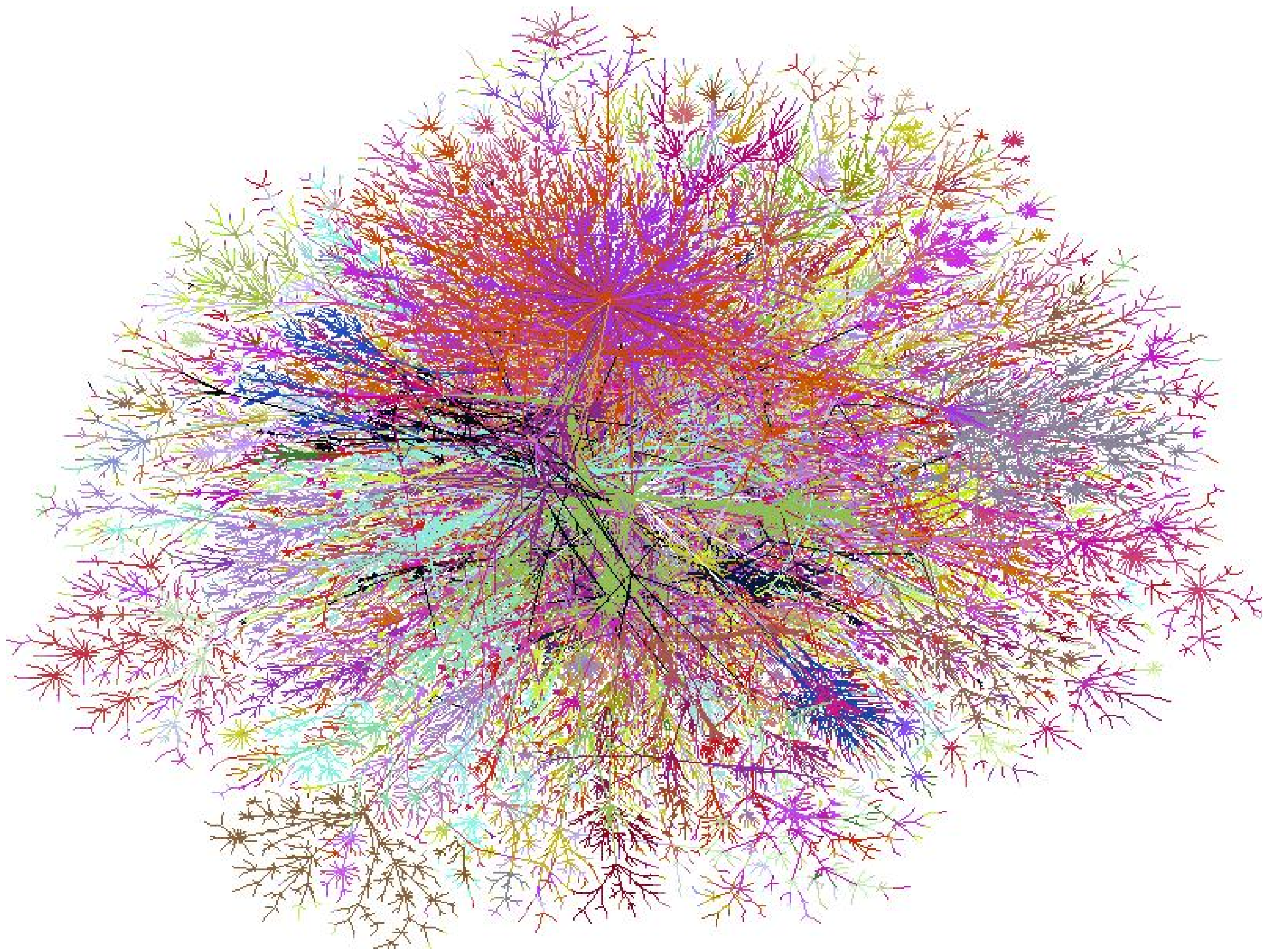
ECS 253 / MAE 253, Lecture 4

April 7, 2016



"Power laws, network robustness, and Small World Networks"

Announcements

- Office hours. morning or afternoon? :
  - Andrew Smith, Tues/Thurs
  - Haochen Wu, M/W

- Multi-dimensional Networks Symposium, May 20-22, 2016.

- Project ideas added under “resources” tab of smartsite.

Network models studied so far

- Erdős-Rényi random graphs, $G(N, p)$
  - Initialized with N isolated nodes
  - Edges arrive in discrete time process with uniform prob.
  - Poisson degree distribution
  - No clustering
  - Emergence of a giant component

- Preferential attachment graphs
  - Initialized with one (or a small set) of seed nodes
  - Nodes arrive and attach with m edges choosing “parent” with prob proportional to degree, $q_{k,t} = k/2mt$.
  - Power law deg dist with $\gamma = 3$
  - Clustering tuned by setting m
  - Fully connected network by construction

PA analyzed via rate equation approach , for p_k

– Gain an edge with probability proportional to current degree.

– Probability a node of degree k gains attachment is $d_k / \sum_k d_k$.

– Let $n_{k,t}$ denote the *expected number* of nodes of degree k at time t.

- For $k > m$: $\quad n_{k,t+1} = n_{k,t} + \frac{m(k-1)}{2mt} n_{k-1,t} - \frac{mk}{2mt} n_{k,t}$

- For $k = m$: $\quad n_{m,t+1} = n_{m,t} + 1 - \frac{m^2}{2mt} n_{m,t}$

Yields: $\boxed{p_k = \frac{2m(m+1)}{(k+2)(k+1)k}}$

For $k \gg 1$

$$\boxed{p_k \sim k^{-3}}$$

Summary of kinetic theory / rate eqn approach

- A stochastic, discrete time process for an evolving graph $G(t)$.
- Approximation 1: Study the average random graph.
- Let $n_{k,t}$ denote the *expected (i.e. average)* number of nodes of degree k at time t into the process. (So $n_{k,t}$ is a real number, not an integer.)
- Write $n_{k,t+1}$ in terms of the $n_{k,t}$'s, accounting for the rates at which node degree is expected to change.
- Translate from $n_{k,t}$ to $p_{k,t} = n_{k,t}/n_t$. For PA $n_t = t$.
- Approximation 2: Assume steady state $p_{k,t} \to p_k$.
- Solve for a recurrence relation for the p_k's. For PA, $p_k = k^{-3}$ for large k.
- Still need to show *convergence (Approx 1) and concentration (Approx 2)*

Difference between ER and PA is not due to edge versus node arrival

- **Edge-arrival PA graph**

  K.-I. Goh, B. Kahng, D. Kim, *Phys. Rev. Lett.* **87**, (2001).

  F. Chung and L. Lu, *Annals of Combinatorics* **6**, 125 (2002). *

  – Initialized with N isolated nodes, labeled $i \in \{1, 2, ..., N\}$, where each node i has a weight $w_i = (i + i_0 - 1)^{-\mu}$.

  – Two vertices (i, j) selected with probability $w_i / \sum_k w_k$ and $w_j / \sum_k w_k$ respectively and connected by an edge.

  – Yields $p_k = Ak^{-\gamma}$ with $\gamma = \mu = -1/(\gamma - 1)$.

  – (Master eqn analysis: Lee, Goh, Kahng and Kim, Nucl. Phys. B 696, 351 (2004).)

* "Chung-Lu" model used extensively to generate graphs.

Difference between ER and PA is not due to edge versus node arrival

- **Erdős-Rényi-like process with node arrival**

  Callaway, Hopcroft, Kleinberg, Newman, Strogatz.
  Phys Rev E **64** (2001).

  – At each discrete time step a new node arrives, and with probability δ a new randomly selected edge arrives.

  – Emergence of giant component only if $\delta \geq 1/8$.

  – Infinite order phase transition. (Kosterlitz Thouless transition.)

  – (That "giant" is finite even as $n \to \infty$).

  – Positive degree-degree correlations (higher degree by virtue of age).

Preferential Attachment and "Scale-free networks"
Why a power law is "scale-free"

- Power law for "x", means "scale-free" in x:

$$p(bx) = (bx)^{-\gamma} = b^{-\gamma} p(x)$$

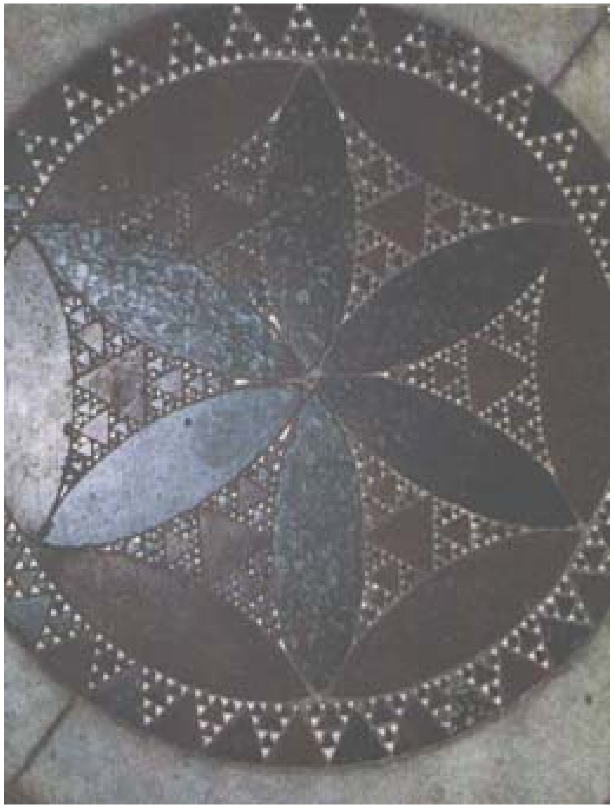
$\boxed{\frac{p(bk)}{p(k)} = b^{-\gamma}}$ regardless of k.

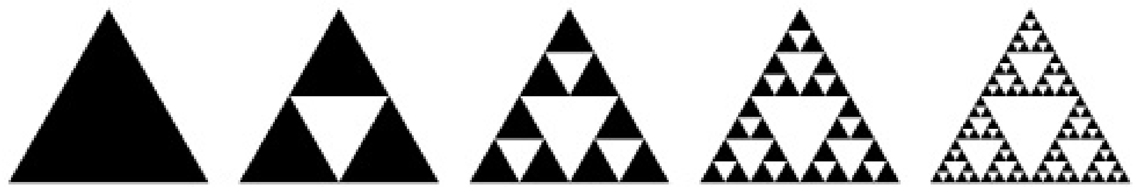
In contrast consider: $p(k) = A \exp(-k)$.

So $p(bk) = A \exp(-bk)$.

$\boxed{\frac{p(bk)}{p(k)} = \exp[-k(b-1)]}$ dependent on k

Self-similar/scale-free fractal structures





Sierpinski Sieve/Gasket/Fractal, $N \sim r^d$.

When r doubles, N triples: $3 = 2^d$

$d = \log N / \log r = \log 3 / \log 2$

Power law degree distribution $\neq$ "scale-free network"

- Power law for "x", means "scale-free" in x.
- BUT only for that aspect, "x". May have a lot of different structures at different scales.
- **More precise: "network with scale-free degree distribution"**

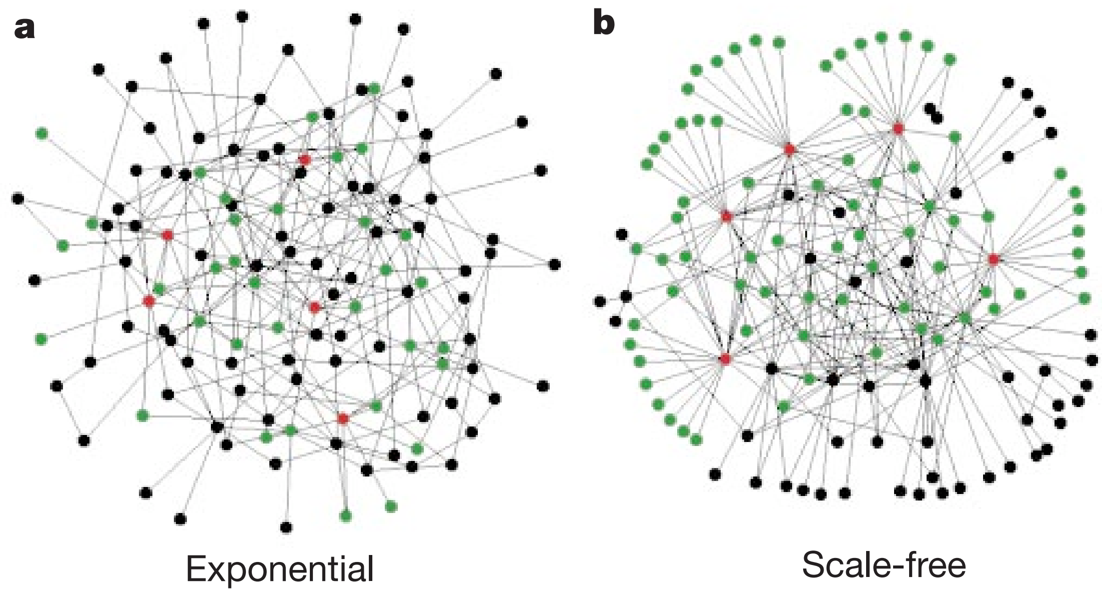
Power Law Random Graph (PLRG)

Robustness of a network

- Robustness/Resilience: A network should be able to absorb disturbance, undergo change and essentially maintain its functionality despite failure of individual components of the network.

- Often studied as maintaining connectivity despite node and edge deletion.

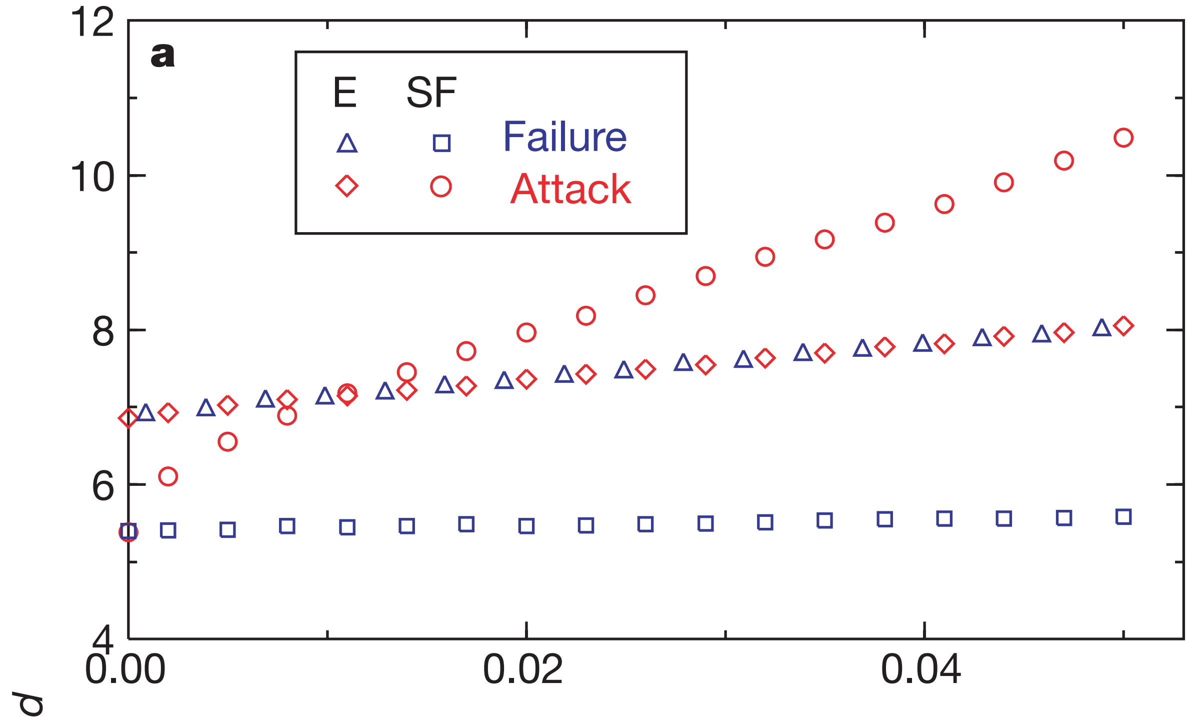
Albert, Jeong and Barabasi, “Error and attack tolerance of complex networks”, Nature, **406** (27) 2000.



N=130, E=215, Red five highest degree nodes; Green their neighbors.

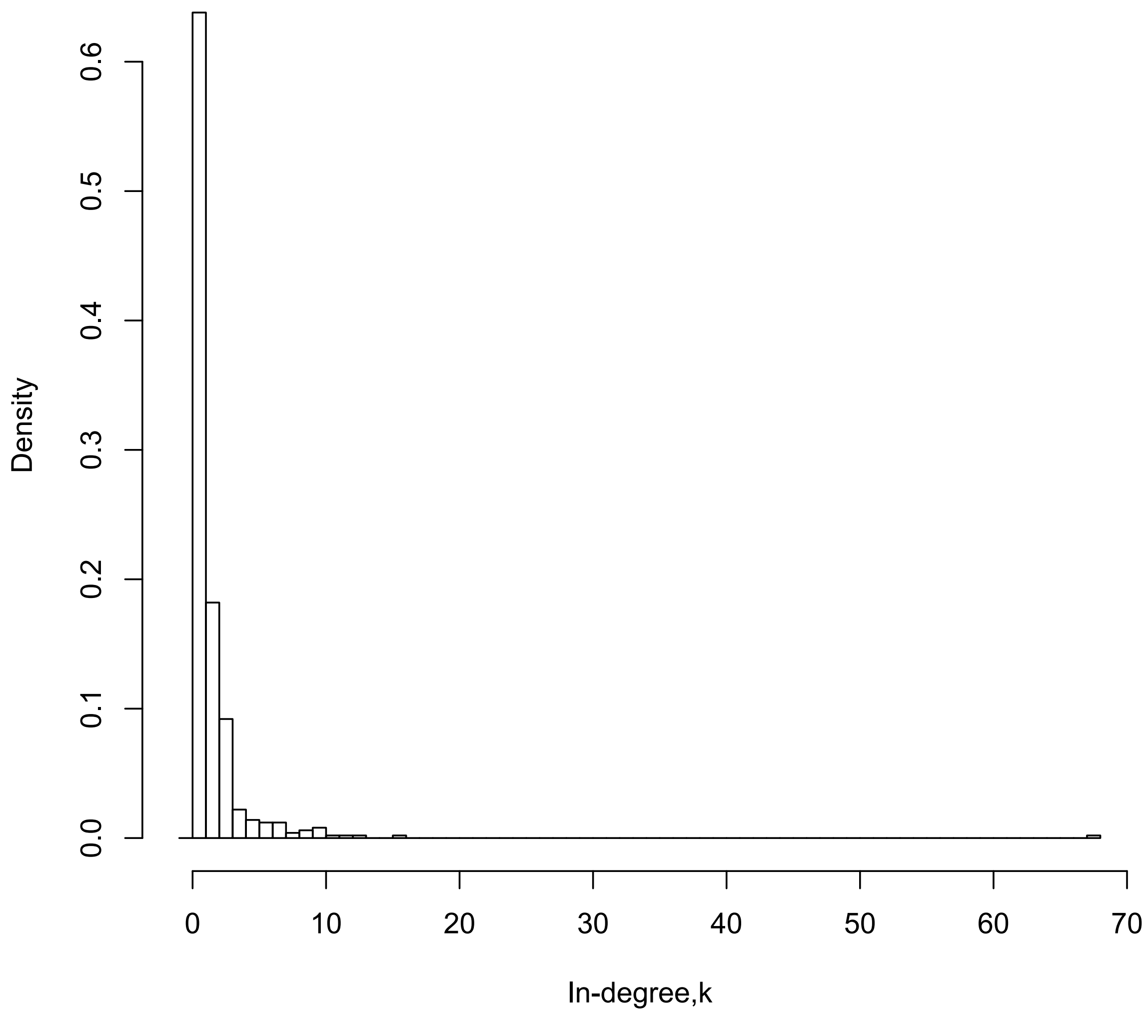
- Exp has 27% of green nodes, SF has 60%.
- PLRG: Connectivity extremely robust to random failure.
- PLRG: Connectivity extremely fragile to targeted attack (removal of highest degree nodes).

Exponential vs scale-free: Robustness



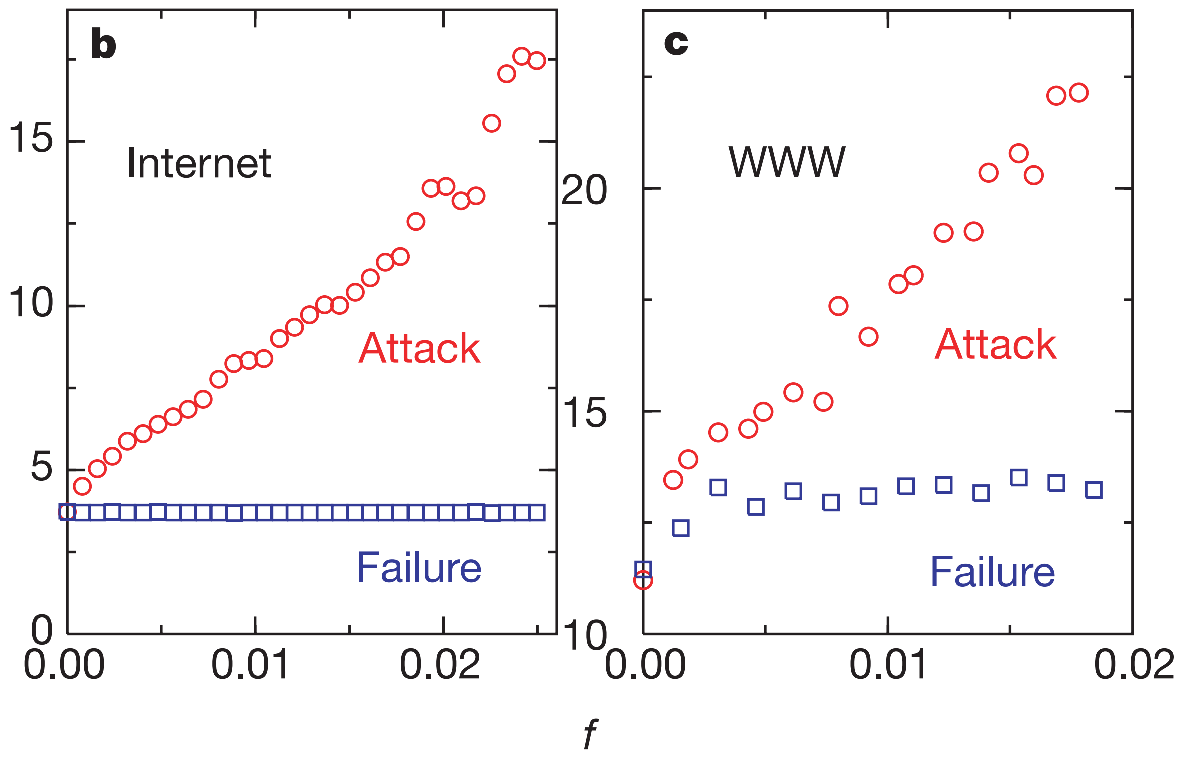
- (Remember, bigger diameter is worse.)
- SF are extremely robust to random failure (blue squares). Remove fraction of nodes at random, and no change in diameter.
- SF are very fragile to targeted attack (removal of highest degree nodes).

Histogram of a typical PA run
Degree distribution (Here N=500)



- Choosing node at random overwhelmingly leads to low degree node

Degree-targeted removal on real sample topologies



- Used the topological map of the Internet, containing 6,209 nodes and 12,200 links $< k >= 3.4$), collected (in 1999 or 2000) by the National Laboratory for Applied Network Research http://moat.nlanr.net/Routing/rawdata/

- World-Wide Web data measured on a sample containing 325,729 nodes and 1,498,353 links, such that $< k >= 4.59$.

Albert, Jeong and Barabasi, Nature, 406 (27) 2000

“The Achilles Heel of the Internet”

- “How robust is the Internet?” Yuhai Tu,
  Nature (New and Views) **406** (27) 2000.

- “Scientists spot Achilles heel of the Internet”,
  CNN, July 26, 2000.

Percolation theory to show the similar results follow in an analytic mathematical formulation

- R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, "Resilience of the Internet to Random Breakdowns", *Phys. Rev. Lett.* 85, 4626 (2000).

- Callaway, Duncan S.; M. E. J. Newman, S. H. Strogatz and D. J. Watts, "Network Robustness and Fragility: Percolation on Random Graphs". *Phys. Rev. Lett.* 85: 546871 (2000).

- $\langle k \rangle$ finite, but $\langle k^2 \rangle \to \infty$ for PLRG with $2 < \gamma < 3$, the cornerstone for the arguments.

Results from Callaway et al

- Degree dist, $p_k \sim k^{-\gamma} e^{-k/C}$ (power law with cutoff w $C \to \infty$).

- Let q be probability that a vertex is "active"/"infected".
  For simplicity assume independent of k.

- Then $p_k q$ is probability of having degree k and being infected.

- Calculate $\langle s \rangle$, the mean cluster size of active nodes. Find (via generating functions ... details later in the course) that

$$\langle s \rangle = q + \frac{q^2 \langle k \rangle}{1 - (q \langle k^2 \rangle / \langle k \rangle)}$$

- $\langle s \rangle \to \infty$ when denominator $1 - q \left\langle k^2 \right\rangle / \langle k \rangle = 0$, i.e.,

$$\boxed{q_c = \frac{\langle k \rangle}{\langle k^2 \rangle}}$$

Infinite cluster even if probability $\to$ 0, when $p_k \sim k^{-\gamma}$ for $2 < \gamma < 3$) .

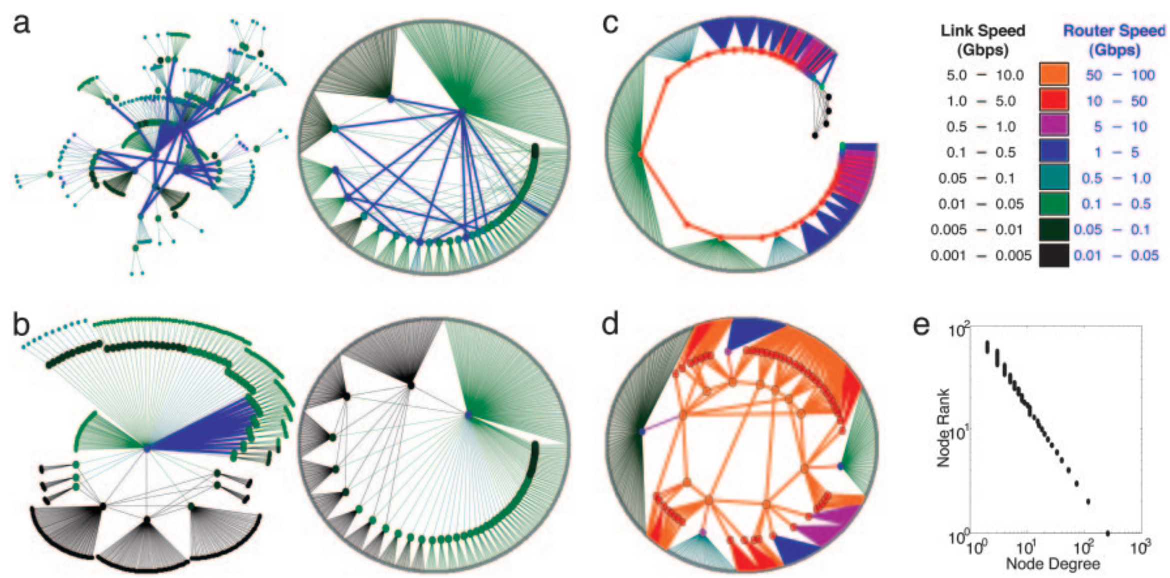
Does the ensemble of random graphs really model engineered or biological systems?

(Is the Internet a random scale-free graph?)

Random vs engineered vs evolved (e.g. biological) systems

- REDUNDANCY!!! a key principle in engineering (and evolution?).

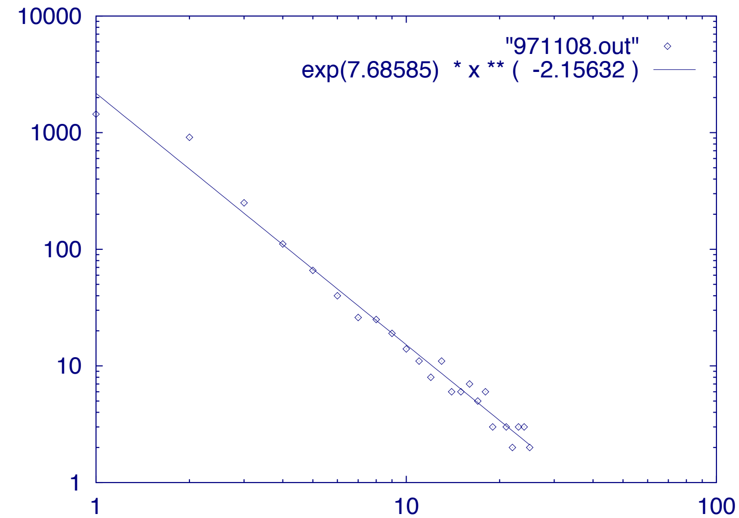
- The 'robust yet fragile' nature of the Internet
  Doyle, Alderson, Li, Low, Roughan, Shalunov, Tanaka, Willinger, PNAS **102** (4) 2005.



- Degree distribution is not the whole story.

Wikipedia entry on "scale-free networks"

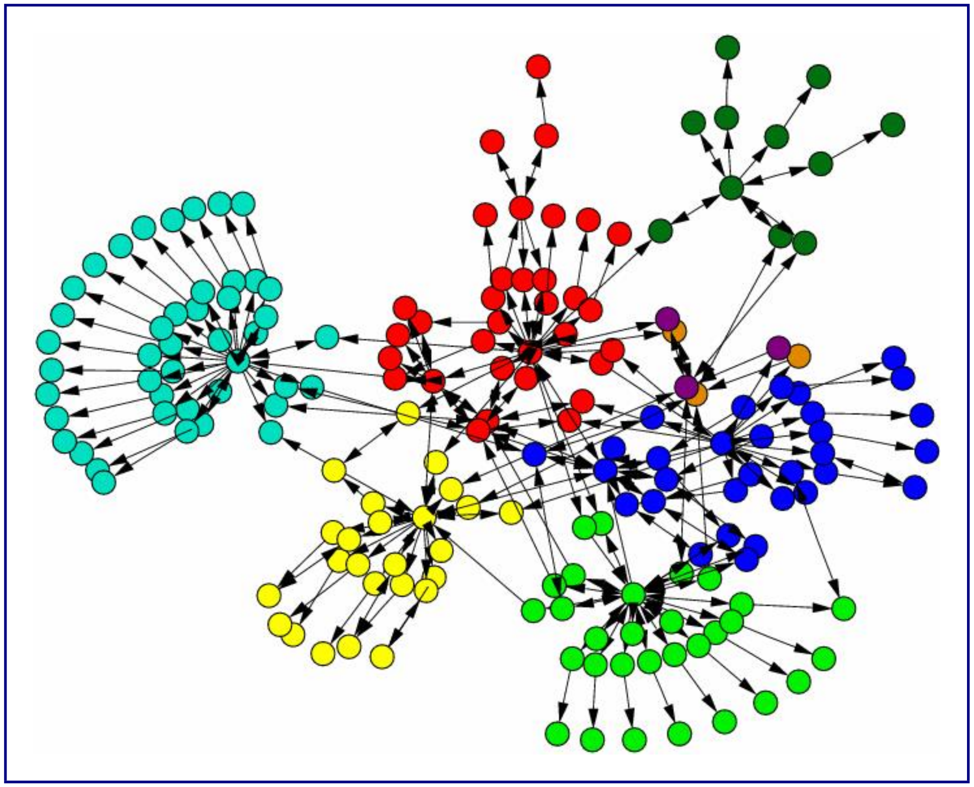
- Good discussion of the history and controversy
  - Faloutsos SIGCOMM 1999 paper on power law in Internet based on trace route sampling.

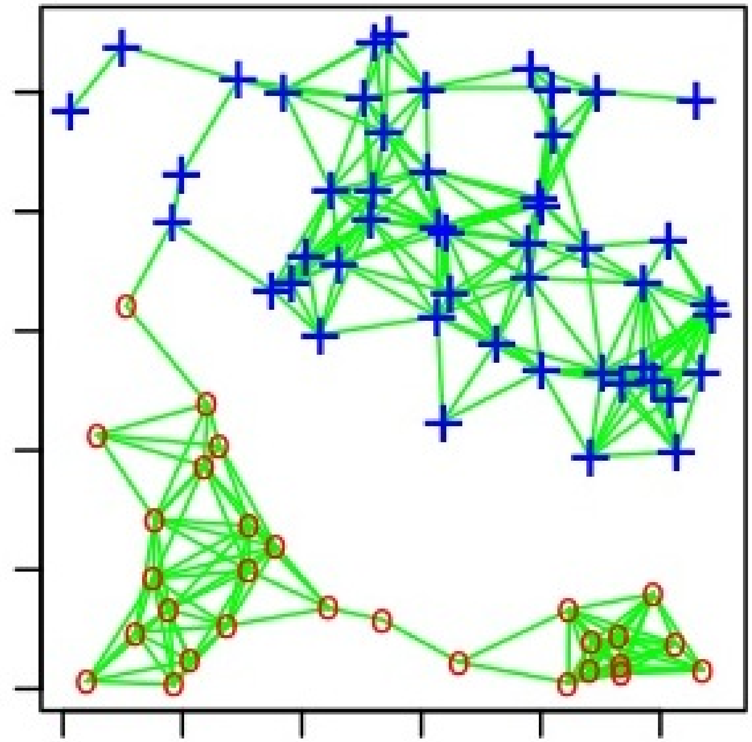


(a) Int-11-97

  - Although many real-world networks are thought to be scale-free, the evidence often remains inconclusive, primarily due to the developing awareness of more rigorous data analysis techniques.

Effectively breaking up different networks





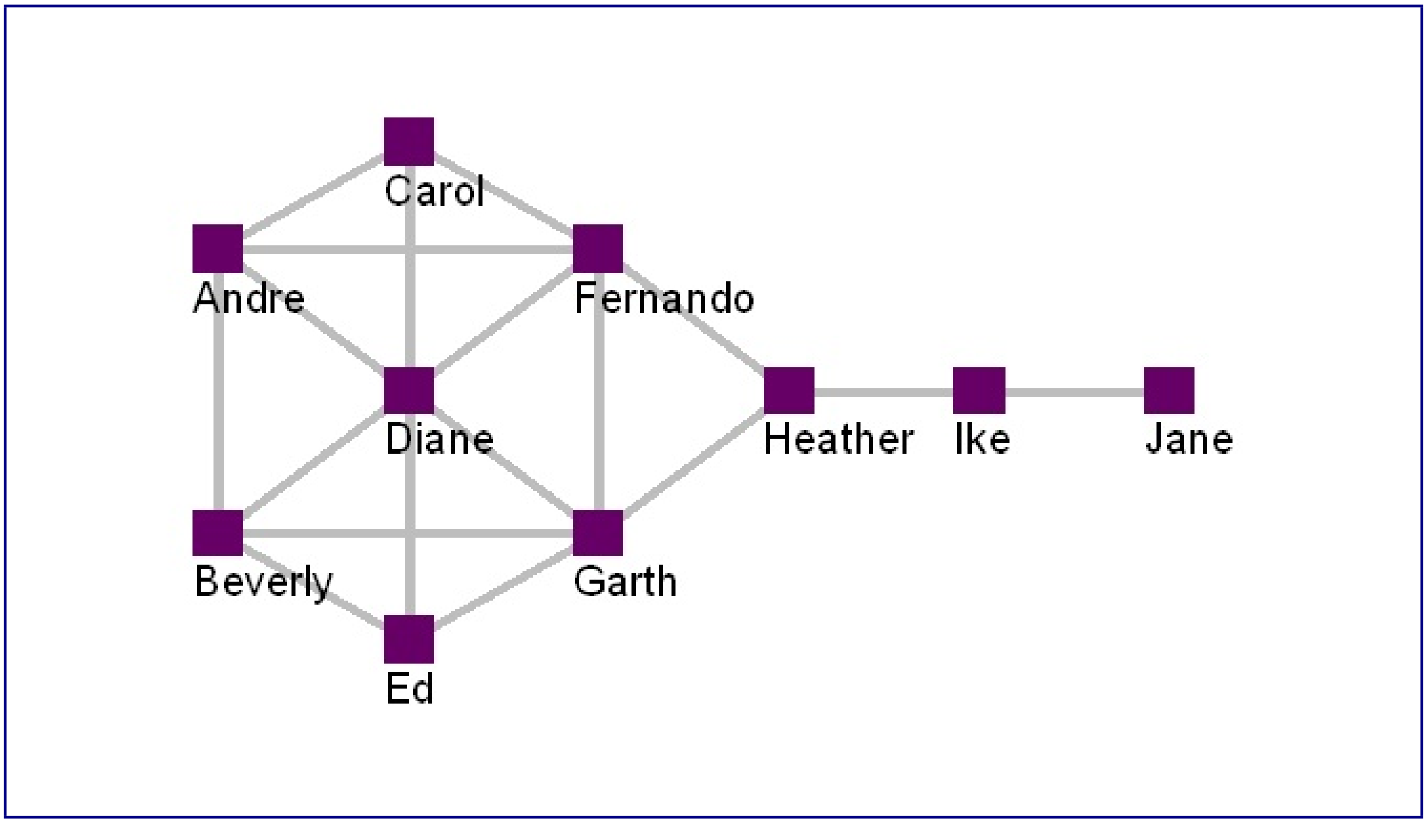
What other types of nodes play key roles?

Other types of important nodes

A classic example from Social Network Analysis (SNA)

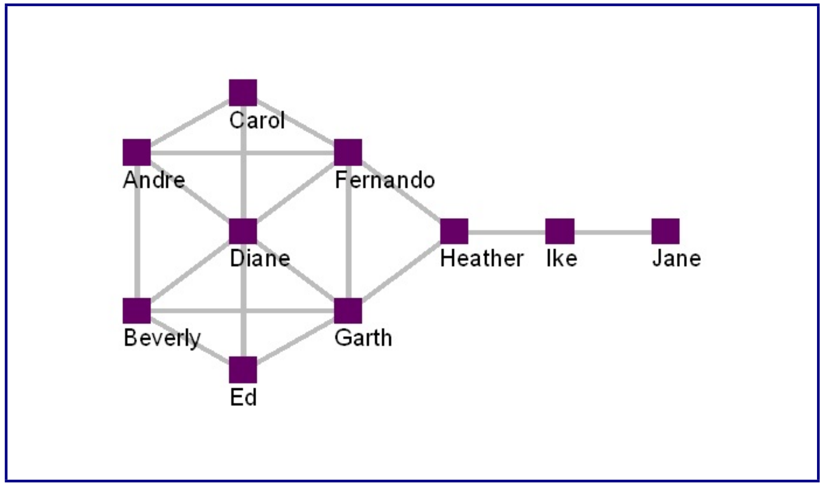
[http://www.fsu.edu/∼spap/water/network/intro.htm]

The "Kite Network"



Who is important and why?

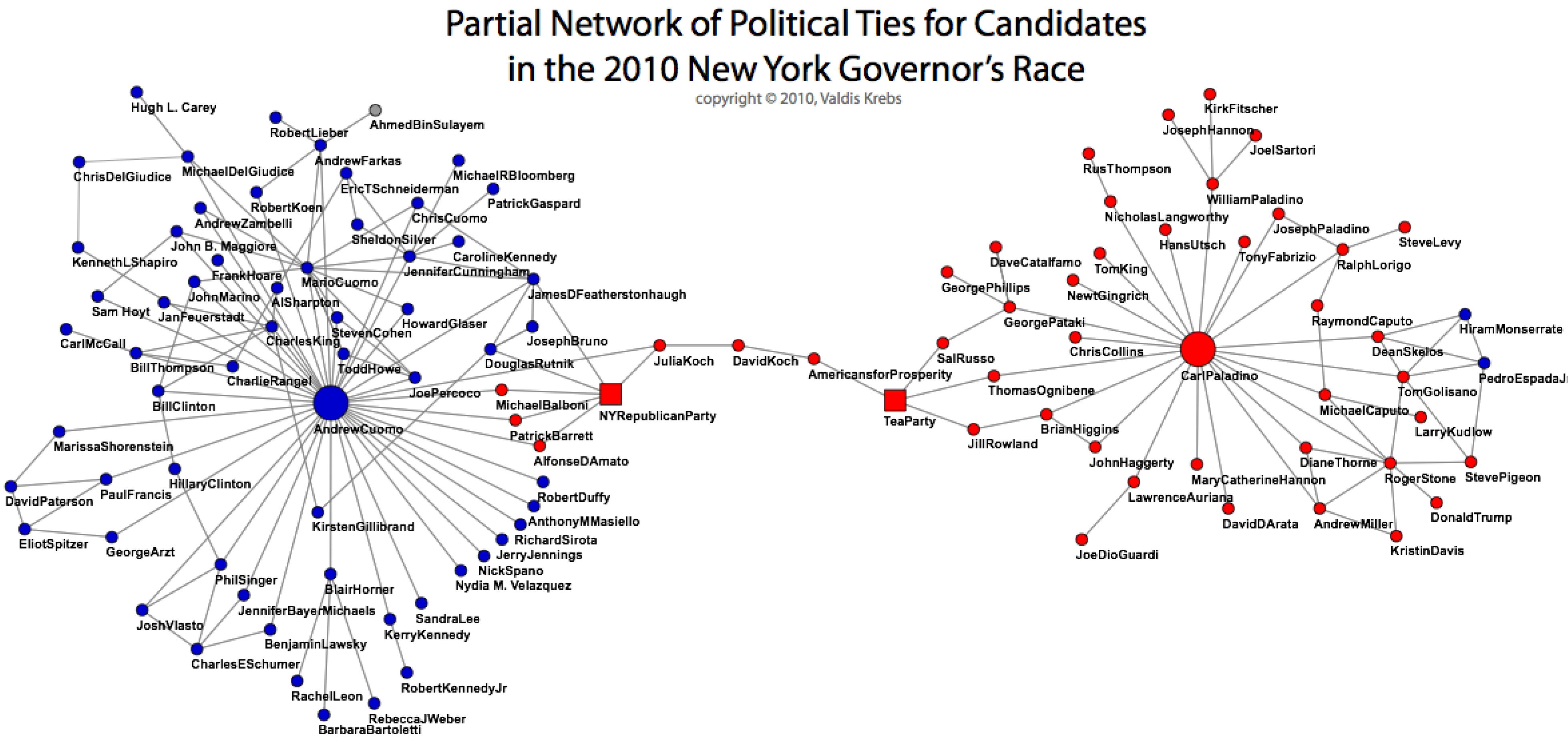
The Kite Network



- Degree – Diane looks important (a "hub").
- Betweenness – Heather looks important (a "connector"/"broker").
- Closeness – Fernando and Garth can access anyone via a short path.
- Boundary spanners – as Fernando, Garth, and Heather are well-positioned to be "innovators".
- Peripheral Players – Ike and Jane may be an important resources for fresh information.

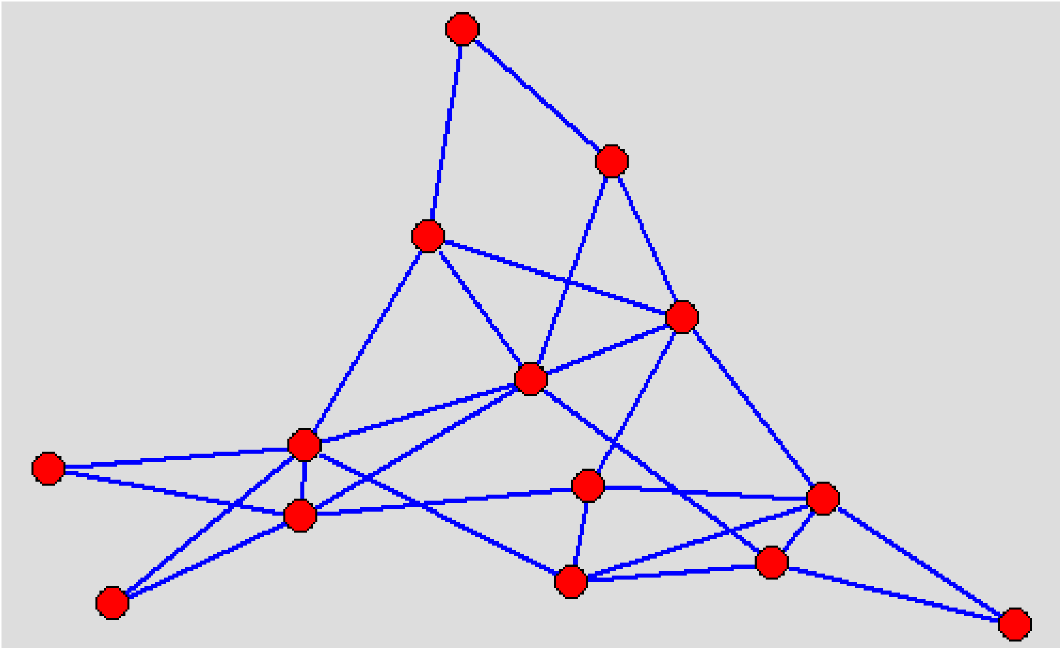
A contemporary social network

(Taken from http://www.thenetworkthinkers.com/)



Betweenness Centrality

[Freeman, L. C. "A set of measures of centrality based on betweenness." *Sociometry* **40** 1977]



A measure of how many shortest paths between all other vertices pass through a given vertex.

Betweenness (formal definition)

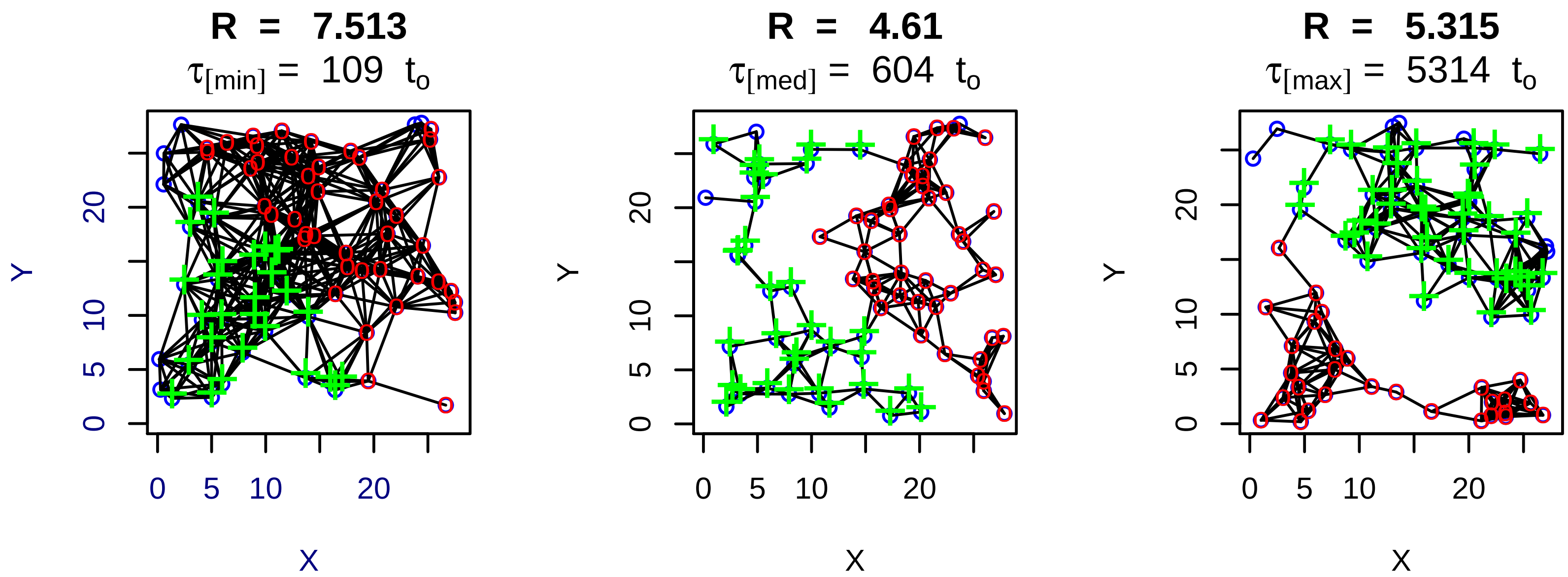
For a given vertex i:

$$\boxed{B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}}$$

- Where σ_{st} is the number of shortest geodesic paths between s and t.
- And $\sigma_{st}(i)$ are the number of those passing through vertex i.

(Calculating shortest paths efficiently ... http://en.wikipedia.org/wiki/Dijkstra's_algorithm)

Betweenness and eigenvalues (bottlenecks)



- Bottlenecks have large betweenness values.
- In social networks betweenness is a measure of a nodes “centrality” and importance (could be a proxy for influence).
- In a road network, high betweenness could indicate where alternate routes are needed.
- Also a measure of the resilience of a network (next page).

Targeted attack by different metrics

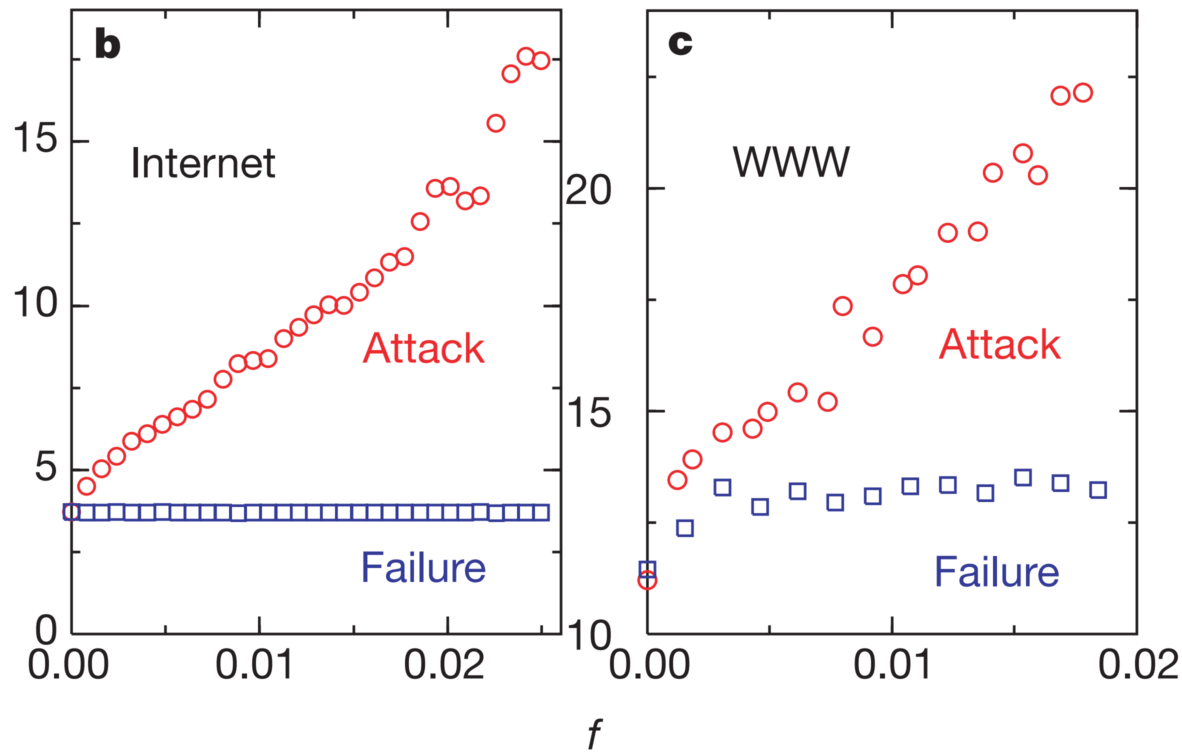
Holme P, Kim BJ, Yoon CN, Han SK (2002) "Attack vulnerability of complex networks". *Phys. Rev. E* **65**:056109

- Degree centrality
- Betweeness centrality

Typically (but not always) high degree are high betweeness.

High betweeness the more effective strategy to break up a network's connectivity.

But back to Albert, Jeong and Barabasi



So why did Albert, Jeong and Barabasi find that their sample of the internet topology was vulnerable to degree targeted attack?

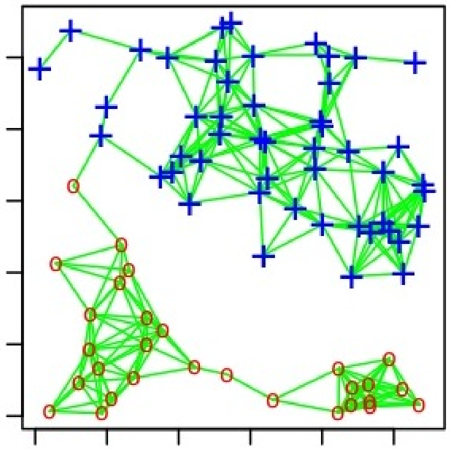
How to measure the structure of the Internet?

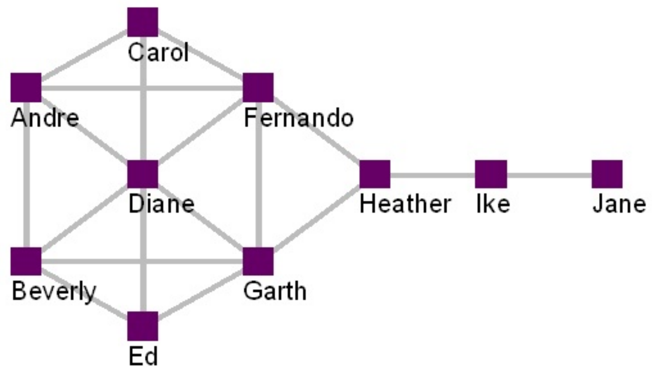
The focus of the next lecture (Lecture 5)

Summary

- **"Error and attack tolerance of complex networks"**
  Random networks with power law degree distribution show:
  – Fragility to degree-targeted removal
  – Robustness to random node removal
  (This is in the context of keeping the full network connected.)

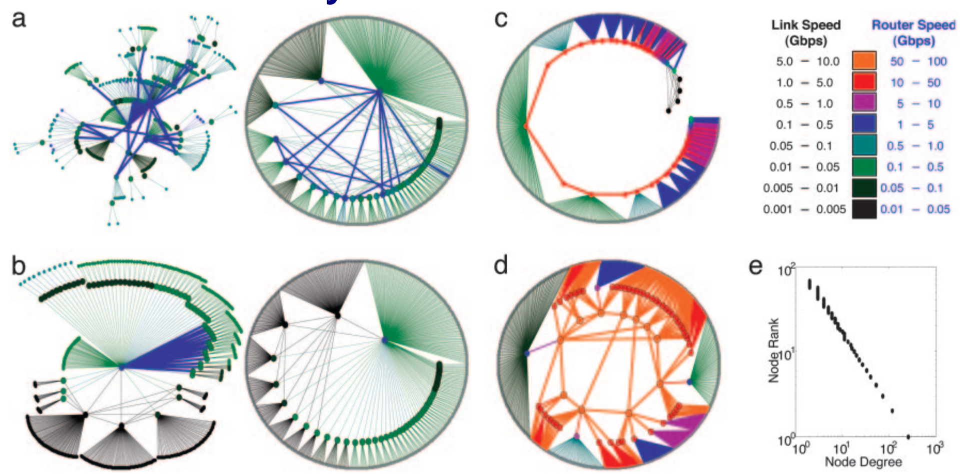
- **Important nodes beyond degree**
  – Betweeness centrality (shortest paths)
  (Are their local ways to detect this?)
  – Boundary spanners / peripheral players / weak-ties





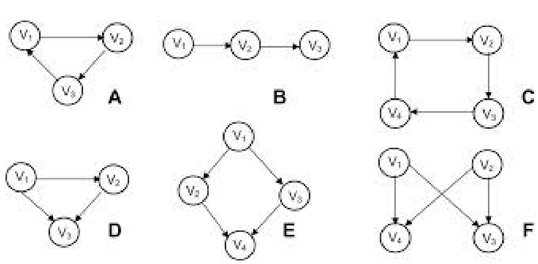
Structure beyond degree distribution

- Power law degree distribution actually a weak constraint on network structure:

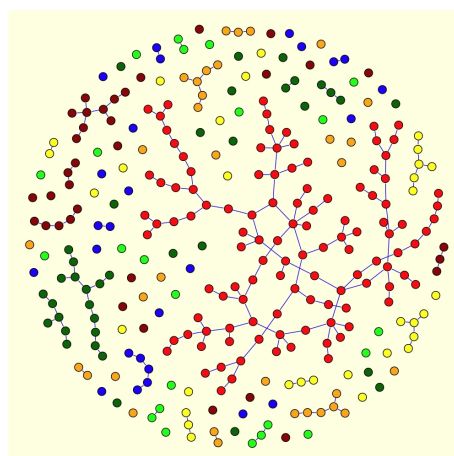


- Additional properties include:

Motifs



Components



Communities