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 Erdös & 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

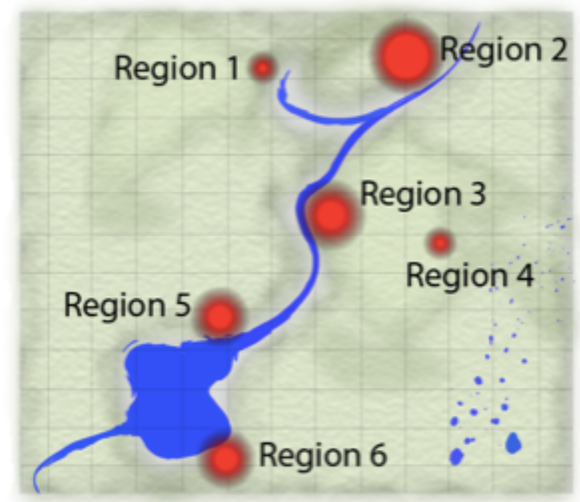
Types of network dynamics

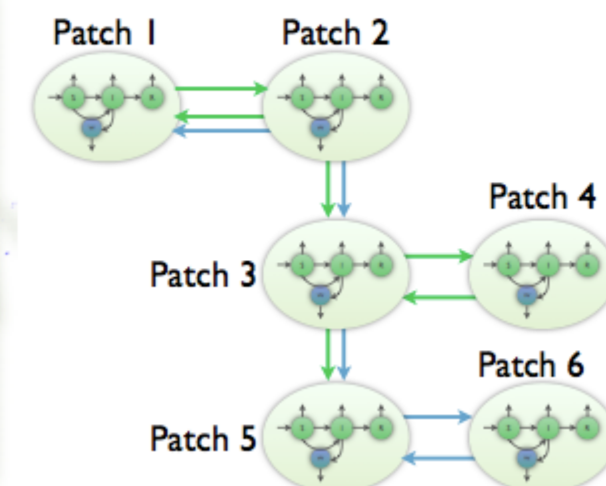
- **Dynamics on networks**: models where the processes of interest occur over a fixed network structure
- **Dynamics of networks**: models of the dynamic changes over time of the network topology itself
- **Adaptive networks**: models looking at the interplay of the two (both the processes on the network,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

Sayama, Chp. 16

Networks

- Network (graph)
  = nodes & edges



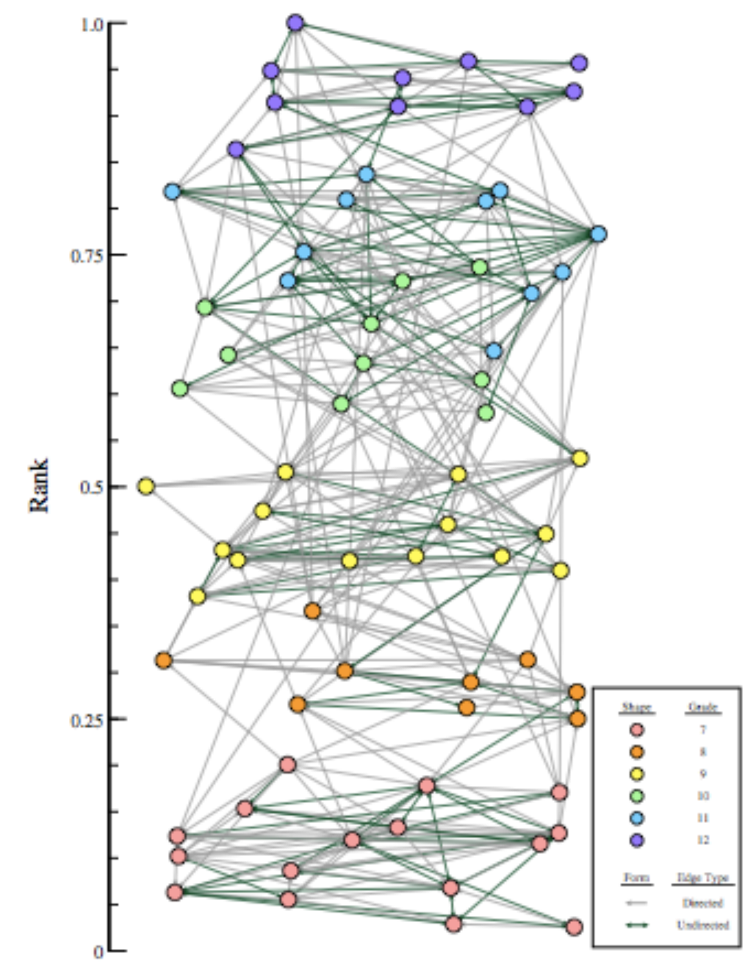


- Node (vertex) - an object,
  can be people, communities, locations, water sources, signaling molecules, genes, etc.

- 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

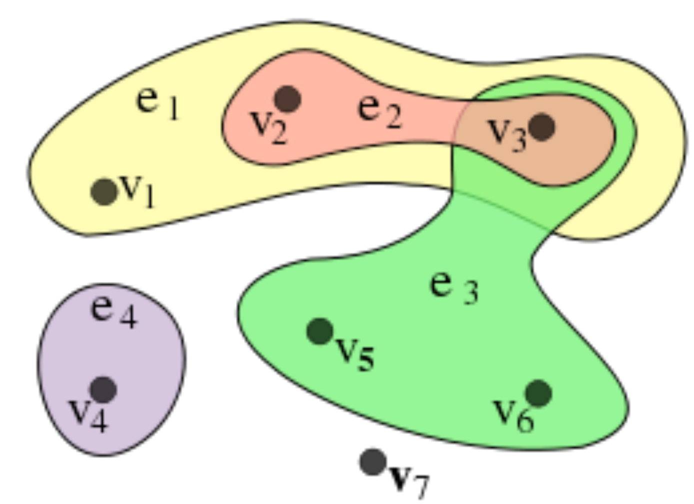


Types of Networks

- **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

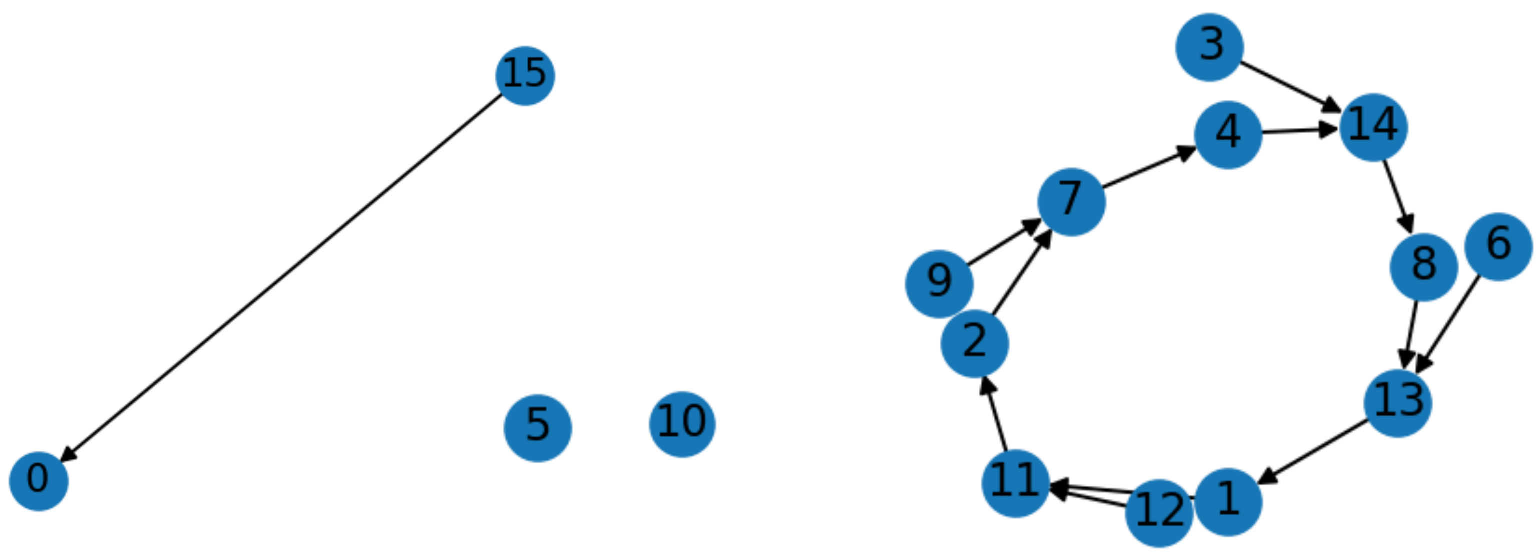
Types of Networks

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



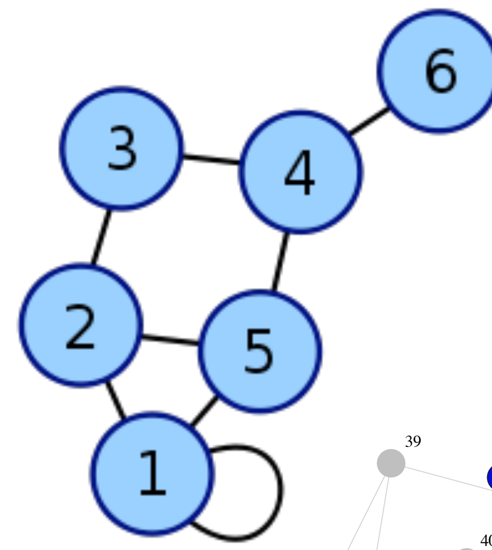
Key definitions/vocab

- 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

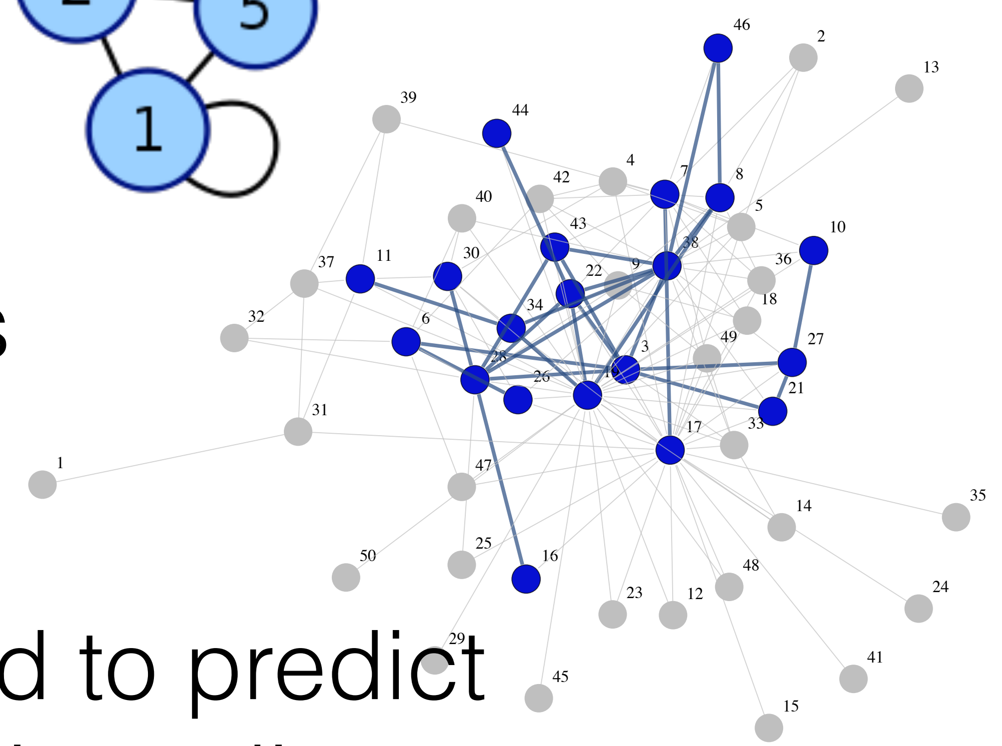


Key definitions/vocab

- **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



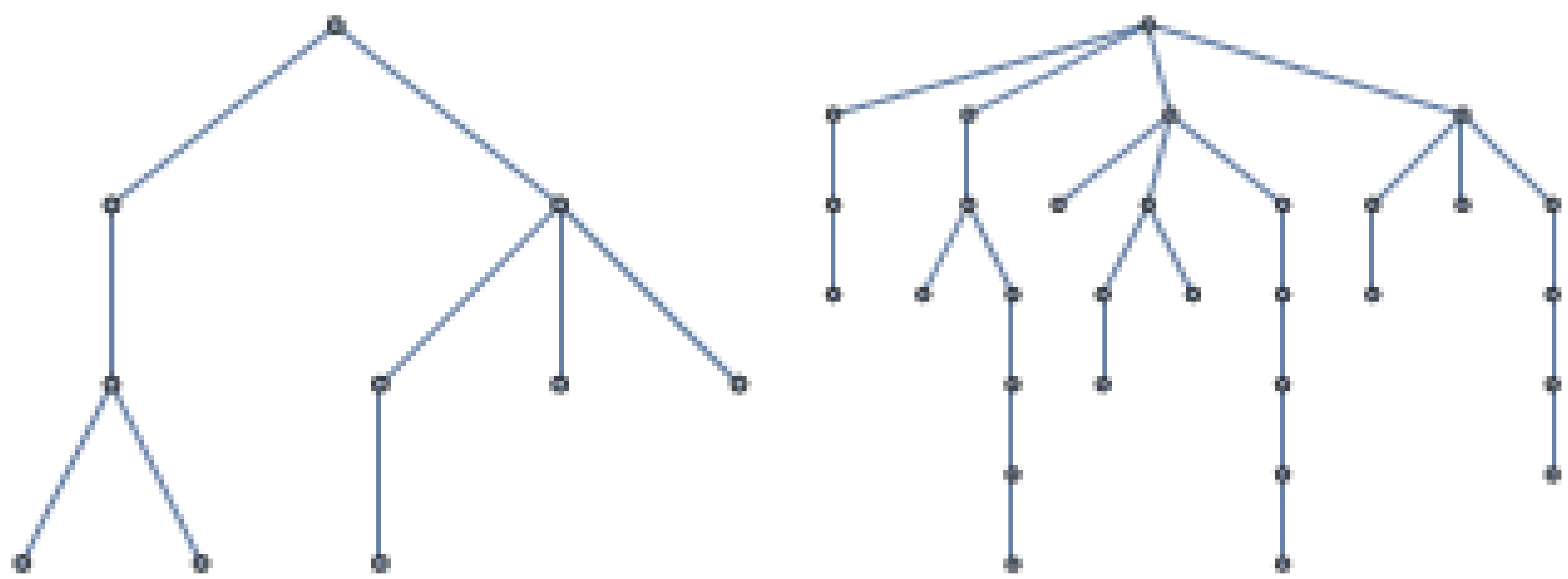
  - Often want to use neighborhood to predict effects on individual, e.g. infectious disease, behavioral influence

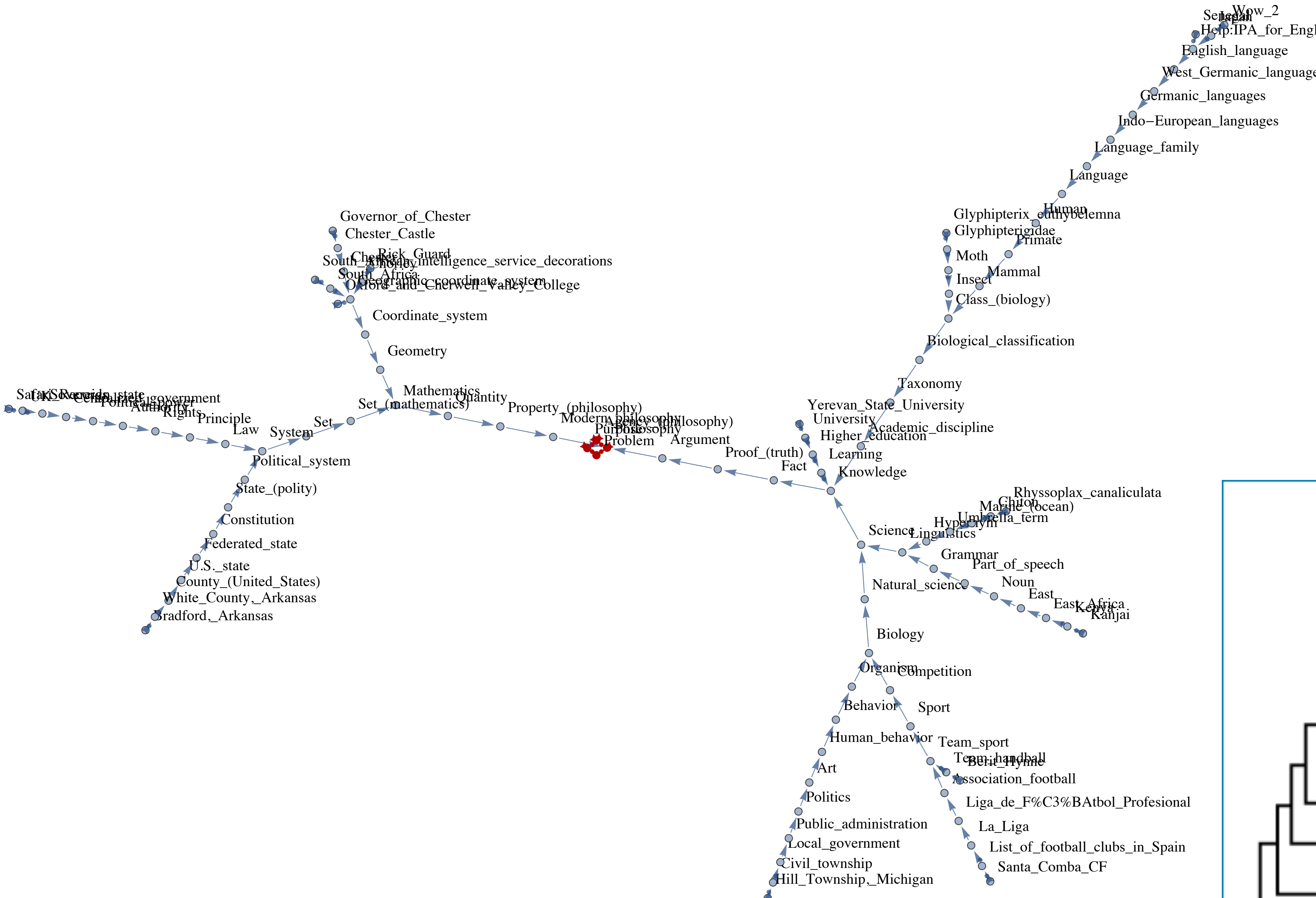
Key definitions/vocab

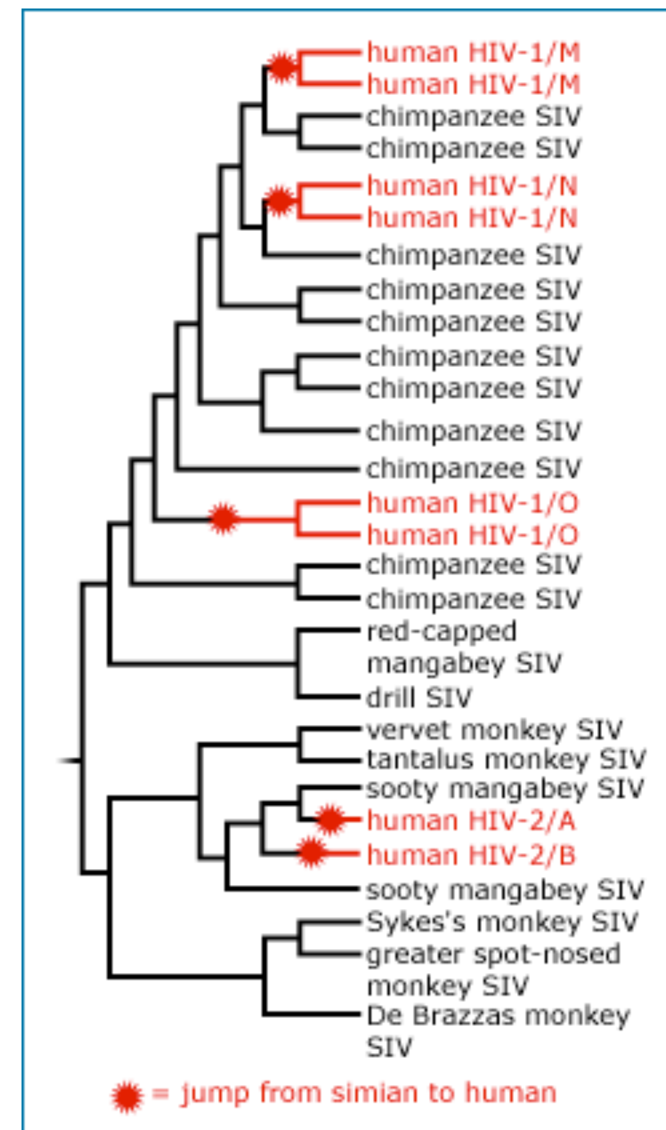
- **Walk** - List of edges sequentially connected to form a continuous route
- **Path** - Walk that doesn’t visit any node twice
- **Cycle** (sometimes called a **circuit**) - Walk that starts and ends at same node (called a **simple cycle** if no repeated nodes)

Trees & Forests

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

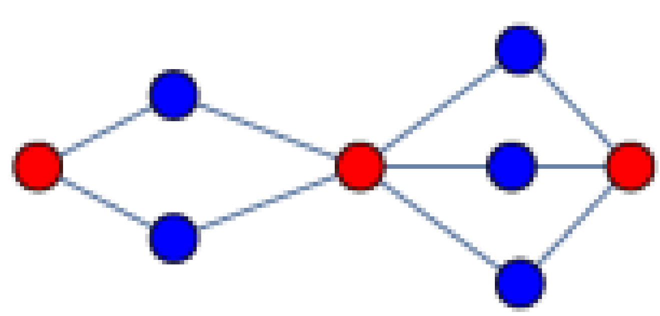
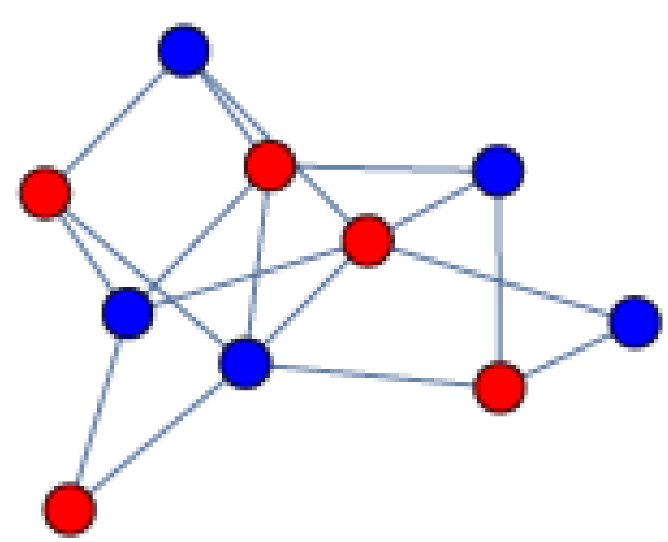
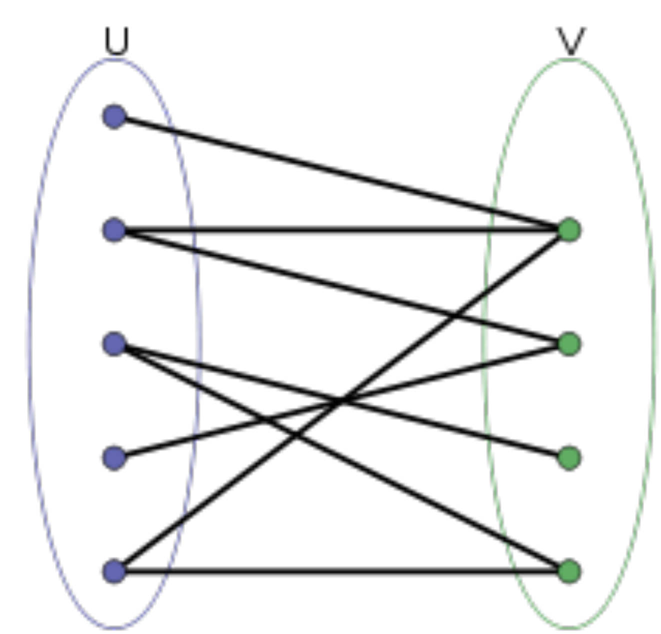



English_language
West_Germanic_language
Germanic_languages
Indo−European_languages
Language_family
Language
Glyphipterigidae
Primate
Moth
Mammal
Insect
Class_(biology)
Biological_classification
Taxonomy
Yerevan_State_University
University
Academic_discipline
Learning
Knowledge
Proof_(truth)
Fact
Argument
Problem
Property_(philosophy)
Quantity
Mathematics
Set_(mathematics)
Set
System
Law
Principle
Political_system
State_(polity)
Constitution
Federated_state
U.S._state
County_(United_States)
White_County,_Arkansas
Bradford,_Arkansas
Governor_of_Chester
Chester_Castle
Coordinate_system
Geometry
Rhyssoplax_canaliculata
Science
Grammar
Part_of_speech
Noun
East
Kanjai
Natural_science
Biology
Competition
Behavior
Sport
Human_behavior
Team_sport
Association_football
Art
Politics
Liga_de_F%C3%BAtbol_Profesional
La_Liga
Public_administration
List_of_football_clubs_in_Spain
Local_government
Civil_township
Santa_Comba_CF
Hill_Township,_Michigan


human HIV-1/M
human HIV-1/M
chimpanzee SIV
chimpanzee SIV
human HIV-1/N
human HIV-1/N
chimpanzee SIV
chimpanzee SIV
chimpanzee SIV
chimpanzee SIV
chimpanzee SIV
chimpanzee SIV
chimpanzee SIV
human HIV-1/O
human HIV-1/O
chimpanzee SIV
chimpanzee SIV
red-capped mangabey SIV
drill SIV
vervet monkey SIV
tantalus monkey SIV
sooty mangabey SIV
human HIV-2/A
human HIV-2/B
sooty mangabey SIV
Sykes's monkey SIV
greater spot-nosed monkey SIV
De Brazzas monkey SIV
= jump from simian to human

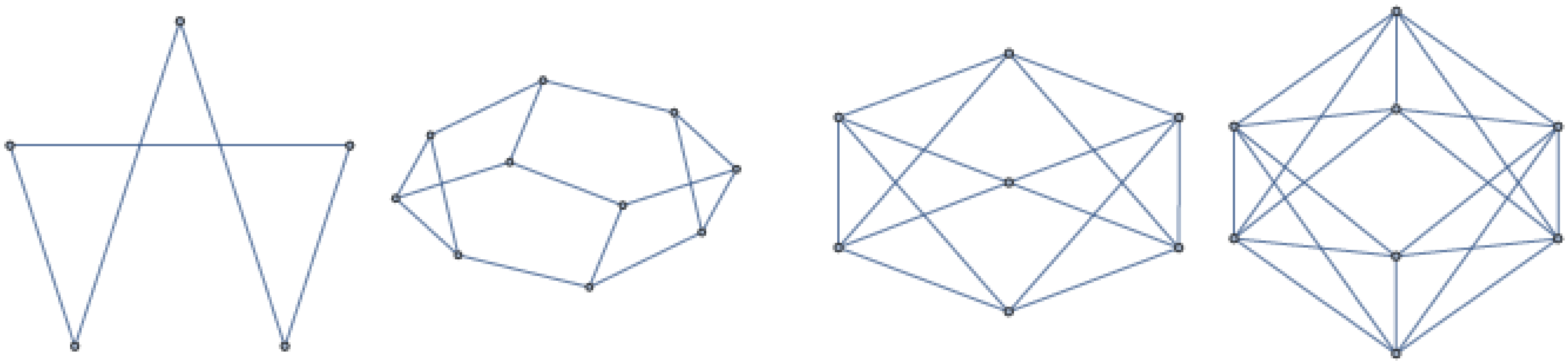
Bipartite (n-partite) graph

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

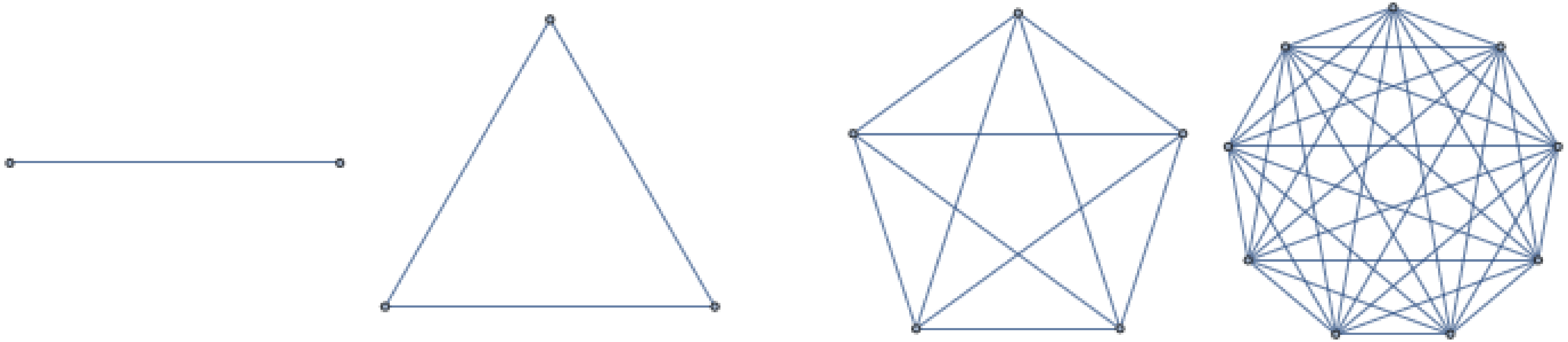
Regular graph

- All nodes have the same degree



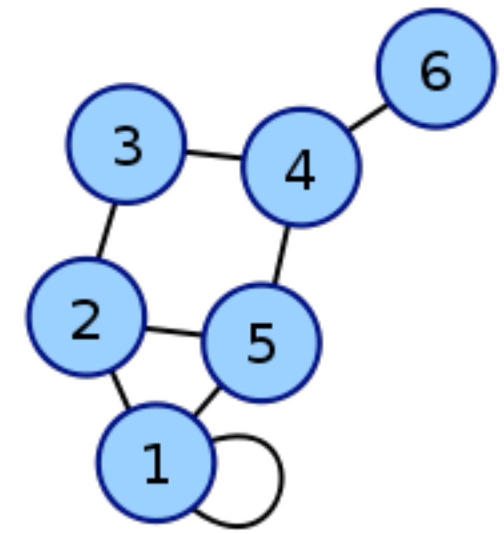
Complete graph

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



Adjacency Matrix

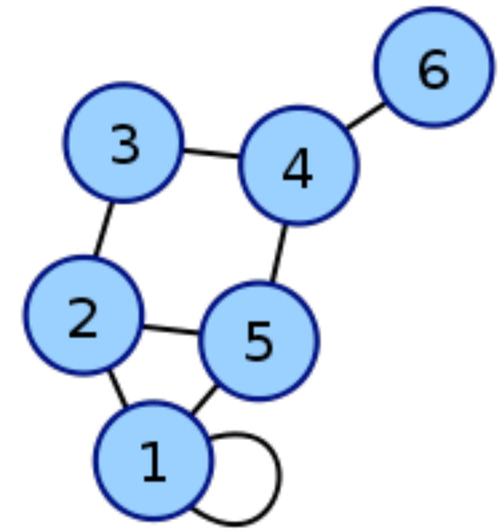
- Matrix representing the graph structure
- Can reconstruct the graph from the matrix & vice versa
- Pattern, eigenvalues, etc. of adjacency matrix can often tell you about the graph



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

Adjacency Matrix

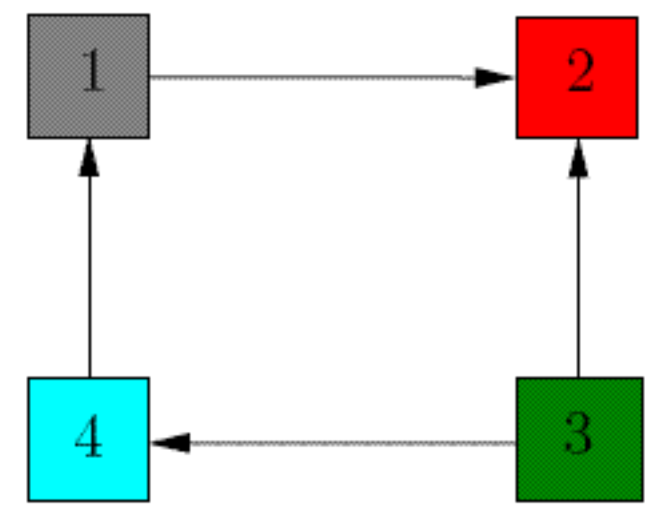
- Undirected graph - adjacency matrix is symmetric
- Directed graph - asymmetric
- Weighted graph - takes non 0/1 values to match edge weights



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

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 G1 and G2 are ‘the same’ (isomorphic) if $A_1 = P A_2 P^{-1}$

- Can use to find number of connected components, bipartite-ness, etc.

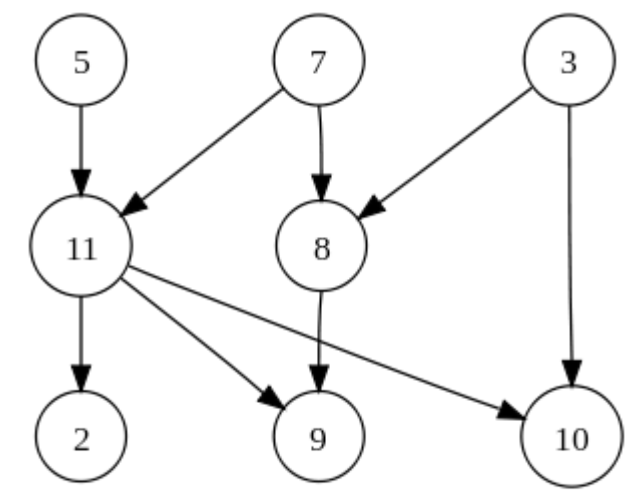
Network Metrics

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

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


Nellie_Mae_Rowe
Georgia_(U.S._state)
Southeastern_United_States
Southeast_(disambiguation)
Santa_Comba_CF
List_of_football_clubs_in_Spain
La_Liga
Hill_Township,_Michigan
Civil_township
Local_government
Public_administration
Politics
Association_football
Team_sport
Sport
Competition
Behavior
Biology
Physics
Natural_science
East_Africa
Science
Hypernym
Umbrella_term
Knowledge
Learning
Fact
Proof_(truth)
Higher_education
Academic_discipline
University
Yerevan_State_University
Argument
Taxonomy
Biological_classification
Class_(biology)
Insect
Mammal
Moth
Glyphipterigidae
Human
Person
Order_of_chivalry
Language
Chinese_language
Language_family
Robert_Thom_(translator)
Indo-European_languages
Germanic_languages
West_Germanic_language
English_language
Help:IPA_for_English
Senegal
Japan
Wow_2
Safari_Records
UK
Sovereign_state
Centralized_government
Political_power
Authority
Rights
Principle
Law
Political_system
Set
Set_(mathematics)
Mathematics
Quantity
Property
Geometry
Coordinate_system
Oxford_and_Cherwell_Valley_College
Geographic_coordinate_system
Process_(philosophy)
Abstraction
Entity
Political_organization
Political_party
Major_party
Two_party_system
Democratic_Party_(United_States)
Rhode_Island_Democratic_Party

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 it's neighbors' centralities. We can write that like this:

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

- 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:

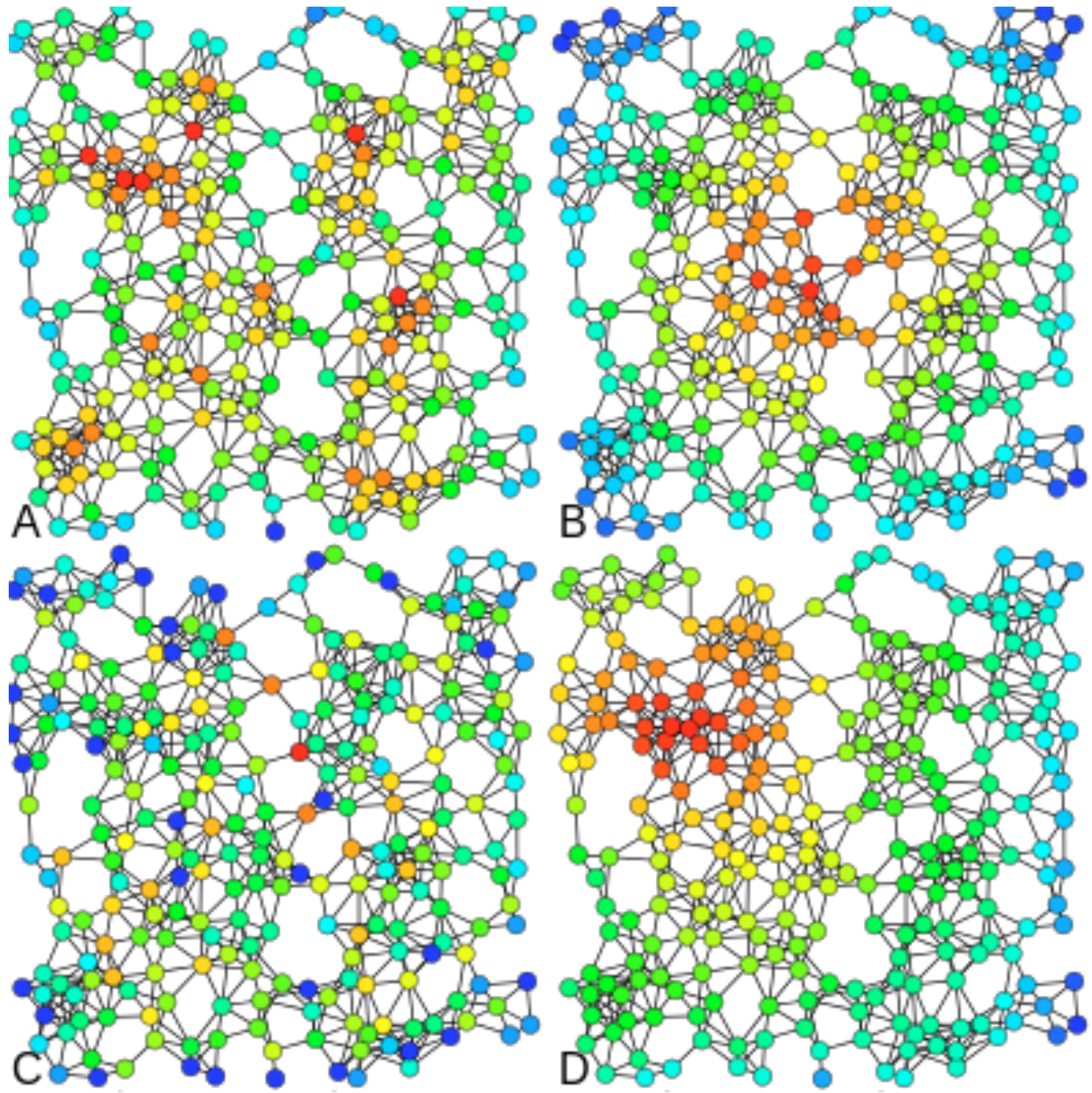
$$A\mathbf{x} = \lambda\mathbf{x}$$

Eigenvector centrality

- Eigenvector centrality of a node x_i is the ith 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)

Degree

Closeness

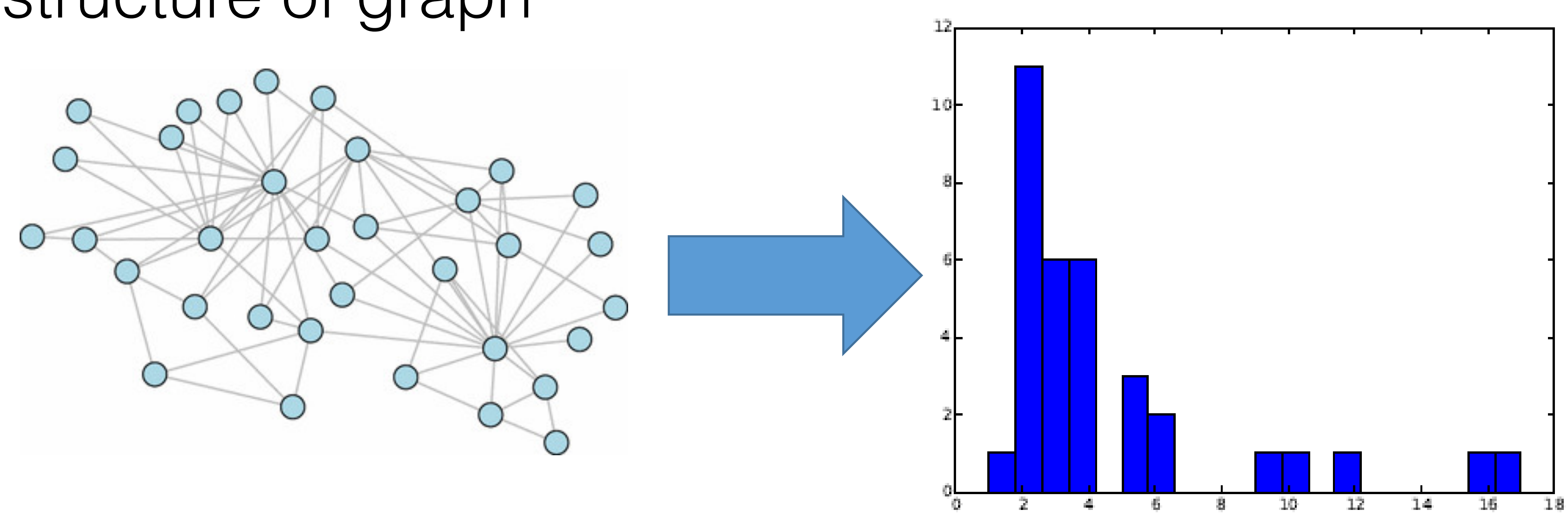


Betweenness

Eigenvector

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



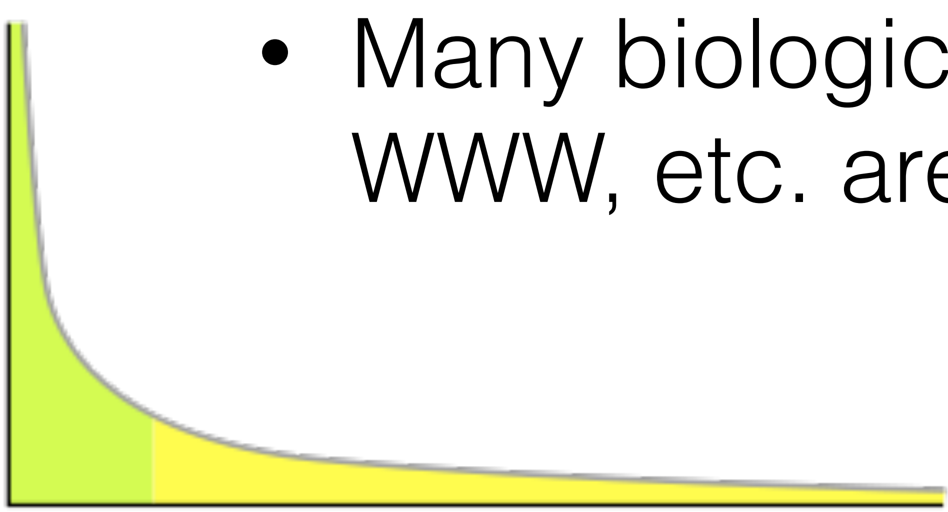
Power Law degree distribution

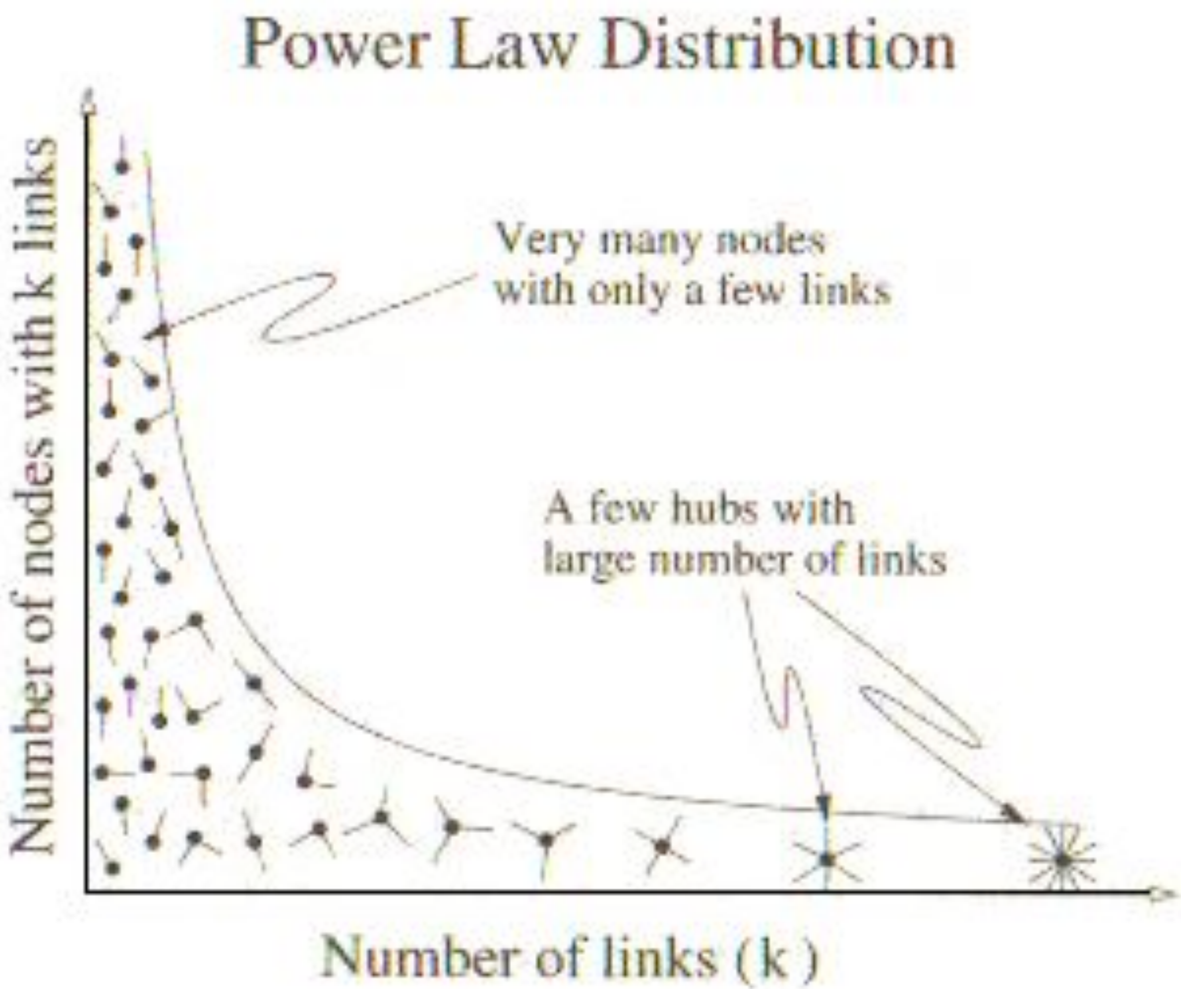
- **Scale free networks** - power law degree distribution

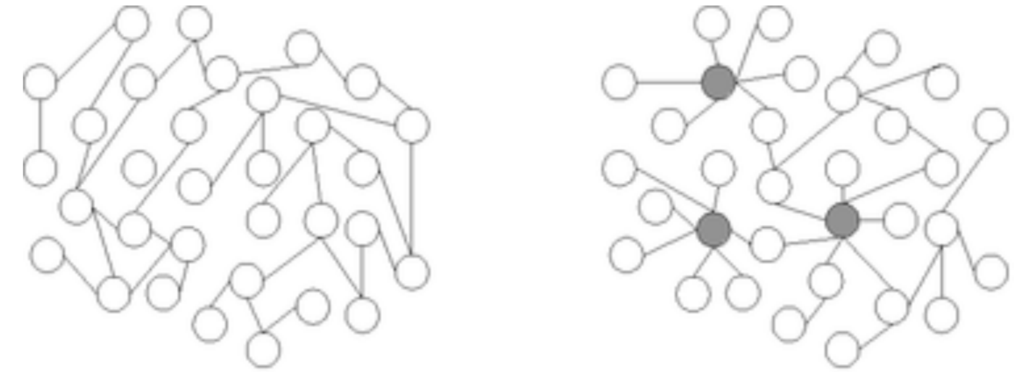
$$P(k) \sim k^{-\gamma}$$

  - Long tail results in both very sparse nodes and hub nodes

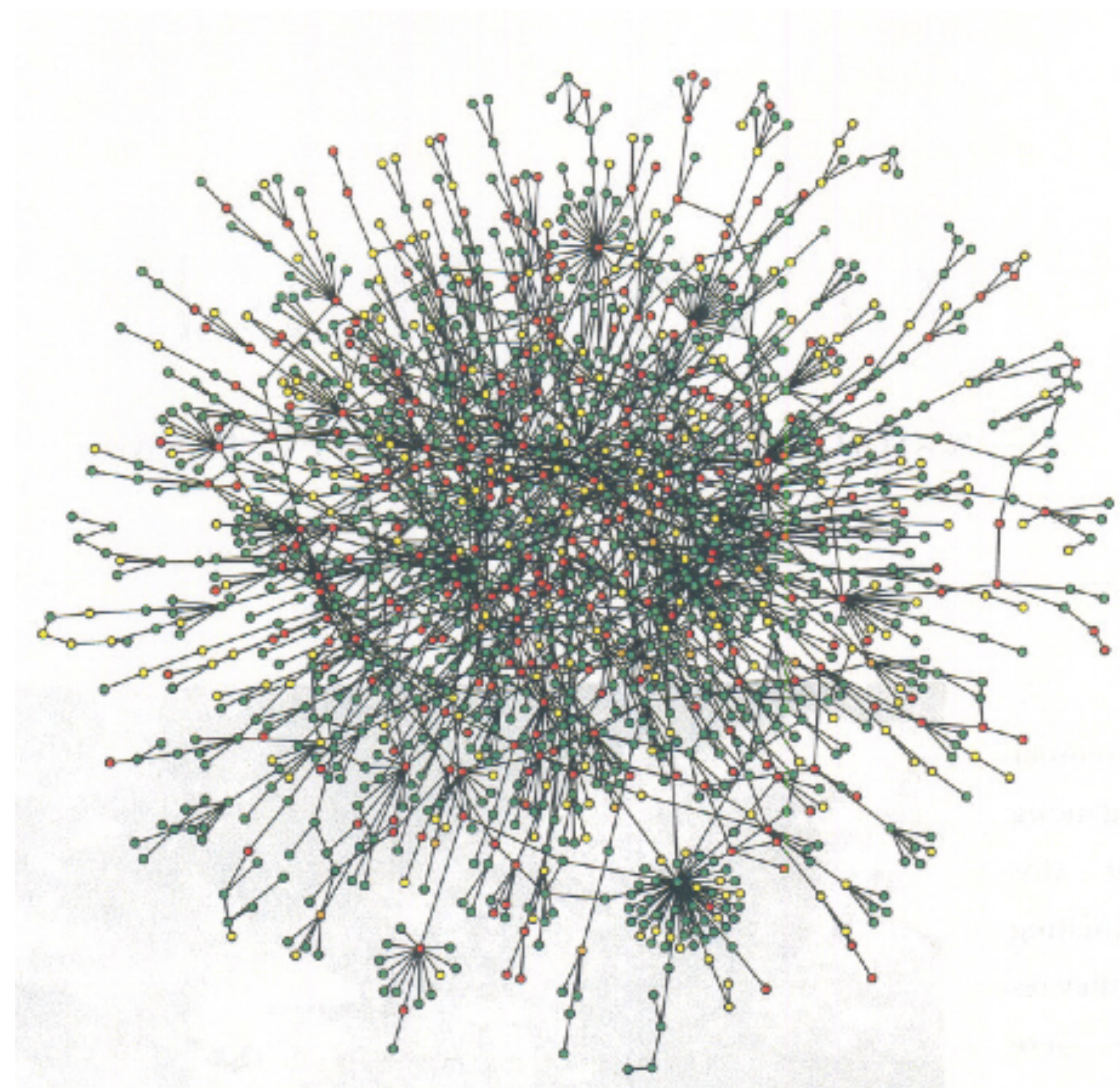
  - Many biological networks, social networks, WWW, etc. are scale free



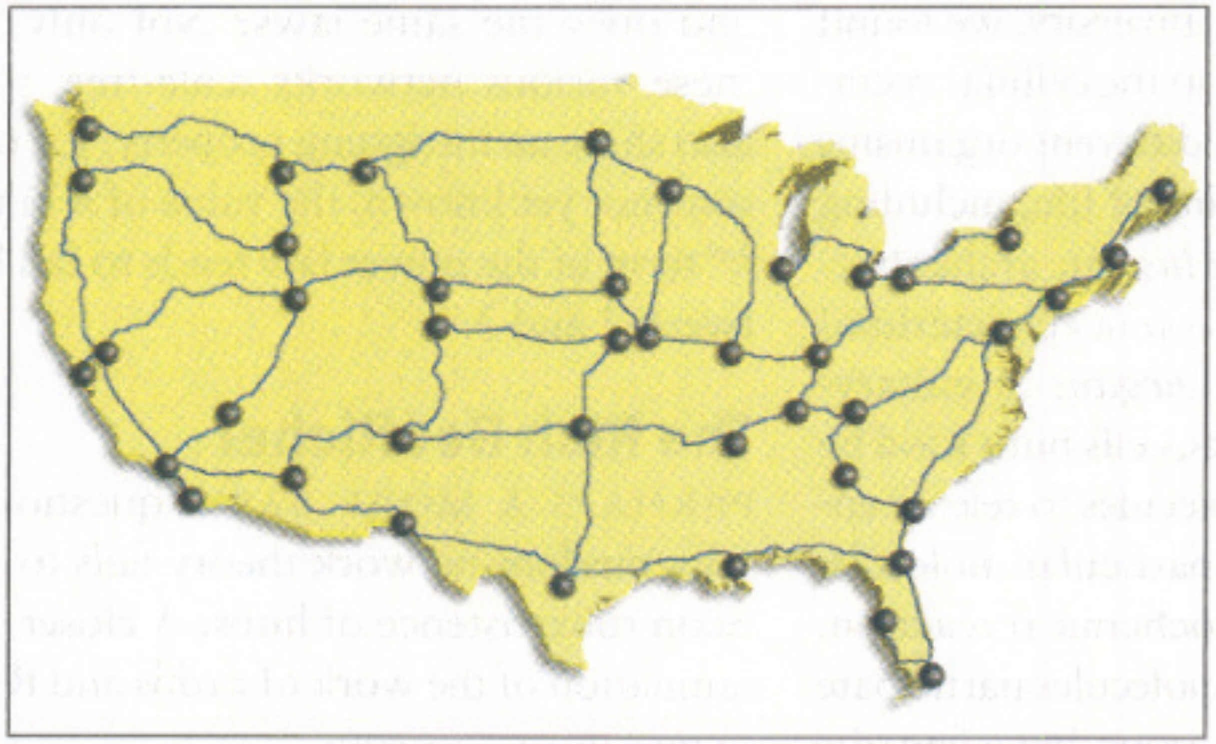




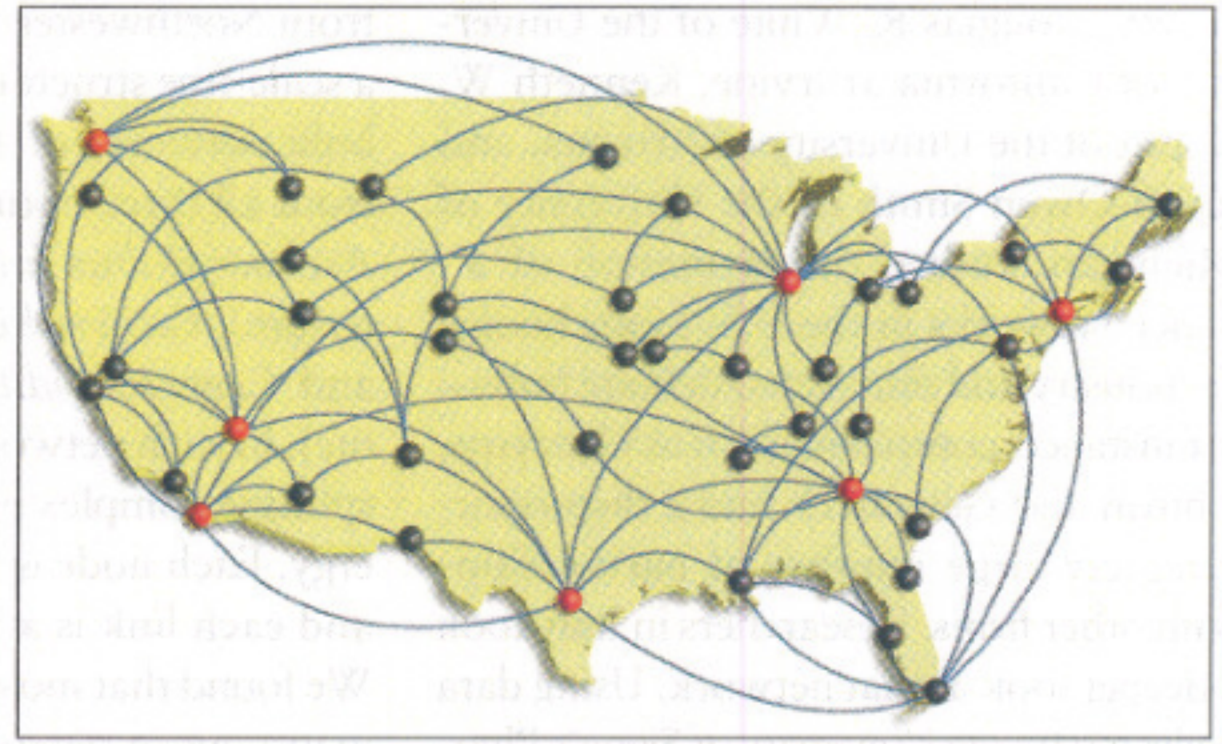
(a) Random network (b) Scale-free network



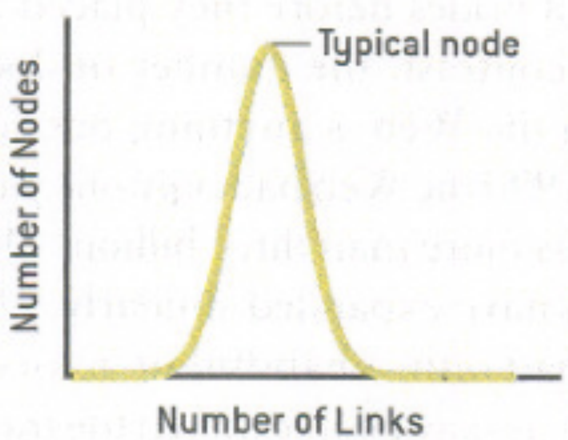
Random Network



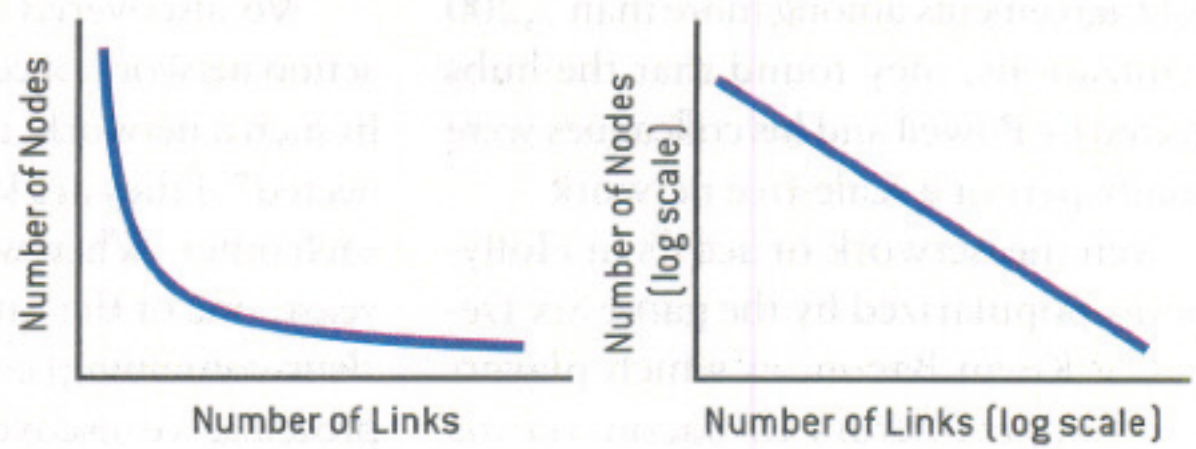
Scale-Free Network



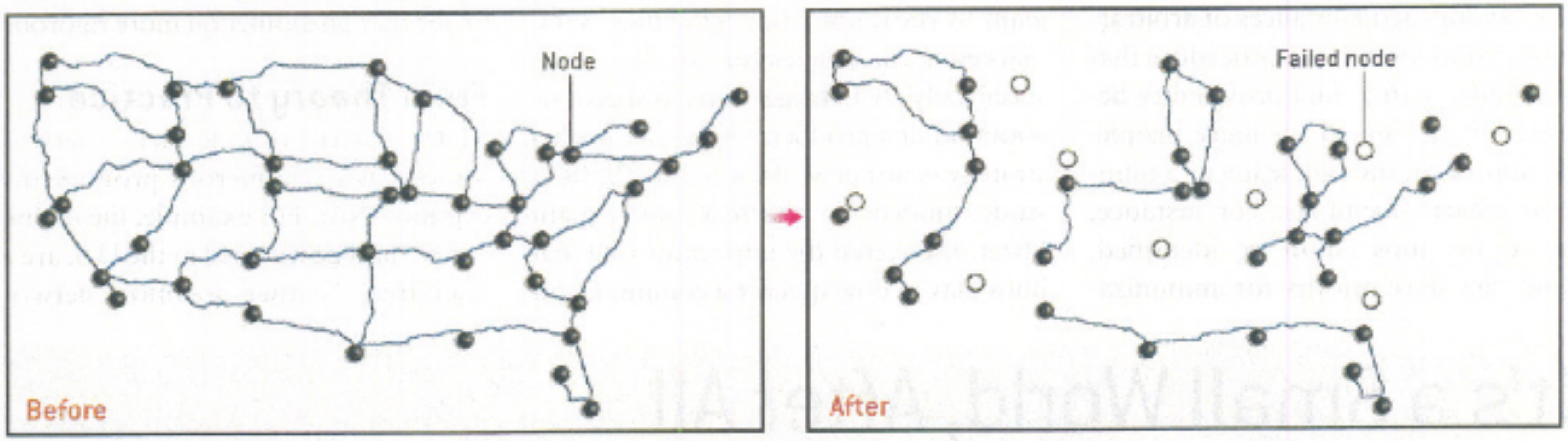
Bell Curve Distribution of Node Linkages



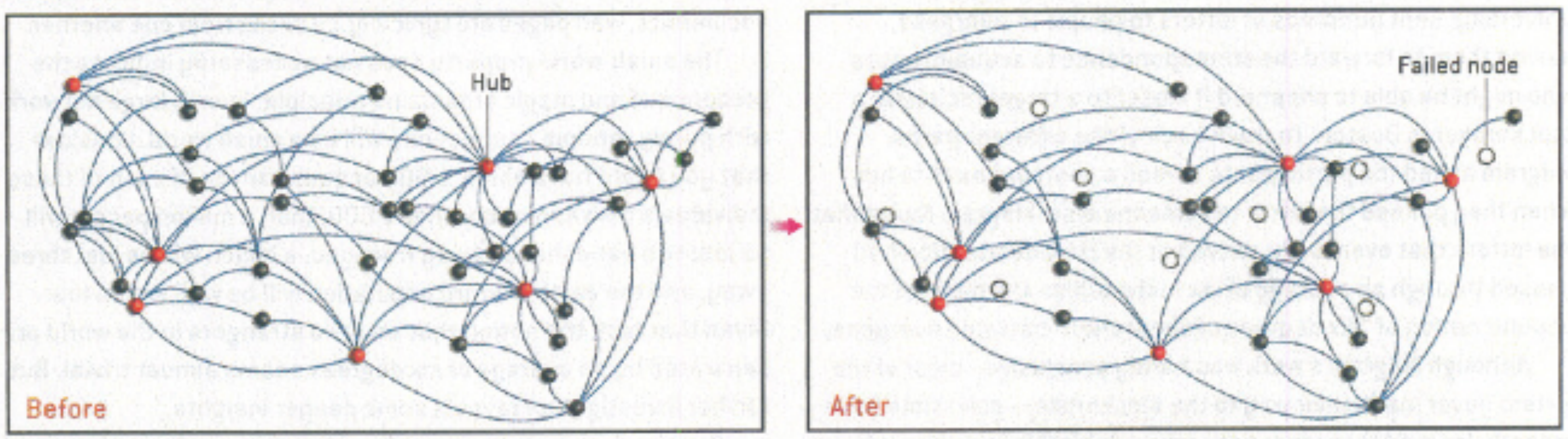
Power Law Distribution of Node Linkages



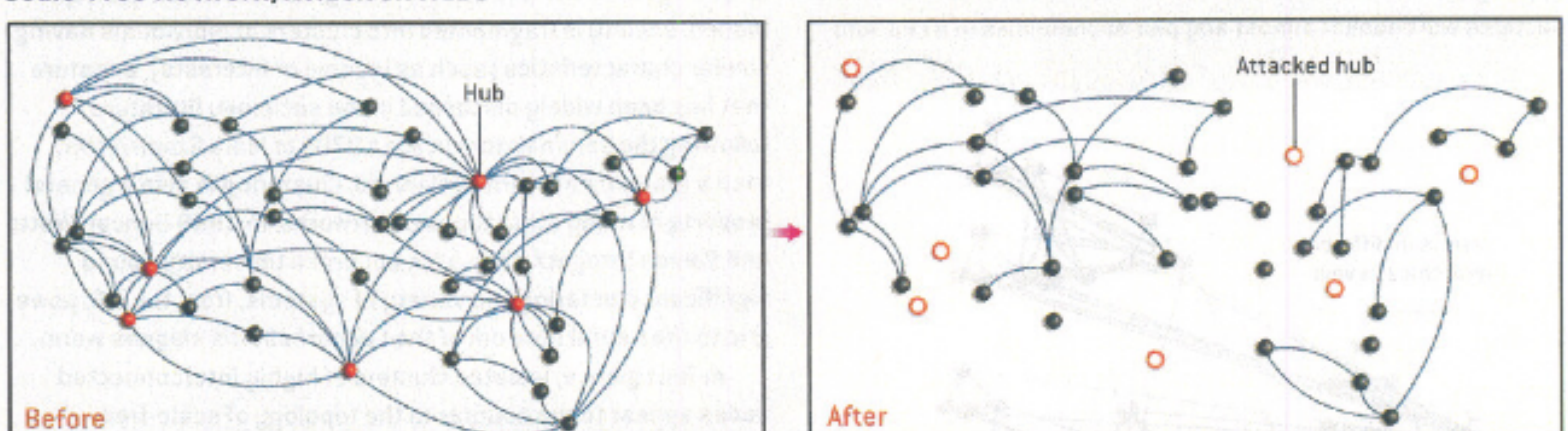
Random Network, Accidental Node Failure



Scale-Free Network, Accidental Node Failure



Scale-Free Network, Attack on Hubs

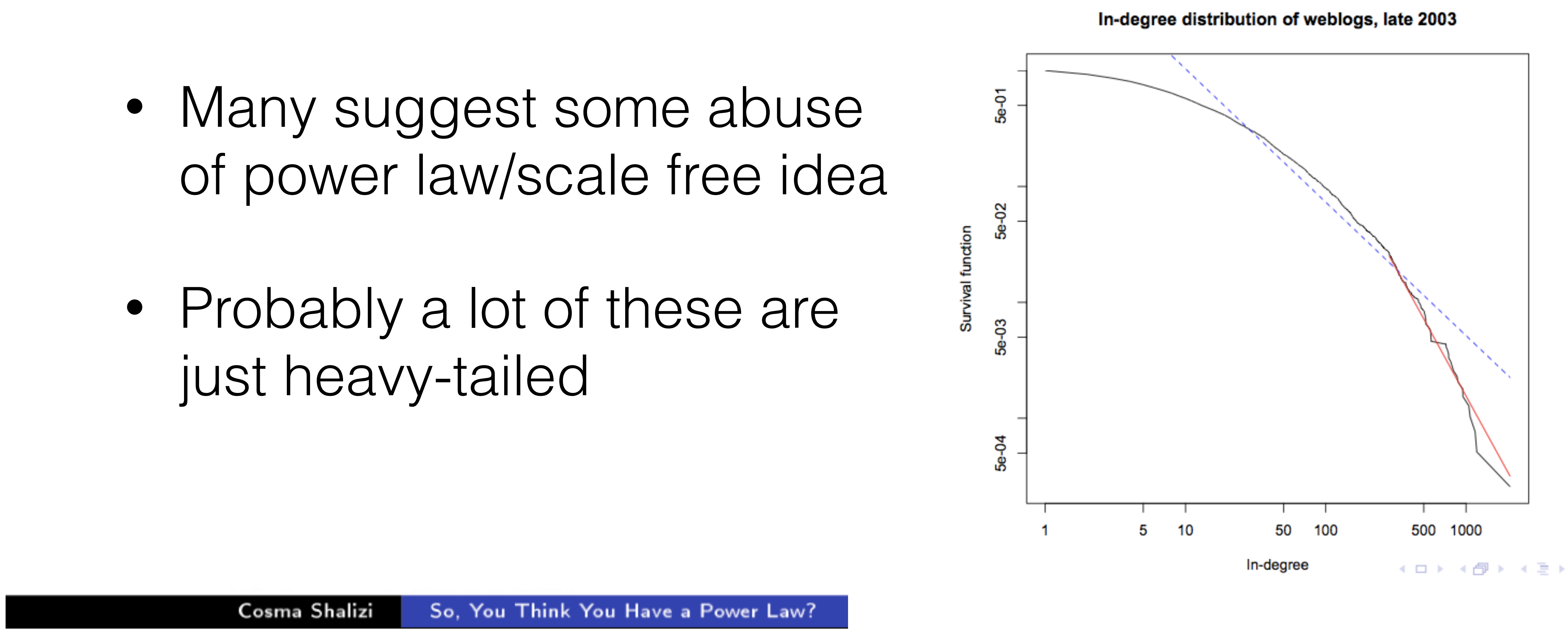


Scale Free Networks

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

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

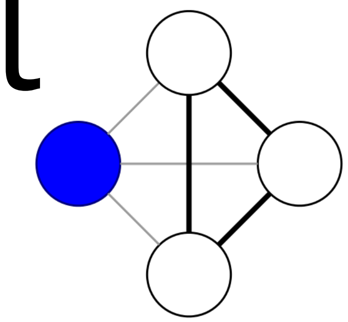


Clustering in networks

Clustering in networks

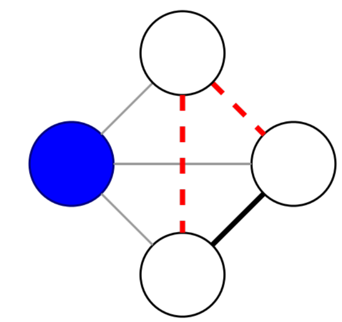
- Many different ways to look at clustering
- How do node traits (degree, covariates) cluster based on edges? E.g. do smokers tend to be friends with other smokers? Do individuals cluster by popularity?
- Community detection - finding clusters (groups) of nodes that are highly connected within the group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

Clustering Coefficient



c = 1

- Based on the number of triangles in the network

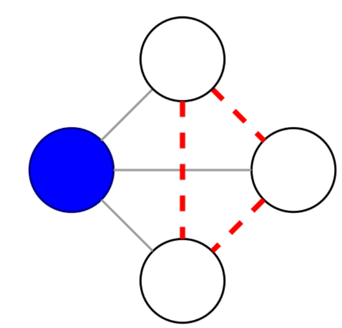


- How many of my friends are also friends?

c = 1/3

- Global clustering coefficient

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



- Local clustering coefficient

c = 0

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

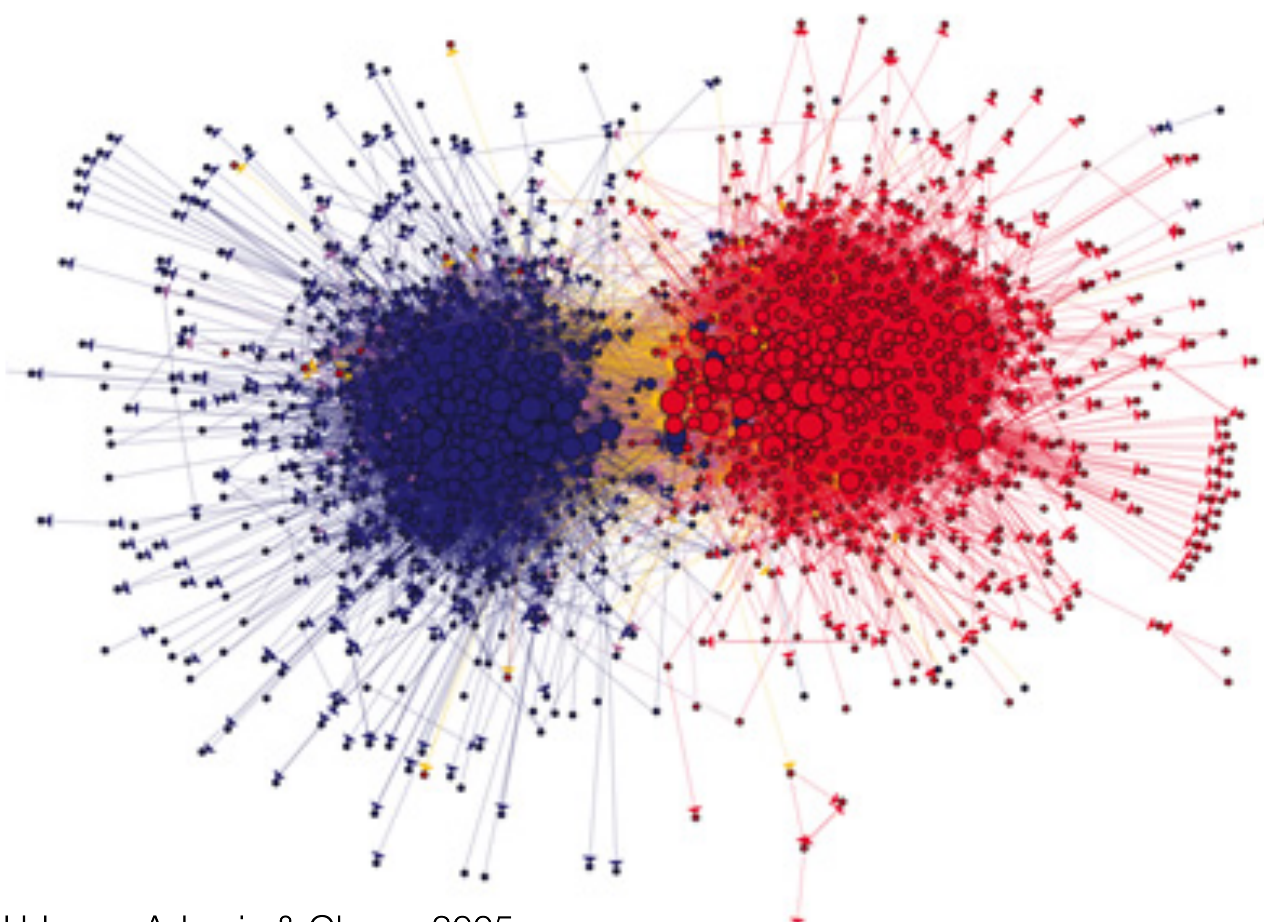
Assortativity

- **Assortativity** - measures network-level tendency for nodes to 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

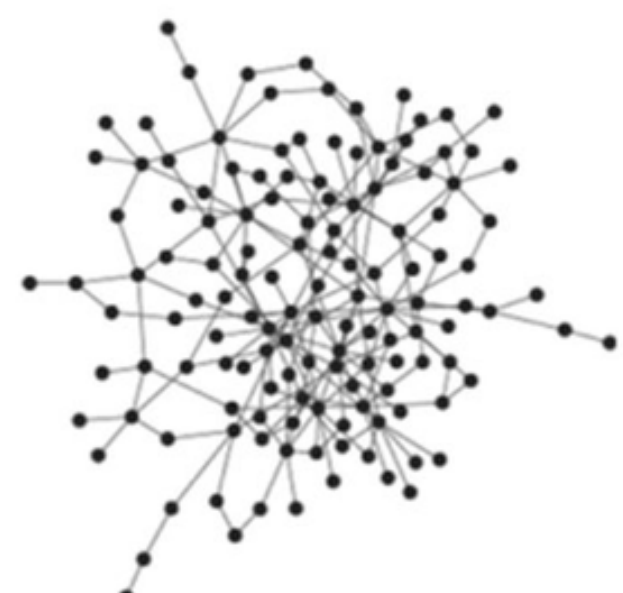
M. E. J. Newman, Mixing patterns in networks Physical Review E, 67 026126, 2003

Assortativity

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



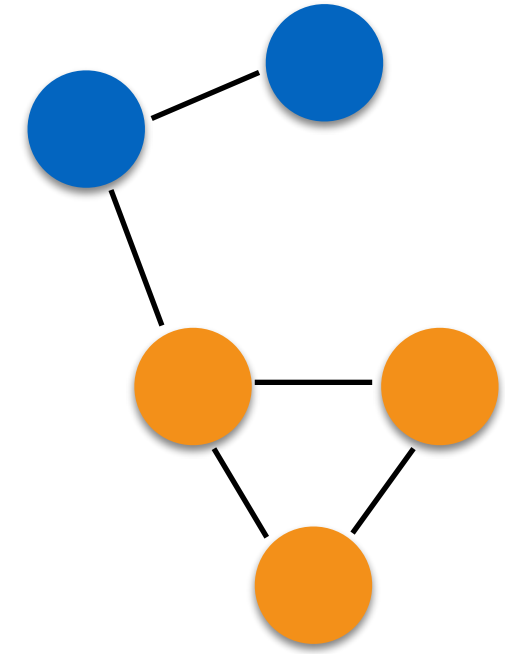
Political blogs - Adamic & Glance 2005



assortative

disassortative

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 $\sum_{ij} e_{ij} = 1, \quad \sum_j e_{ij} = a_i, \quad \sum_i e_{ij} = b_j$

Assortativity

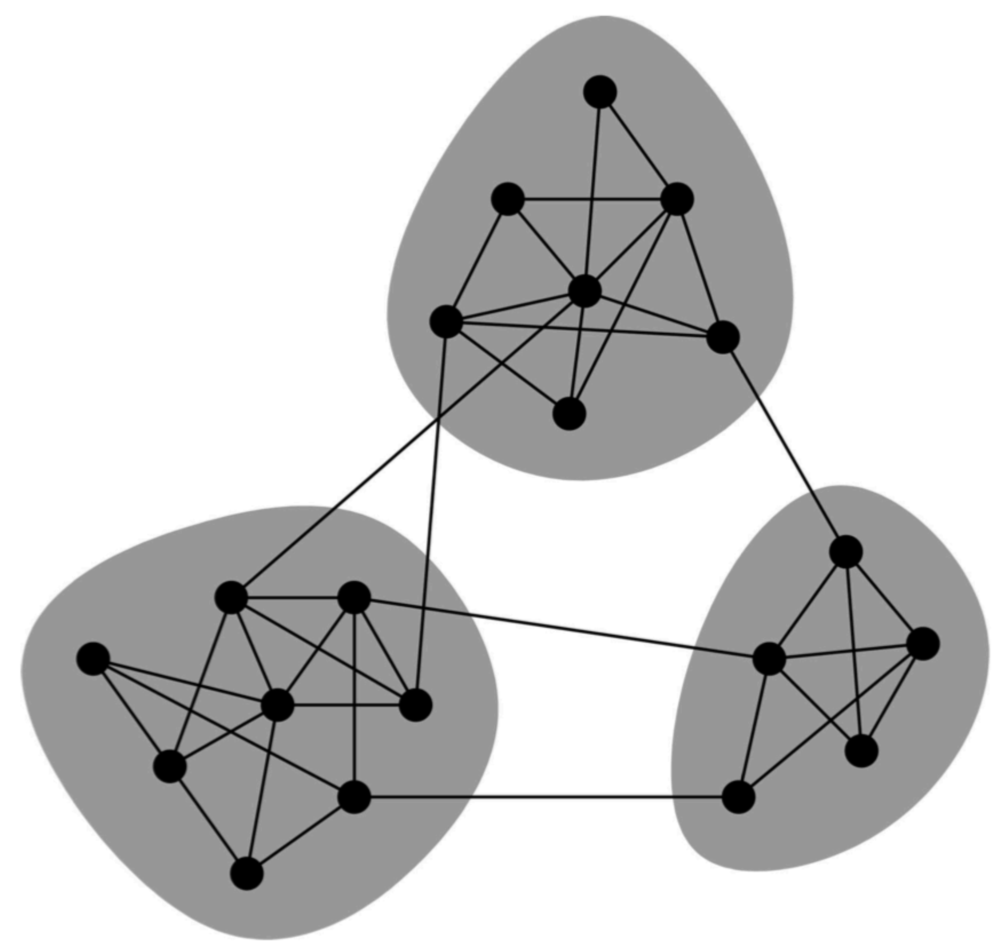
- Defined based on a mixing matrix - entries are the fraction of edges in a network linking type i to type j

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

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

Modularity

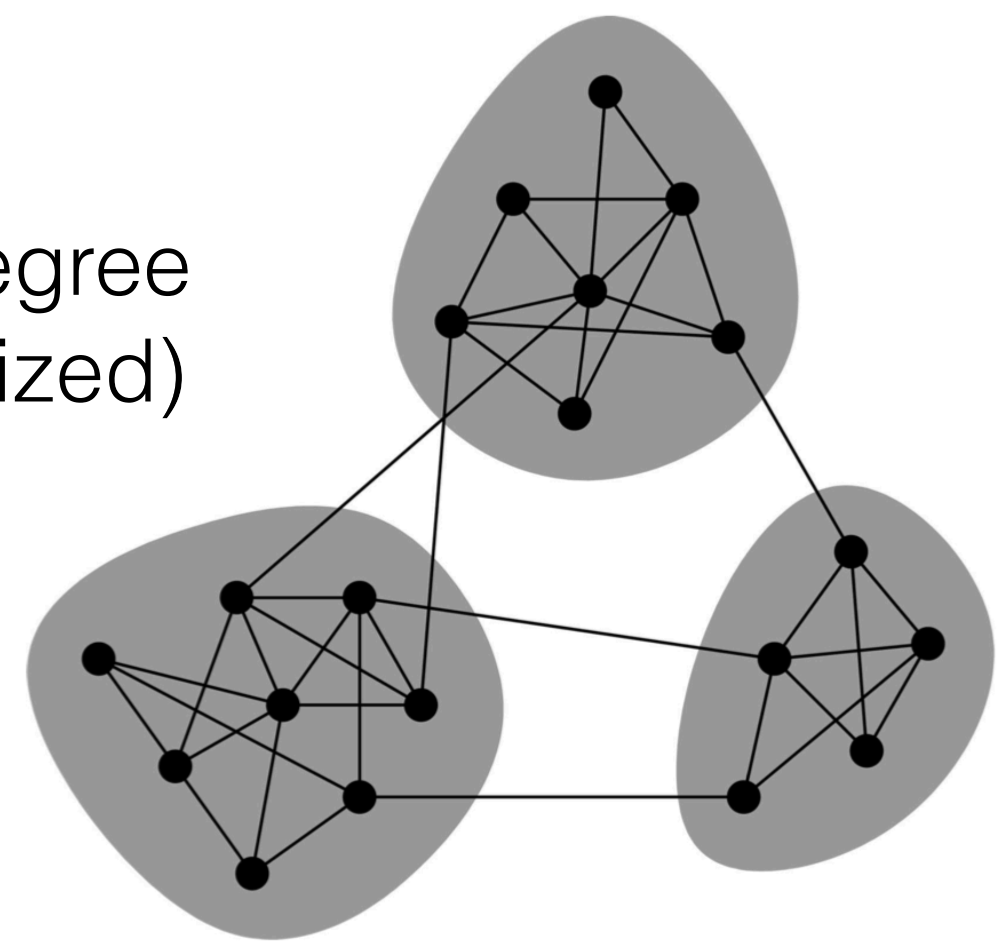
- How to decide communities (clusters) in a network?
- We want communities to have more in-group edges than between-group edges
- We could minimize between group edges, but this would lead to just putting all nodes in one community



Newman, Mark EJ. "Modularity and community structure in networks." *PNAS* 103.23 (2006): 8577-8582.

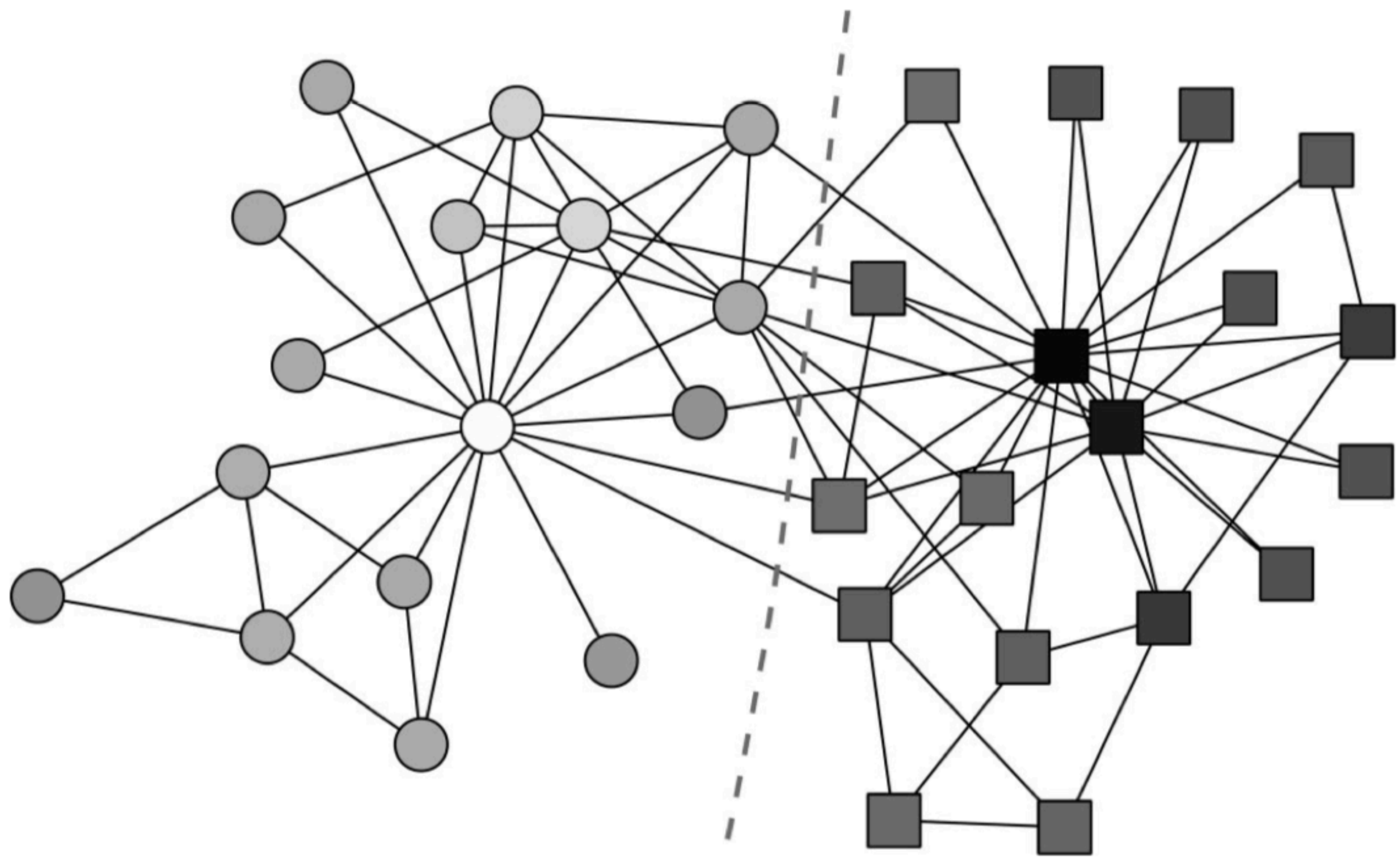
Modularity

- 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



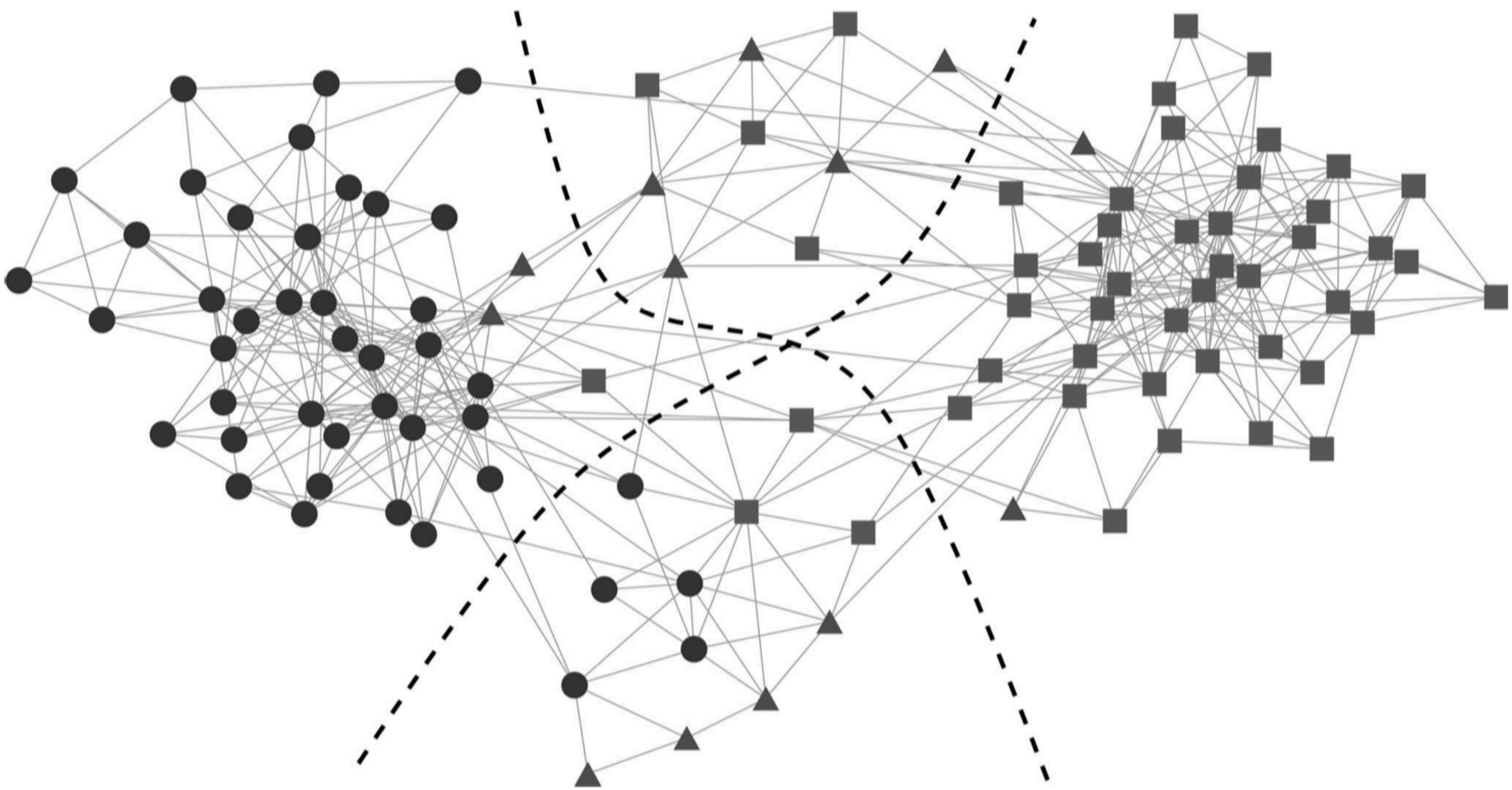
Newman, Mark EJ. "Modularity and community structure in networks." *PNAS* 103.23 (2006): 8577-8582.

Karate club example



Newman, Mark EJ. "Modularity and community structure in networks." *PNAS* 103.23 (2006): 8577-8582.

Political books



Newman, Mark EJ. "Modularity and community structure in networks." *PNAS* 103.23 (2006): 8577-8582.

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

- Since someone brought up in class—idea is to 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/1FMJatlYt0es1XbghNJH4h-kOEgiloH6i?authuser=1#scrollTo=QIoVhQdoBwkP

For next time…

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