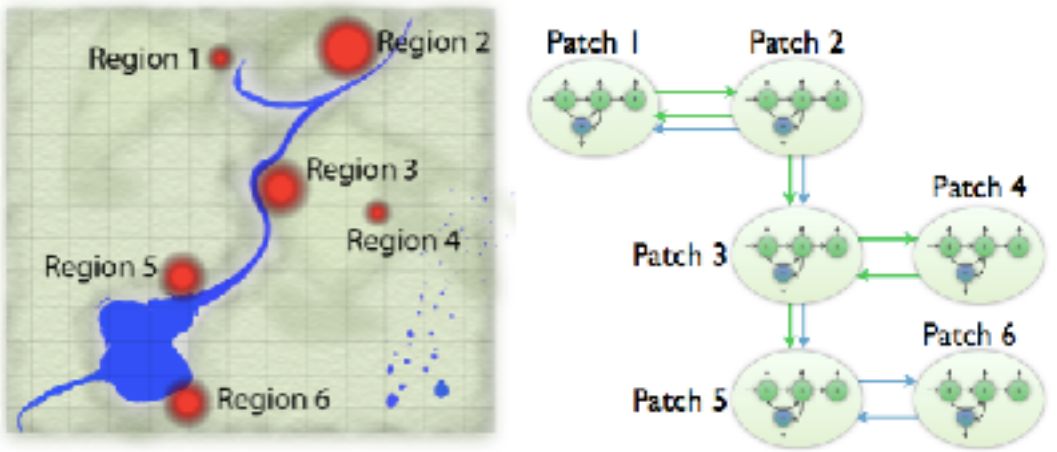
Networks

Marisa Eisenberg

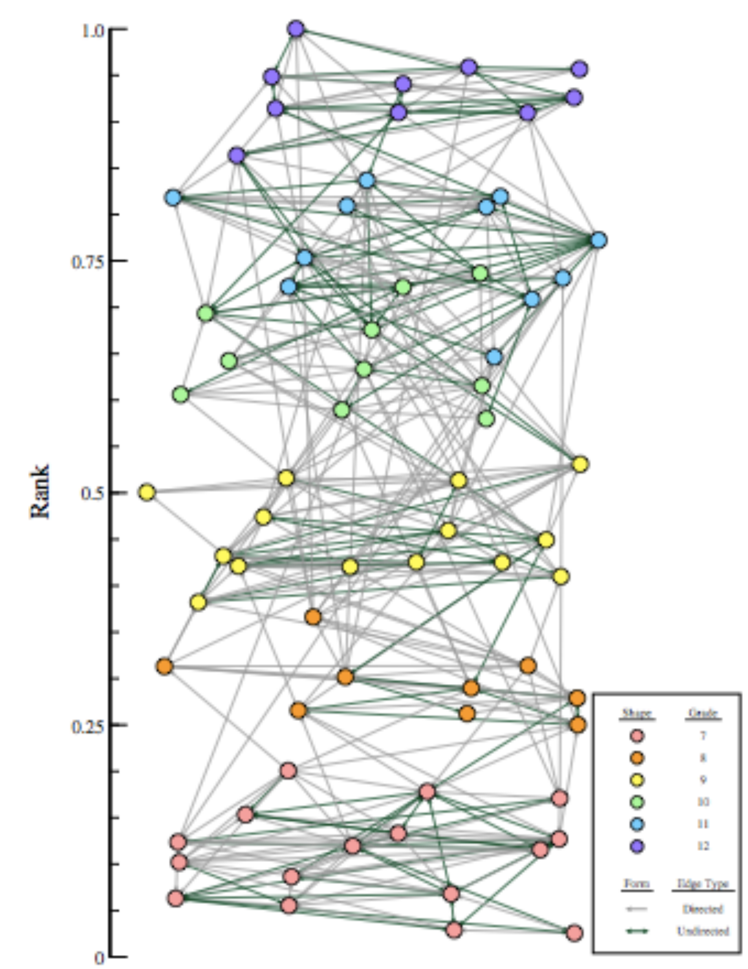
Network Modeling



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

Types of Graphs (Networks)

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

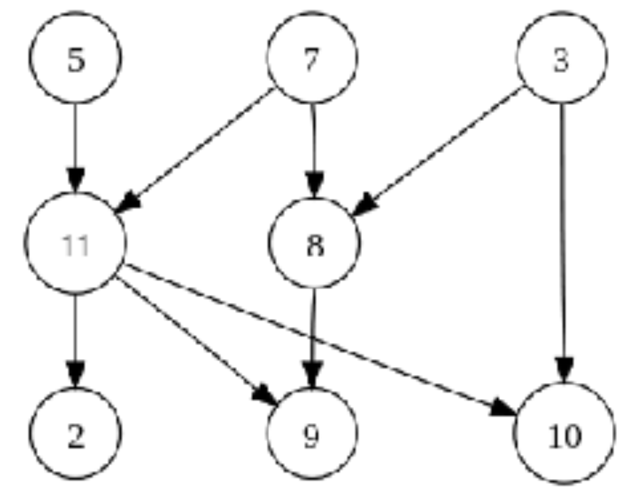


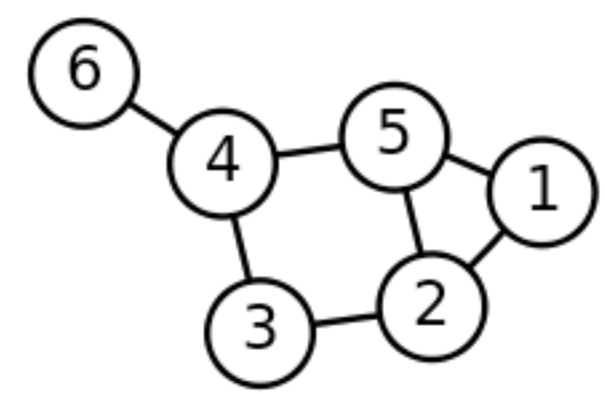
Types of Graphs (Networks)

- **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

Key definitions/vocab

- **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

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)},$$

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

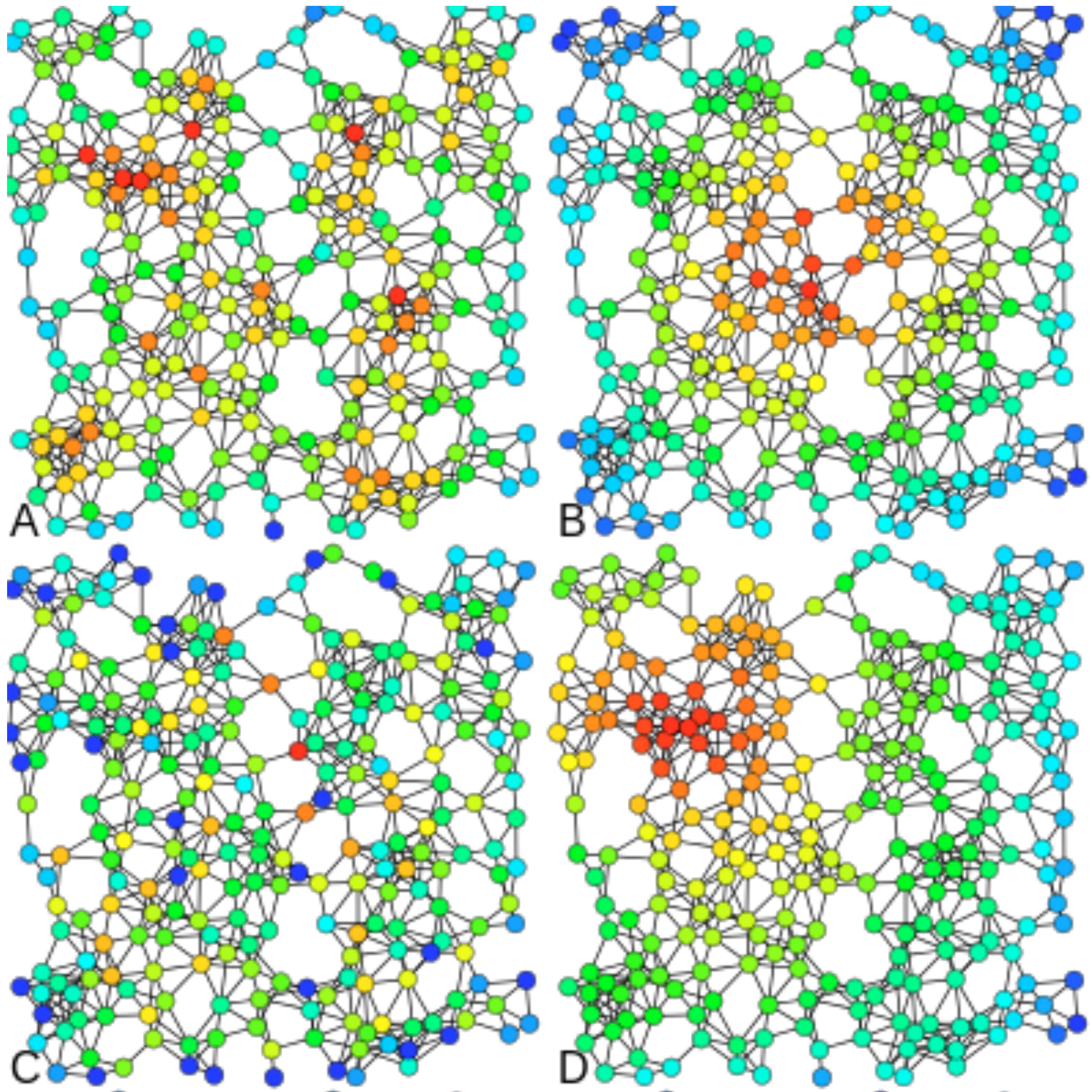

Nellie_Mae_Rowe
Georgia_(U.S._state)
Southeastern_United_States
Southeast_(disambiguation)
Santa_Comba_CF
La_Liga
Hill_Township,_Michigan
Civil_township
Local_government
Public_administration
Association_football
Compass_direction
Sport
Behavior
Biology
Physics
Natural_science
Kenjai
East_Africa
Hypernym
Umbrella_term
Knowledge
Learning
Fact
Proof_(truth)
Argument
Higher_education
Academic_discipline
University
Yerevan_State_University
Taxonomy
Biological_classification
Class_(biology)
Insect
Moth
Mammal
Glyphipterigidae
Human
Person
Language
Chinese_language
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
System
Set
Set_(mathematics)
Mathematics
Quantity
Property_(philosophy)
Geometry
Coordinate_system
Oxford_and_Cherwell_Valley_College
Process_(philosophy)
Abstraction
Entity
Organization
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)
- Google pagerank
- This works out to be the eigenvector of the largest eigenvalue of the adjacency matrix

Degree

Closeness



Betweenness

Eigenvector

Network Features & Properties

- **Degree** of nodes
- **Degree sequence** - List of degrees for all nodes in a graph
- Often use this to infer the **degree distribution** (more on this later when we talk about random graphs)
- Degree sequence/distribution can tell you a lot about structure of graph

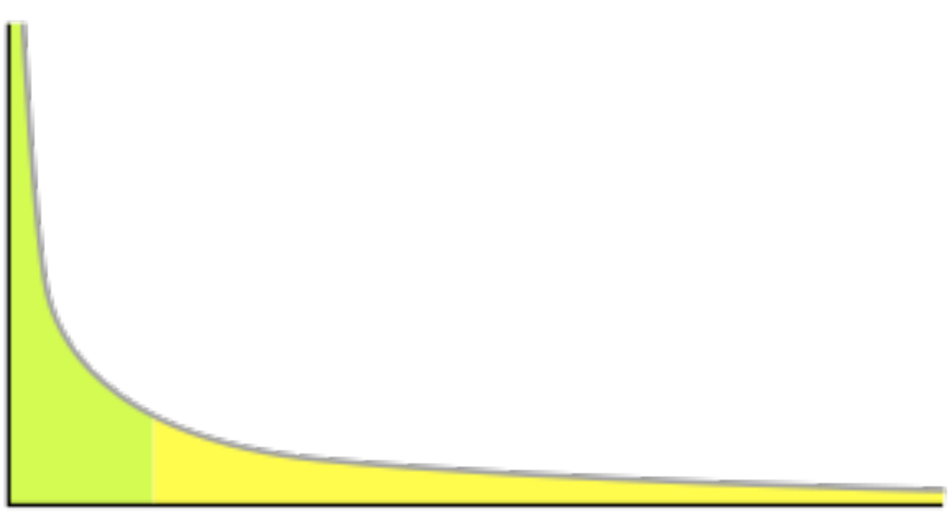
Power Law degree distribution

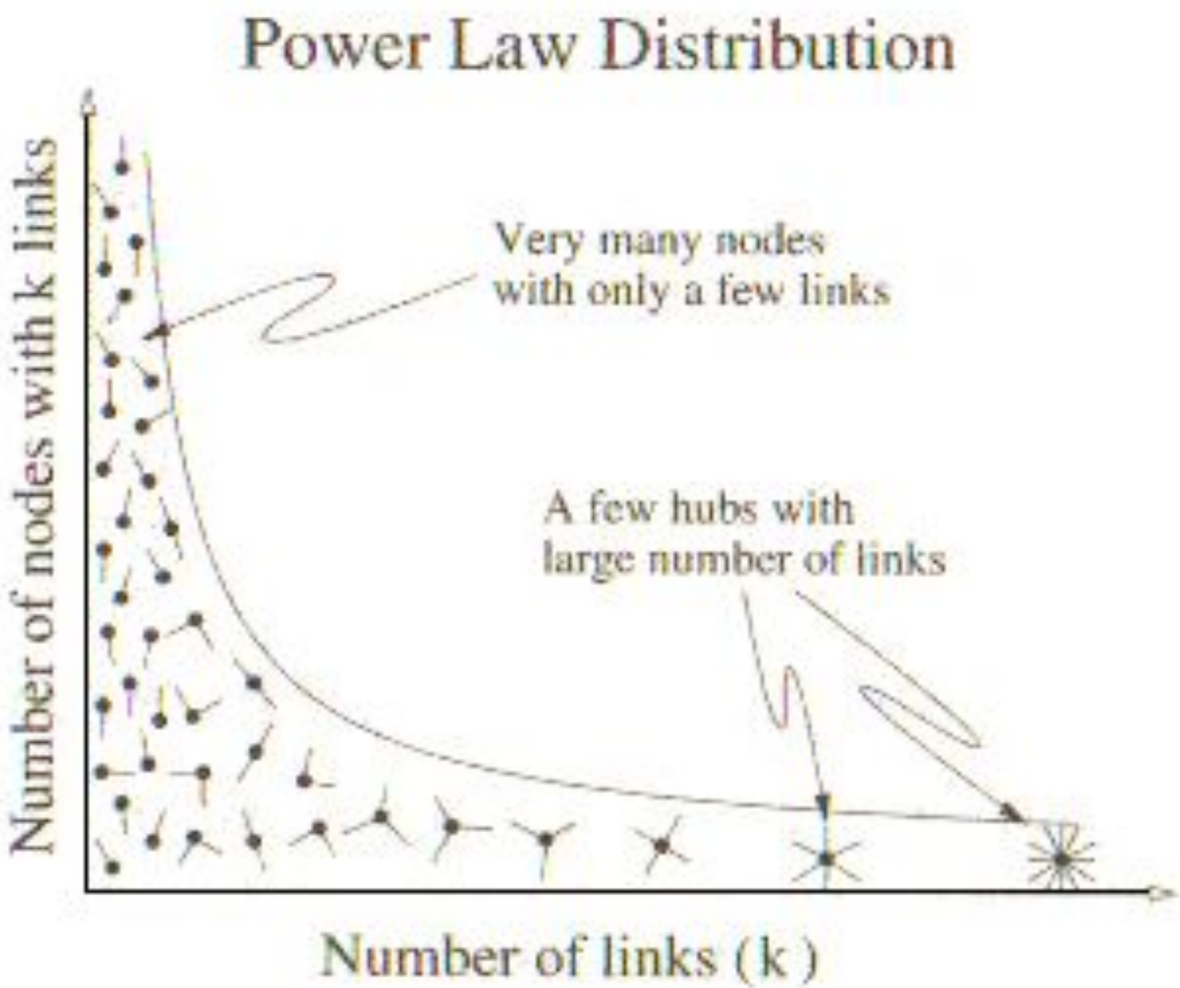
- **Scale free networks** - power law degree distribution

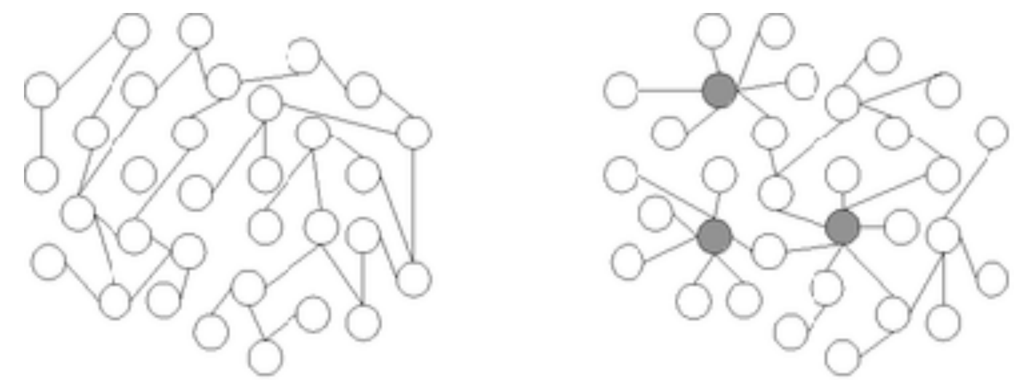
$$P(k) \sim k^{-\gamma}$$

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

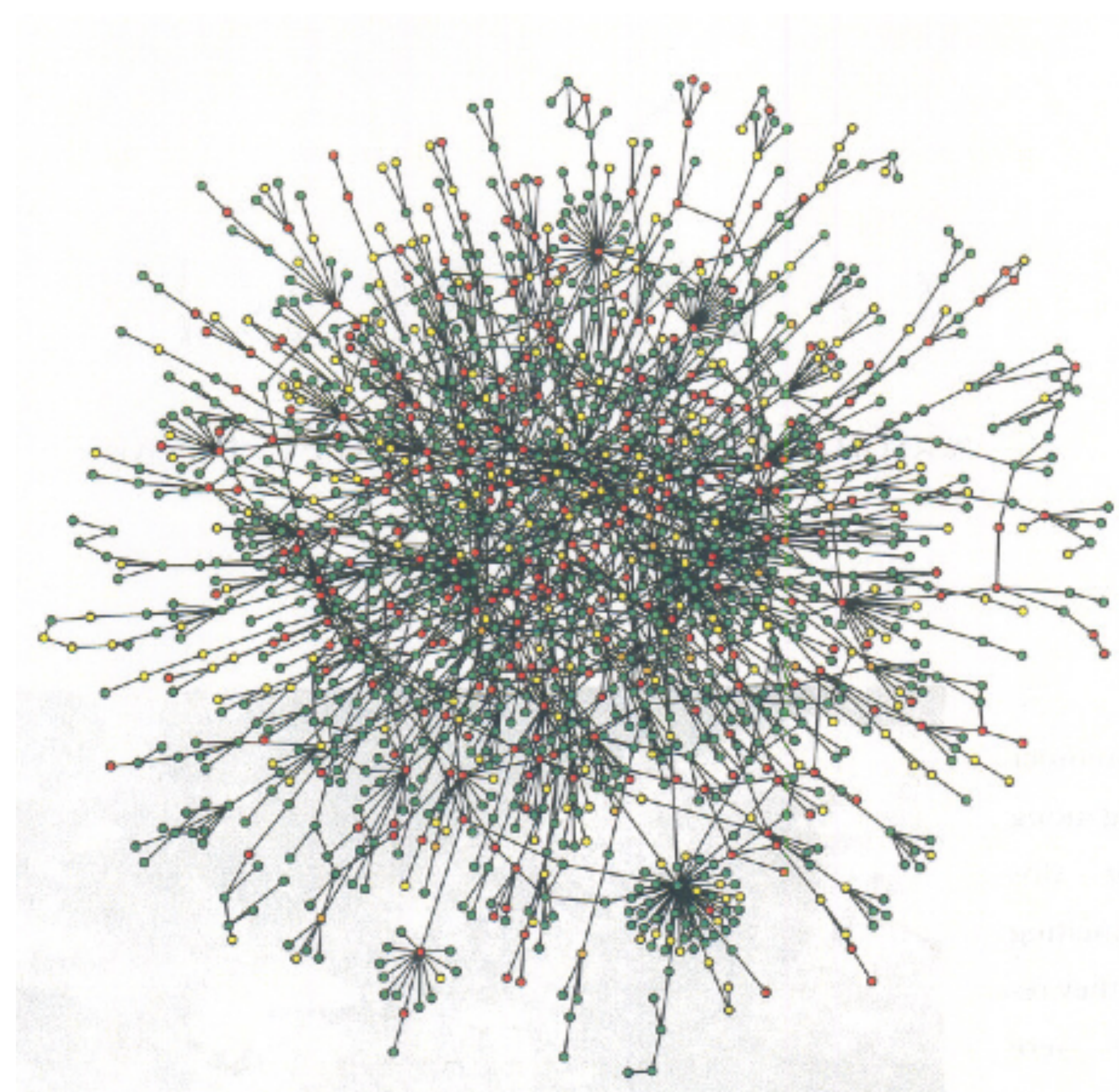
  - Many biological networks, social networks, WWW, etc. are scale free

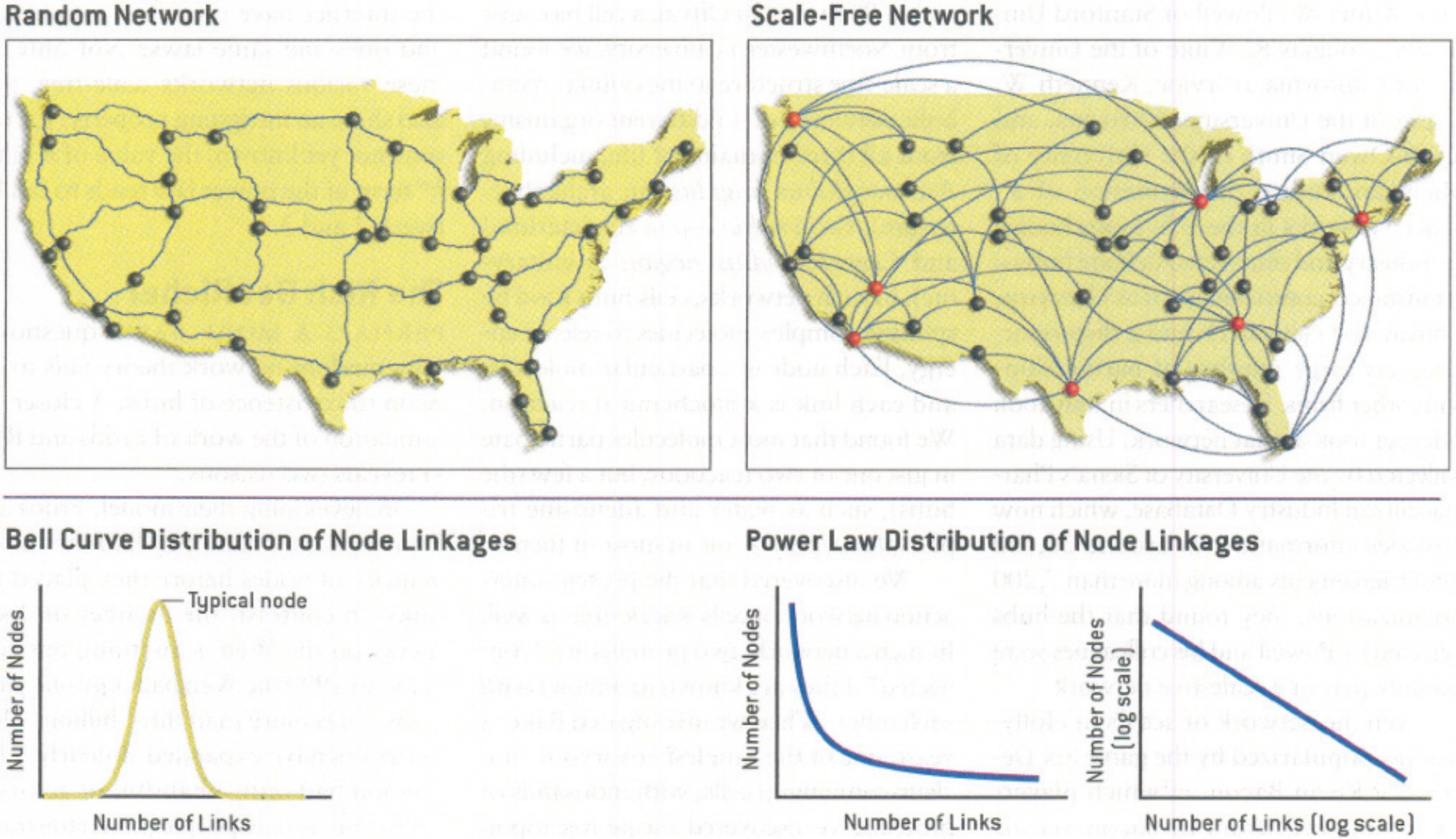


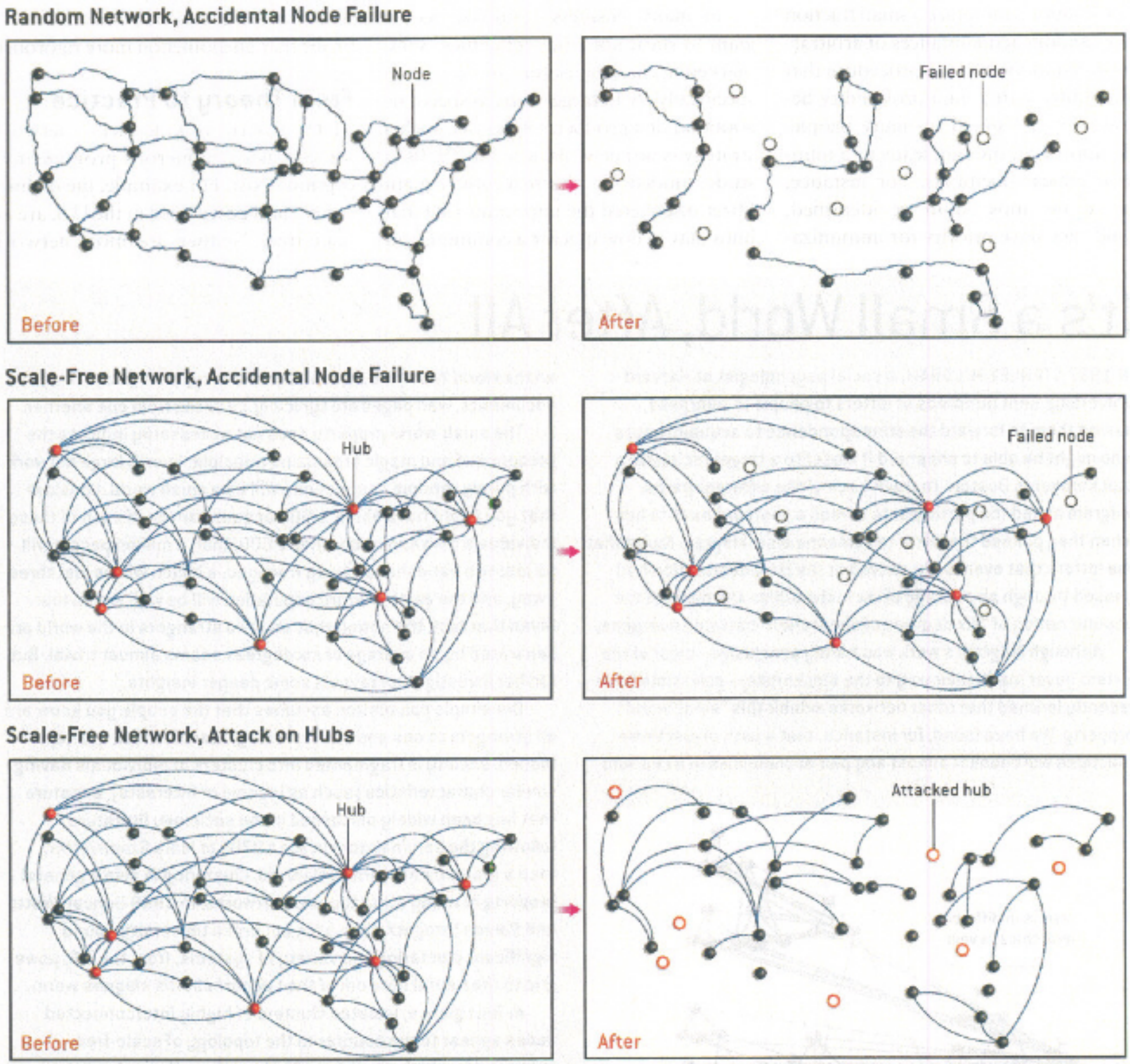




(a) Random network (b) Scale-free network




Random Network
Scale-Free Network
Bell Curve Distribution of Node Linkages
Typical node
Number of Nodes
Number of Links
Power Law Distribution of Node Linkages
Number of Nodes
Number of Links
Number of Nodes (log scale)
Number of Links (log scale)

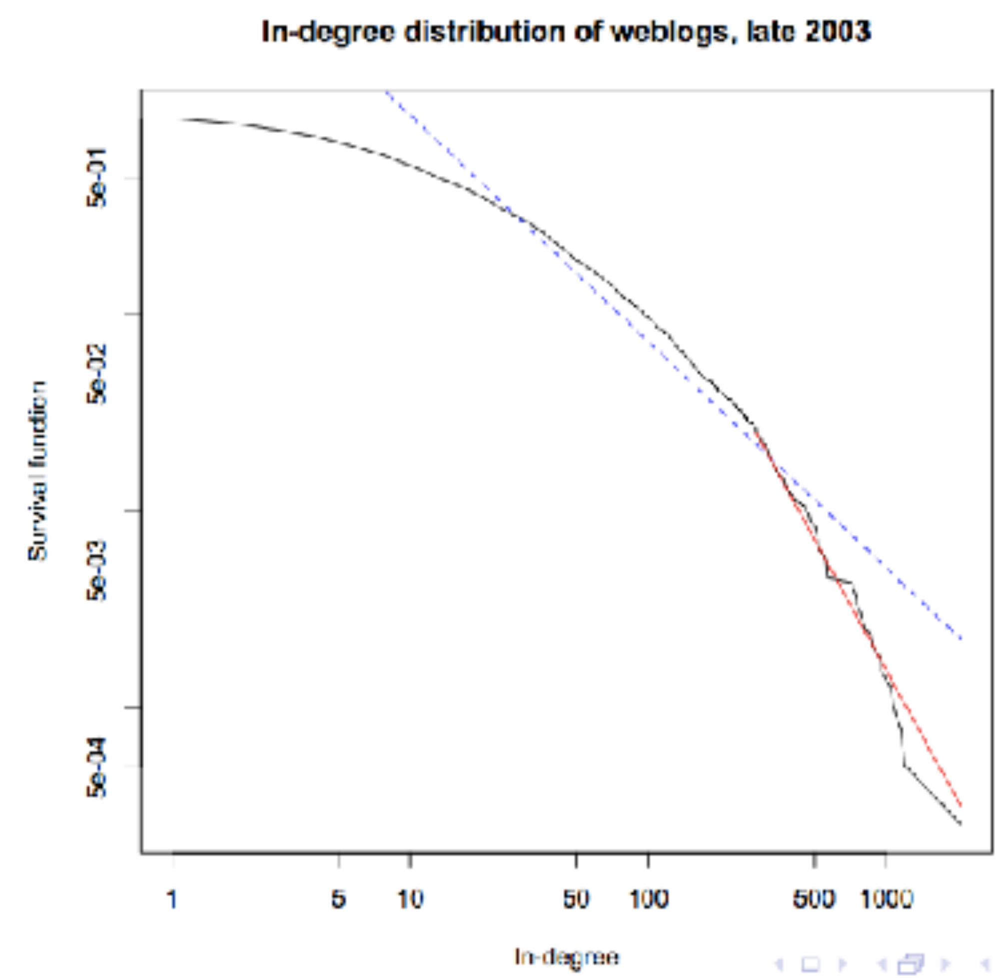

Random Network, Accidental Node Failure
Node
Before
Failed node
After
Scale-Free Network, Accidental Node Failure
Hub
Before
Failed node
After
Scale-Free Network, Attack on Hubs
Hub
Before
Attacked hub
After

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



Random Networks

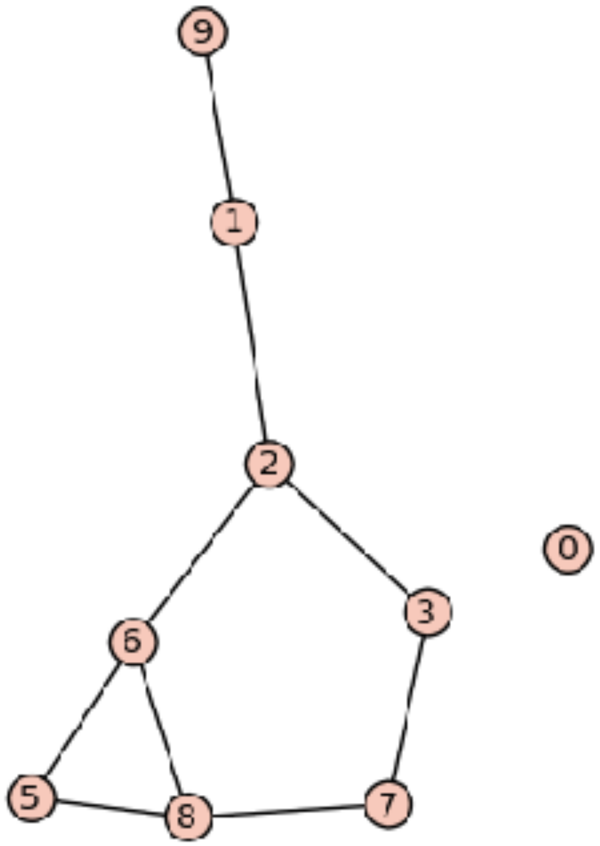
Random Networks

- Why would you want to do this?
  - Often want to simulate network effects
  - May not know exact network
  - But often do know some general features of the network (e.g. degree distribution)
  - So: simulate random networks with those features conserved

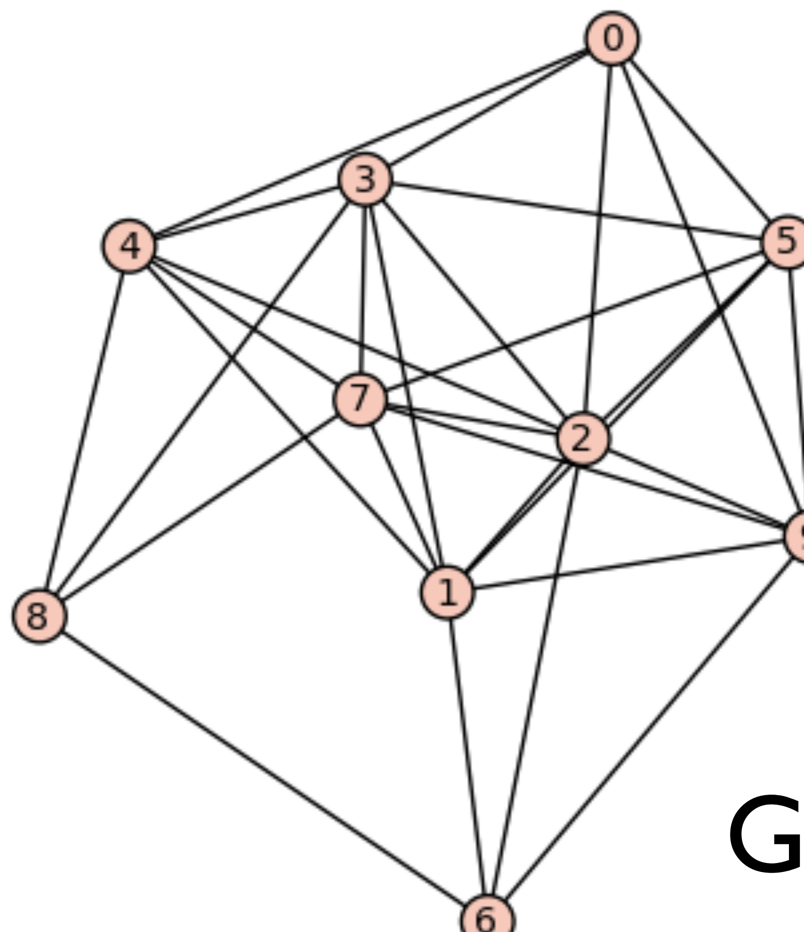
Erdös-Rényi Networks

- **Erdös–Rényi** (also Gilbert) Network - two forms:
  - *G(n,p)* - network on *n* nodes with each edge having probability *p* of existing
  - *G(n,M)* - network on *n* nodes with *M* edges chosen randomly
  - Often called a “random network” even though all of the networks here are also random

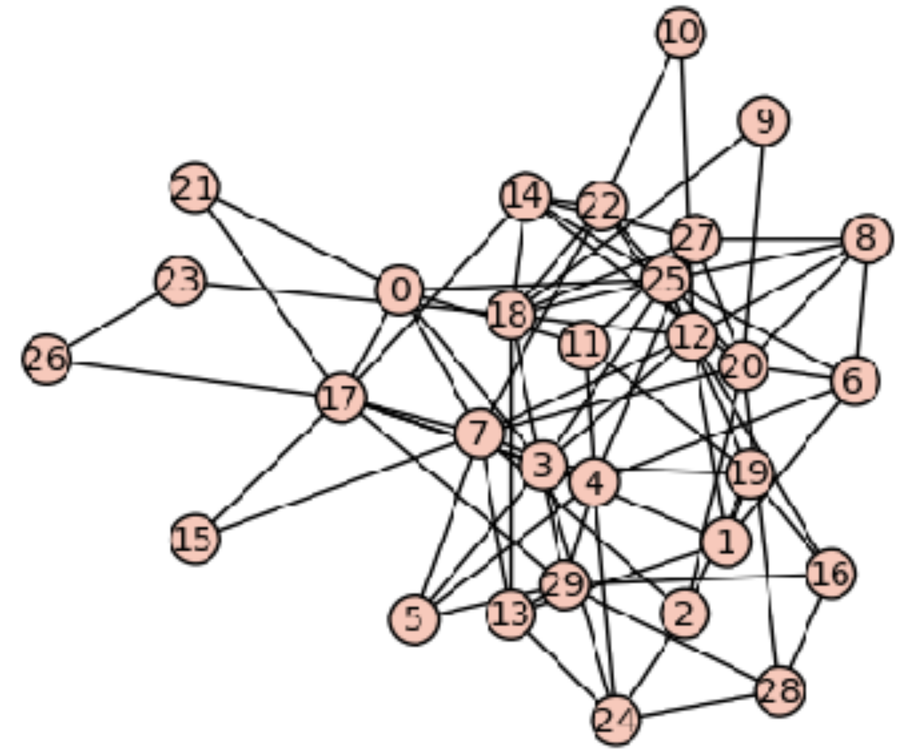
G(n,p)



G(10,0.2)

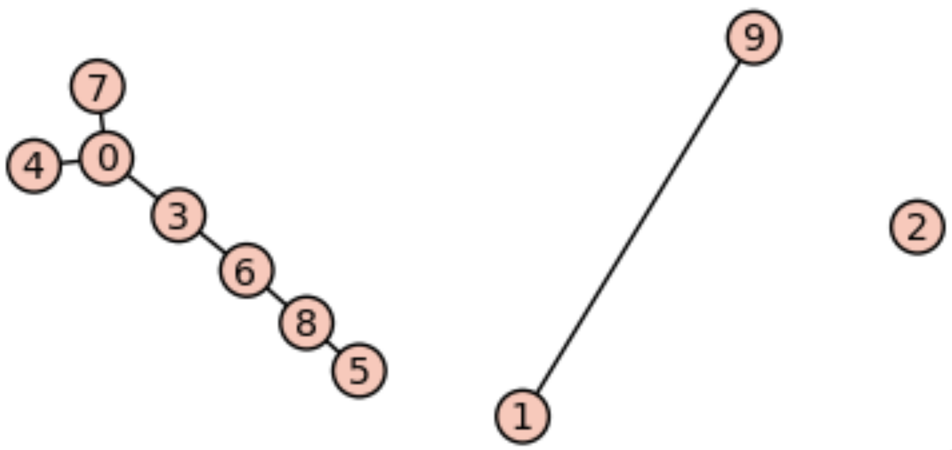


G(10,0.6)

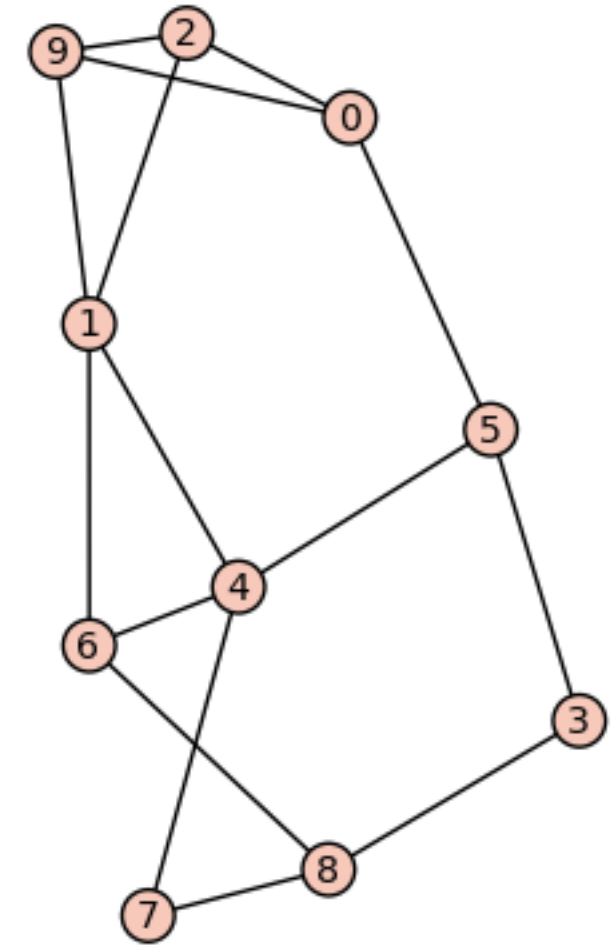


G(30,0.3)

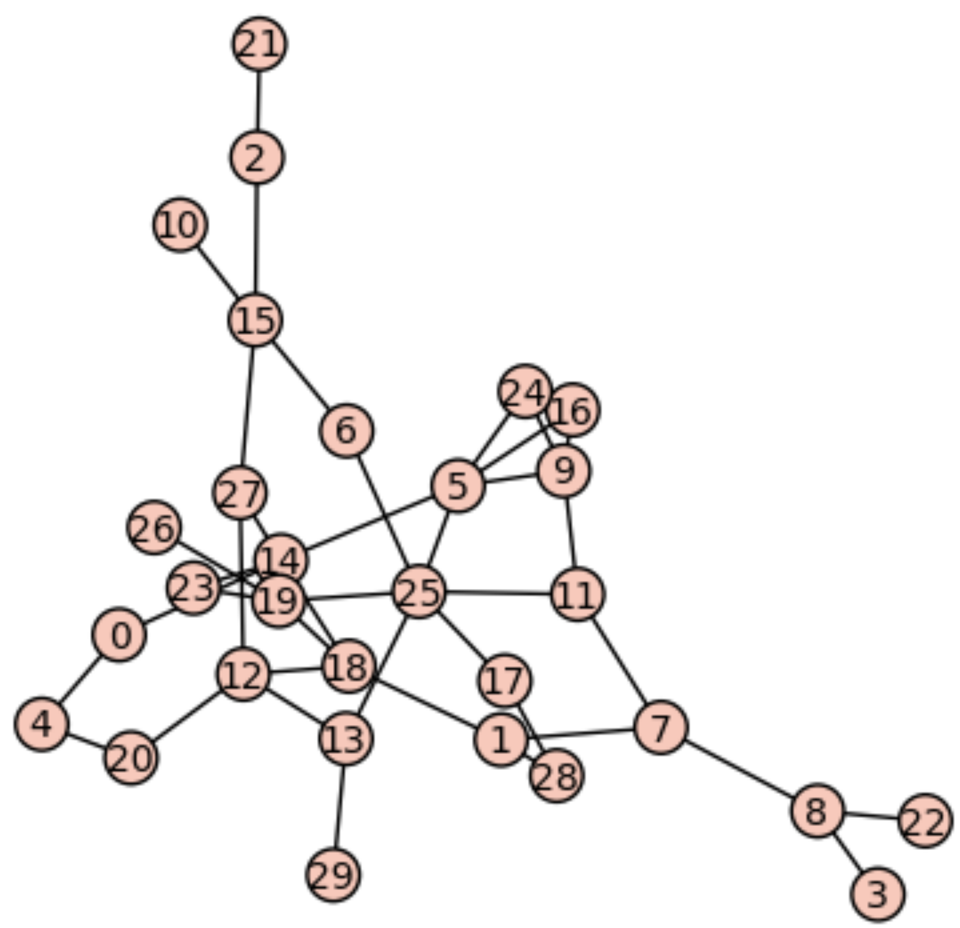
G(n,M)



G(10,7)



G(10,15)



G(30,40)

Erdös-Rényi Networks

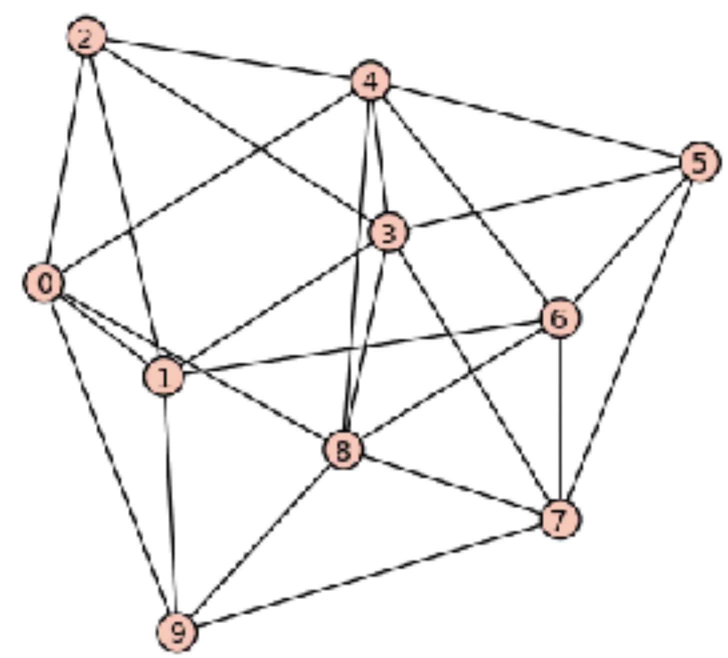
- Not so realistic for lots of things (e.g. social networks, many gene/protein/biological networks)
- But, often handy as a test case
- And useful for making analogs of homogeneous mixing (e.g. from SIR or linear compartmental models)

Preferential Attachment Networks

- **Barabasi-Albert** algorithm
- Add new nodes to the network sequentially, preferentially connecting them to high-degree nodes
- Generates random scale free networks

Small World Networks

- Most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops or steps
- Average distance L between two nodes is proportional to $\log n$ (where n is the number of nodes)



Small World Network

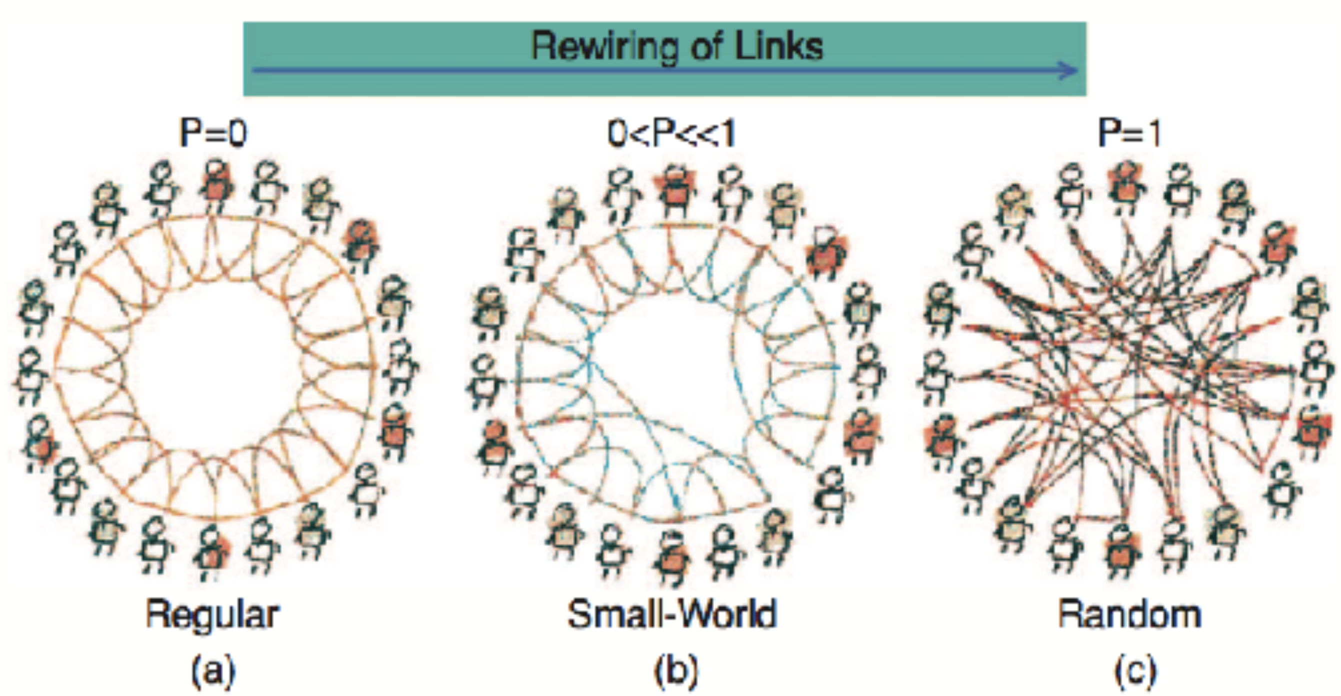
- Creates the “what a small world!” effect: two nodes will tend to have a mutual friend (adjacent node)

- Can be similar to scale free in that they tend to produce hubs as well as sparsely connected individuals

    - Network can be both small-world and scale-free

Small World Networks

- How to generate random small world network?
- Newman-Watts-Strogatz Algorithm
- However, this tends to produce more similar degrees for nodes rather than scale free



Configuration models

- Given a degree sequence, generate random network with that sequence
- Random graphs, but with the advantage that the degree sequence can be chosen realistically
- E.g. useful when we have egocentric data!
- Algorithm: generate ‘stubs’ with the correct degree, then connect pairs of stubs

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)

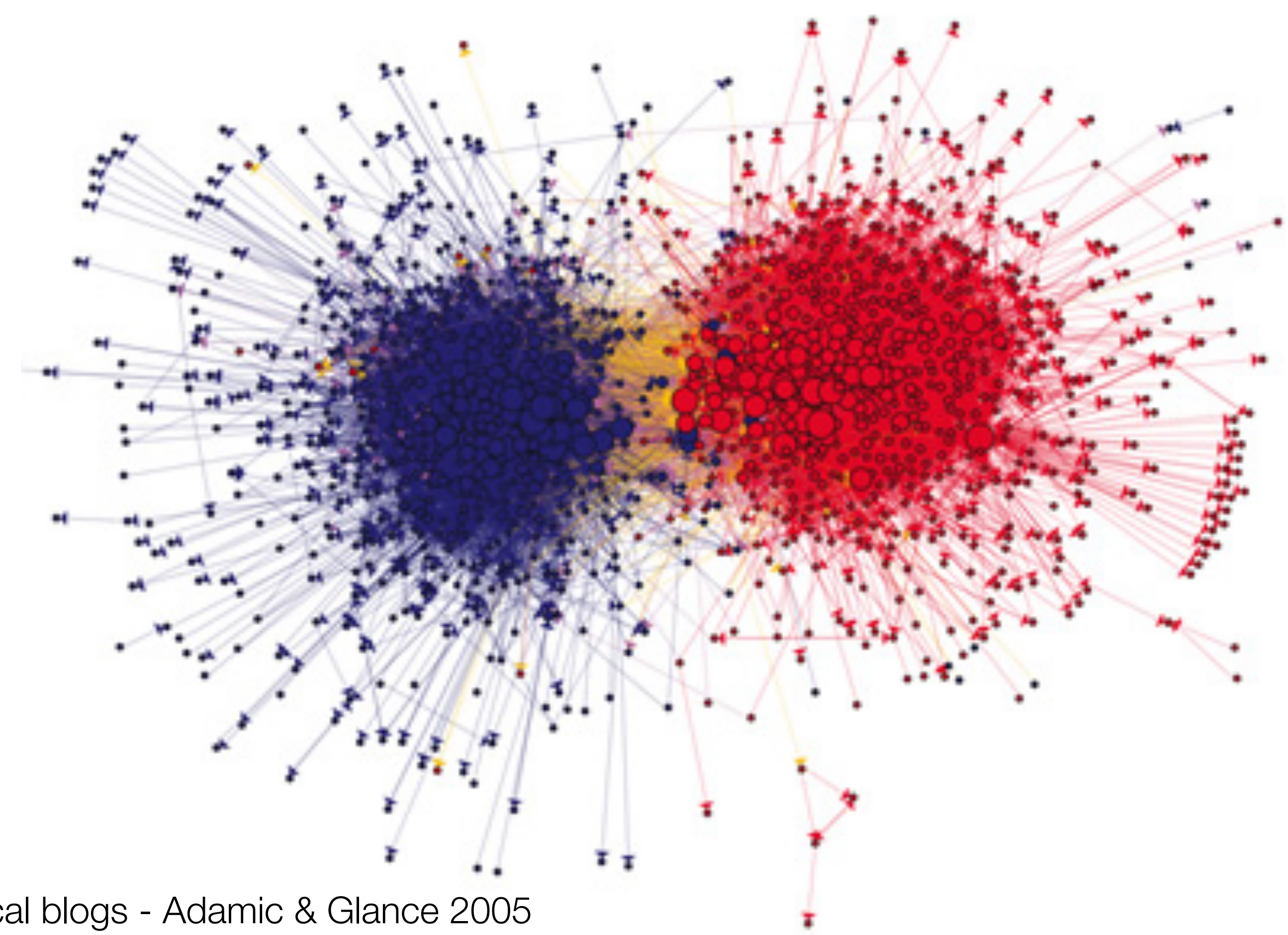
Assortativity

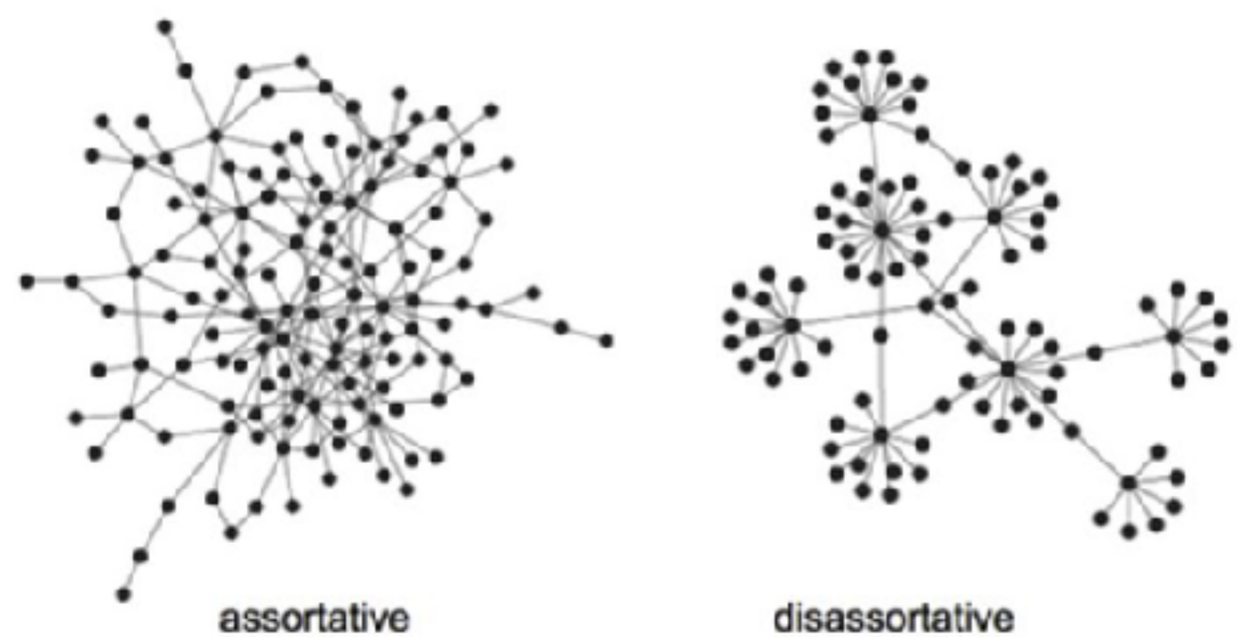
- Assortativity - measures network-level tendency for nodes to to attach to similar nodes
  - Similarity can be defined by node attributes, degree, etc.
- Calculate fraction of edges between nodes of the same type/ value, compare to what would be expected from a random network
- Ranges from -1 (dissassortative)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

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

Assortativity

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

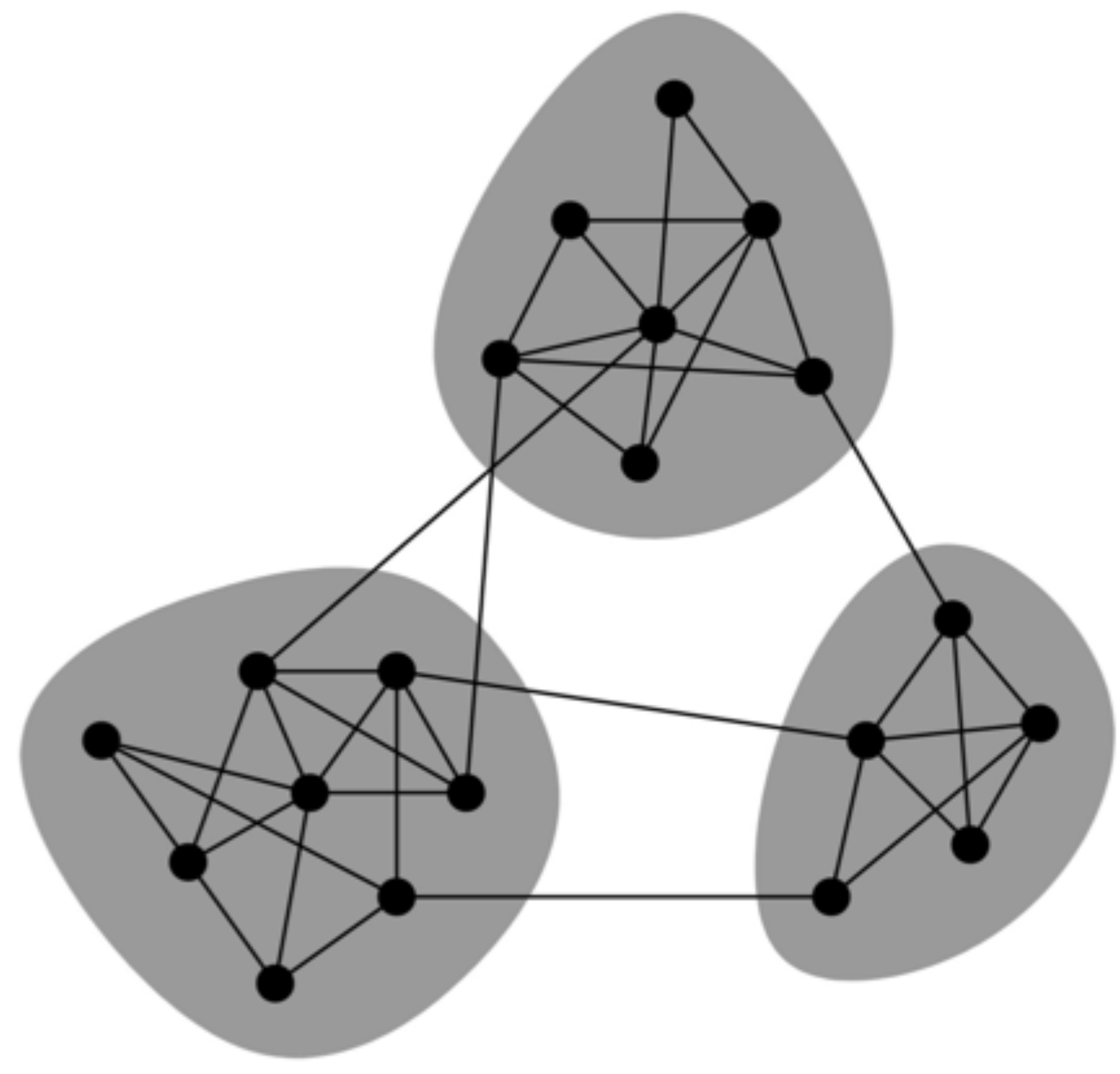
$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} = \frac{\mathrm{Tr}\,\mathbf{e} - ||\mathbf{e}^2||}{1 - ||\mathbf{e}^2||},$$

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

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

Modularity

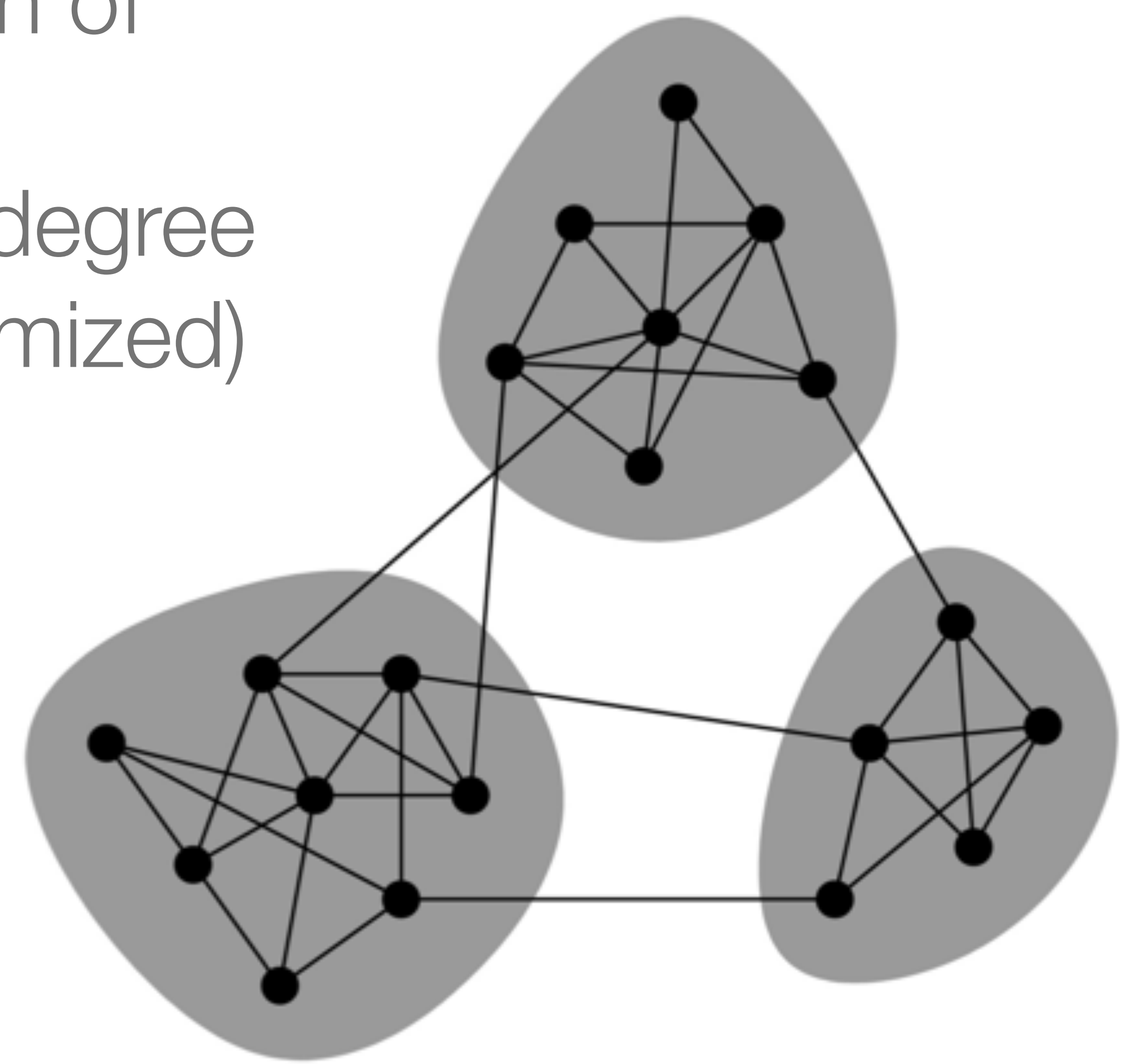
- How to decide communities (clusters) in a network?
- We want communities to have more in-group edges than between-group edges
- We could minimize between group edges, but this would lead to just putting all nodes in one community



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

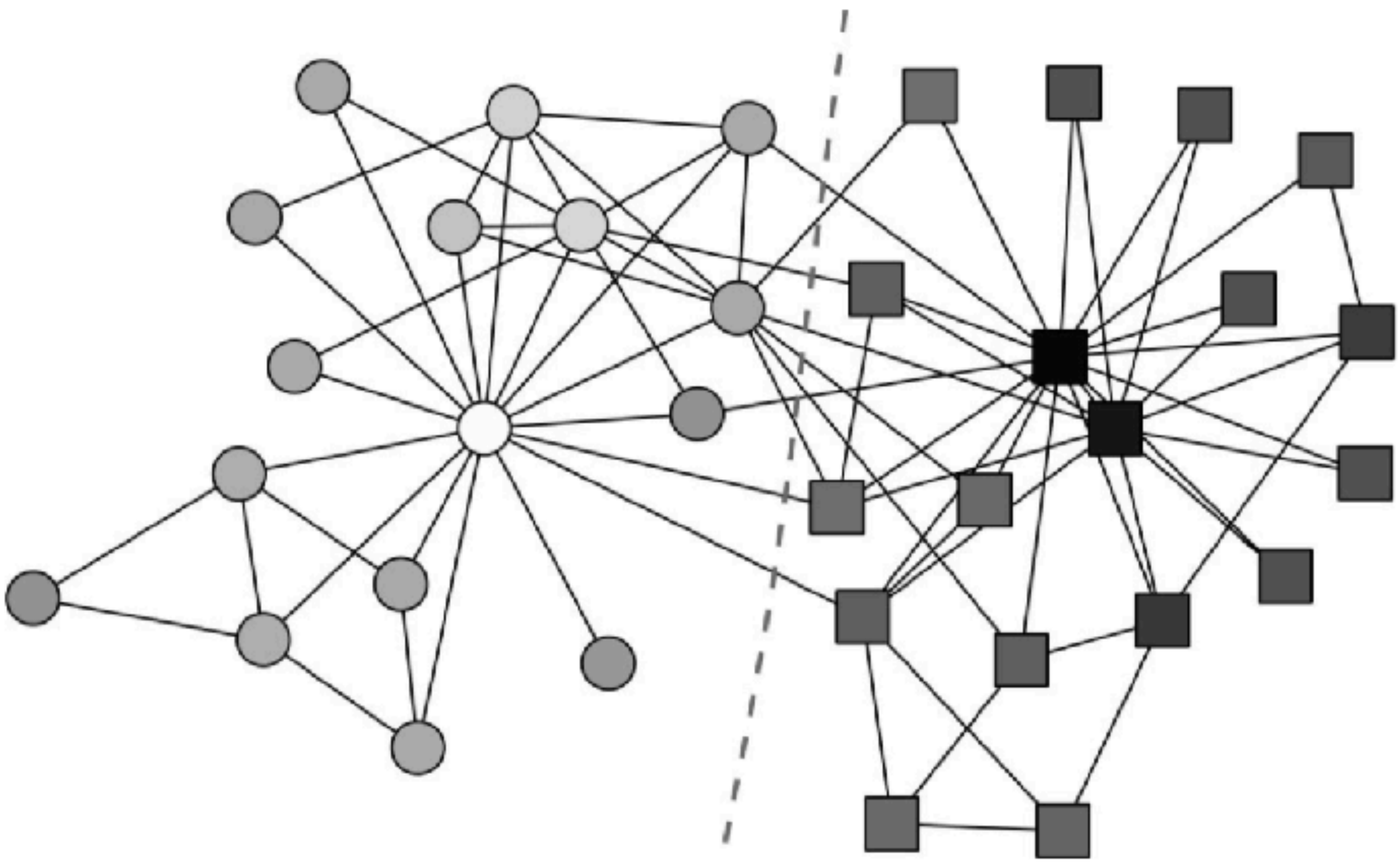
Modularity

- Modularity compares observed community edges to what would be expected at random
- Modularity is the number 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

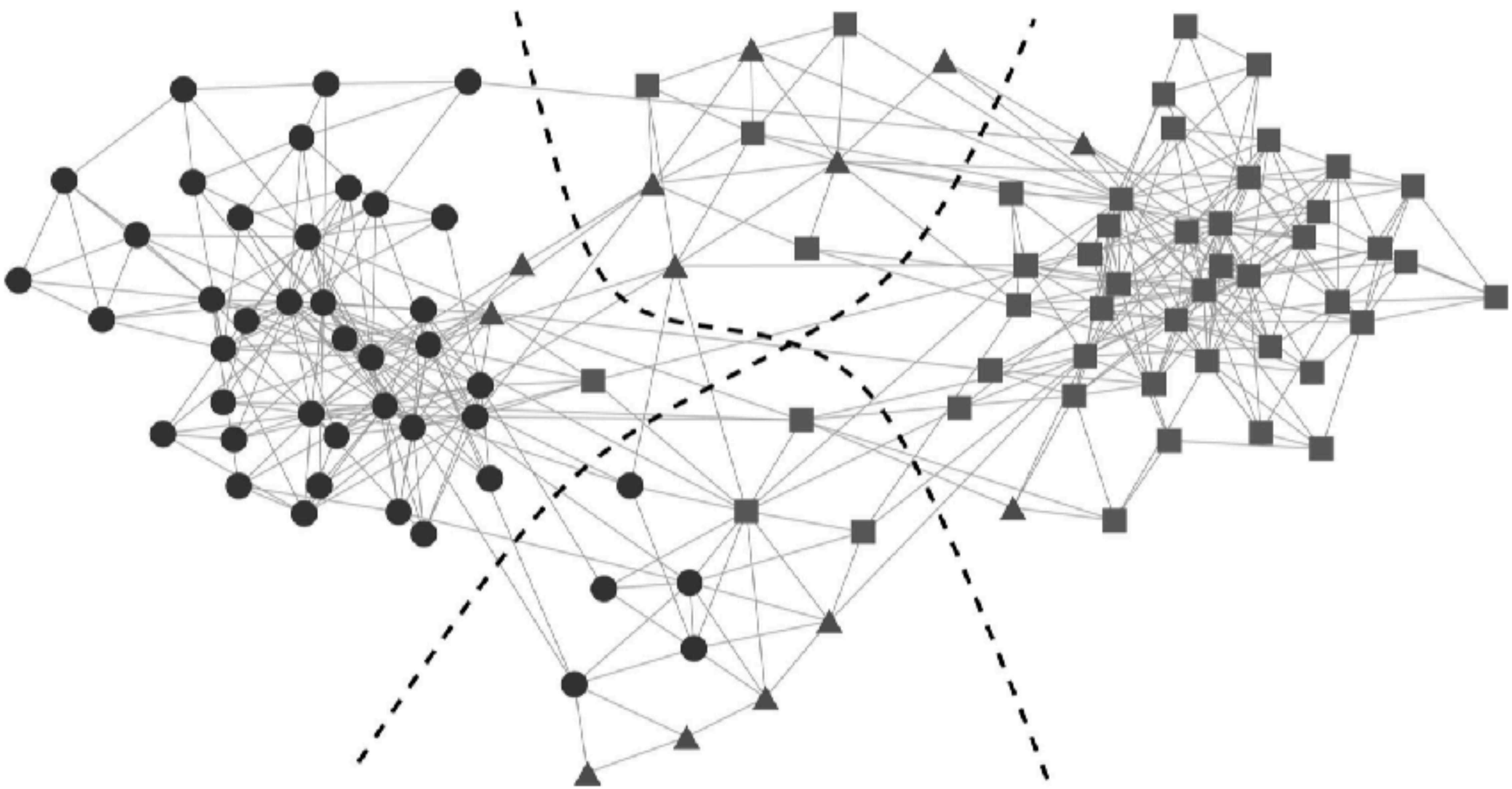


Newman, Mark EJ. "Modularity and community structure in networks." *PNAS* 103.23 (2006): 8577-8582.

Karate club example



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