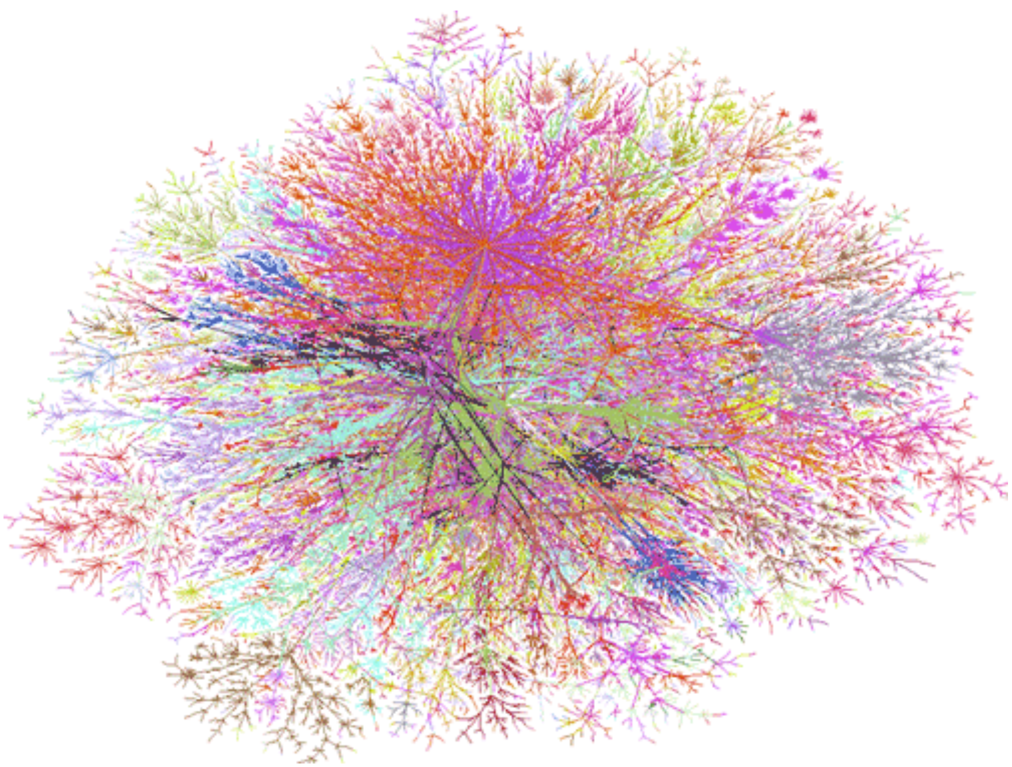
The Machinery of Complex Network Theory:

From Network Measures to Network Robustness



Michelle Girvan

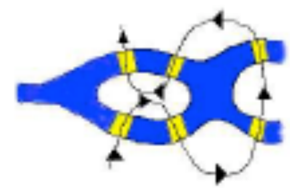
University of Maryland

Traditional vs. Complex Systems Approaches to Networks

Traditional Questions:

Social Networks:
Who is the most important person in the network?

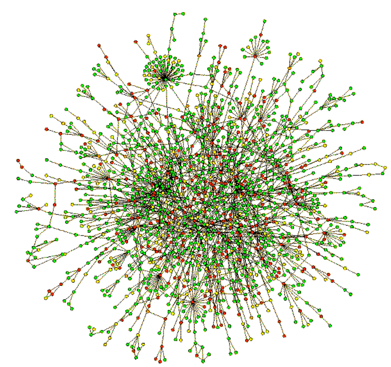
Graph Theory:
Does there exist a cycle through the network that uses each edge exactly once?



Complex Systems Questions:

What fraction of edges have to be removed to disconnect the graph?

What kinds of structures emerge from simple growth rules?



The field of complex networks exists at the intersection of…

- Graph Theory
- Statistical Physics
- Computer Science: Algorithms & Networks
- Nonlinear Dynamics
- Social Network Theory
- Complex Systems

Areas of Network Research

Structural Complexity

- The wiring diagram could be an intricate tangle, far from perfectly regular or perfectly random.
- The network could include different classes of nodes
- The edges could be heterogeneous with different weights, directions and signs.

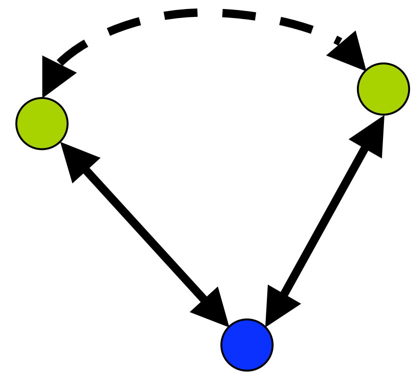
Dynamical Complexity

- Dynamics on the network: processes could be taking place on the fixed network.
- Dynamics of the network: the network itself could be evolving in time.

Network Measures

- Average path lengths
- Clustering coefficients
- Degree distributions
- Degree correlations
- Community structure
- Hierarchy
- And many more…

Clustering

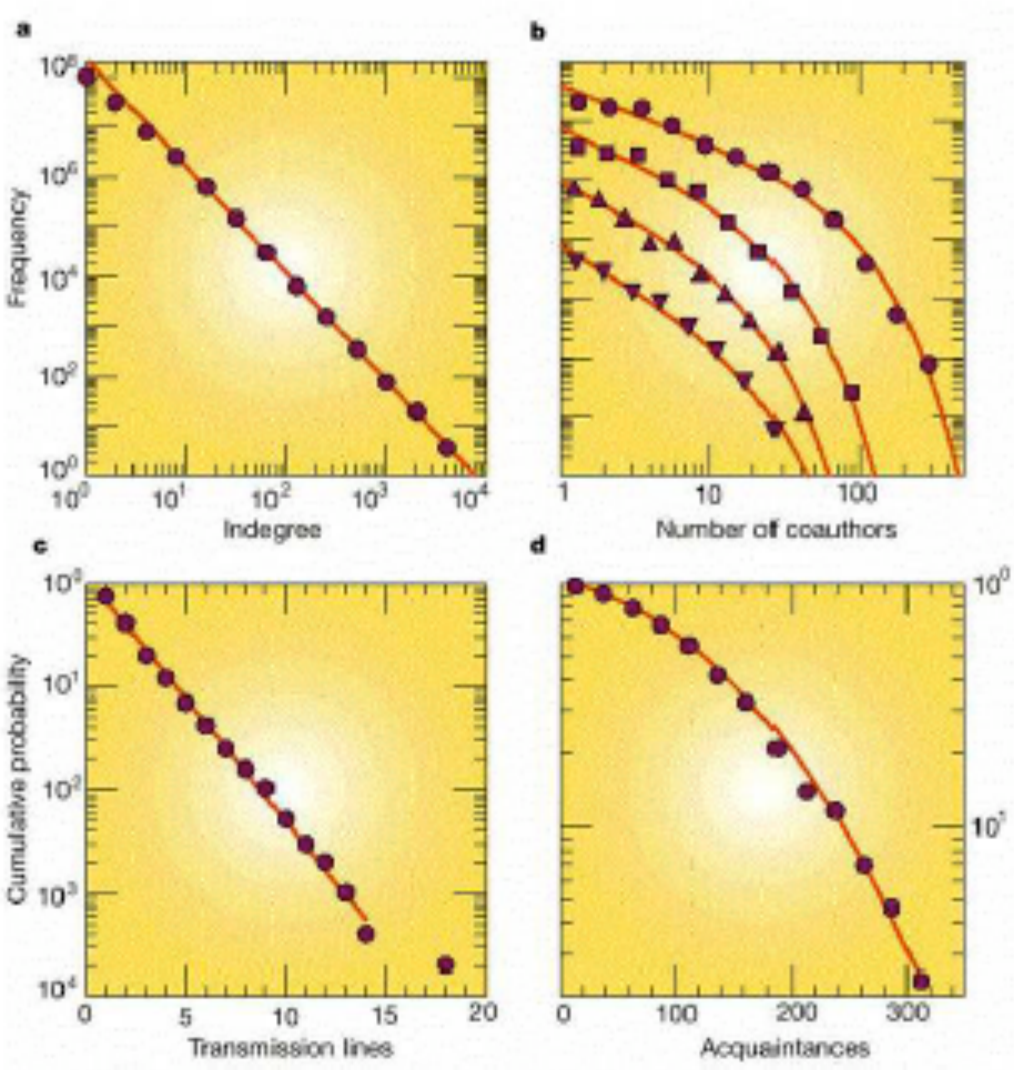


C = Probability that two of a node's neighbors are themselves connected

In a random graph: $C_{rand} \sim 1/N$ (if the average degree is held constant)

Network	N	ℓ	C	C_{rand}
movie actors	225 226	3.65	0.79	0.00027
neural network	282	2.65	0.28	0.05
power grid	4941	18.7	0.08	0.0005

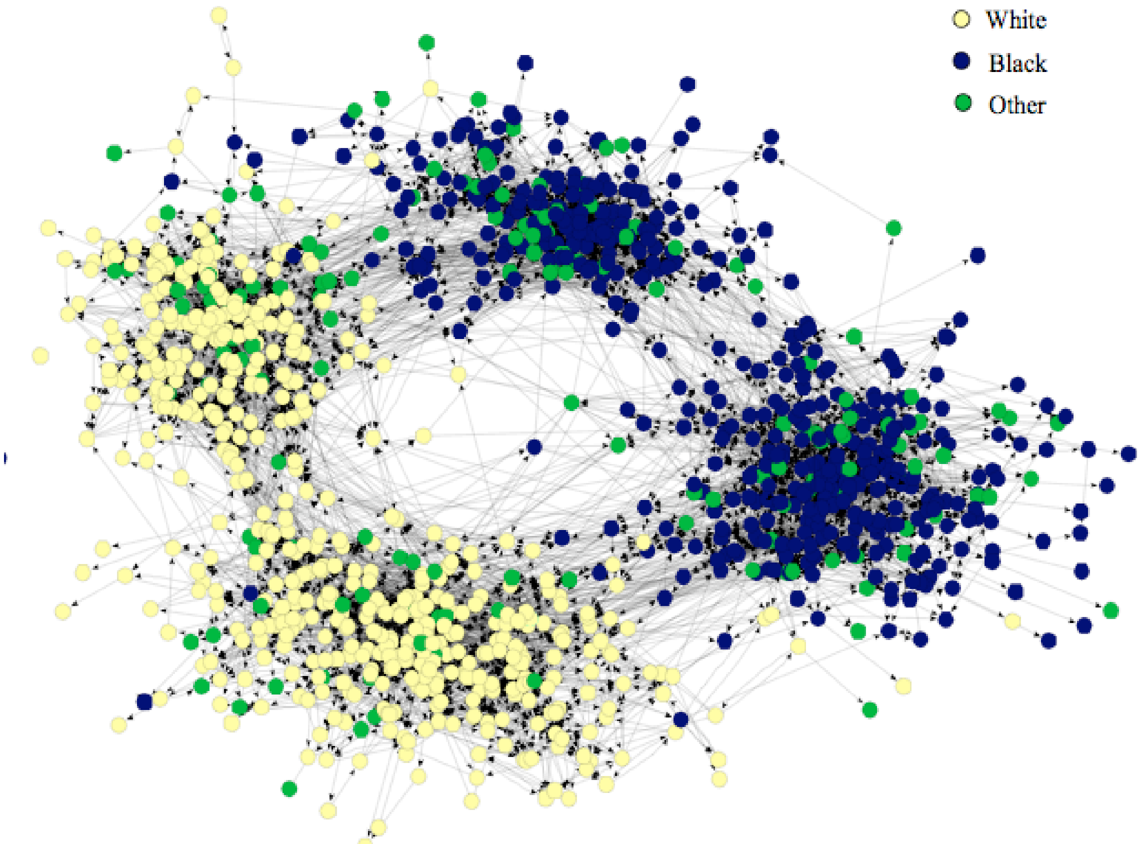
Degree distributions for various networks



(a) World-Wide Web
(b) Coauthorship networks: computer science, high energy physics, condensed matter physics, astrophysics
(c) Power grid of the western United States and Canada
(d) Social network of 43 Mormons in Utah

Assortative Mixing

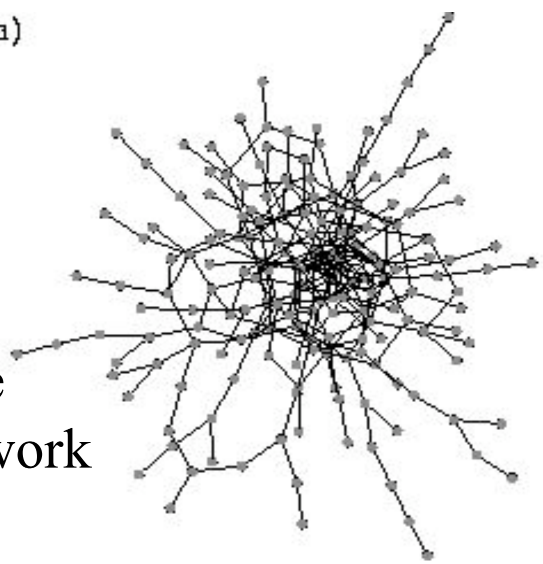
In assortatively mixed networks, like vertices tend to connect preferentially to one another.



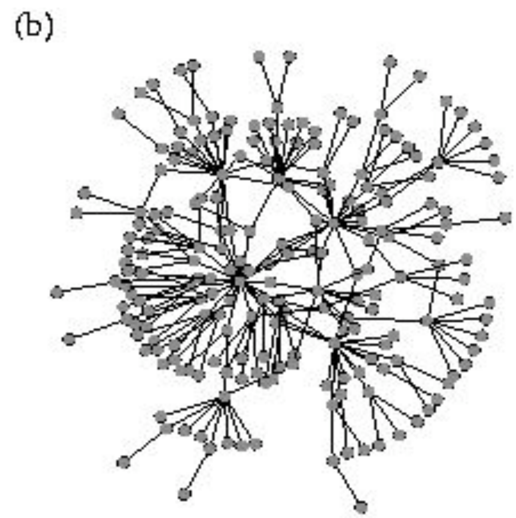
Friendship network of students in a U.S. school. Friendships are determined by asking the participants, and hence are directed, since A may say that B is their friend but not vice versa. Vertices are color coded according to race, as marked, and the split from left to right in the figure is clearly primarily along lines of race. The split from top to bottom reflects a division between middle school and high school students.

Assortative Mixing by Degree

- A network is said to be assortatively mixed by degree if high degree vertices tend to connect to other high degree vertices
- A network is disassortatively mixed by degree if high degree vertices tend to connect to low degree vertices.



Assortative Scale-free network



Disassortative Scale-free network

Measured assortativity for various networks

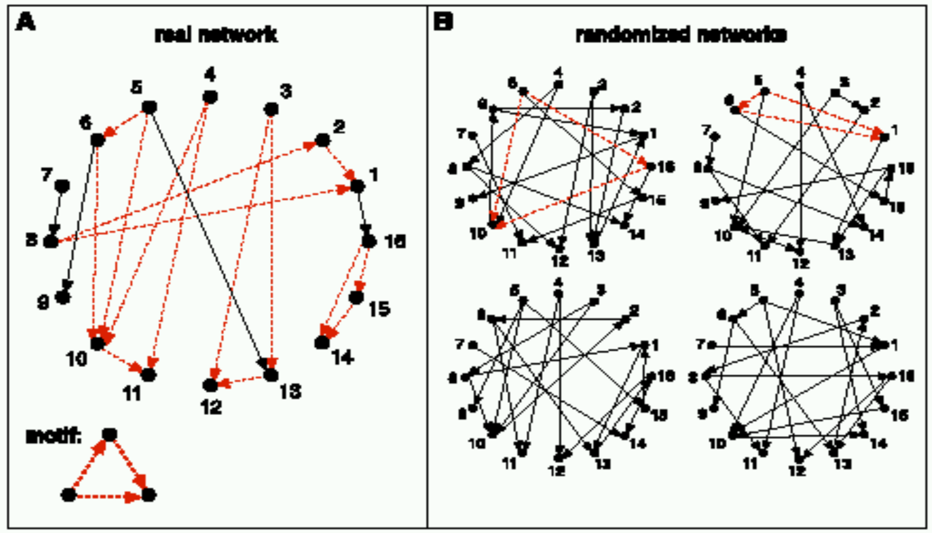
	network	type	size n	assortativity r
social	physics coauthorship	undirected	52 909	0.363
	biology coauthorship	undirected	1 520 251	0.127
	mathematics coauthorship	undirected	253 339	0.120
	film actor collaborations	undirected	449 913	0.208
	company directors	undirected	7 673	0.276
	email address books	directed	16 881	0.092
technol.	Internet	undirected	10 697	−0.189
	World-Wide Web	directed	269 504	−0.067
	software dependencies	directed	3 162	−0.016
biological	protein interactions	undirected	2 115	−0.156
	metabolic network	undirected	765	−0.240
	neural network	directed	307	−0.226
	marine food web	directed	134	−0.263
	freshwater food web	directed	92	−0.326

M.E.J Newman and M. Girvan, *Mixing Patterns and Community Structure in Networks* (2002).

Network Motifs

Motifs

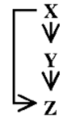
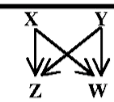
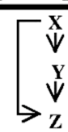
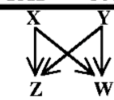
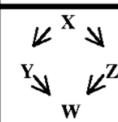
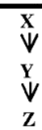
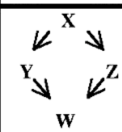
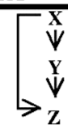
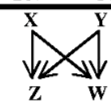
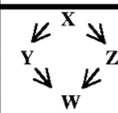
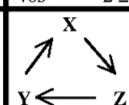
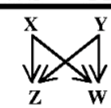
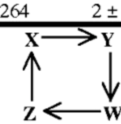
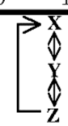
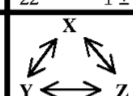
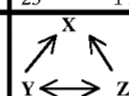
Subgraphs that have a significantly higher density in the observed network than in the randomizations of the same.



Randomized networks:

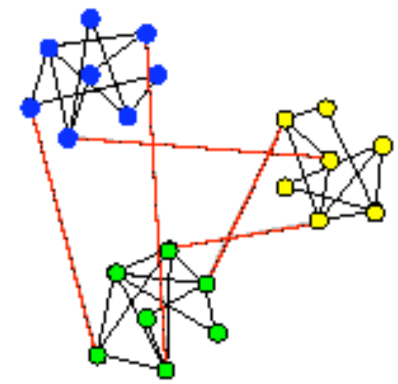
Ensemble of maximally random networks preserving the degree distribution (or some other feature(s)) of the original network.

R. Milo et al., Science **298**, 824 (2002)

Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)					**Three-node feedback loop**			**Bi-fan**			**Four-node feedback loop**
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web					**Feedback with two mutual dyads**			**Fully connected triad**			**Uplinked mutual dyad**
nd.edu§	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4±4e2	15,000	1.2e6	1e4 ± 2e2	5000

R Milo et al., *Science* **298**, 824-827 (2002).

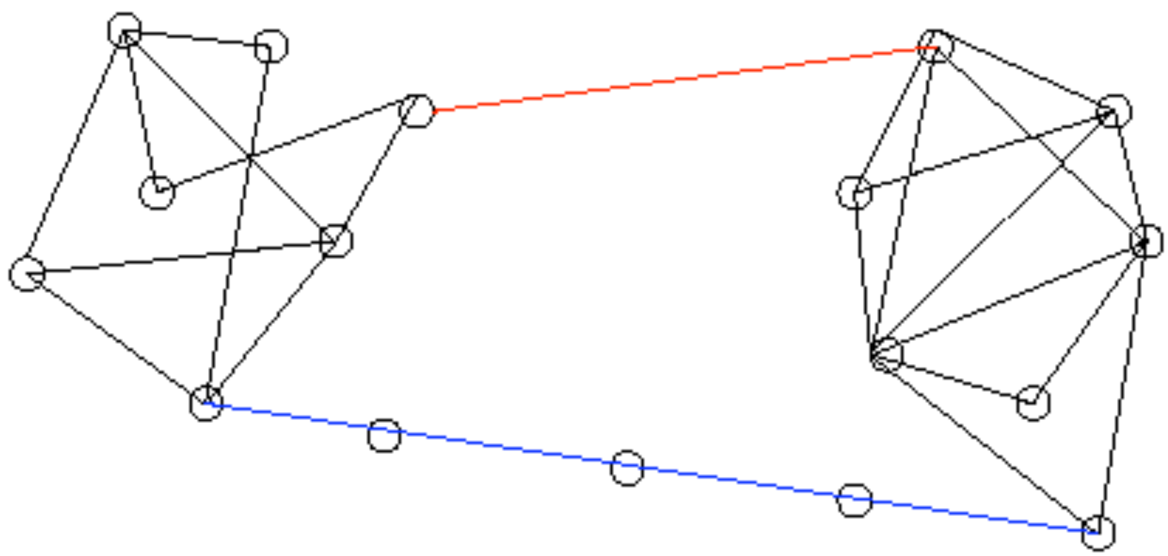
Community Structure



- Community detection based on centrality indices:
  - Node betweenness -- The betweenness centrality of a vertex *i* is the number of shortest paths between pairs of other vertices which run through *i.*
  - Edge betweenness -- Similarly, the betweenness of an edge *j* is the number of shortest paths between pairs of nodes which run along *j.*

Algorithm for Detecting Communities

1. Calculate the betweenness for all edges in the network.
2. Remove the edge with the highest betweenness.
3. Recalculate betweennesses for all edges affected by the removal.
4. Repeat from step 2 until no edges remain.



Quantifying the Community Structure

- Each level of the hierarchical tree represents a division of the network into a certain number of communities.
- Let e_{ij} be the fraction of edges that connect a node in community *i* to a node in community *j*. For a system of *S* communities, *e* is an *S* x *S* matrix.
- Define $a_i = \Sigma_j e_{ij}$, *i.e.* a_i is the fraction of edges attached to vertices in community *i*.
- The level of community structure can then be quantified with a modularity coefficient *Q:*

$$Q = \Sigma_i e_{ii} - \Sigma_i a_i^2$$

which has the following properties:

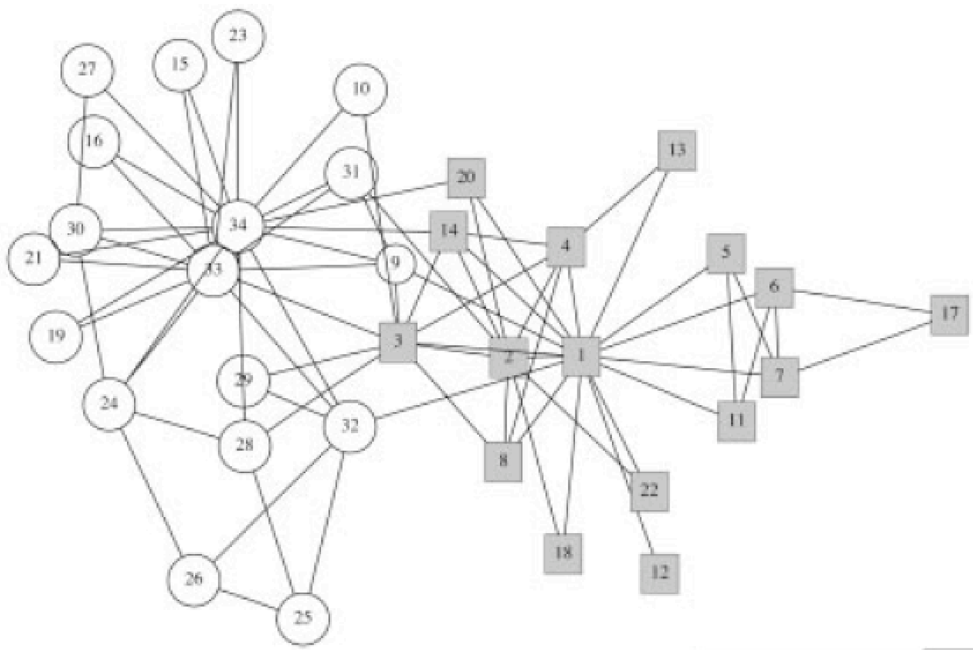
- $Q = 0$ when there is no community structure.

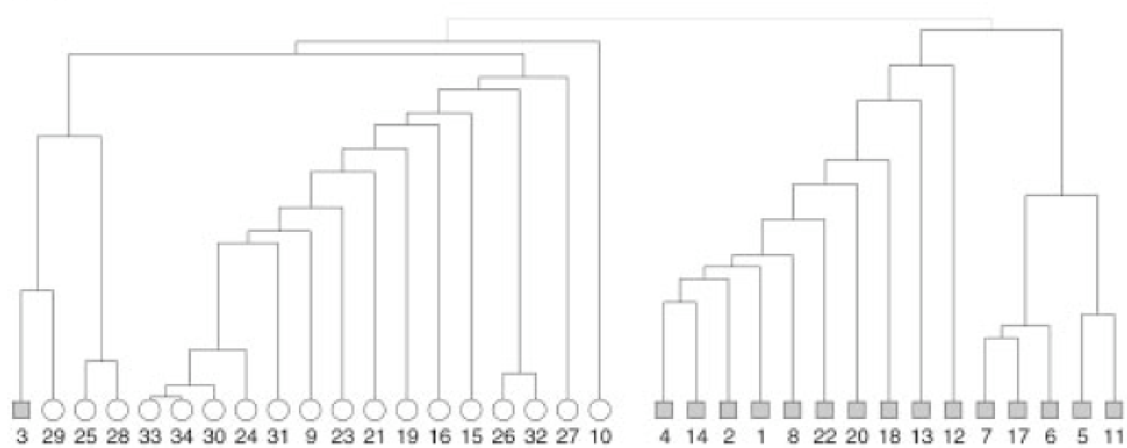
- *Q* approaches *1* as *S* gets large and the network displays perfect modularity. $Q = 1 - 1/S$ for a perfectly modular network of *S* equally sized communities.

- When the network is divided into a certain number of communities, *Q* is largest when the communities are equal in size.

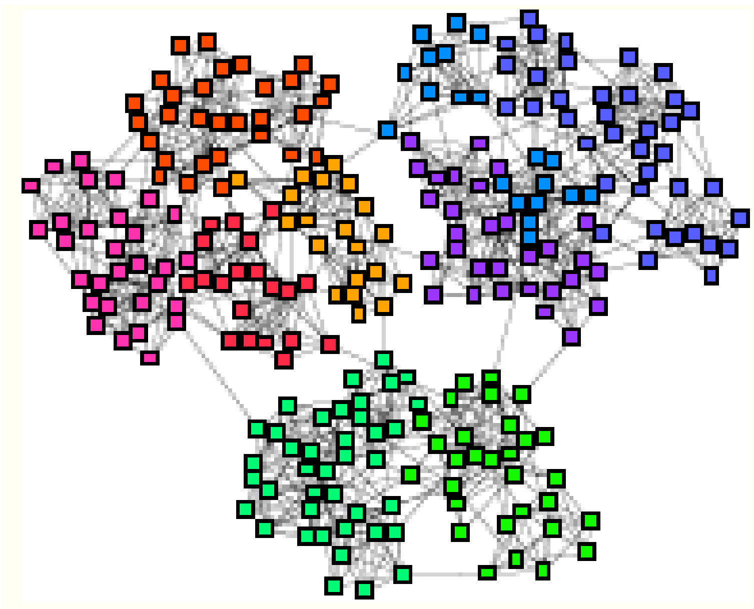
- The minimum value of *Q* likes in the range $-1 < Q < 0$

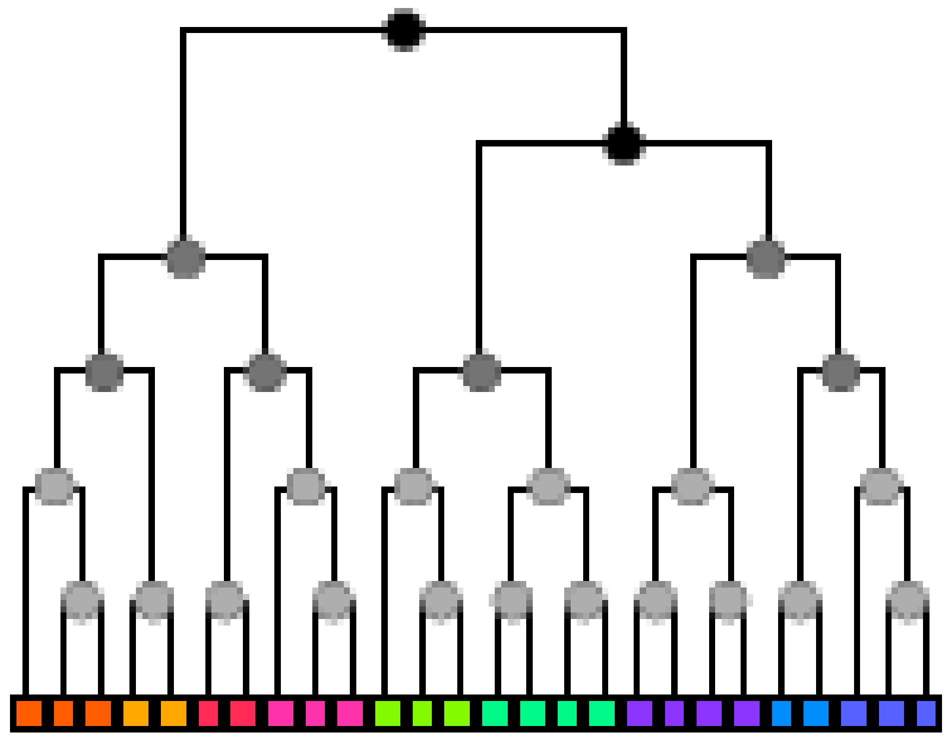
Zachary's Karate Club





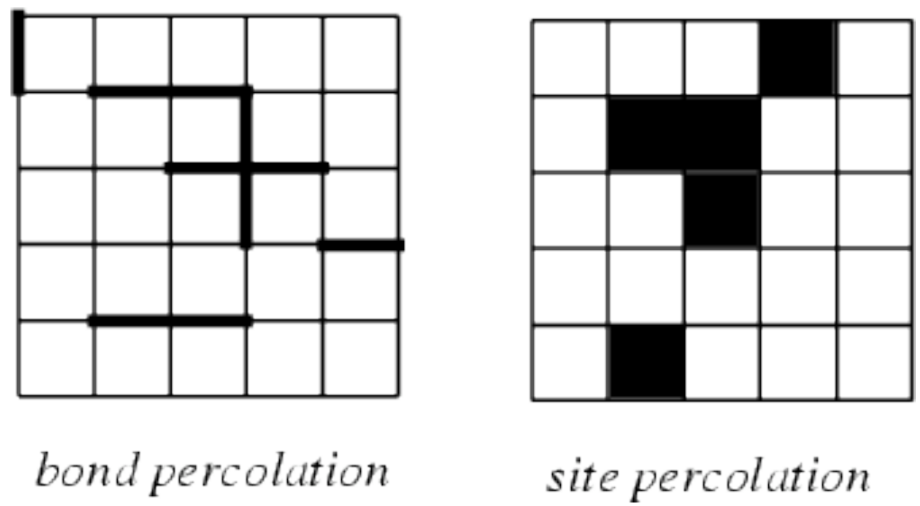
Hierarchy in Graphs





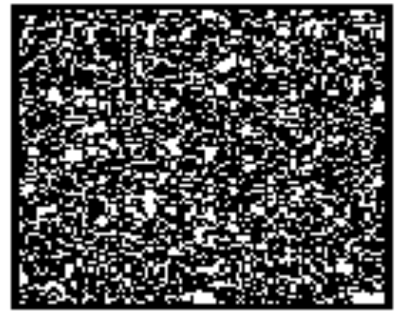
Clauset, Moore, Newman, *submitted*

Network Robustness and Resilience: Percolation as a Starting Point

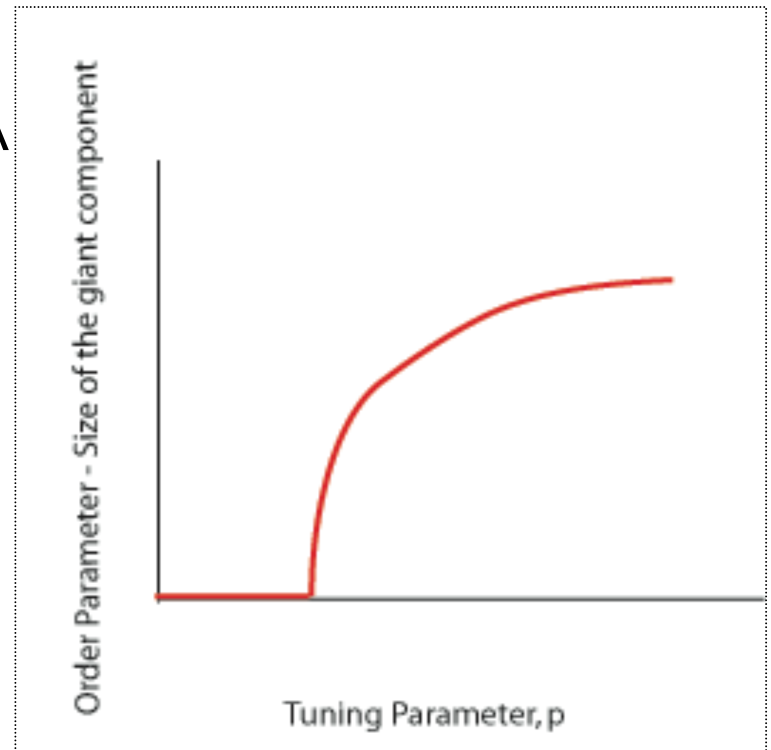


Ordinary Percolation on Lattices: Fill in each link (bond percolation) or site (site percolation) with probability p and ask questions about the sizes of connected components.

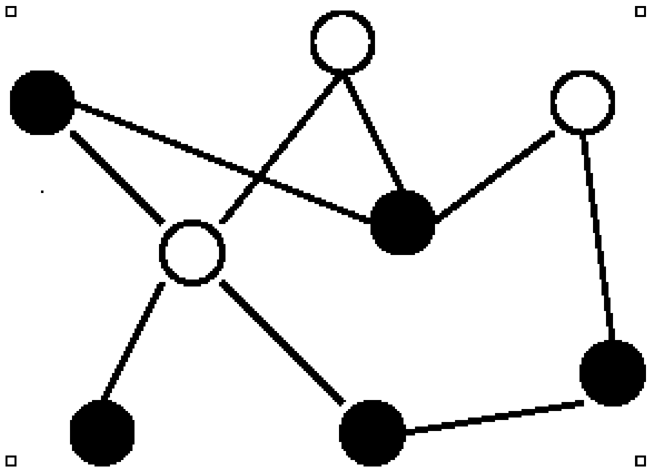
Q: What happens as we increase the probability, *p*, of filling in each site?



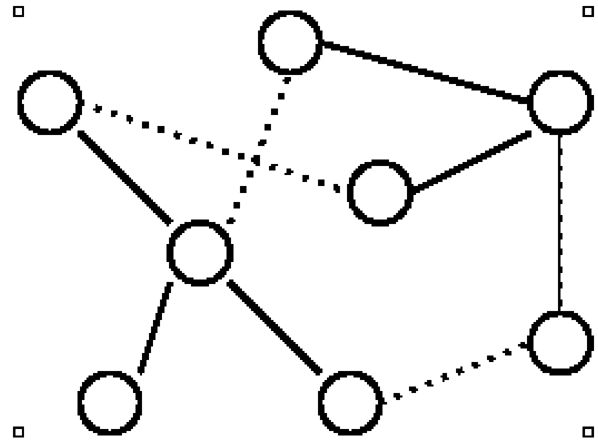
- For low values of p, we see small islands of connected components.
- At a critical value of p, a giant component forms. A giant component is a connected component that occupies a finite fraction of the system, in the limit of infinite system size. At the critical point, there is a power law distribution of the size of connected components.
- Above the critical value, the giant component occupies an increasingly large fraction of the system. If we look at the mean component size excluding the giant component, we observe a characteristic component size.



Percolation on Complex Networks



Site Percolation



Bond Percolation

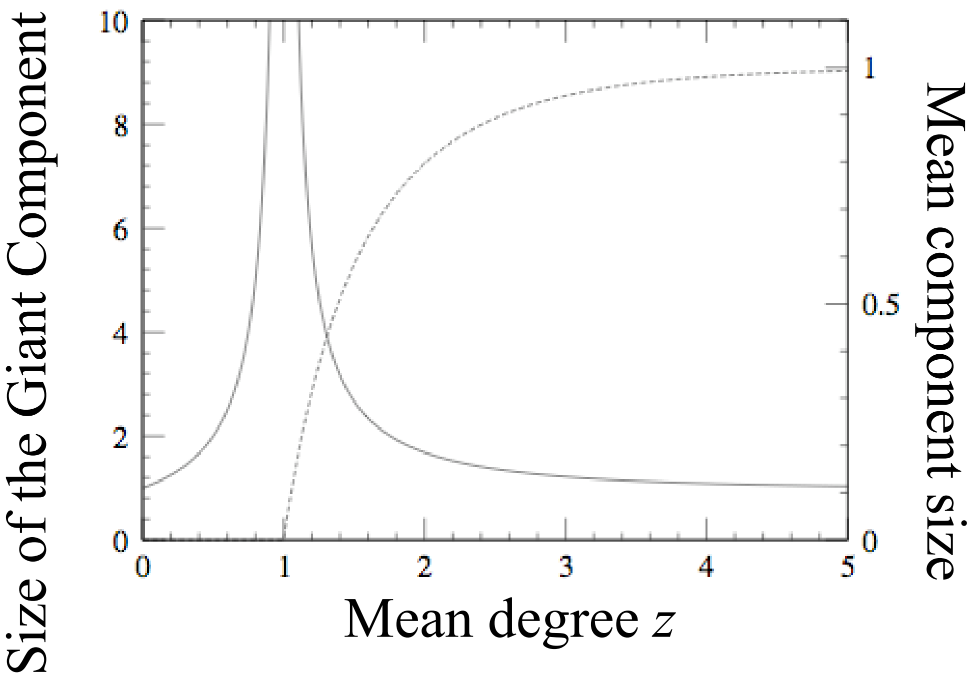
- Percolation can be extended to networks of arbitrary topology.
- We say the network percolates when a giant component forms.

How does percolation relate to network resilience?

- We consider the resilience of the network to the removal of its vertices (site percolation) or edges (bond percolation).
- As vertices (or edges) are removed from the network, the average path length will increase.
- Ultimately, the giant component will disintegrate.
- Networks vary according to their level of resilience to vertex (or edge) removal.

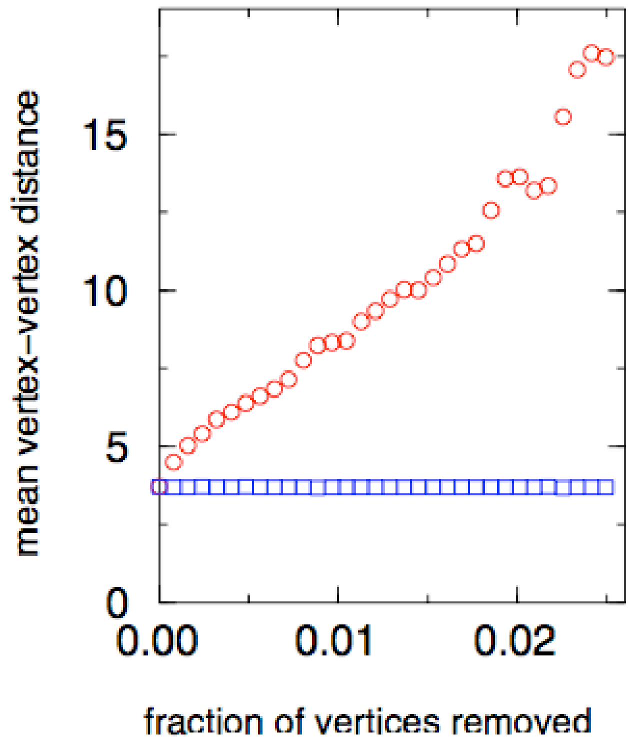
Percolation in Poisson Random Graphs

We can study the percolation process either by varying occupation probability, or some other parameter of the networked system. Here, we consider Poisson (Erdos-Renyi) random graphs in which our tuning parameter is the average degree of the system.



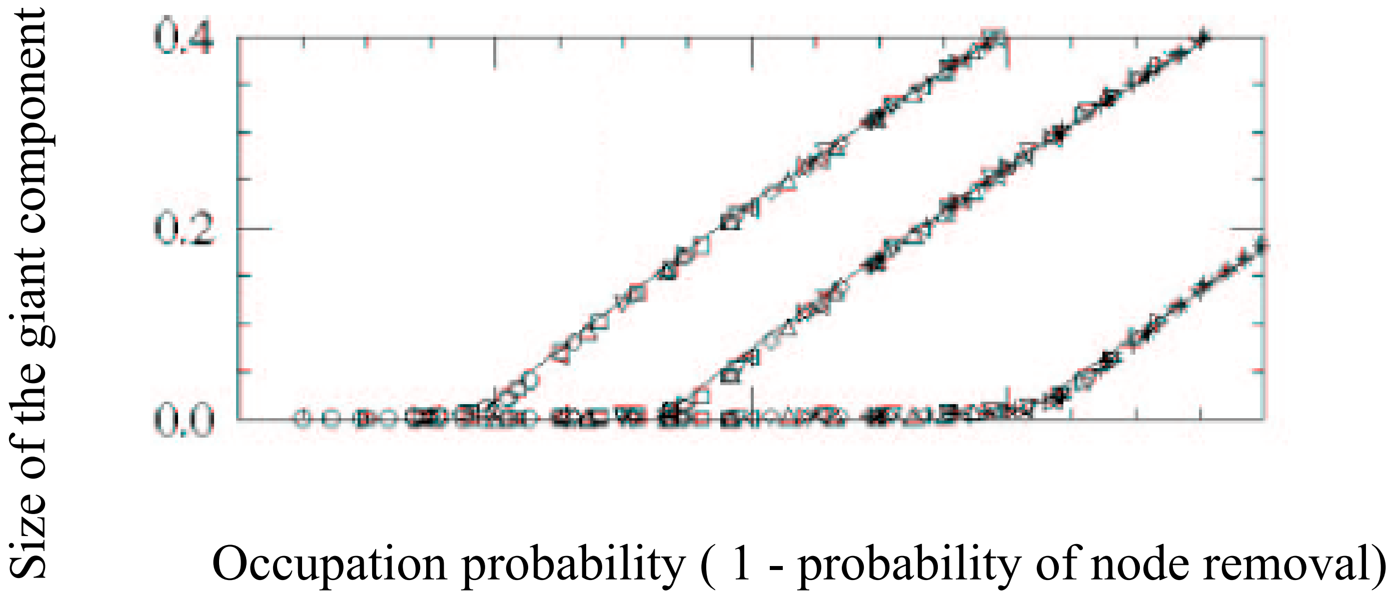
Robustness and fragility of scale free networks

Mean vertex–vertex distance on a graph representation of the Internet at the autonomous system level, as vertices are removed one by one. If vertices are removed in random order (squares), distance increases only very slightly, but if they are removed in order of their degrees, starting with the highest degree vertices(circles), then distance increases sharply. We say the network is resilient to random removal of vertices, but sensitive to targeted removal.



R. Albert, H. Jeong, and A.-L. Barabasi, *Attack and error tolerance of complex networks*, Nature, 406 (2000), pp. 378–382.

Resilience of skewed networks to random removals



Size of the giant component as a function of the occupation probability for three different degree distributions, decreasing skewness from left to right.

D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Network robustness and fragility: Percolation on random graphs*, Phys. Rev. Lett., 85 (2000), pp. 5468–5471.

Calculating the phase transition for networks with arbitrary degree distributions that are otherwise random

- Let p_k be the probability density function for the degrees of nodes in the network.
- Let q_k be the probability density function for the degree of a node at the end of a randomly chosen edge: $q_k = k\, p_k / \langle k \rangle$.
- Let f be the occupation probability for each edge.

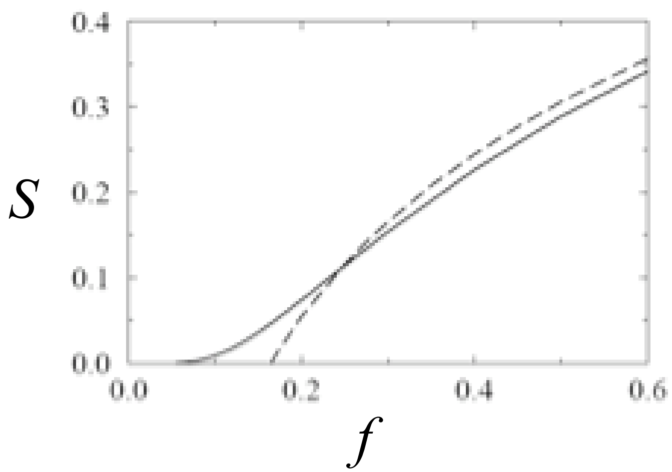


Figure taken from Pastor-Satorras and Vespignani, 2002)

- Assume that we start with one node and we want to find the size of the component connected to that node.
- Let z_n be the number of neighbors reachable in n steps.
- $z_{n+1} = z_n$ x the average excess degree of nodes reachable in n steps (the expectation value of $(k-1)$ over the distribution q_k) x the occupation probability, f

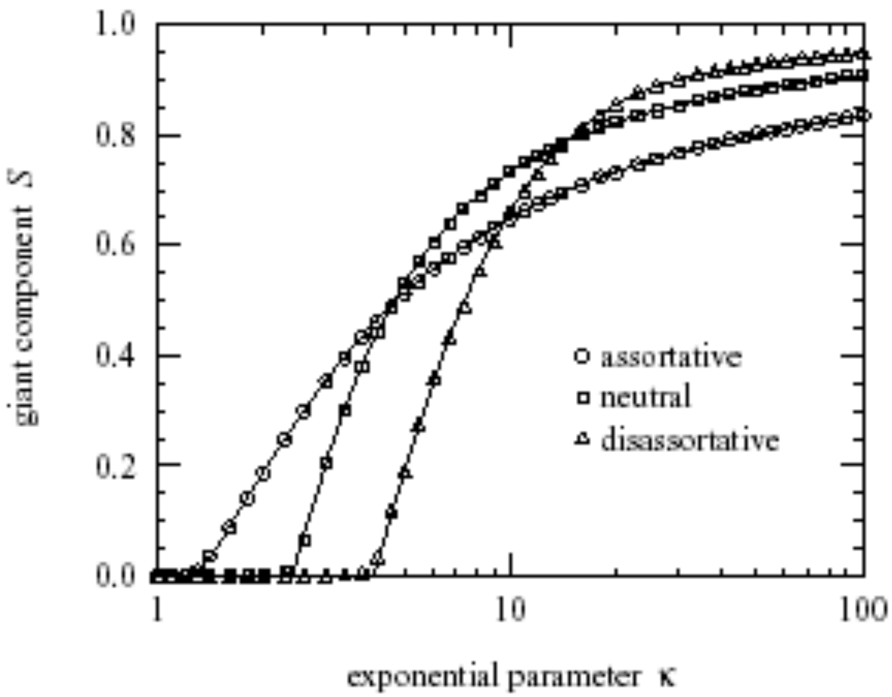
$$z_{n+1} = f\, z_n (\langle k^2 \rangle - \langle k \rangle)/\langle k \rangle$$

- Critical value of f:

$$f = \langle k \rangle / (\langle k^2 \rangle - \langle k \rangle)$$

How does assortative mixing impact percolation?

- In assortatively mixed networks, the phase transition occurs sooner as the tuning parameter is increased.
- When the tuning parameter is well past the critical point, the size of the giant component is largest for disassortative networks.



Some cautionary tales of complex networks…

Barabasi et al. showed that the internet was fragile to a targeted attack of high degree nodes. This is true, but slightly misleading. The network they looked at was on the level of autonomous systems, groups of singly managed computers each consisting of thousands of machines. The cost of disabling high degree nodes in the network is expected to be much higher than the cost of disabling a random node.

Poor sampling approaches can also give us misleading results

Traceroute sampling makes random graphs appear to have power law degree distributions

Aaron Clauset* and Cristopher Moore†,*
Computer Science Department and †Department of Physics and Astronomy,
University of New Mexico, Albuquerque NM 87131
`{aaron,moore}@cs.unm.edu`
(Dated: June 13, 2005)

The topology of the Internet has typically been measured by sampling *traceroutes*, which are roughly shortest paths from sources to destinations. The resulting measurements have been used to infer that the Internet's degree distribution is scale-free; however, many of these measurements have relied on sampling traceroutes from a small number of sources. It was recently argued that sampling in this way can introduce a fundamental bias in the degree distribution, for instance, causing random (Erdős-Rényi) graphs to appear to have power law degree distributions. We explain this phenomenon analytically using differential equations to model the growth of a breadth-first tree in a random graph $G(n, p = c/n)$ of average degree c, and show that sampling from a single source gives an apparent power law degree distribution $P(k) \sim 1/k$ for $k \lesssim c$.

A **Clauset**, C **Moore**, - Arxiv preprint cond-mat/0312674, 2003 - arxiv.org

Outlook for the Field

- We have focused primarily on developing measures to characterize these complex networks, we need to move away from a purely structural point of view to a more dynamical point of view.
- Current approaches in complex network theory generally work in the large network limit. We need to expand our toolbox to hand smaller networks with the same rigor.