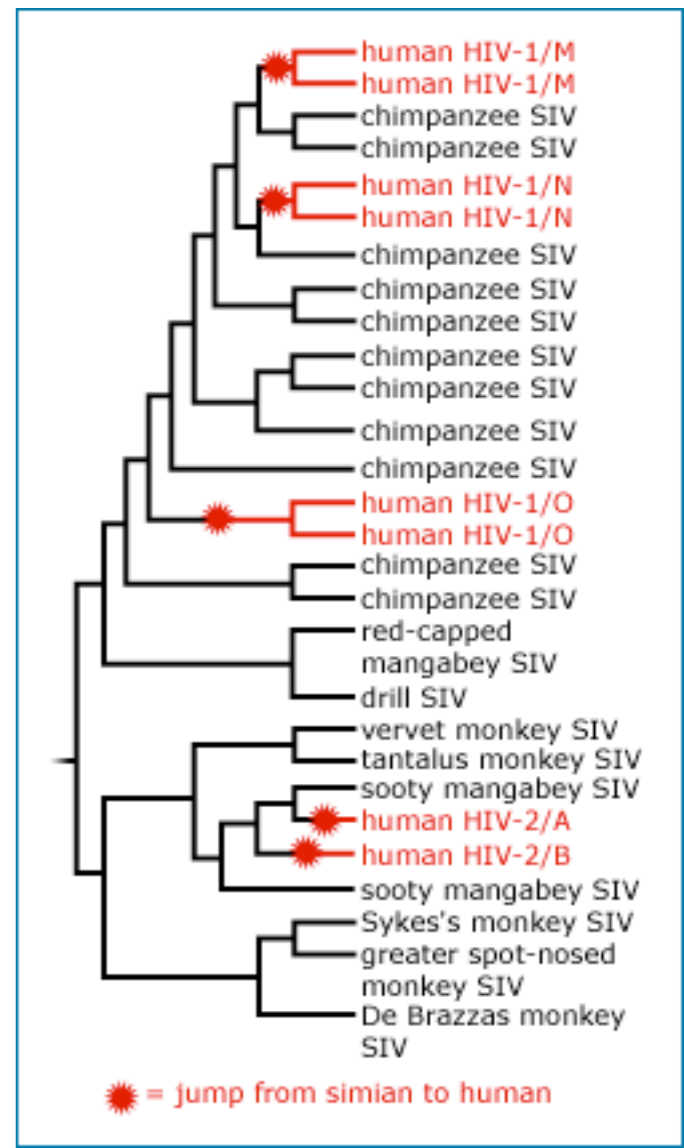
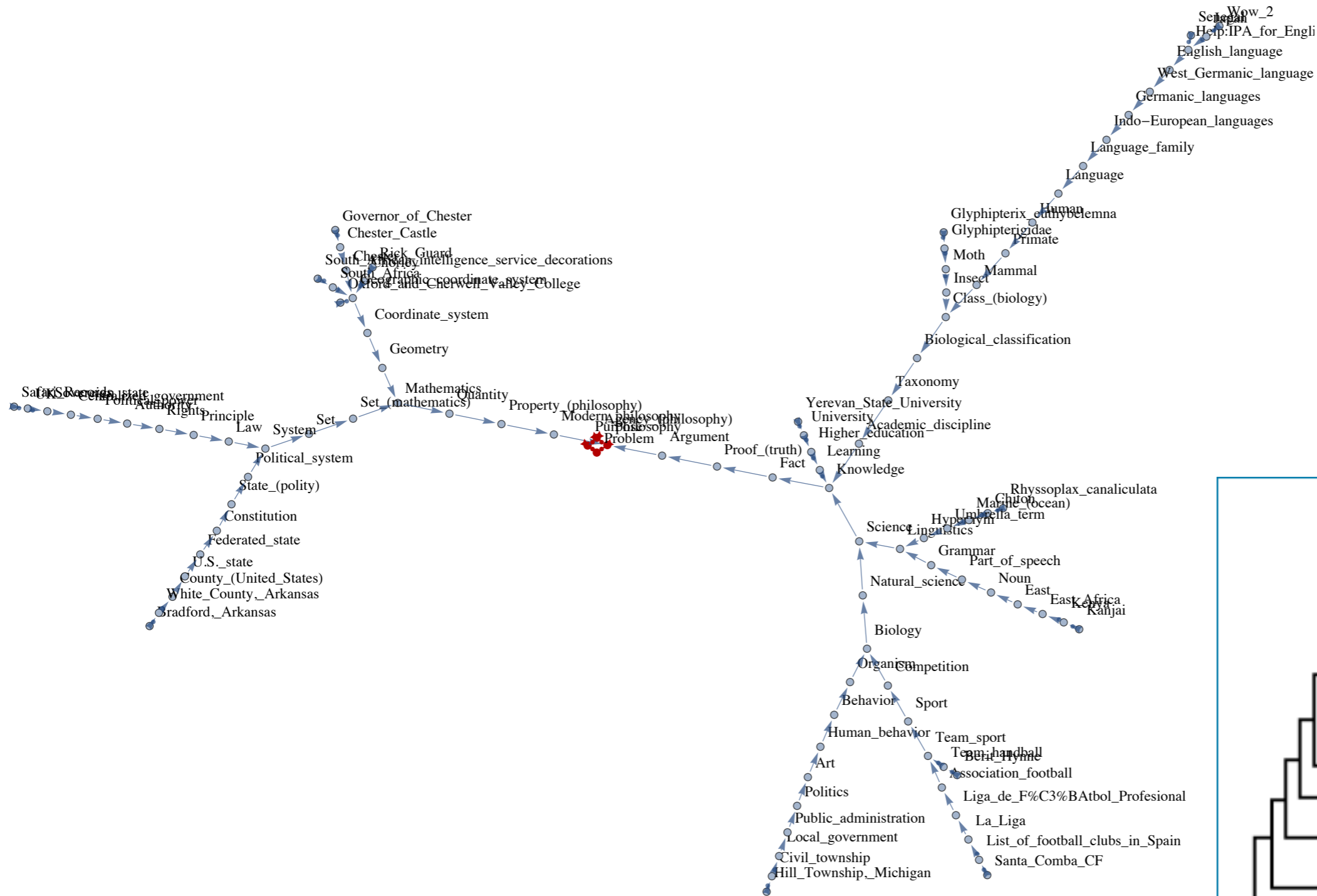
Key definitions/vocab

- **Walk** - List of edges sequentially connected to form a continuous route
- **Path** - Walk that doesn't visit any node twice
- **Cycle** (sometimes called a **circuit**) - Walk that starts and ends at same node (called a **simple cycle** if no repeated nodes)

Trees & Forests

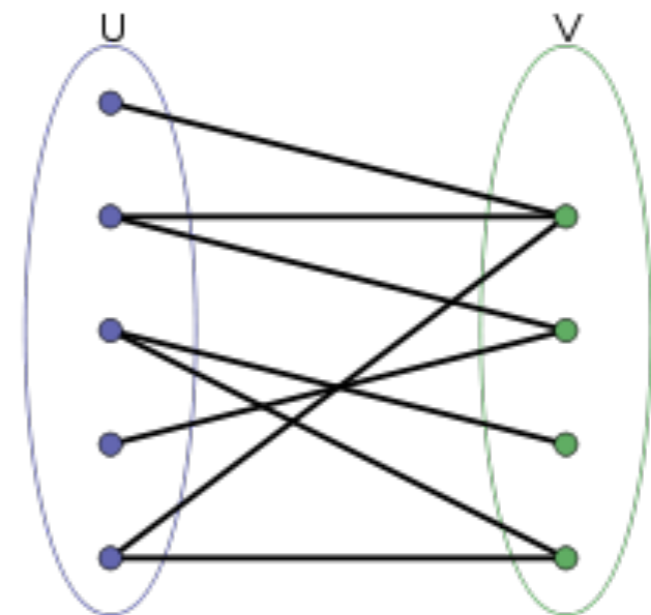
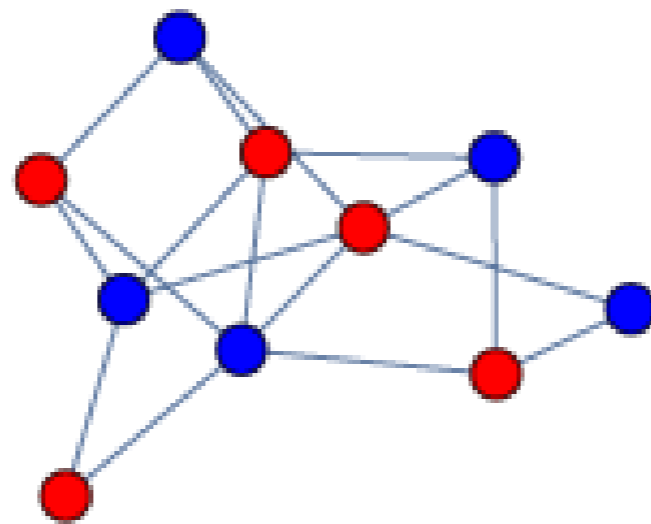
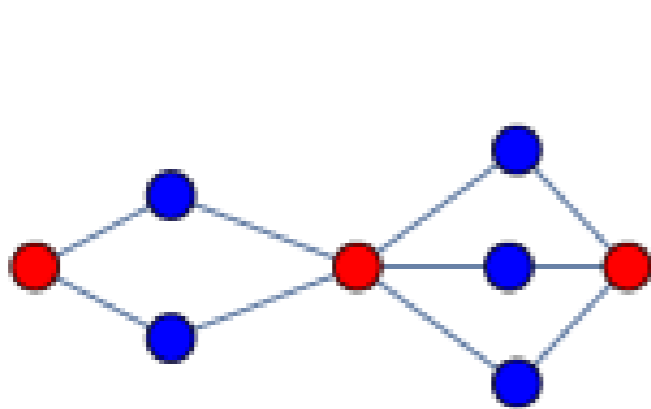
- **Tree** - connected graph with no cycles
- **Forest** - multiple trees





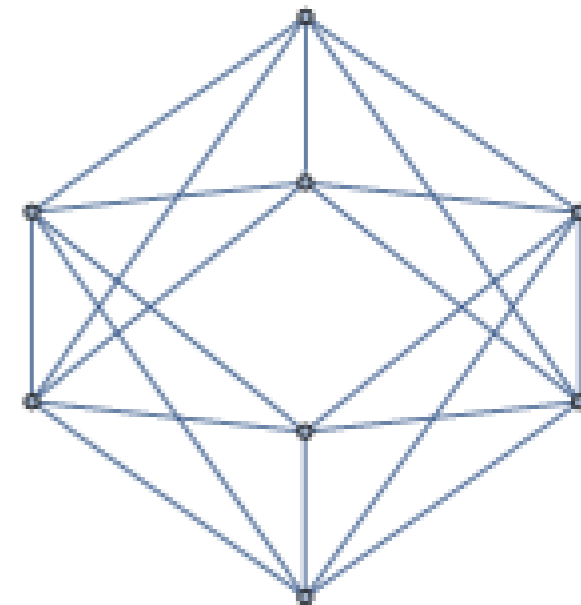
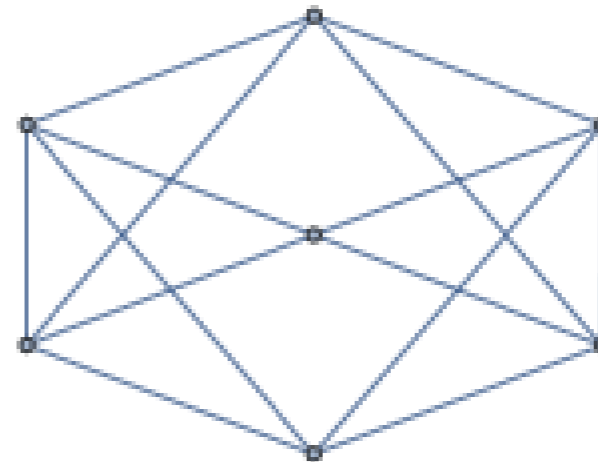
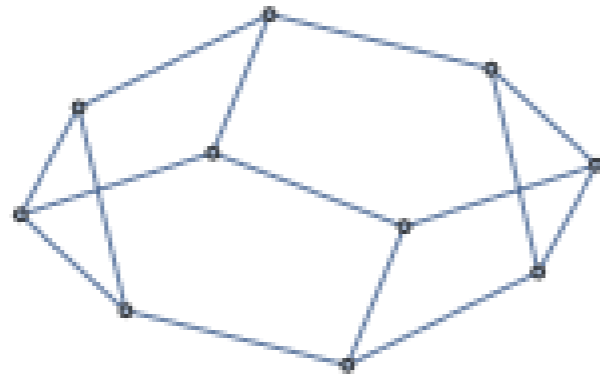
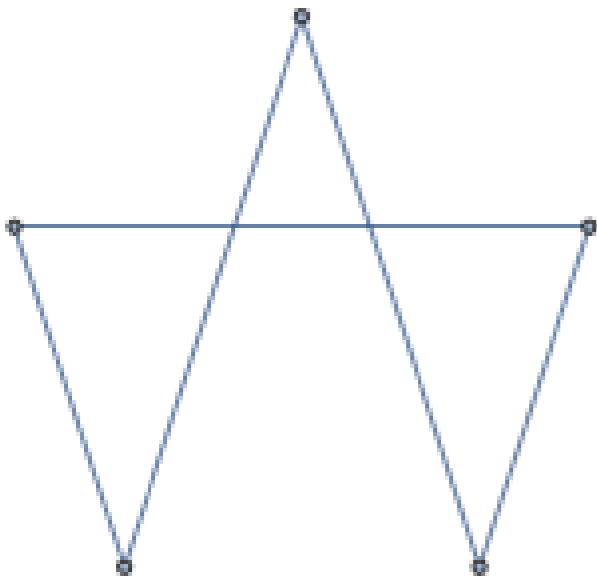
Bipartite (n-partite) graph

- Can be partitioned into two (n) groups, with edges only between the two groups, not within them
- E.g. heterosexual sexual network



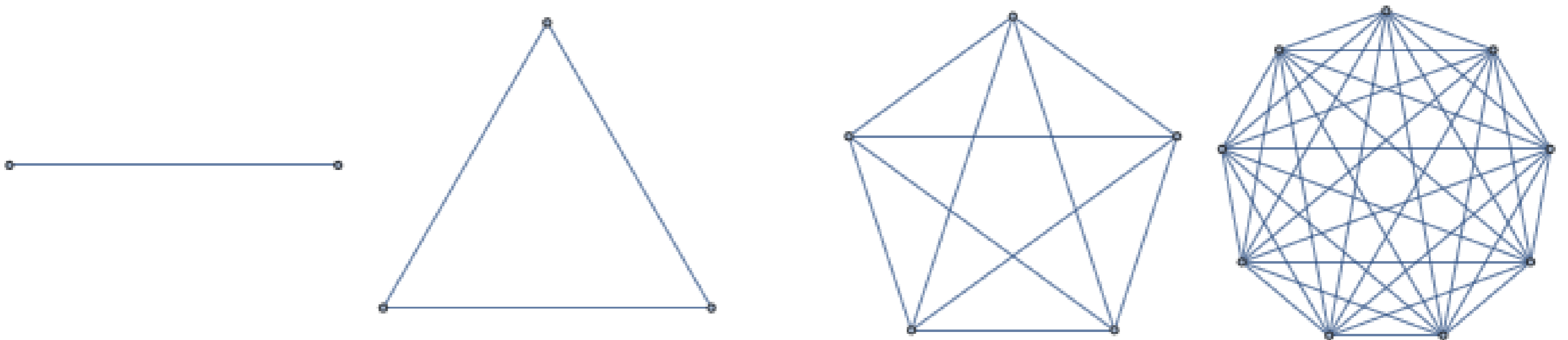
Regular graph

- All nodes have the same degree



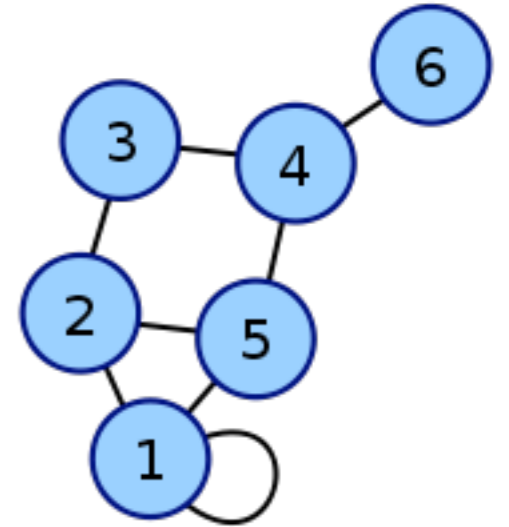
Complete graph

- All-to-all connectivity
- Can sometimes be used to represent homogeneous mixing



Adjacency Matrix

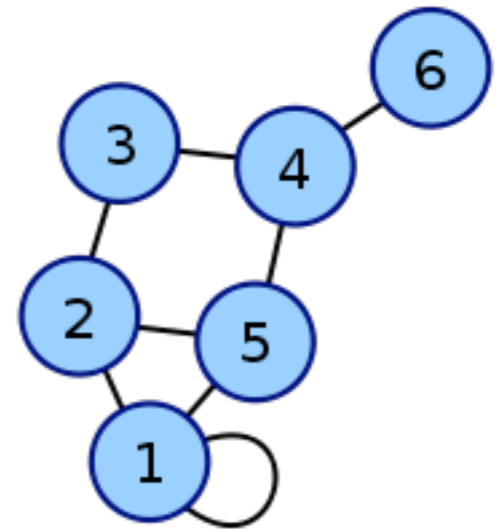
- Matrix representing the graph structure
- Can reconstruct the graph from the matrix & vice versa
- Pattern, eigenvalues, etc. of adjacency matrix can often tell you about the graph



$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Adjacency Matrix

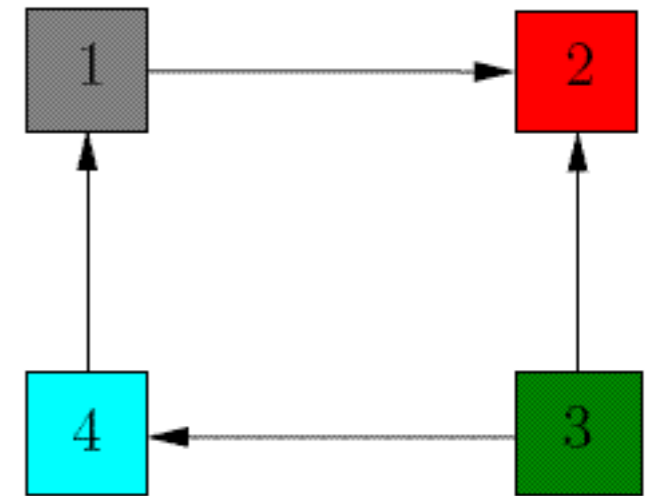
- Undirected graph - adjacency matrix is symmetric
- Directed graph - asymmetric
- Weighted graph - takes non 0/1 values to match edge weights



$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Adjacency Matrix

- Undirected graph - adjacency matrix is symmetric
- Directed graph - asymmetric
- Weighted graph - takes non 0/1 values to match edge weights



$$A = \begin{pmatrix} 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0.75 \\ 0.3 & 0 & 0 & 0 \end{pmatrix}$$

Adjacency Matrix

- Many useful properties - particularly for huge graphs where it's hard to test visually or by checking connectivity
- (i,j) spot of A^k gives paths of length k from i to j
- Two graphs G_1 and G_2 are 'the same' (isomorphic) if $A_1 = P A_2 P^{-1}$
- Can use to find number of connected components, bipartite-ness, etc.

Network Metrics

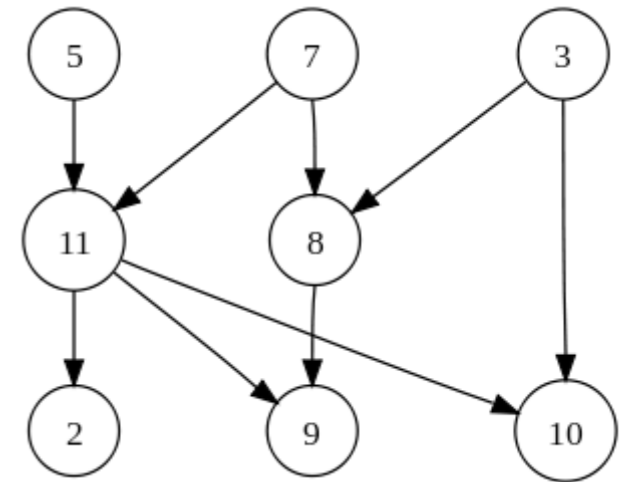
- Wide range of measures, of varying levels of complexity, that are used to characterize networks at both the “micro” (i.e. node and edge) and “macro” (i.e. network) levels

Network Metrics

- **Size** - Number of nodes and edges in a network
- **Density** - Fraction of all realized edges relative to possible edges
 - n = number of nodes
 - m = number of edges
 - $D = 2m/n(n-1)$ (for undirected graph)

Degree

- **Degree** - number of edges attached to a node
- “Egocentric” social network
- **In-degree** - number of incoming edges
- **Out-degree** - number of outgoing edges



Network Centrality

- How central or important is a particular node? How to find “important” nodes?
- Many different approaches & types of centrality
- **Degree centrality** of a node is just the degree (can also use indegree & outdegree)

Closeness Centrality

- **Closeness centrality** of node x - measures shortest paths from x to other nodes
- Idea is that the easier it is to get from one node to all other nodes quickly the more 'central' it is

- $$C(i) = \frac{n}{\sum_j d(i, j)}$$

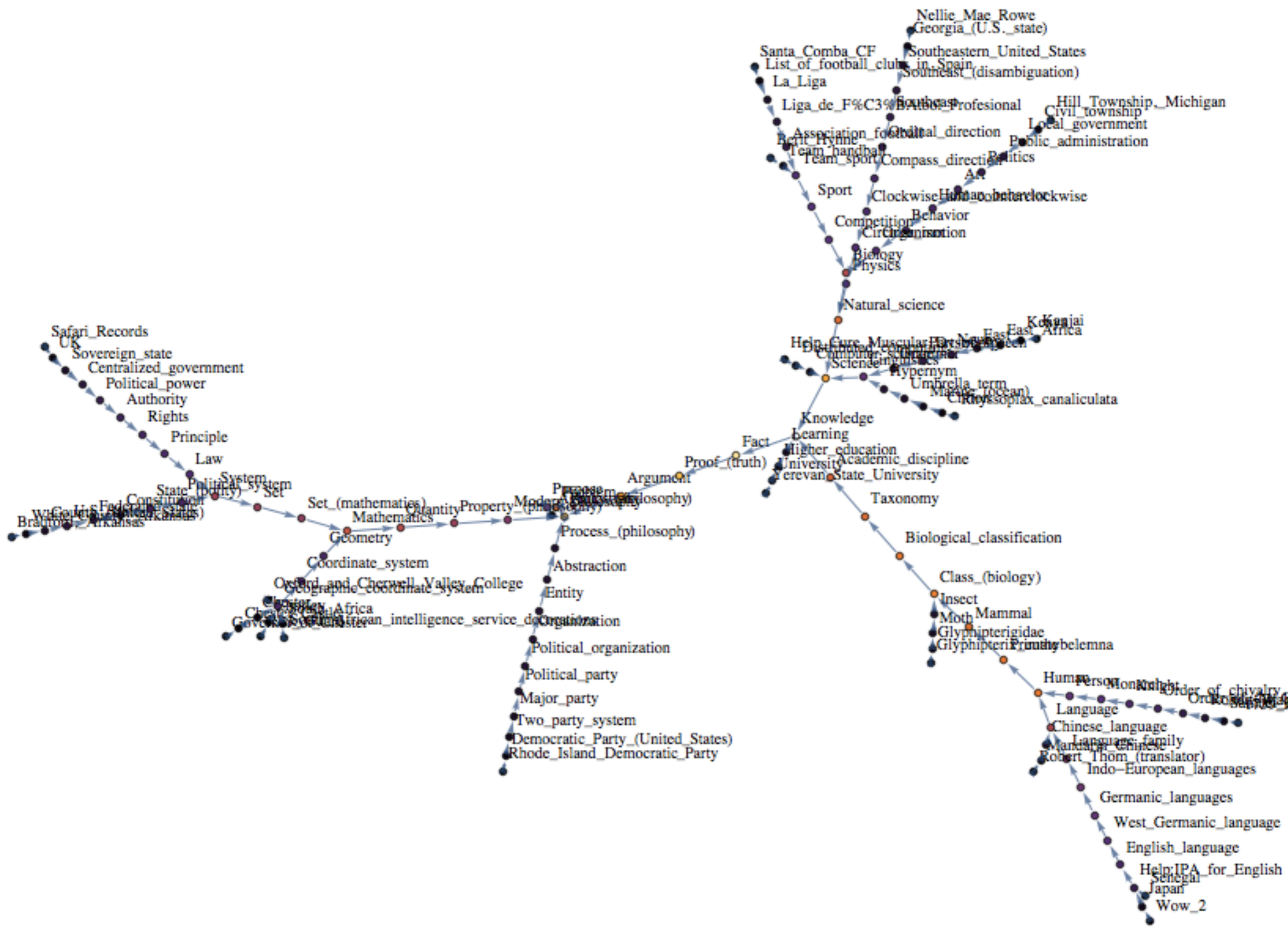
$$C(i) = \frac{n - 1}{\sum_{j \neq i} d(i, j)}$$

Betweenness Centrality

- **Betweenness Centrality** - measures how 'bridge-y' the node is, i.e. if a node is an important bridge from one set of nodes to another, it is more central
- Betweenness centrality of node x - determine how often the shortest path between two nodes uses x

$$B(i) = \sum_{s,t} \frac{n_{st}^i}{g_{st}}$$

where n_{st}^i is the number of shortest paths from s to t that pass through i , and g_{st} is the total number of shortest path routes from s to t



Eigenvector Centrality

- Centrality is based on centrality of your neighbors (connections to highly central individuals increases your centrality)
- Closely related to Google pagerank
- This works out to be the eigenvector of the largest eigenvalue of the adjacency matrix

Why eigenvector?

- Suppose the centrality x for each node i is proportional to the sum of its neighbors' centralities. We can write that like this:

$$x_i = k \sum_j A_{ij} x_j \quad \text{where } A \text{ is the adjacency matrix}$$

- This is just the equation for a matrix times a vector!
We can rewrite as: $\mathbf{x} = kA\mathbf{x}$
- This is the equation for eigenvectors and eigenvalues! If we let $\lambda = 1/k$ then we have:

$$A\mathbf{x} = \lambda\mathbf{x}$$

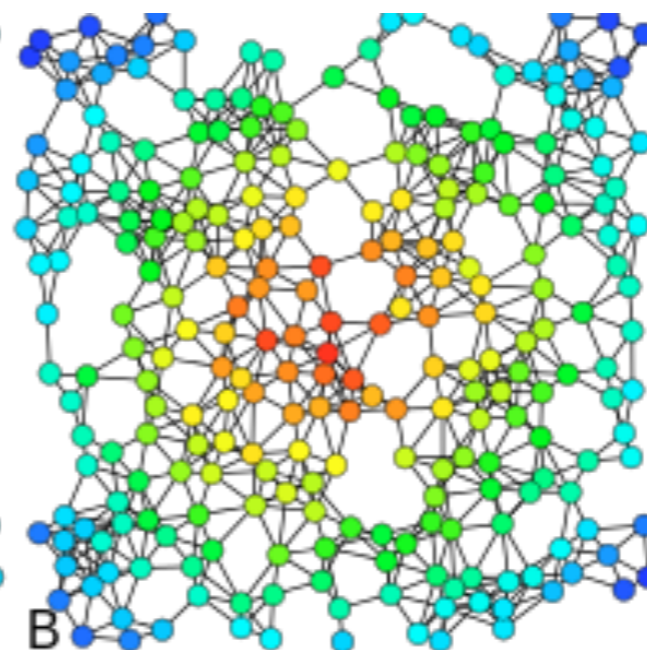
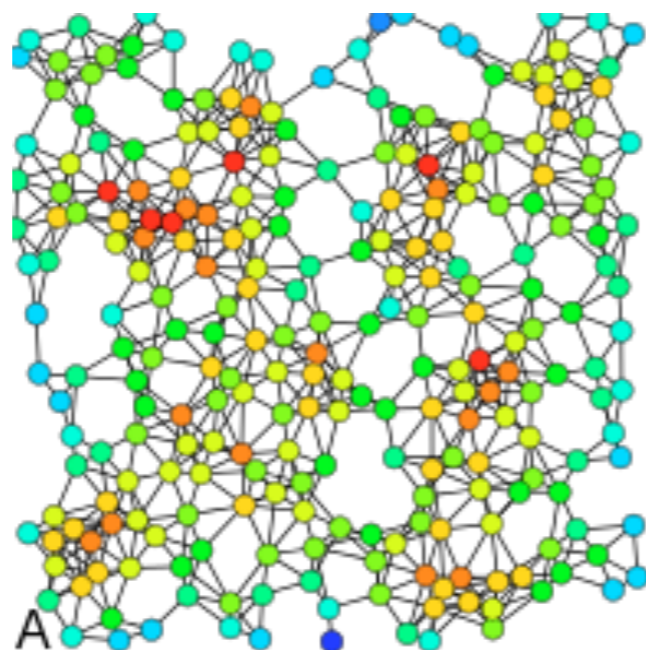
Eigenvector centrality

- Eigenvector centrality of a node x_i is the i th entry of the eigenvector corresponding to the largest eigenvalue
- The eigenvector for the largest eigenvalue ensures that all the eigenvector entries are non-negative (Frobenius–Perron theorem)

Eigenvector centrality

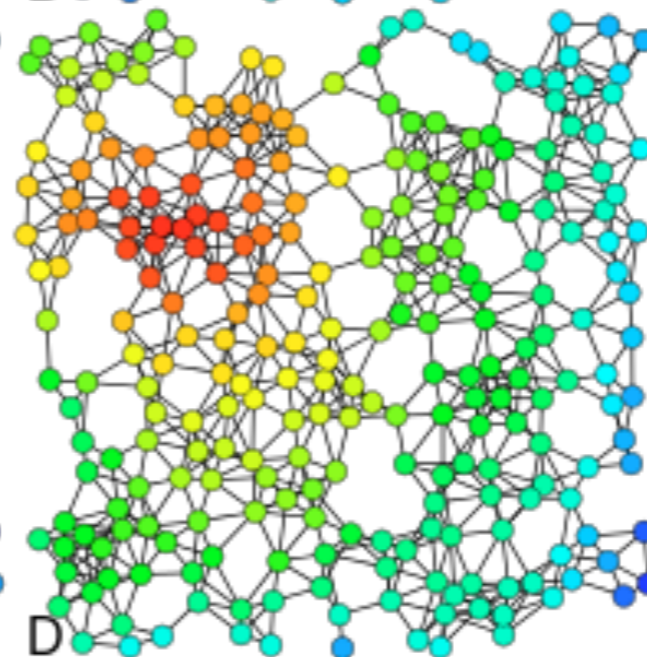
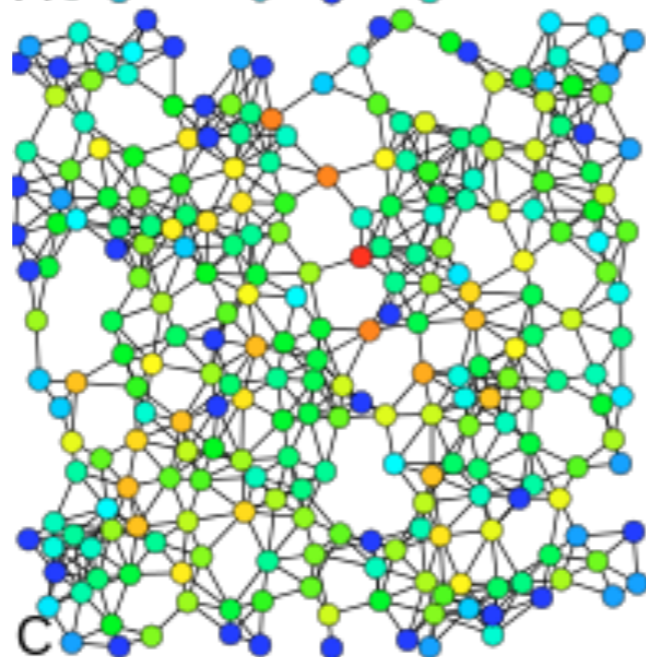
- A normalized version of eigenvector centrality is what is used for Google Pagerank!
- Neuroscience: eigenvector centrality of a neuron in model neural networks has been found to correlate with its relative firing rate
- Eigenvector centrality has been found to correlate with intermarriage networks in political families, etc
- Used to model opinion influence, cooperation, etc.

Degree



Closeness

Betweenness

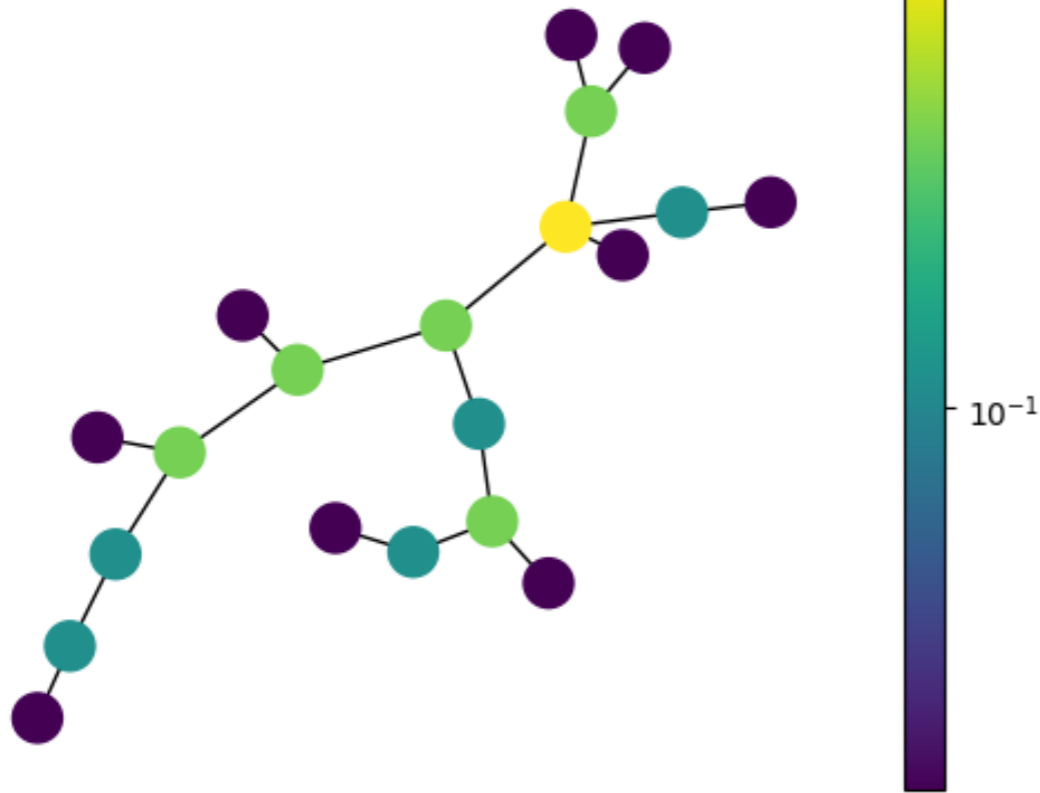


Eigenvector

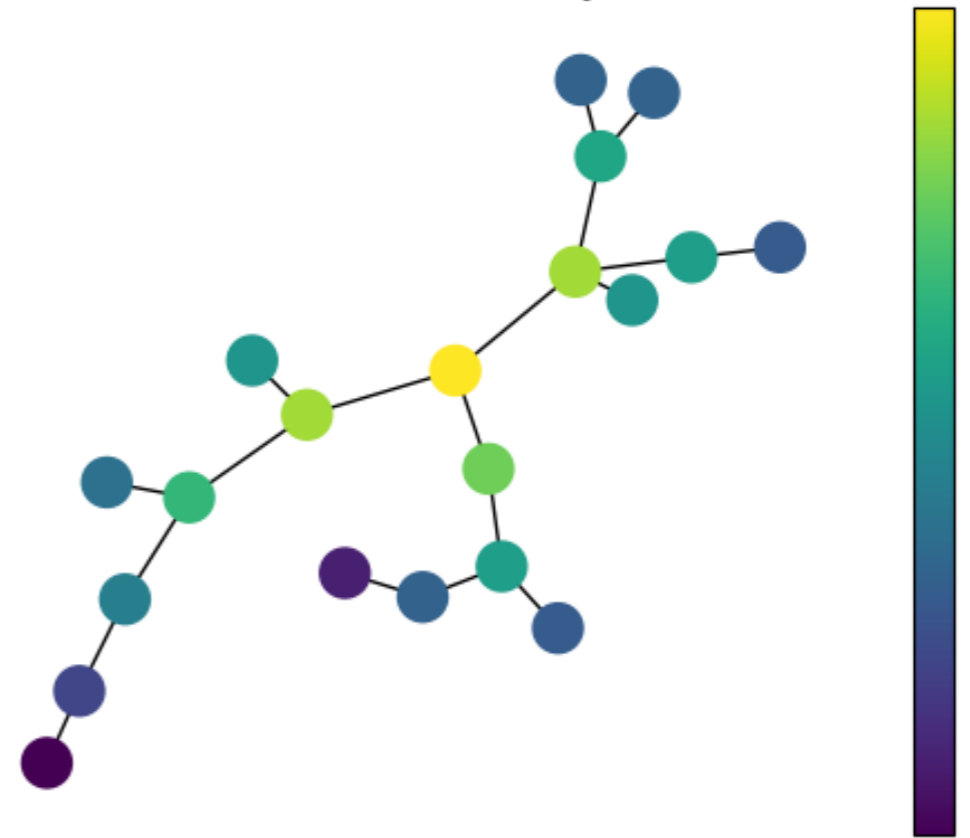
Centrality metrics tend to be correlated

- When they aren't, that often means there's something interesting about the network structure!
- For example:
 - If a node has high betweenness centrality but low degree centrality?
 - If a node has high closeness but low betweenness?

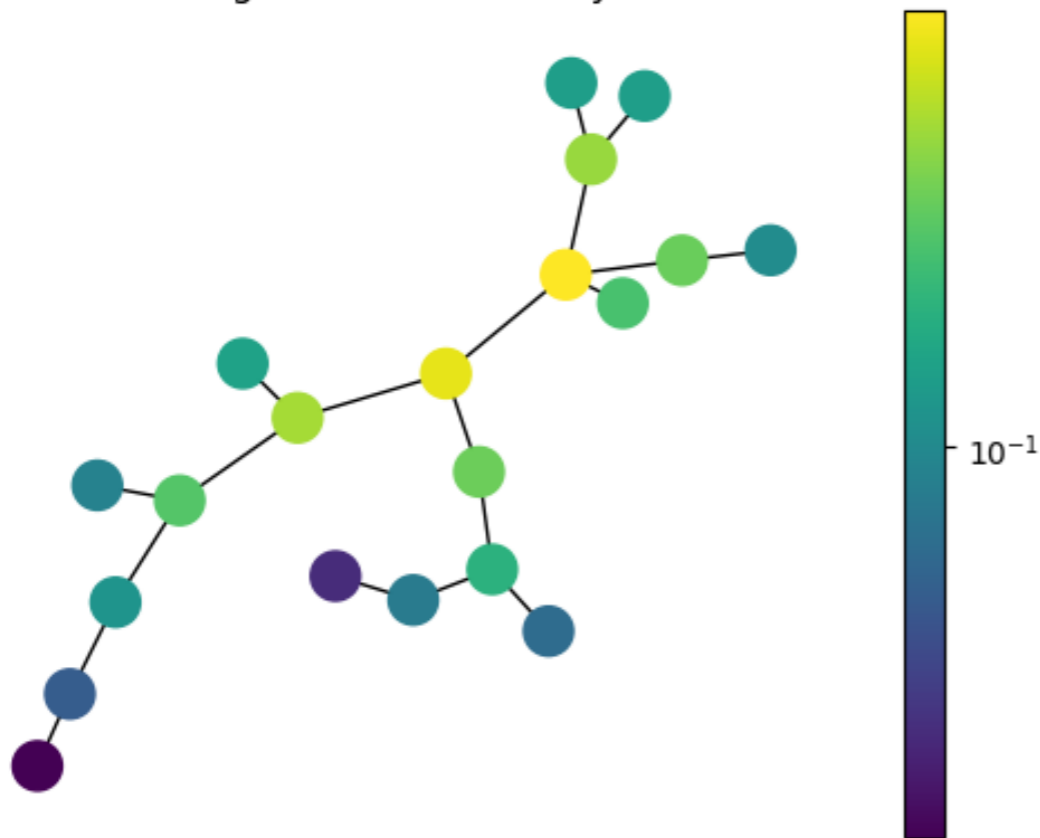
Degree Centrality



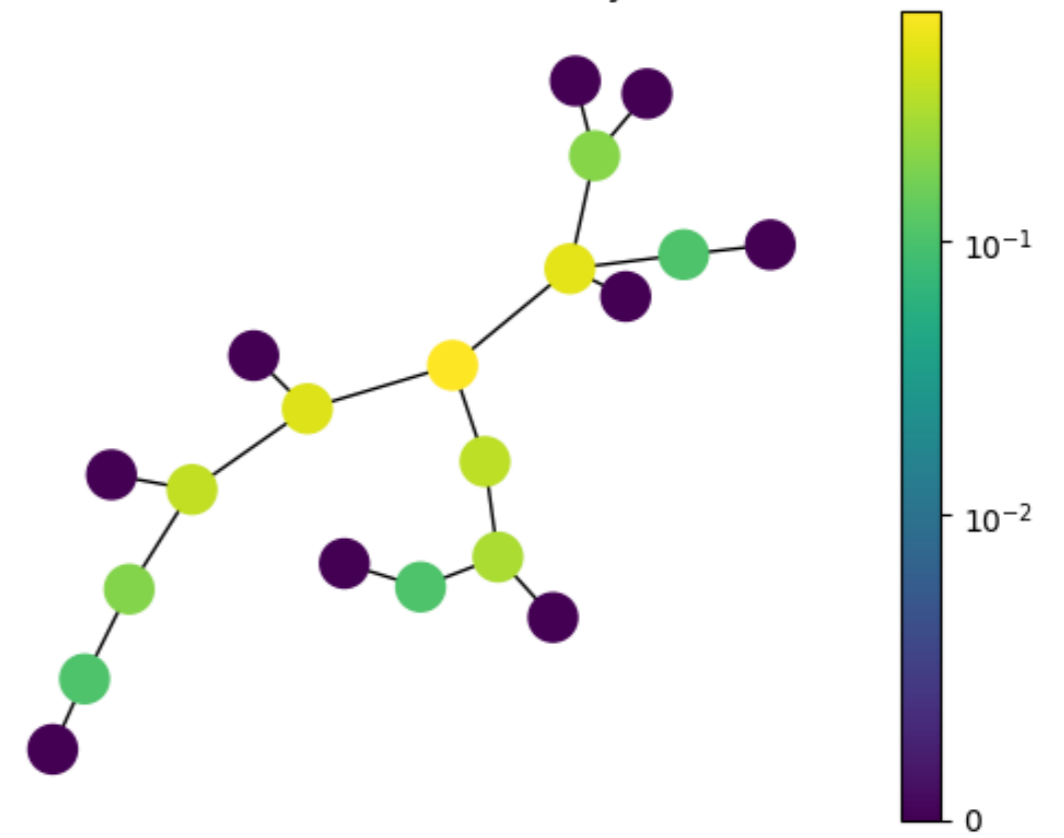
Closeness Centrality



Eigenvector Centrality

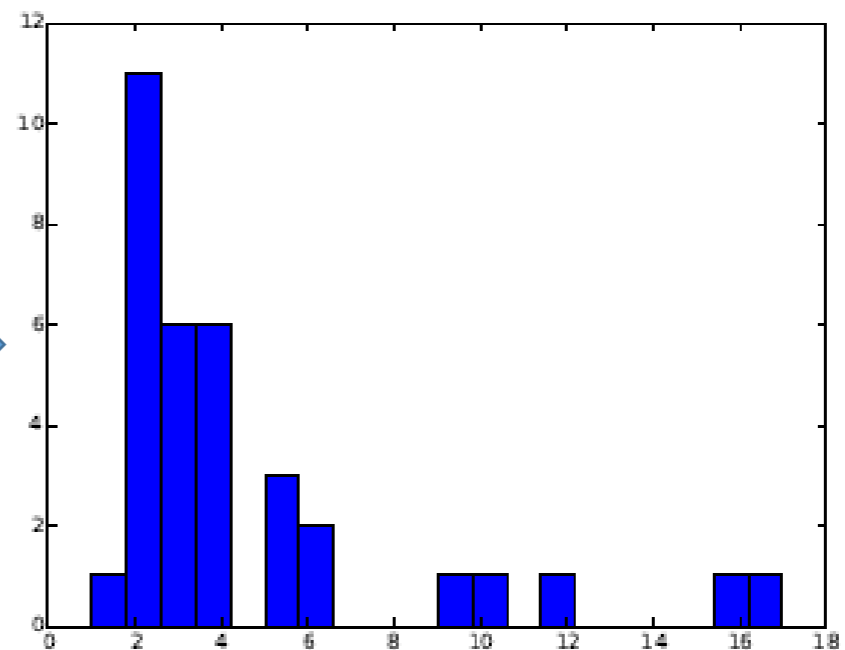
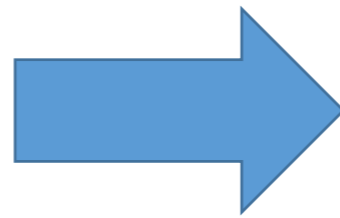
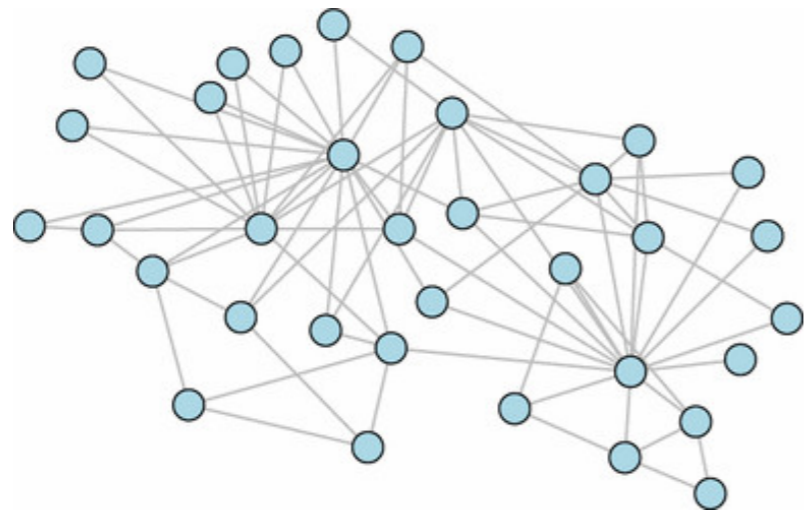


Betweenness Centrality

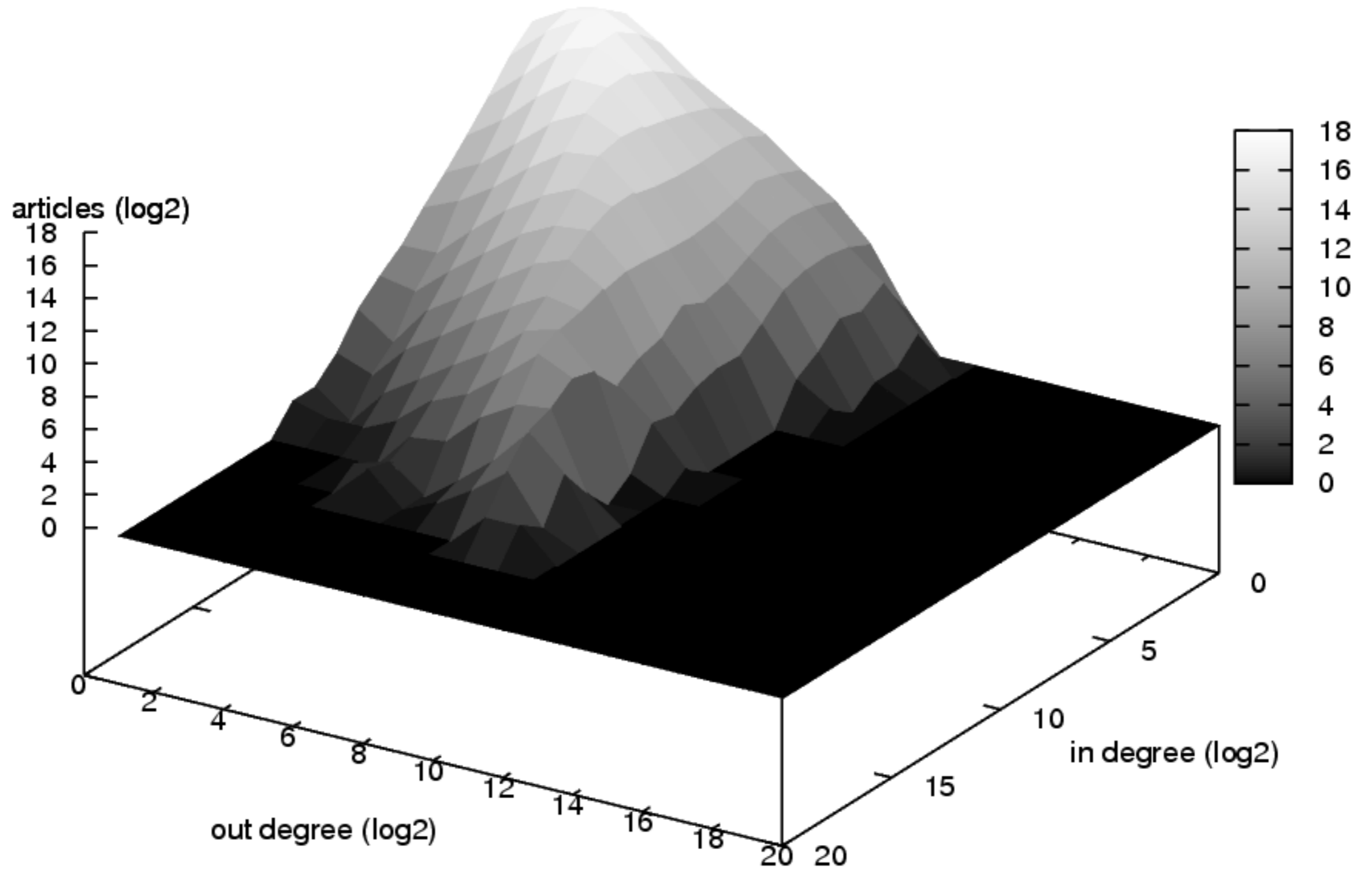


Degree distribution

- **Degree sequence** - List of degrees for all nodes in a graph
- Often use this to determine the **degree distribution** (often these are treated as the same)
- Degree sequence/distribution can tell you a lot about structure of graph



English Wikipedia: in/out degree distribution

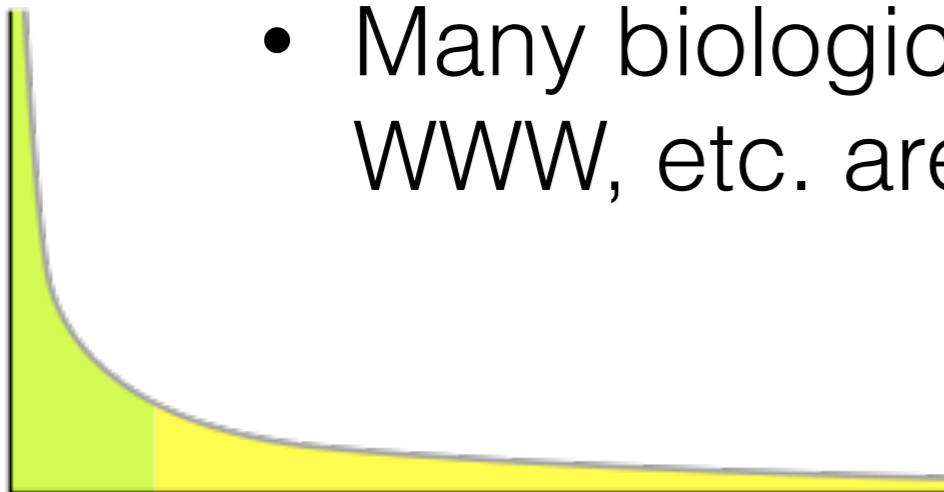


Power Law degree distribution

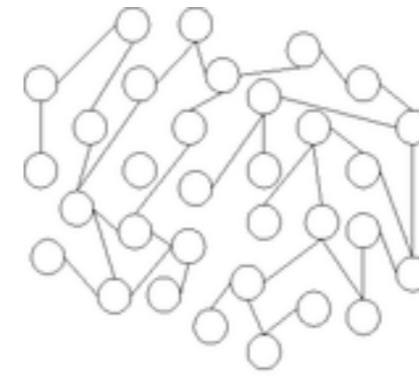
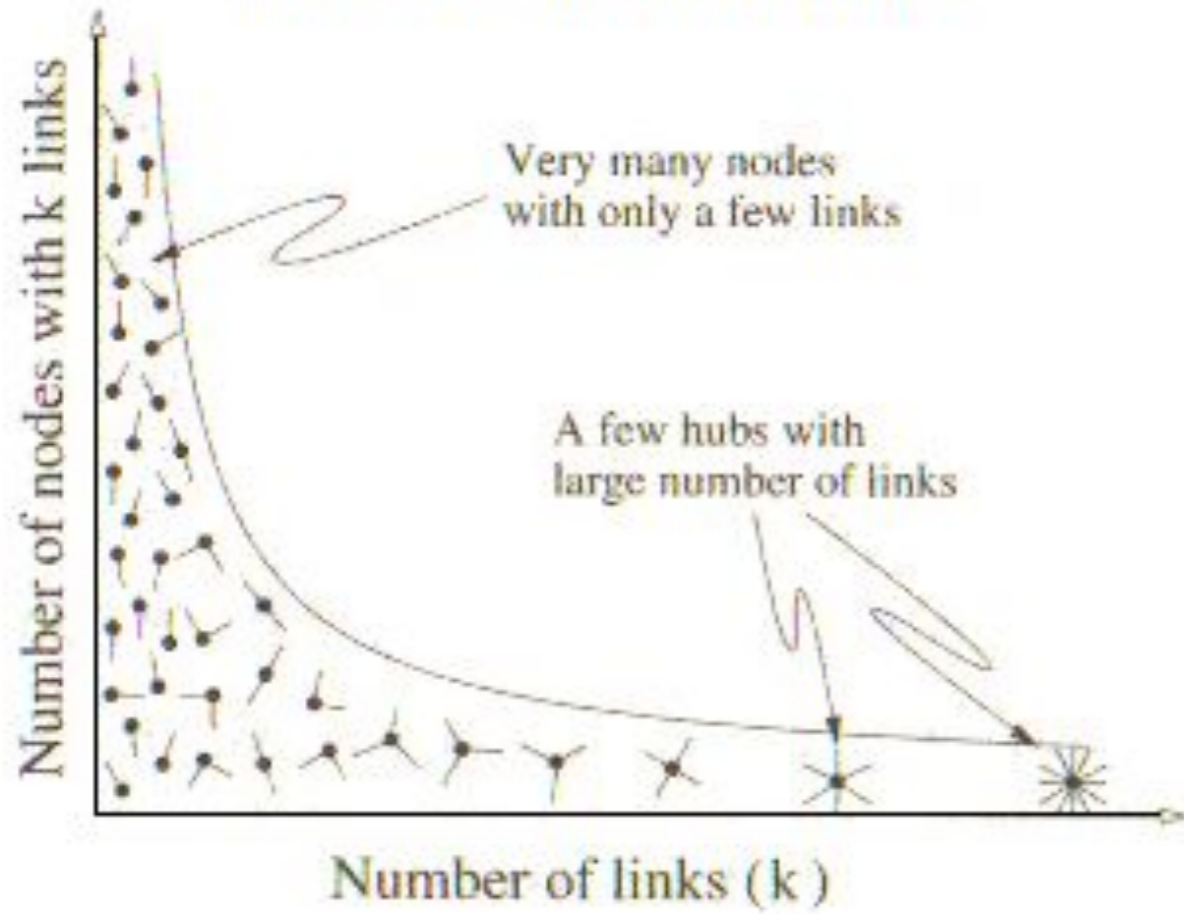
- **Scale free networks** - power law degree distribution

$$P(k) \sim k^{-\gamma}$$

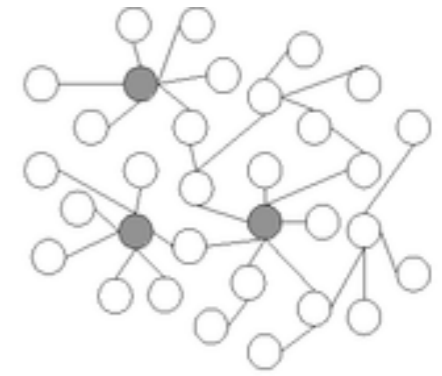
- Long tail results in both very sparse nodes and hub nodes
- Many biological networks, social networks, WWW, etc. are scale free



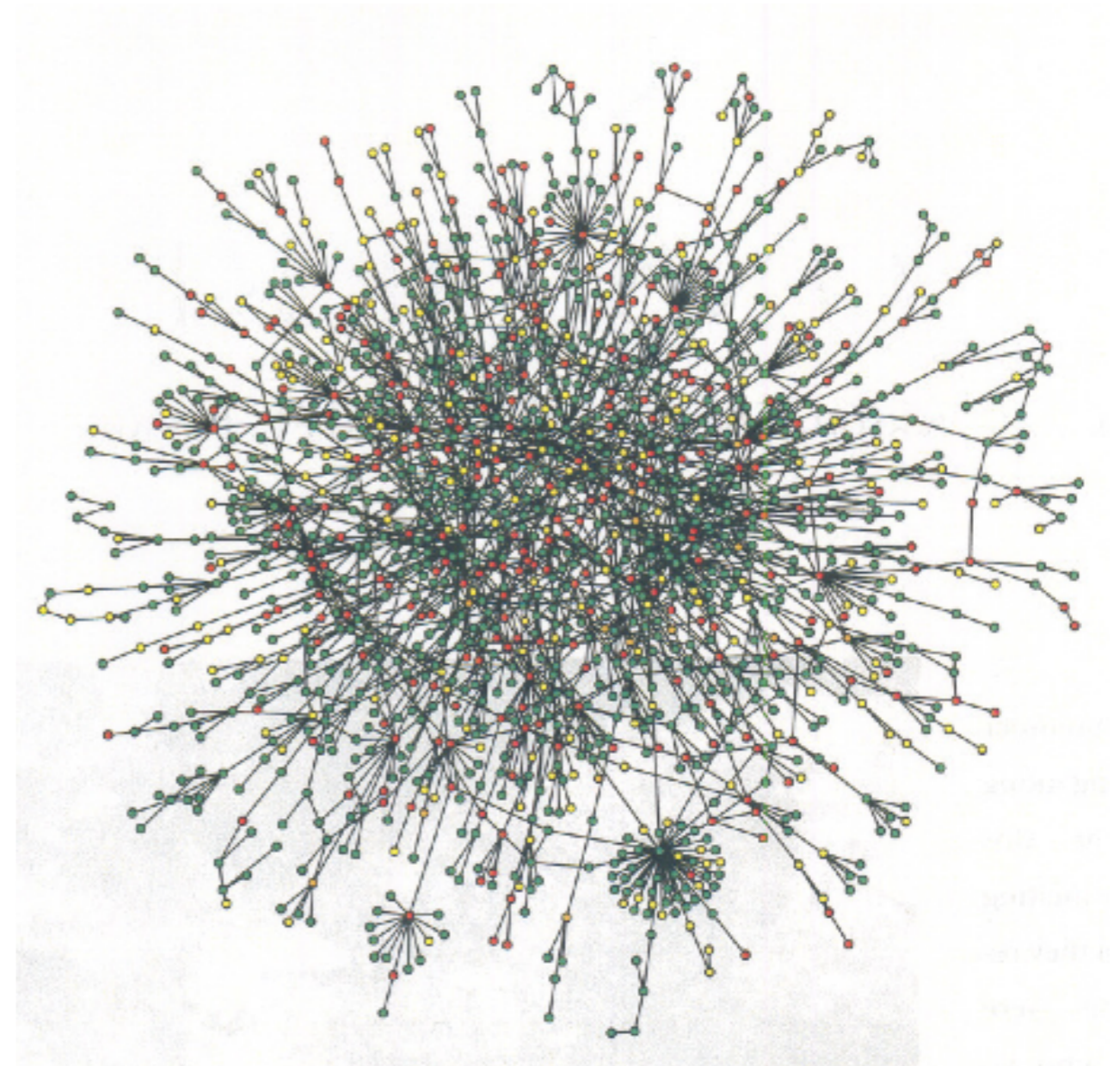
Power Law Distribution



(a) Random network



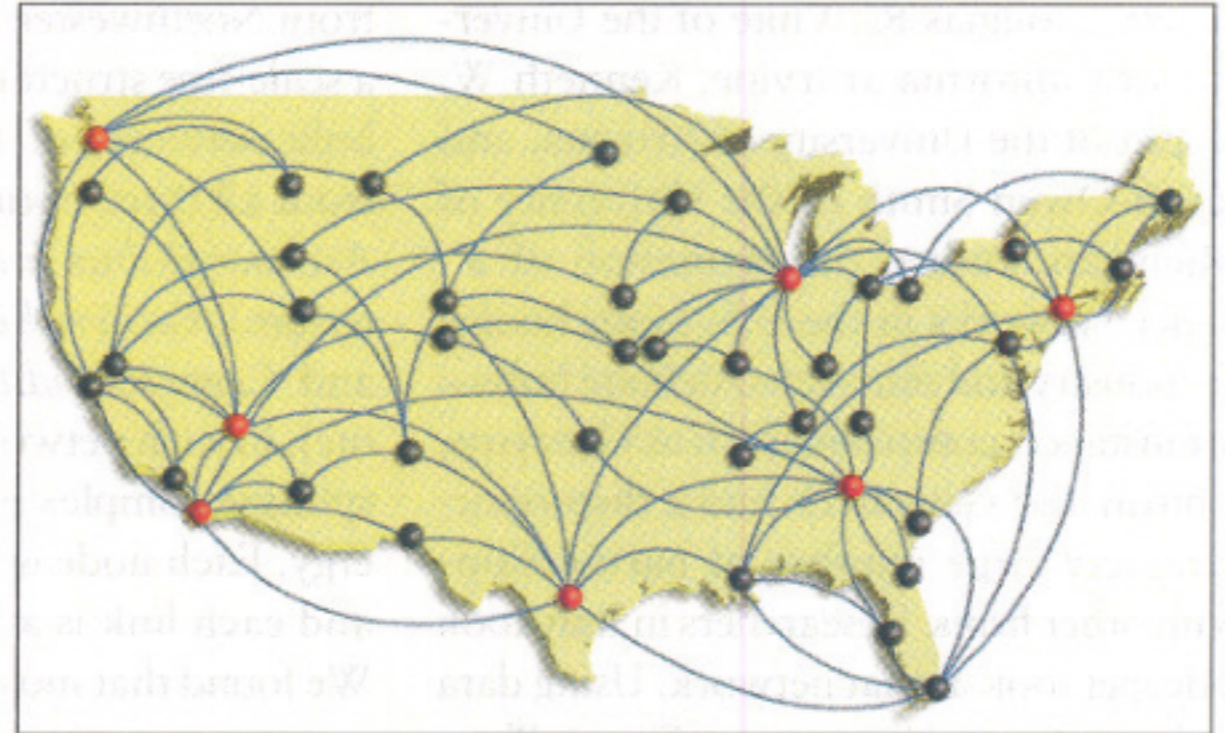
(b) Scale-free network



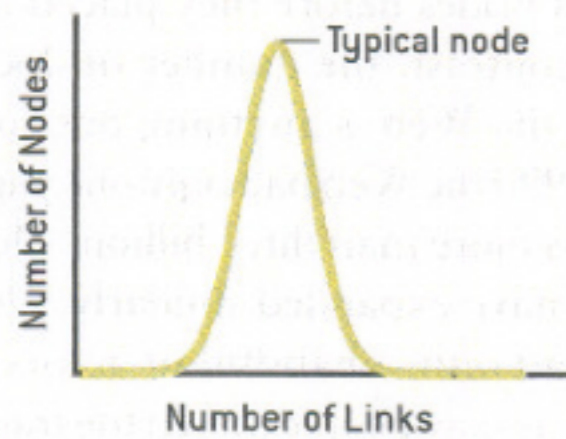
Random Network



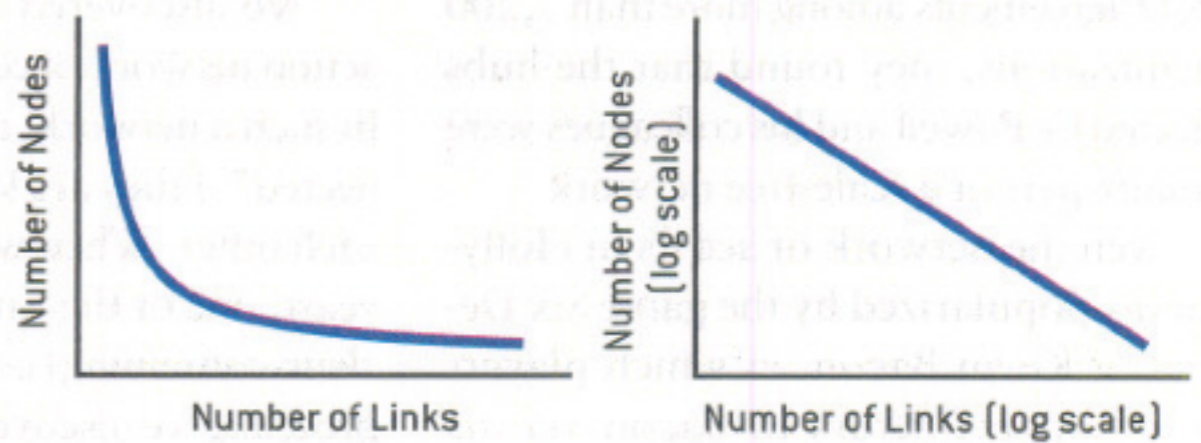
Scale-Free Network



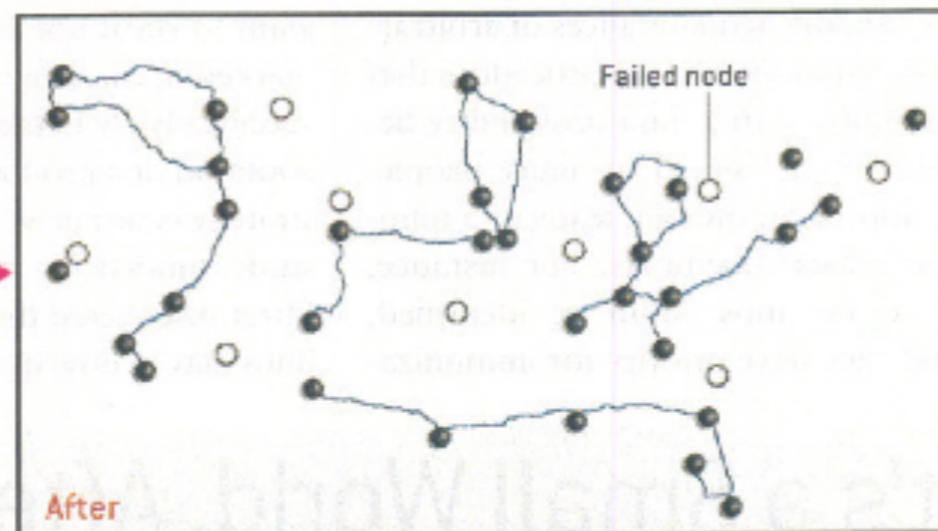
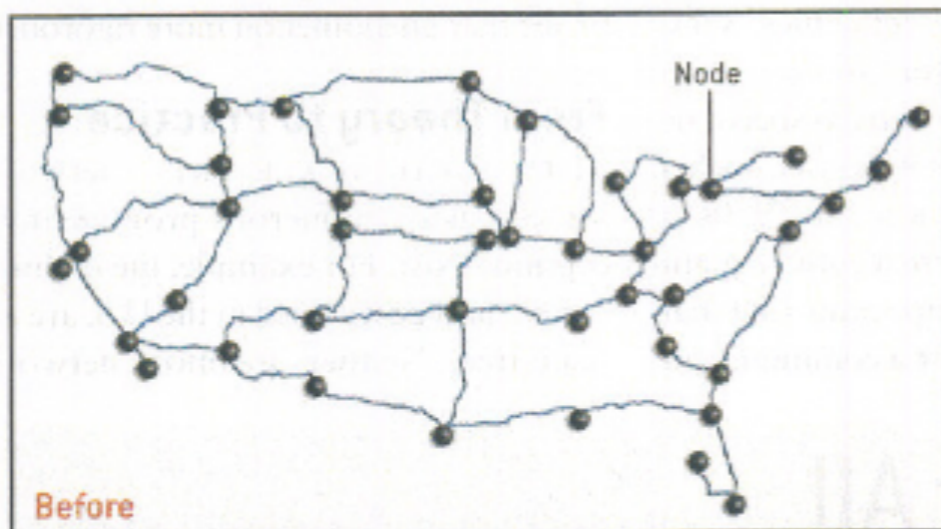
Bell Curve Distribution of Node Linkages



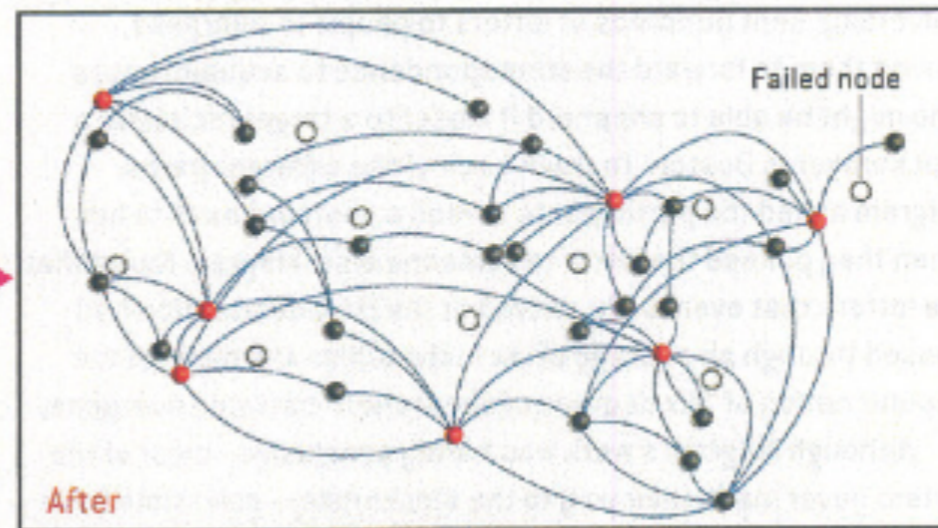
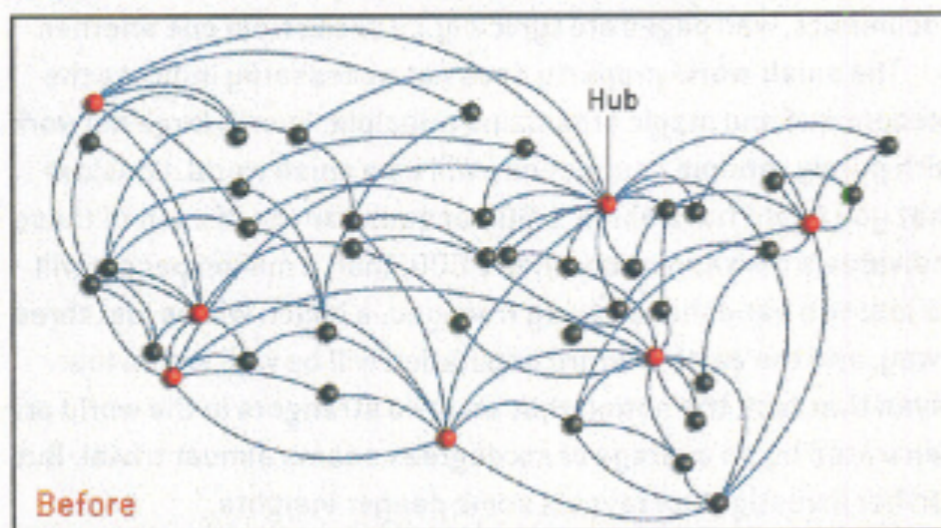
Power Law Distribution of Node Linkages



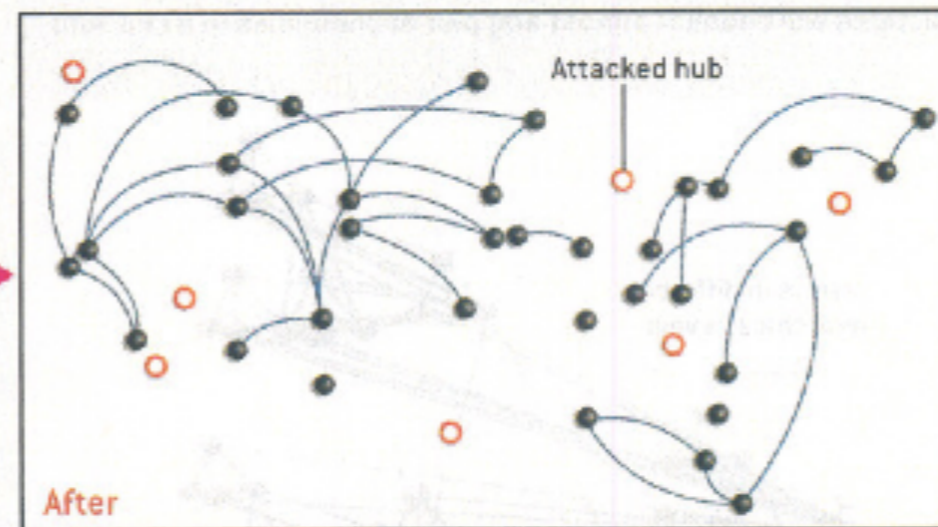
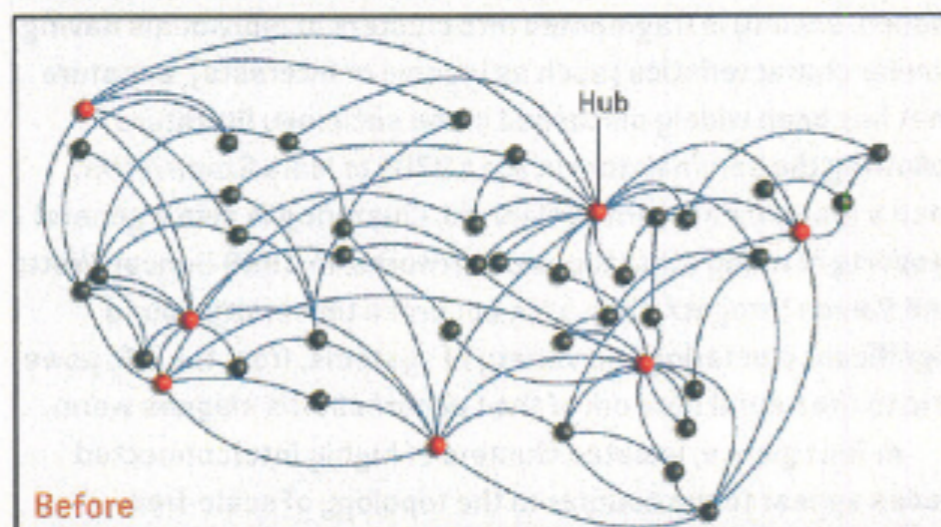
Random Network, Accidental Node Failure



Scale-Free Network, Accidental Node Failure



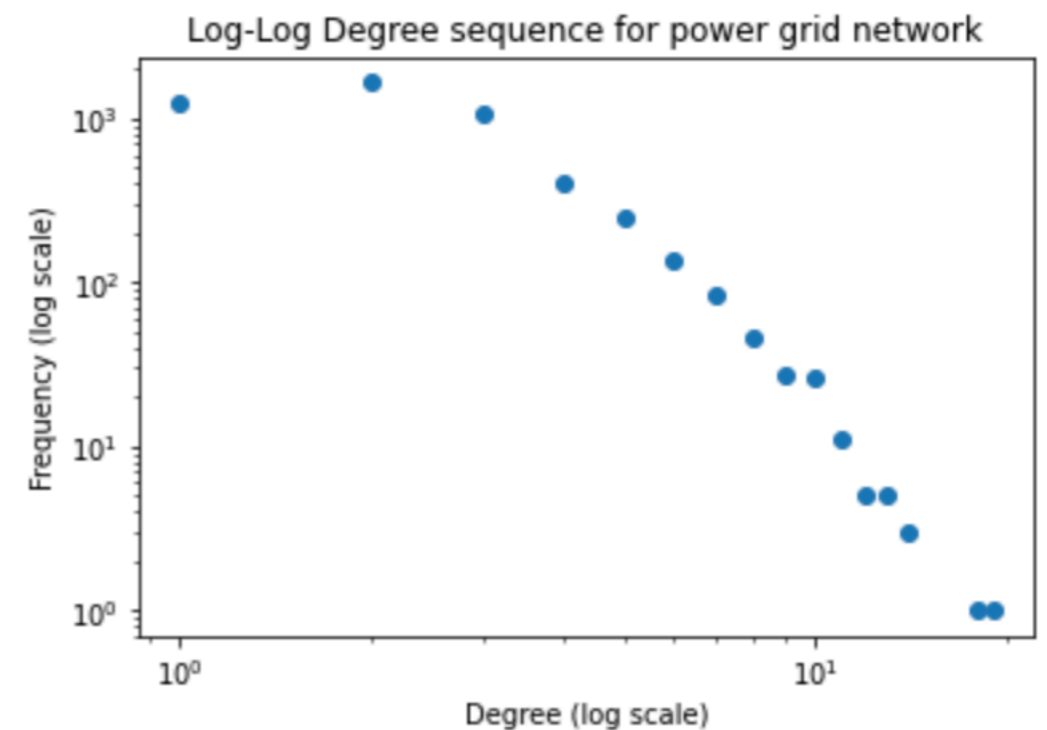
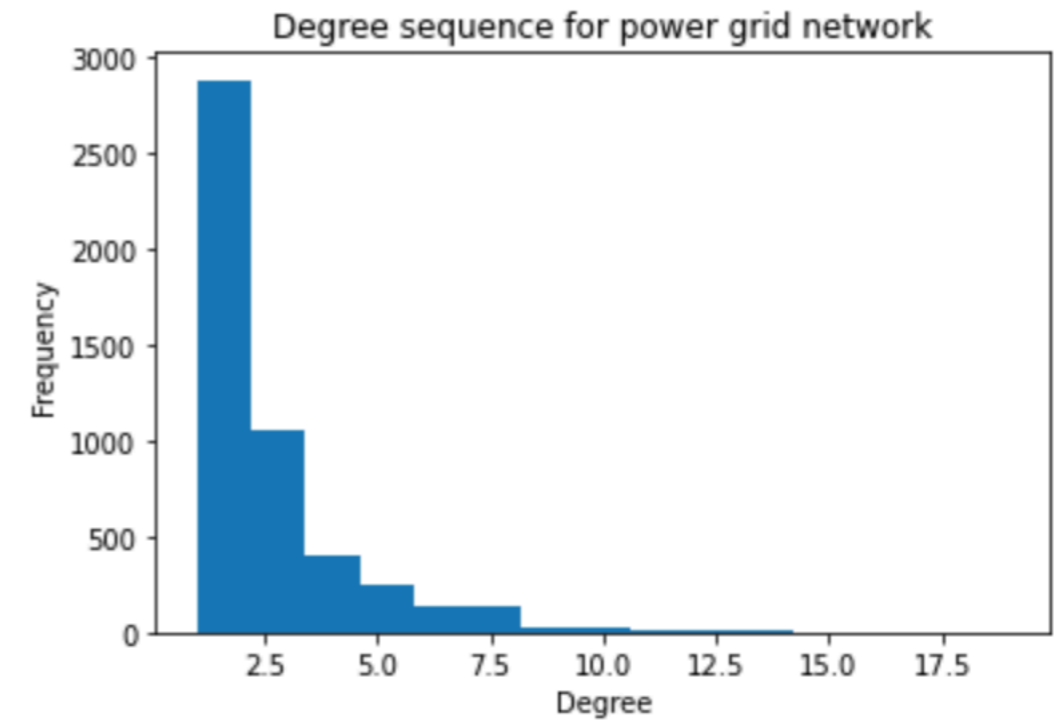
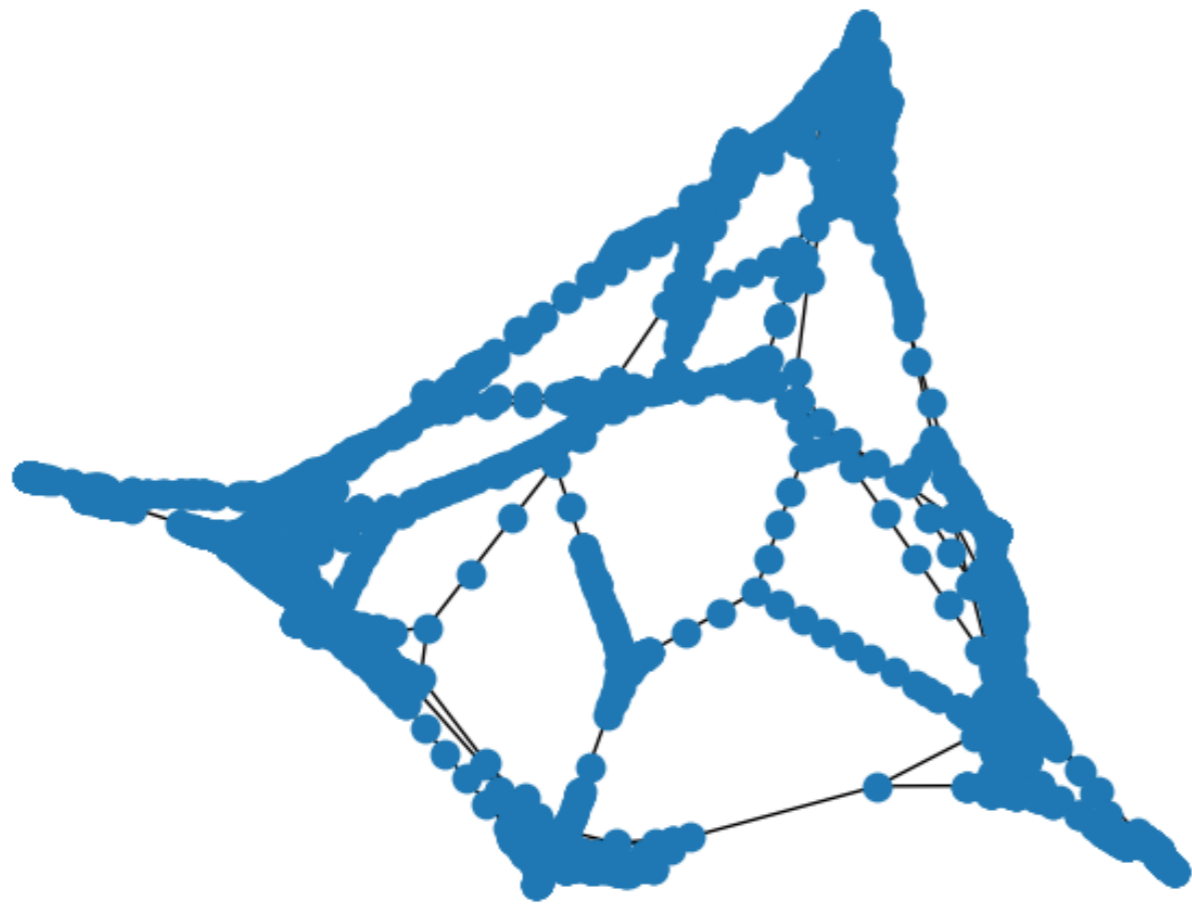
Scale-Free Network, Attack on Hubs



Scale Free Networks

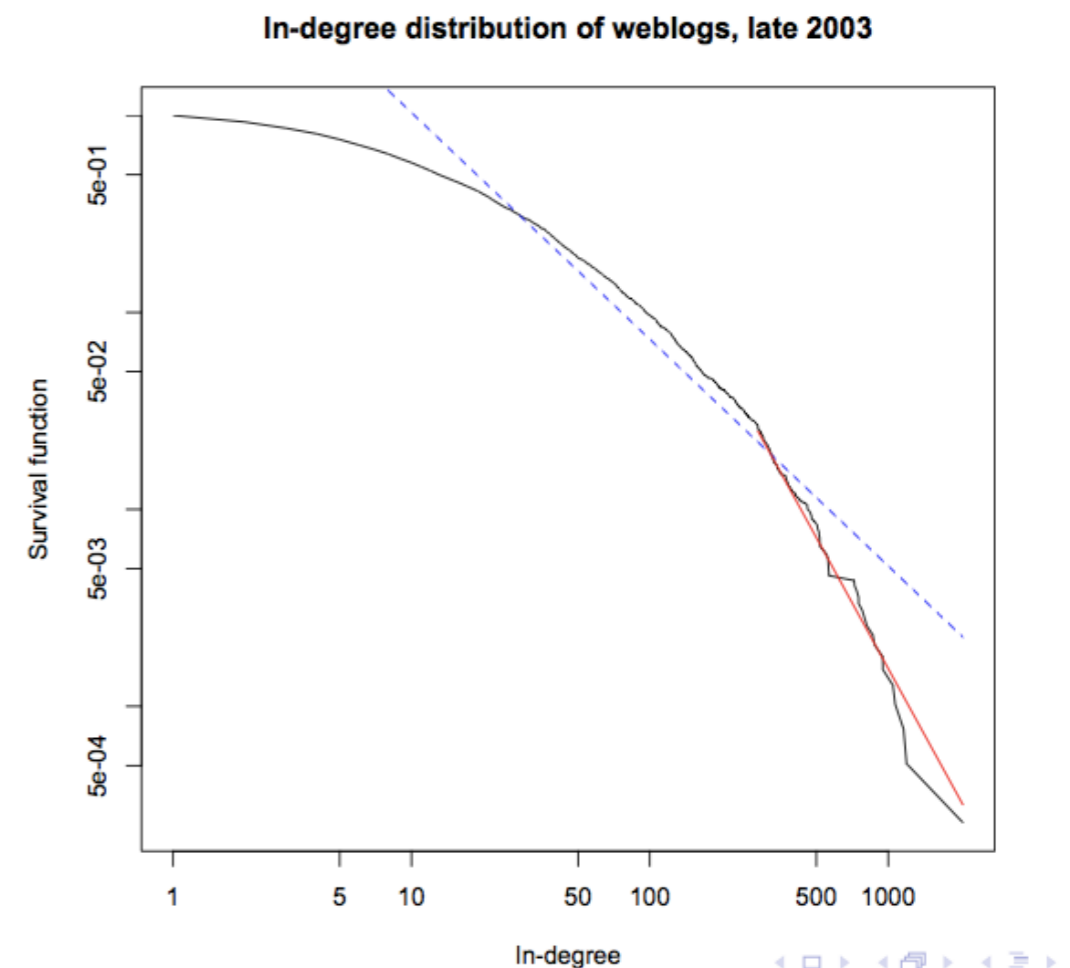
- Scale free networks are robust to random failures (e.g. mutations in a gene)
- However, vulnerable to targeted attacks on hubs

Western US power grid



Scale Free Networks

- However, lots of things look linear-ish on a log-log scale...
- Many suggest some abuse of power law/scale free idea
- Probably a lot of these are just heavy-tailed



Clustering in networks

Clustering in networks

- Many different ways to look at clustering
- How do node traits (degree, covariates) cluster based on edges? E.g. do smokers tend to be friends with other smokers? Do individuals cluster by popularity?
- Community detection - finding clusters (groups) of nodes that are highly connected within the group and less connected between groups (i.e. clustering, where similarity is based on connectivity)

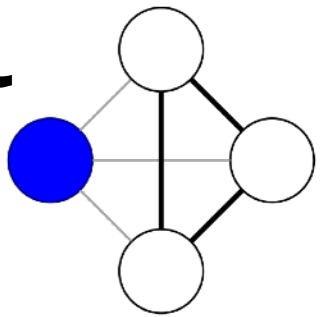
Clustering Coefficient

- Based on the number of triangles in the network
- How many of my friends are also friends?
- Global clustering coefficient

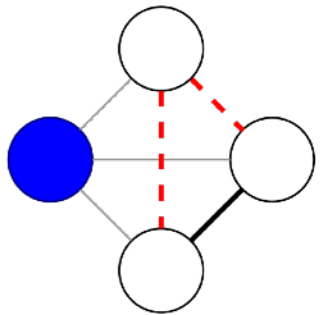
$$C = \frac{\text{number of triangles}}{\text{number of possible triangles}}$$

- Local clustering coefficient

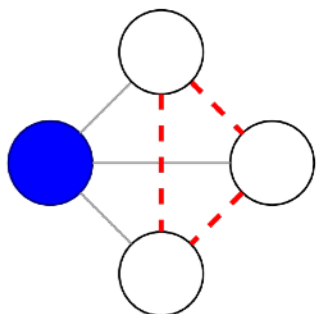
$$C_i = \frac{\text{actual edges between neighbors of } v_i}{\text{possible edges between neighbors of } v_i} = \frac{e_{jk} : v_j, v_k \in N_i | e_{jk} \in E}{|N_i|(|N_i| - 1)/2}$$



$c = 1$



$c = 1/3$



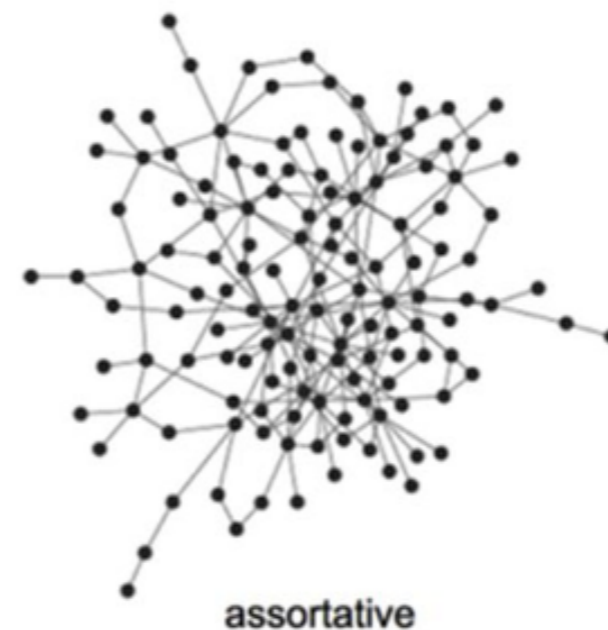
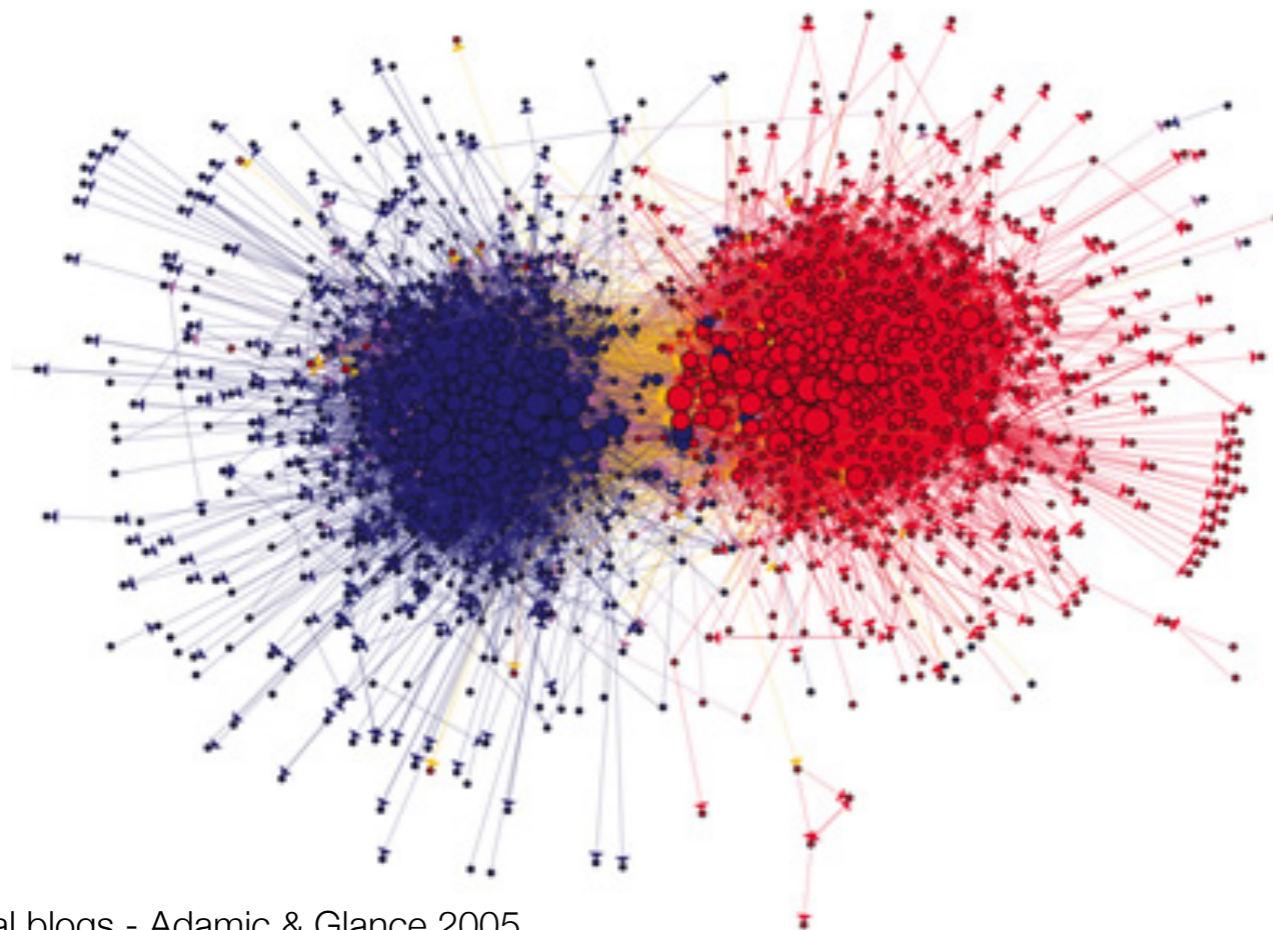
$c = 0$

Assortativity

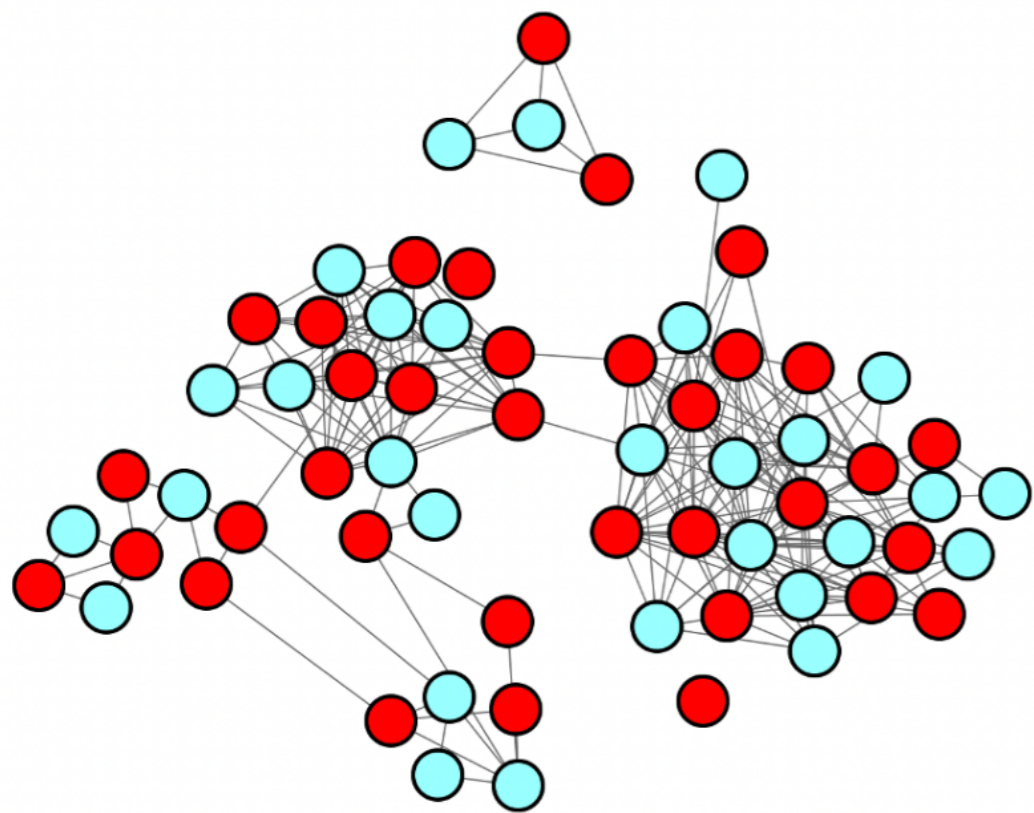
- **Assortativity** - measures network-level tendency for nodes to attach to similar nodes
 - Similarity can be defined by node attributes, degree, etc.
- Calculate fraction of edges between nodes of the same type/value, compare to what would be expected from a random network
- Ranges from -1 (dissassortative) to 1 (assortative)
 - But min value (most dissassortative) is between -1 and 0 depending on the composition of the network

Assortativity

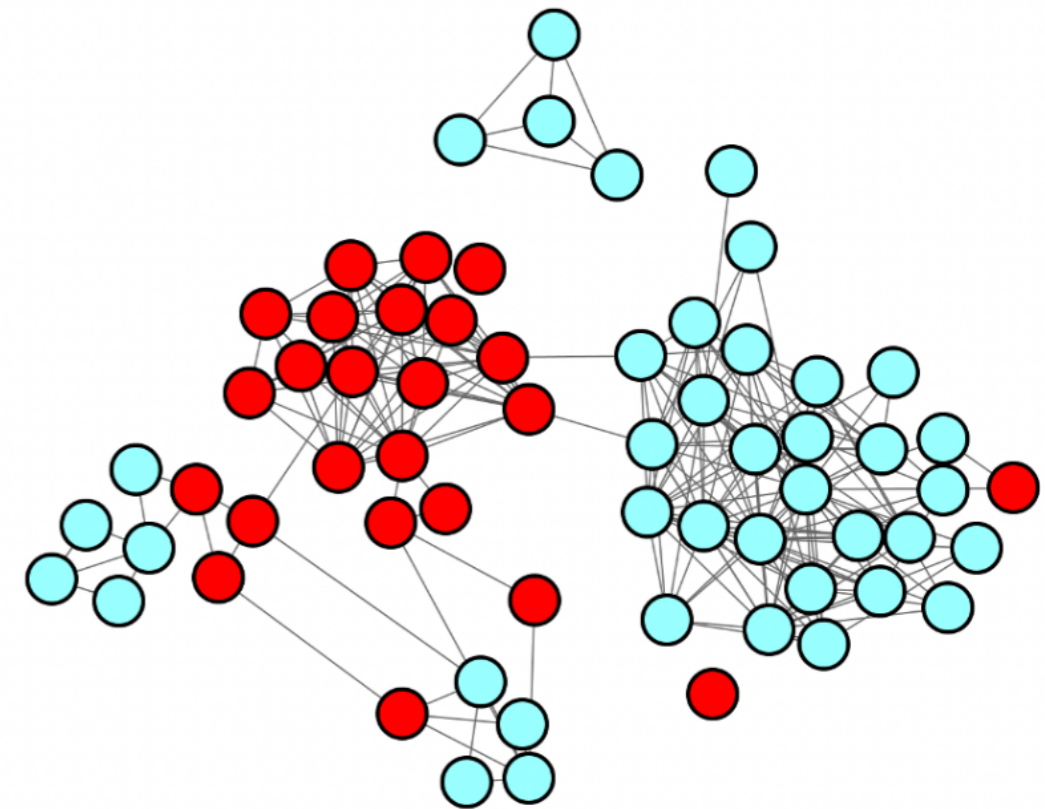
- Heterosexual networks - highly disassortative by gender
- Social/sexual networks often assortative on a range of demographic, degree, behavioral traits - 'birds of a feather flock together'



Assortativity



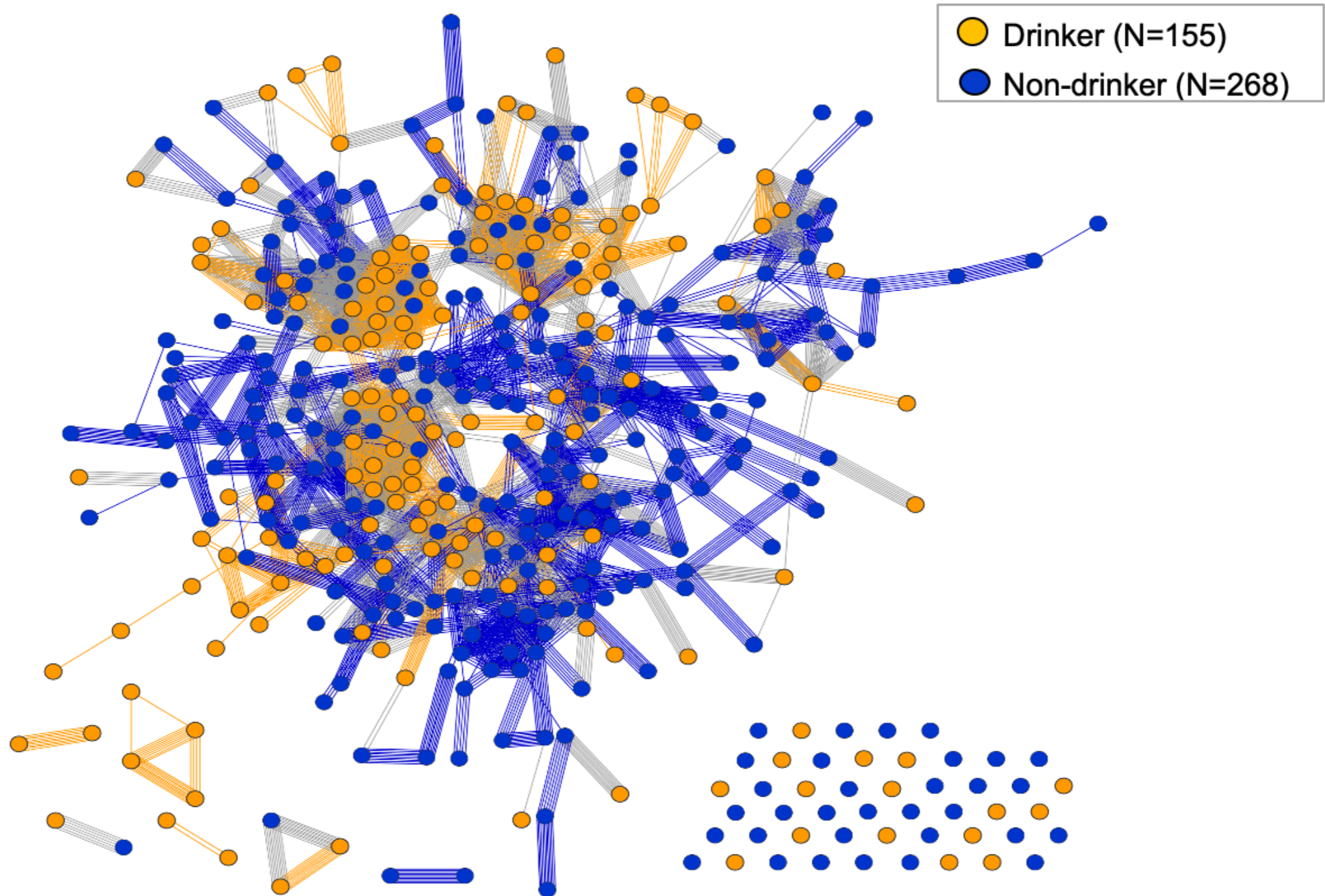
Lower Homophily



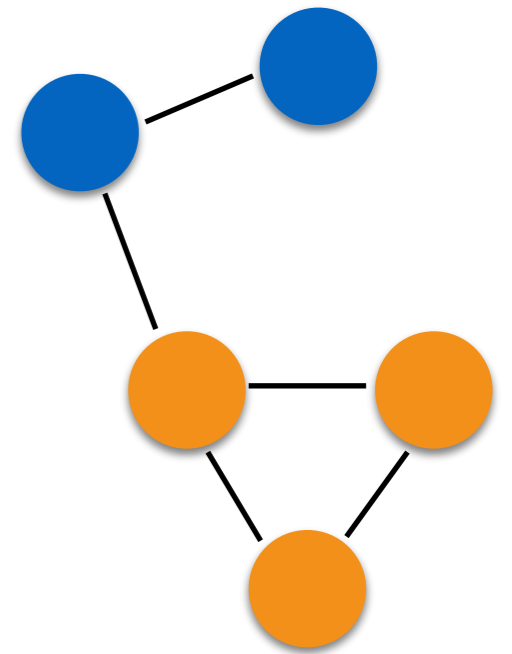
Higher Homophily



Assortativity of alcohol use among college students



Assortativity



- Consider a case where we have discrete characteristics on the nodes
- Define a mixing matrix with entries e_{ij} given by the fraction of the total edges linking type i to type j
- Let a_i and b_i be the total fractions of each end type that we have ($a_i = b_i$ for undirected graphs)
- Note that $\sum_{ij} e_{ij} = 1$, $\sum_j e_{ij} = a_i$, $\sum_i e_{ij} = b_j$

Assortativity

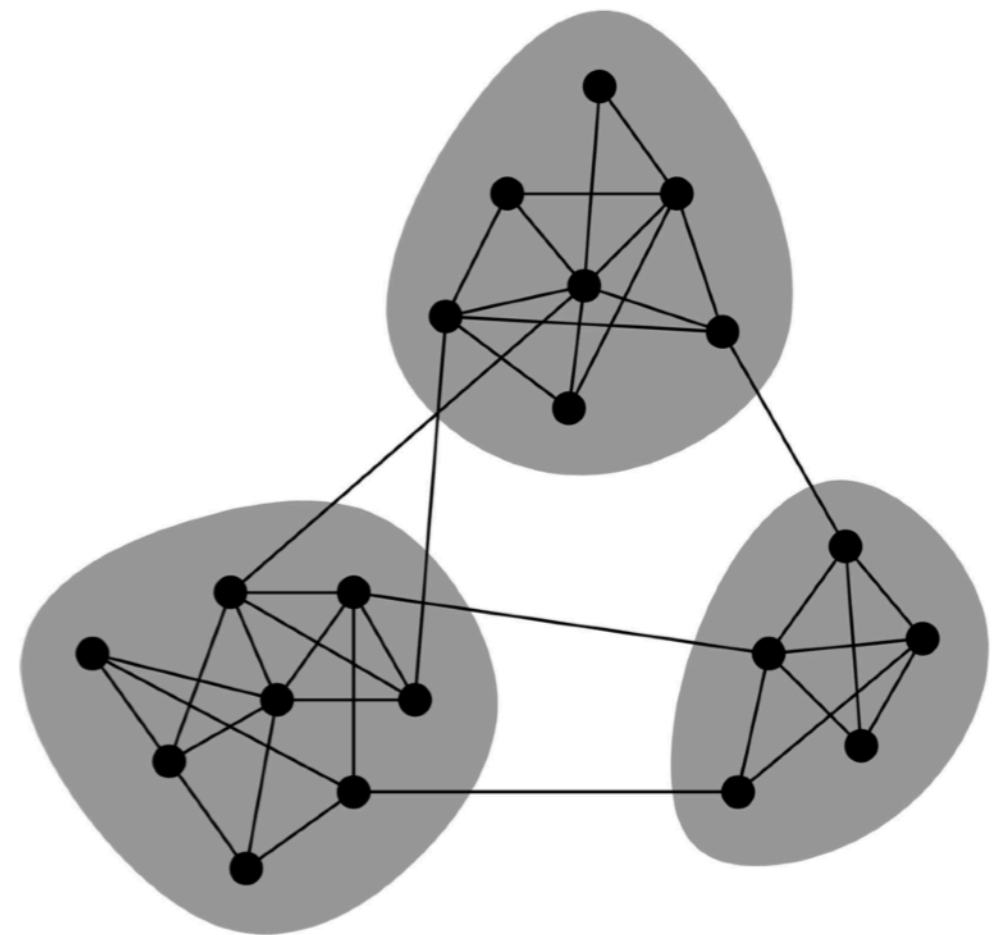
- Defined based on a mixing matrix - entries are the fraction of edges in a network linking type i to type j

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} = \frac{\text{Tr} \mathbf{e} - \|\mathbf{e}^2\|}{1 - \|\mathbf{e}^2\|},$$

- For degree assortativity (and other scalar variables), assortativity is the Pearson correlation coefficient of degree between pairs of linked nodes

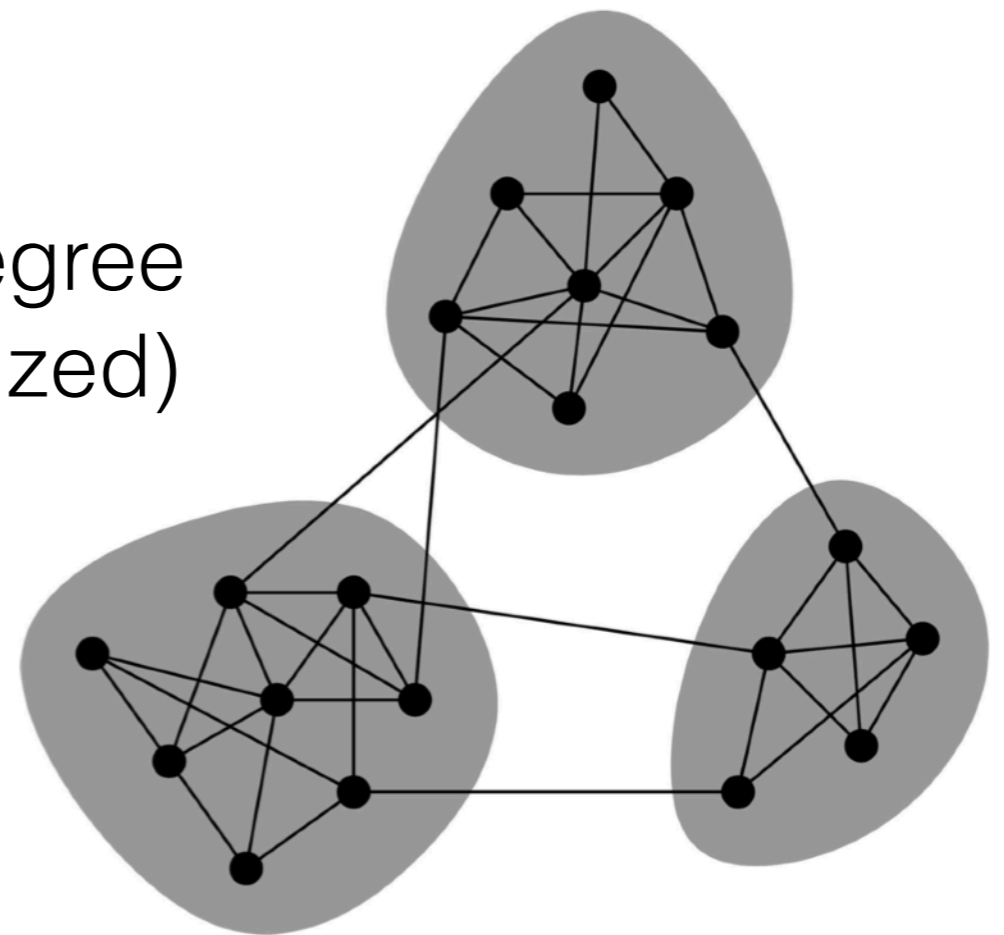
Modularity

- How to decide communities (clusters) in a network?
- We want communities to have more in-group edges than between-group edges
- We could minimize between group edges, but this would lead to just putting all nodes in one community

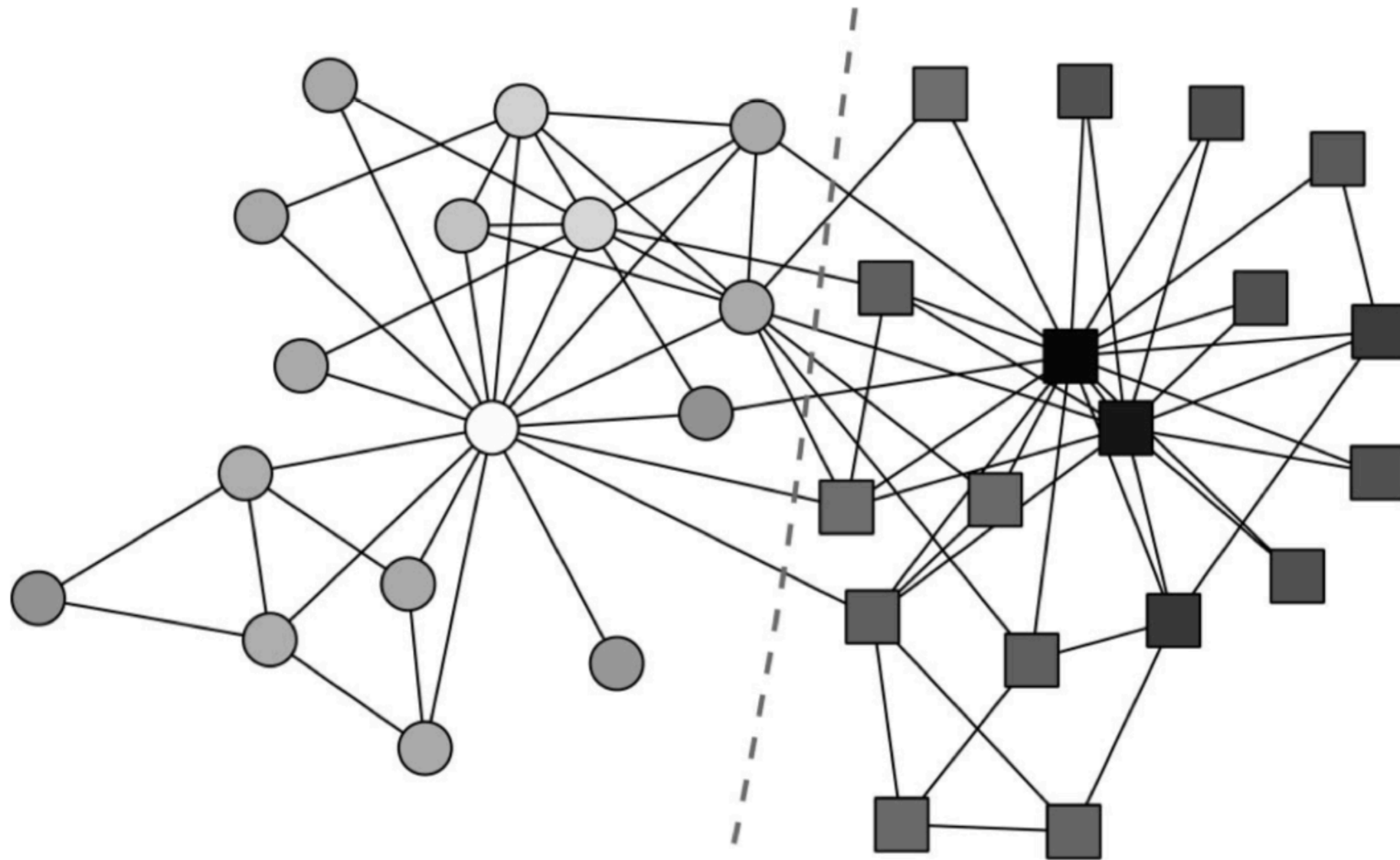


Modularity

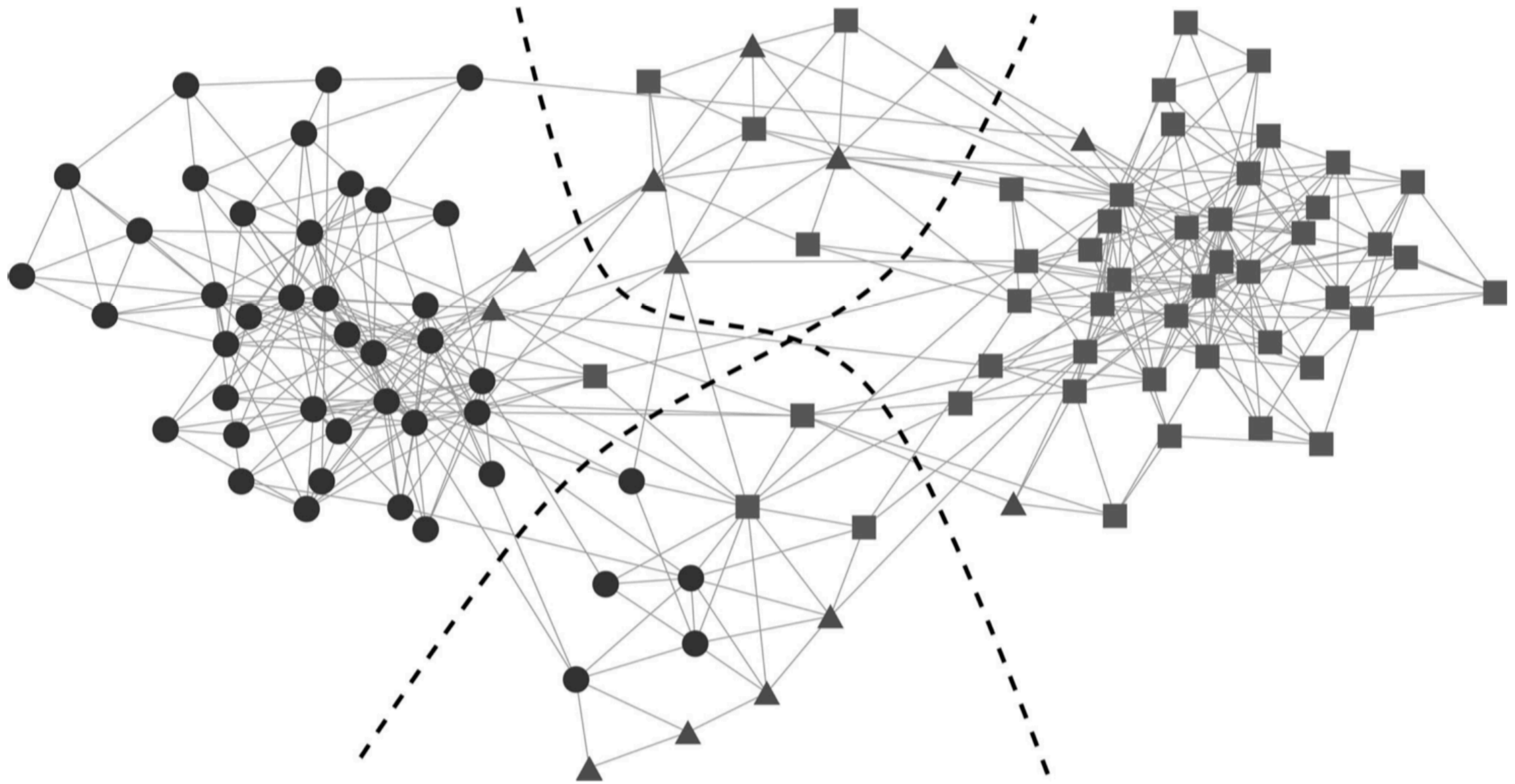
- Modularity compares observed community edges to what would be expected at random
- Modularity is the fraction of within-group edges minus the fraction expected at random (if degree conserved but edges are randomized)
- Modularity-based community detection: find community groupings that maximize modularity



Karate club example



Political books



Modularity

- Can be slow/difficult to maximize—spectral methods have made much faster
- Resolution limit - as the network grows larger, it is harder for modularity-based community detection methods to find small communities

Betweenness Methods for Clustering

- Sometimes people ask—can you form communities by removing highest betweenness centrality nodes one at a time, where community formed when the network becomes disconnected
- Will generate a hierarchical structure of communities—can be advantageous! (also see hierarchical clustering methods)
- But also does not tend to perform as well as modularity or information theory based methods in tests
- <https://colab.research.google.com/drive/1FMJatIYt0es1XbghNJH4h-kOEgiloH6i?authuser=1#scrollTo=QloVhQdoBwkP>

For next time...

- Reading
 - Sayama Chapter 15
 - Sayama Chapter 17
 - Think Complexity Chapter 2