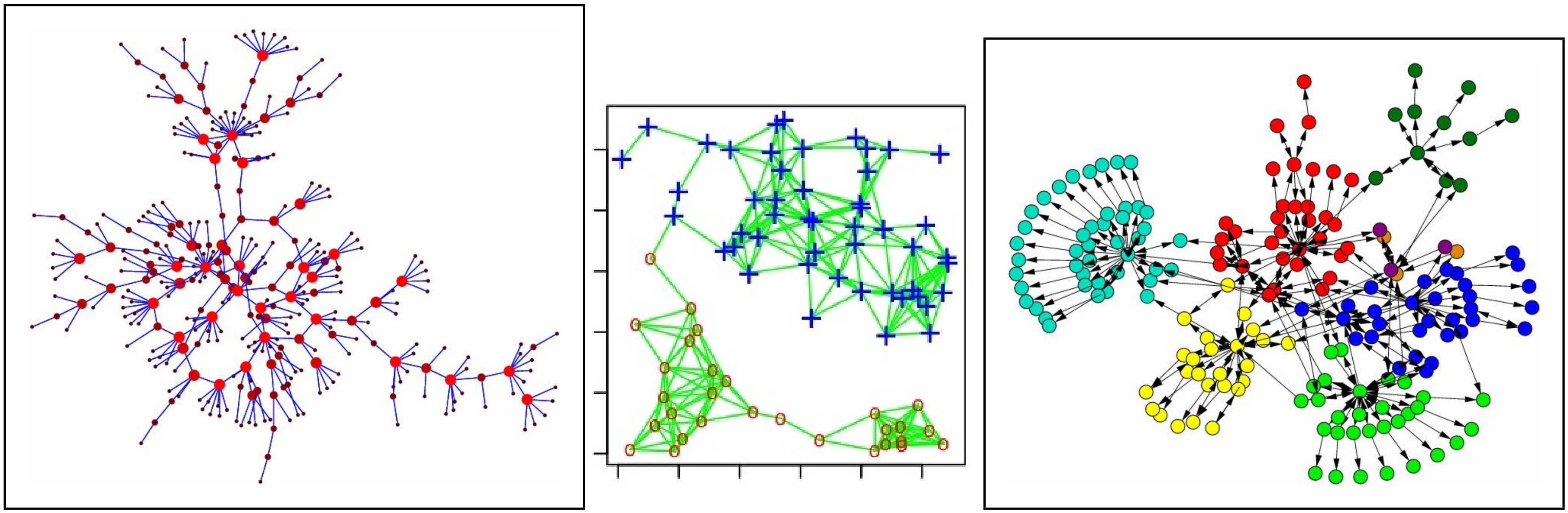
Overview of Network Theory, II



MAE 298, Spring 2009, Lecture 2

Prof. Raissa D'Souza
University of California, Davis

Today

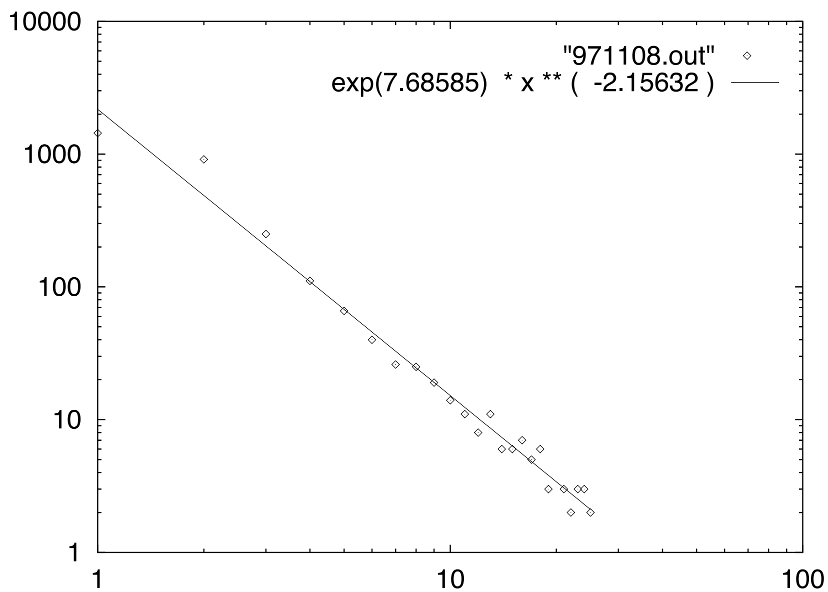
- Robustness of networks
- Optimization and network growth
- Internet overview

Typical distribution in node degree

The "Internet"

Faloutsos[3], *SIGCOMM* 1999

$$p(k) \sim k^{-2.16}$$

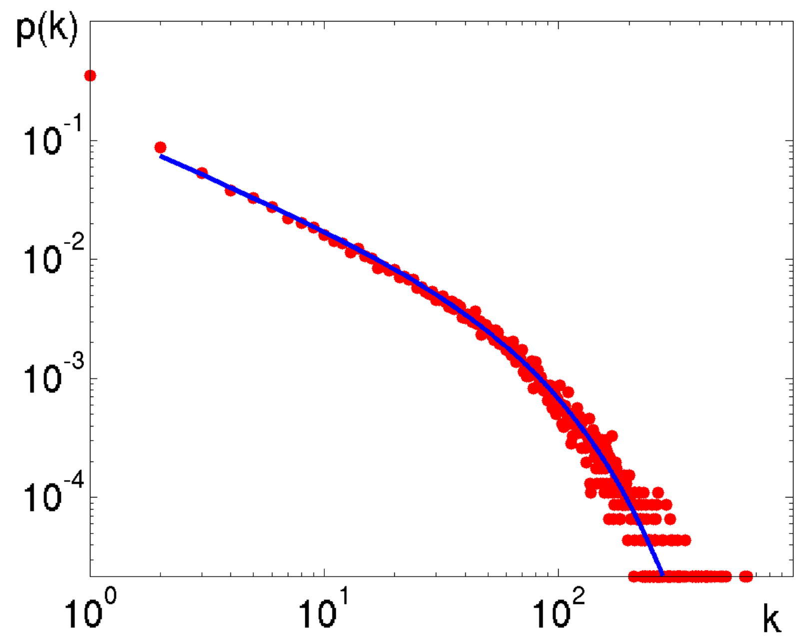


(a) Int-11-97

"Who-is-Who" network

Szendröi and Csányi

$$p(k) = ck^{-\gamma}e^{-\alpha k}$$



- Small data sets, power laws vs other similar distributions?
  - What is the "Internet"/ what level? (e.g., router vs AS)

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

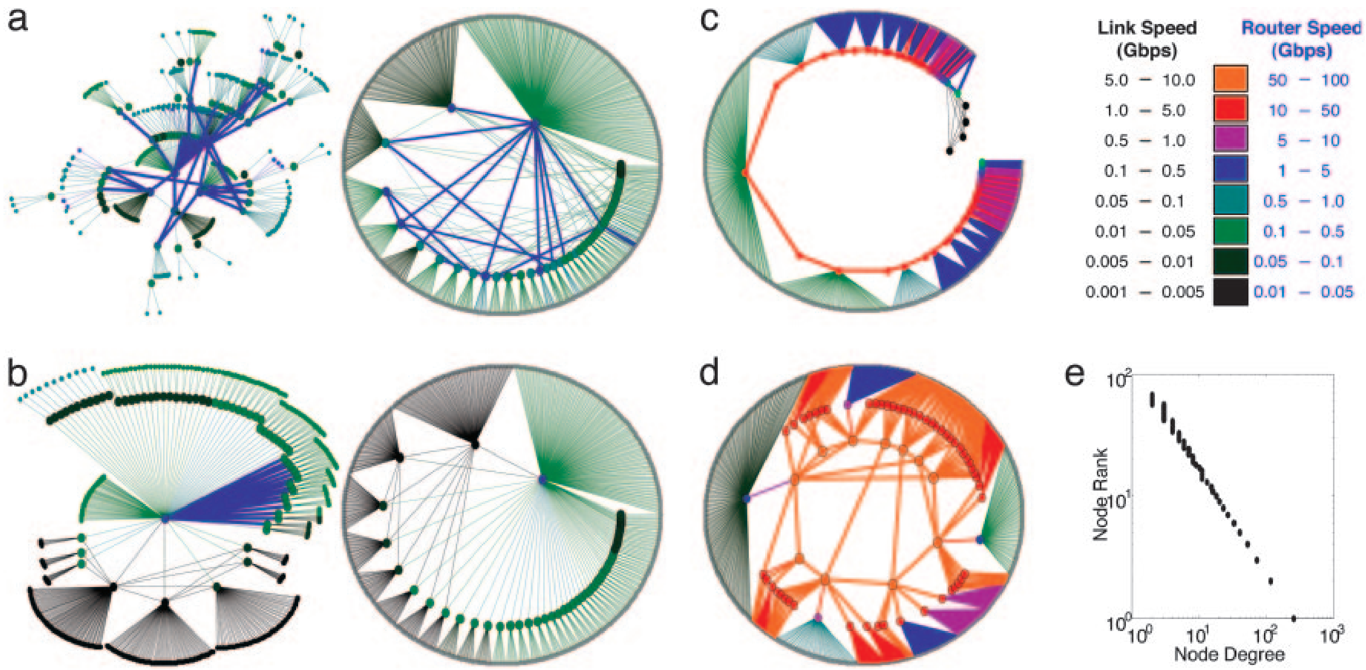
$\boxed{\frac{p(bk)}{p(k)} = \exp[-k(b-1)]}$ dependent on k

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"**

“Scale-rich” networks

- L. Li and D. Alderson and W. Willinger and J. Doyle, *Proceedings of ACM SIGCOMM*, 2004;
- Doyle, Alderson, Li, Low, Roughan, Shalunov, Tanaka, Willinger *PNAS*, 2005.



All these networks have same degree distribution, but very different internal structures.

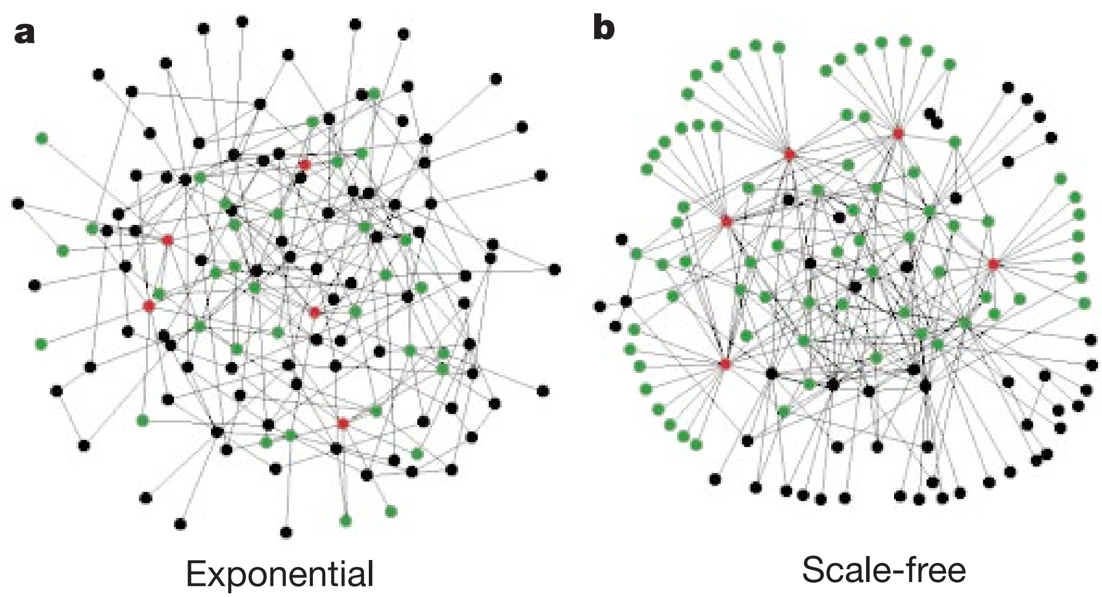
Robustness of a network

- 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.

Robustness of Barabási-Albert random graphs

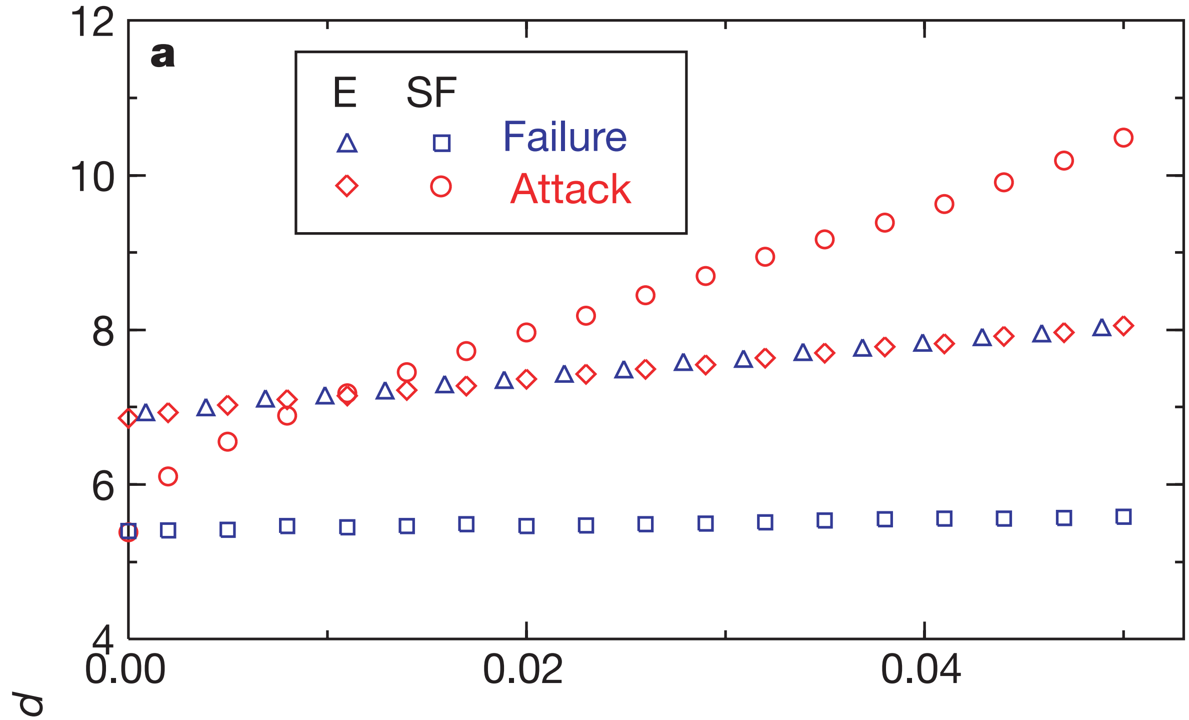
Albert, Jeong and Barabasi, Nature, **406** (27) 2000.



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).

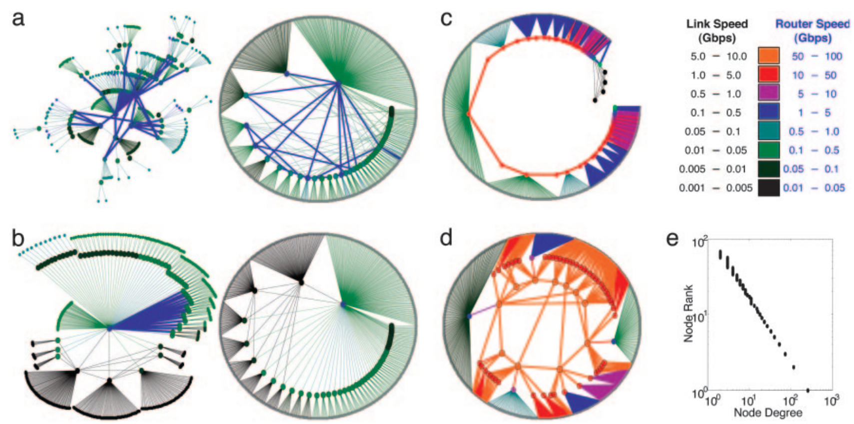
Exponential vs scale-free: Robustness



- (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).

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

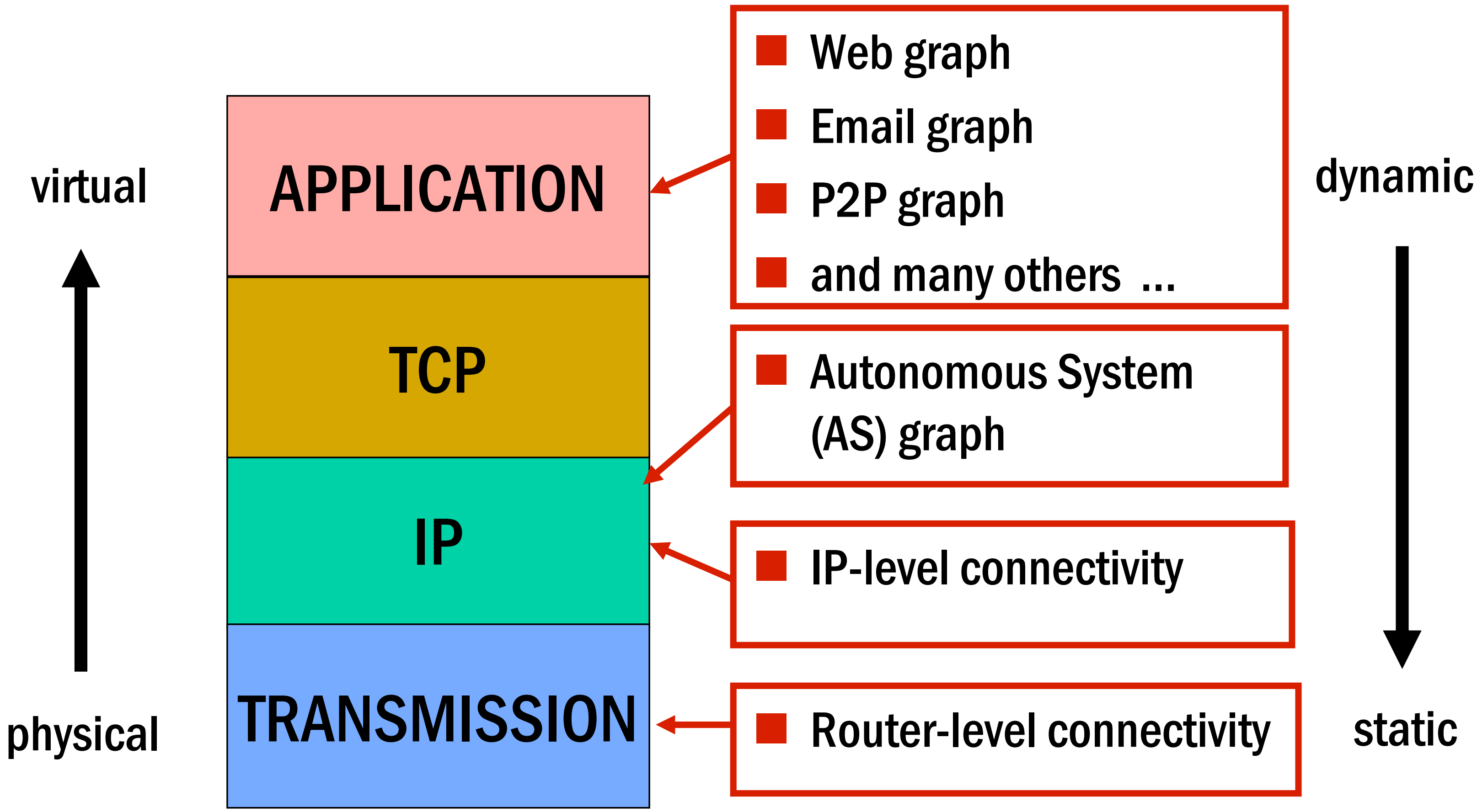
- REDUNDANCY!!! is key principle in engineering.
- The 'robust yet fragile' nature of the Internet
  Doyle, Alderson, Li, Low, Roughan, Shalunov, Tanaka, Willinger, PNAS **102** (4) 2005.



- Degree distribution is not the whole story.
- Also targeted attack by different metrics like betweenness (c.f. Holme P, Kim BJ, Yoon CN, Han SK (2002) "Attack vulnerability of complex networks". *Phys. Rev. E* **65**:056109)

Power Laws in the Internet?
Definition of “node” depends on level of representation

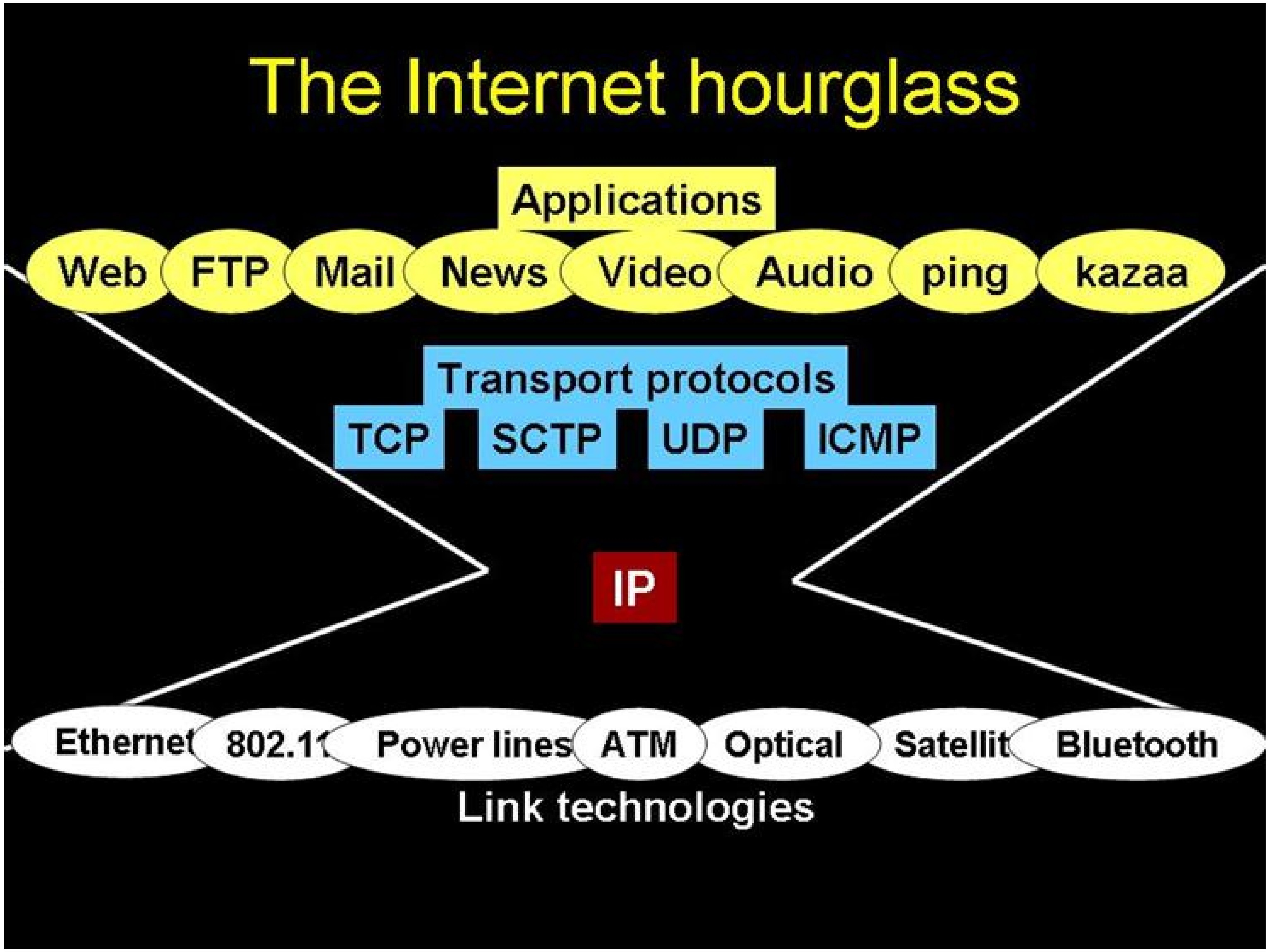
Internet connectivity structures are different at each layer



AS
(BGP routes)

TraceRoute

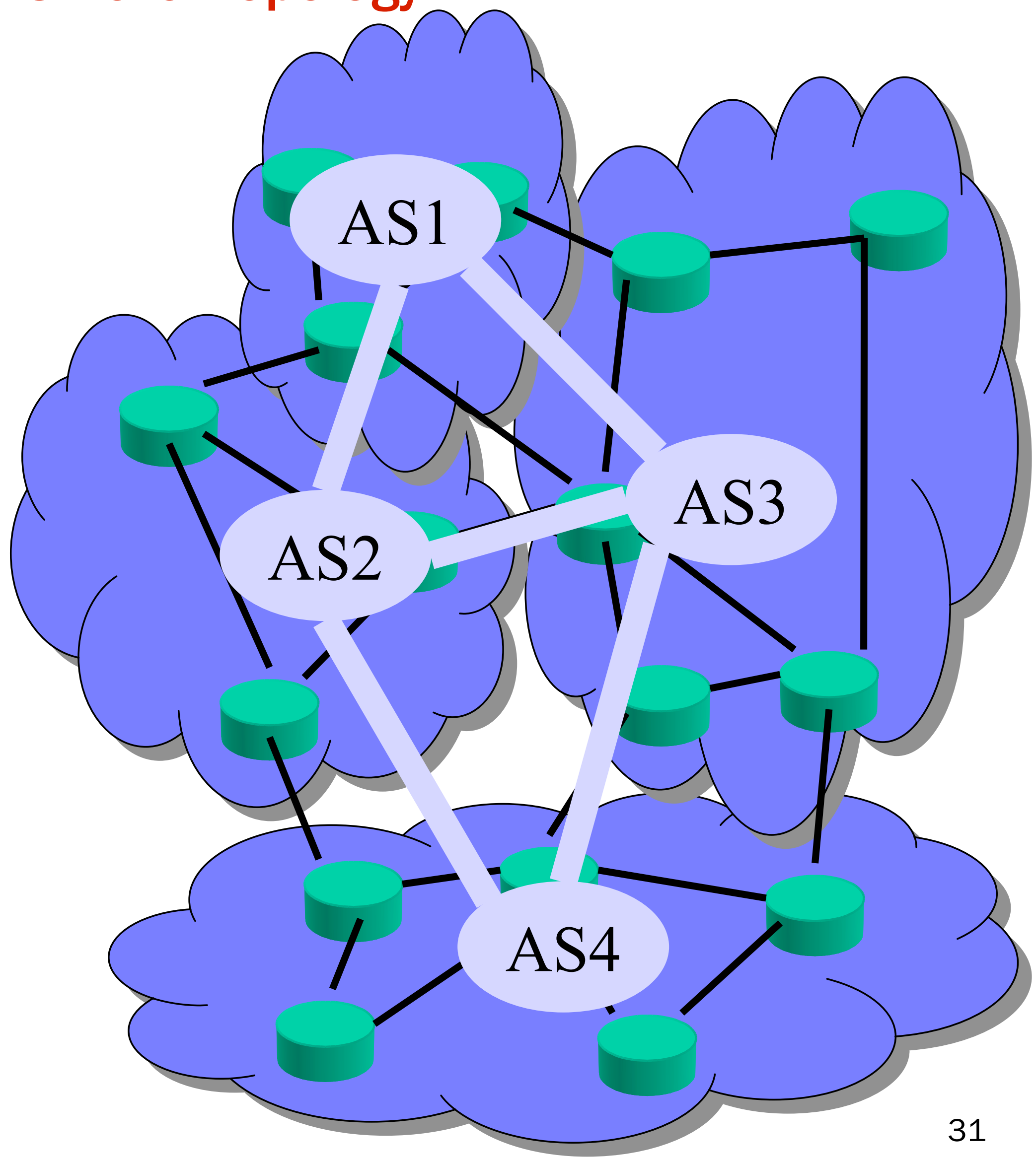
Router level



(picture from David Alderson)

AS-Level Topology

- Nodes = (sets of) entire networks (Autonomous Systems or ASes)
- Links = peering relationships between ASes
- Really a map of economic or business relationships, not of physical connectivity



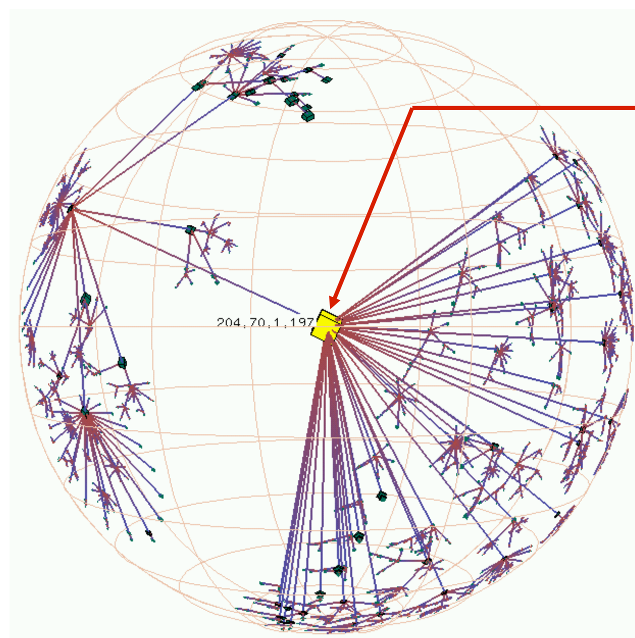
How to measure the structure of the Internet?

CAIDA! (Cooperative Association for Internet Data Analysis, UCSD)

- Traceroute
- BGP tables
- "Whois" data

Known issues:

- Traceroute, s-d sampling bias, makes even ER random graph appear to have power law:
  - Lakhina, Byers, Crovella, Xie *INFOCOM*, 2003.
  - Achlioptas, Clauset, Kempe, Moore *STOC*, 2005.
- Hidden subgraphs:



- www.savvis.net
- managed IP and hosting company
- founded 1995
- offering "private IP with ATM at core"

This "node" is an entire network! (not just a router)

http://www.caida.org/tools/measurement/skitter/

Degree distribution and Network Growth Models

- **Heterogeneity** in real networks.
- Concentrated, **Poisson Distribution in Erdös-Rényi**:
  - Probability to connect to k nodes is p^k.
  - Probability to be disconnected from remaining $(n-k)$ is $(1-p)^{(n-k)}$.
  - Probability for a vertex to have degree k follows a binomial distribution:

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}.$$

- **Seek alternate mechanisms...**
  – Preferential Attachment: $p_{ij} \propto d_j$

  (Probability a new nodes attaches to existing node j is proportional to current degree d_j).

An alternate view, Mandelbrot, 1953: optimization

(Information theory of the statistical structure of language)

- **Goal:** Optimize information conveyed for unit transmission cost
- Consider an alphabet of d characters, with n distinct words
- Order all possible words by length (A,B,C,....AA,BB,CC....)
- “Cost” of j-th word, $C_j \sim \log_d j$
- Ave information per word: $H = -\sum p_j \log p_j$
- Ave cost per word: $C = \sum p_j C_j$
- Minimize: $\frac{d}{dp_j}\left(\frac{C}{H}\right) \implies p_j \sim j^{-\alpha}$

Optimization versus Preferential Attachment origin of power laws

Mandelbrot and Simon's heated public exchange

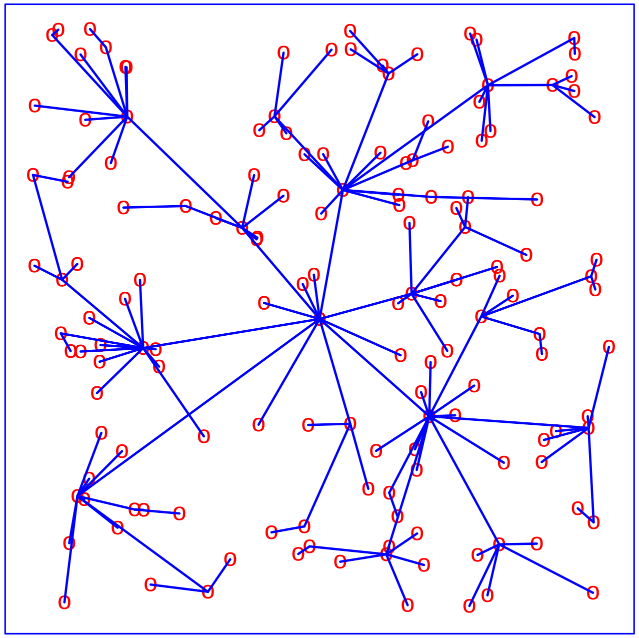
- A series of six letters between 1959-61 in *Information and Control*.

Optimization on hold for many years, but recently resurfaced:

- Calson and Doyle, "HOT" (PRE 1999, PRL 2000, PNAS 2002).
- Fabrikant, Koutsoupias, and Papadimitriou (ICALP 2002).
- Valverde, Ferrer Cancho, and Solé (Europhys. Lett. 2002).

FKP (Fabrikant, Koutsoupias, and Papadimitriou, 2002)

- Nodes arriving sequentially at random in a unit square.
- Upon arrival, each node connects to an already existing node that minimizes "cost": $\alpha d_{ij} + h_j$



Tempered Preferential Attachment

[Berger, Borgs, Chayes, D'Souza, Kleinberg, *ICALP* 2004.]

[Berger, Borgs, Chayes, D'Souza, Kleinberg, *CPC*, 2005.]

[D'Souza, Borgs, Chayes, Berger, Kleinberg, *Proc Natn Acad Sci*, **2007.]**

- **Optimization**
  Like FKP, start with linear tradeoffs, but consider a scale-free metric. (Plus will result in local model.) Gives rise to:

$\longrightarrow$ **PA**

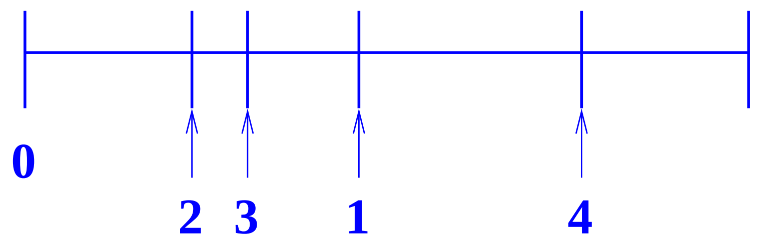
$\longrightarrow$ **Saturation**

$\longrightarrow$ **Viability**

(Not all children have equal fertility, not all spin-offs equally fit, etc).

Competition-Induced Preferential Attachment

Consider points arriving sequentially, uniformly at random along the unit line:



Each incoming node, t, attaches to an existing node j (where $j < t$), which minimizes the function:

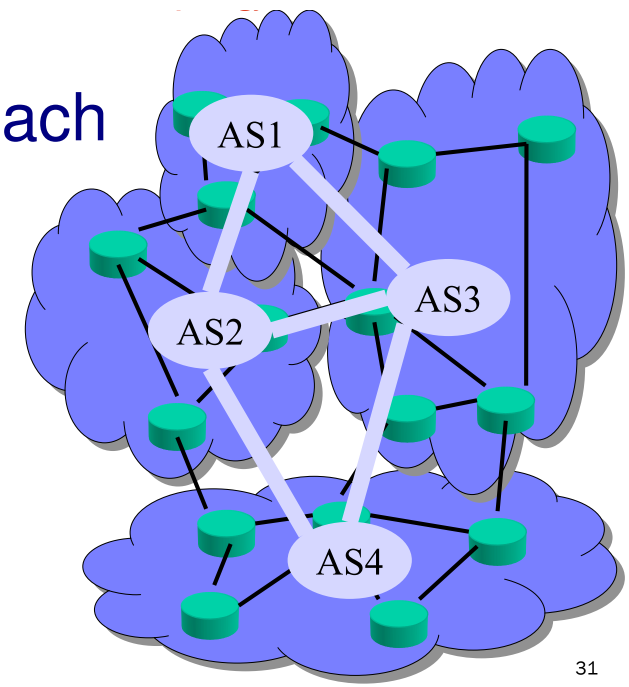
$$F_{tj} = \min_j \left[\alpha_{tj} d_{tj} + h_j\right]$$

Where $\alpha_{tj} = \alpha \rho_{tj} = \alpha n_{tj} / d_{tj}$.

The "cost" becomes: $\boxed{F_{tj} = \min_j \left[\alpha n_{tj} + h_j\right]}$

$$F_{tj} = \min_j \left[\alpha n_{tj} + h_j\right]$$

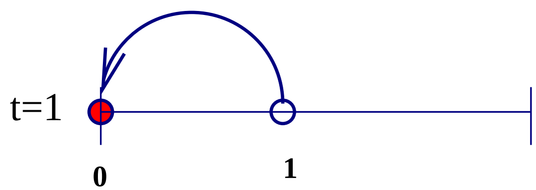
- $\alpha_{tj} = \alpha\rho_{tj}$ geometric cost proportional to local density
- Reduces to n_{tj} — number of points in the interval between t and j
- Minimize "transit domains" required to reach node with strong network centrality (i.e. AS/ISP-transit = BGP and peering).



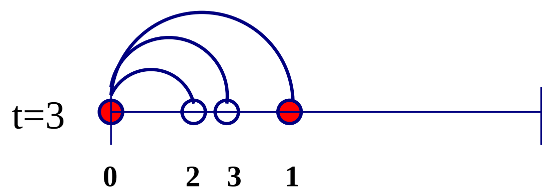
The process on the line (for $1/3 < \alpha < 1/2$)

"Border Toll Optimization Problem" (BTOP)

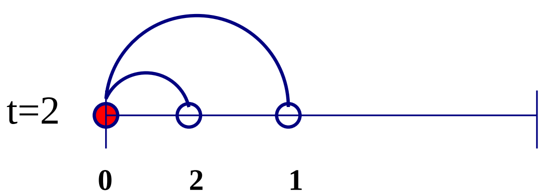
$$F_{tj} = \min_j \left[\alpha n_{tj} + h_j\right]$$



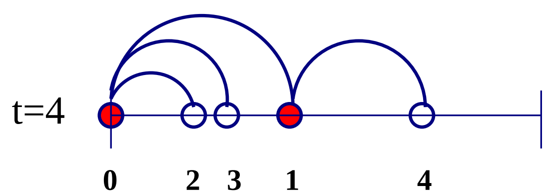
F(10) = 0



F(30) = α
F(32) = 1
F(31) = 1



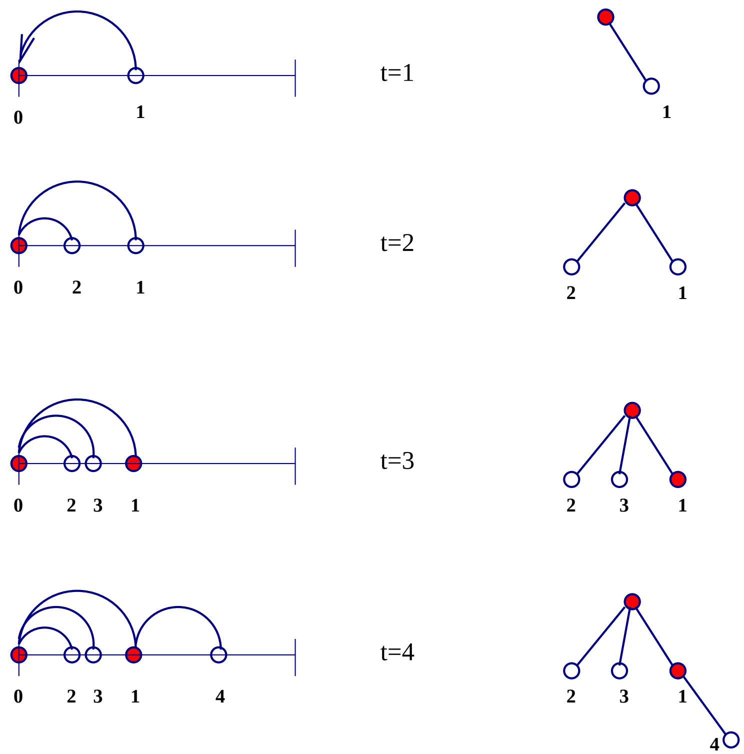
F(20) = 0
F(21) = 1



F(40) = 3 α
F(42) = 1 + 2 α
F(43) = 1 + α
F(41) = 1

(A **local** model – connect either to closest node, or its parent.)

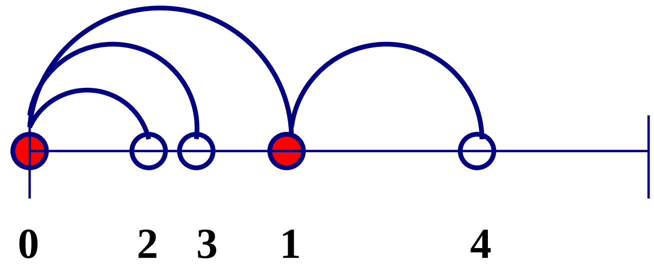
Mapping onto a tree



$$
\begin{aligned}
Pr\left[x_{t+1} \in I_k \left|\pi(t)\right.\right] &= \int Pr\left[x_{t+1} \in I_k \left|\pi(t), \vec{s}(t)\right.\right] dP(\vec{s}(t)) \\
&= \int s_k(t) dP(\vec{s}(t)) = \frac{1}{t+1},
\end{aligned}
$$

i.e., **The probability to land in the k-th interval is uniform over all intervals.**

Preferential attachment with a cutoff



Let $d_j(t)$ equal the degree of fertile node j at time t.

The number of intervals contributing to j's fertility is $\min(d_j(t), A)$.

Probability node $(t+1)$ attaches to node j is:

$$Pr(t+1 \rightarrow j) = \min(d_j(t), A)/(t+1).$$

Standard PA: $Pr(t+1 \rightarrow j) = d_j(t)/(t+1)$.

The process on degree sequence

Let $N_0(t) \equiv$ number of infertile vertices.

Let $N_k(t) \equiv$ number of fertile vertices of degree k (for $1 \leq k < A$).

Let $N_A(t) \equiv$ number of fertile vertices of degree $k \geq A$
(i.e. $N_A(t) = \sum_{k=A}^{\infty} N_k(t)$ "the tail")

Rigorous Proofs for

- Power law for $d < A$, with $1 < \gamma < 3$.

- Exponential decay for $d > A$.

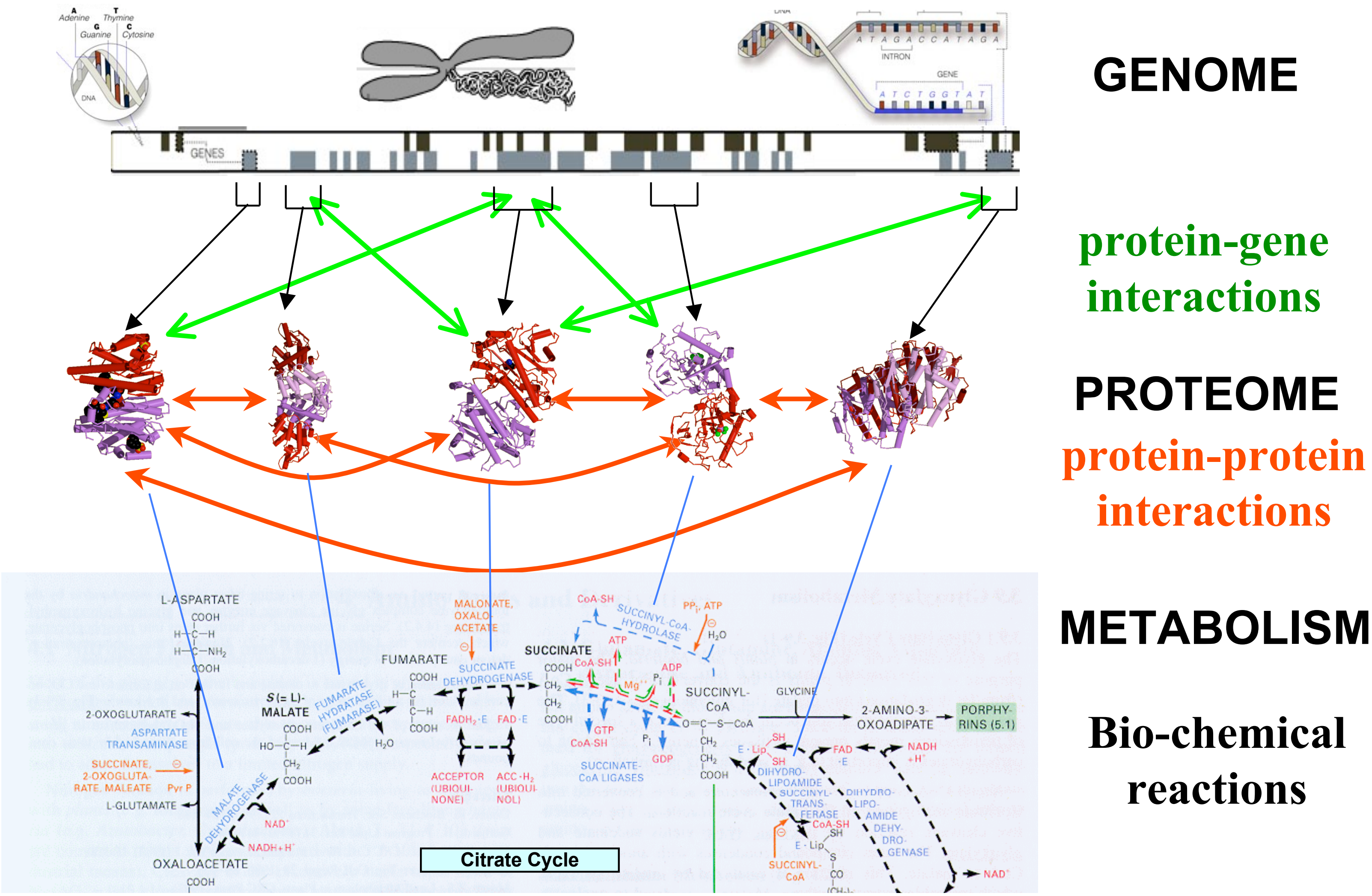
$$p_k = c_1 k^{-\gamma} \quad \text{for} \quad k < A.$$
$$p_k = c_2 \exp[-k/(A+1)] \quad \text{for} \quad k > A.$$

Optimization, Preferential Attachment and Network Growth

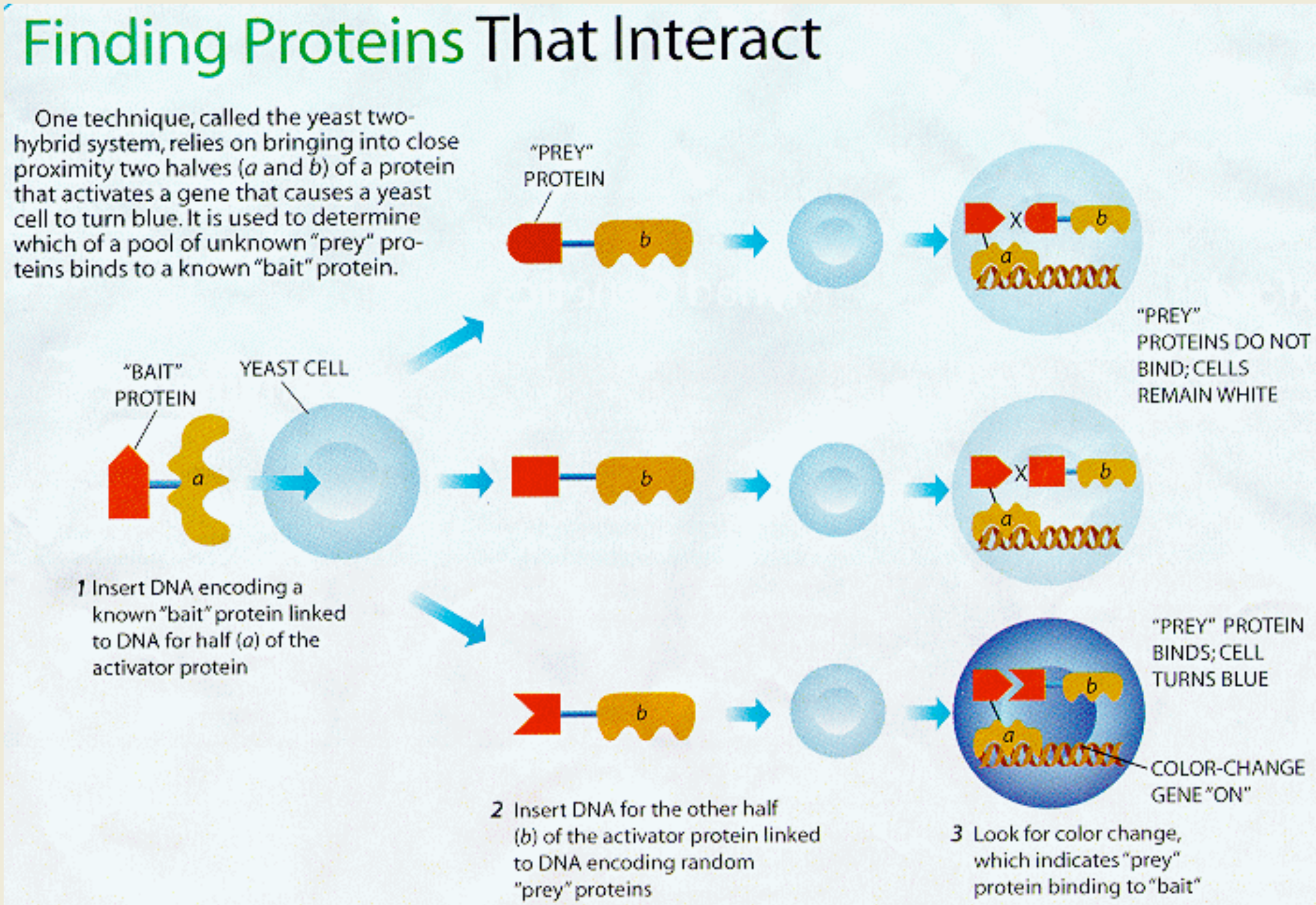
- Optimization can give rise to PA and hence to Power Laws.
- Different cost functions and geometries:
  - Biological choices? (modularity versus efficiency)
  - Open-source software (“systems’ motifs”)
  - Economics/financial trades (trust versus value)
- Gastner and Newman work on road versus airline networks.

  (See MAE 298 Feb 20, 2008 lecture).

Biological networks

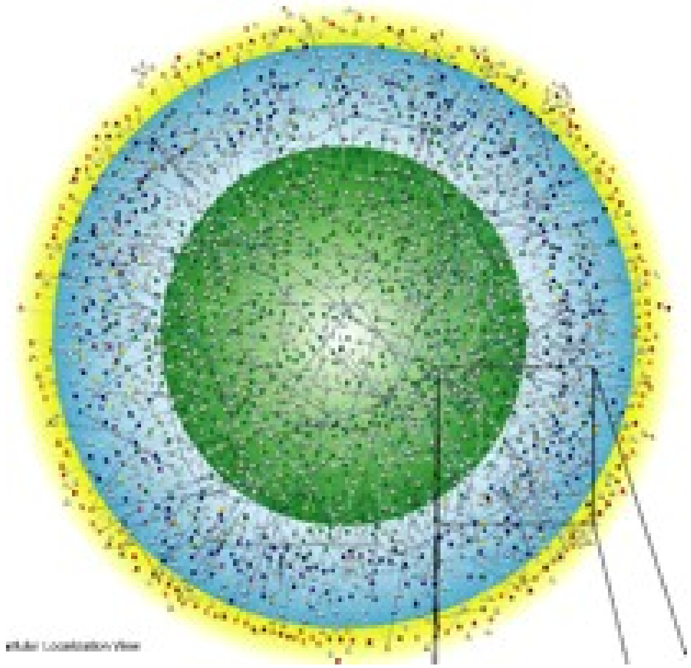


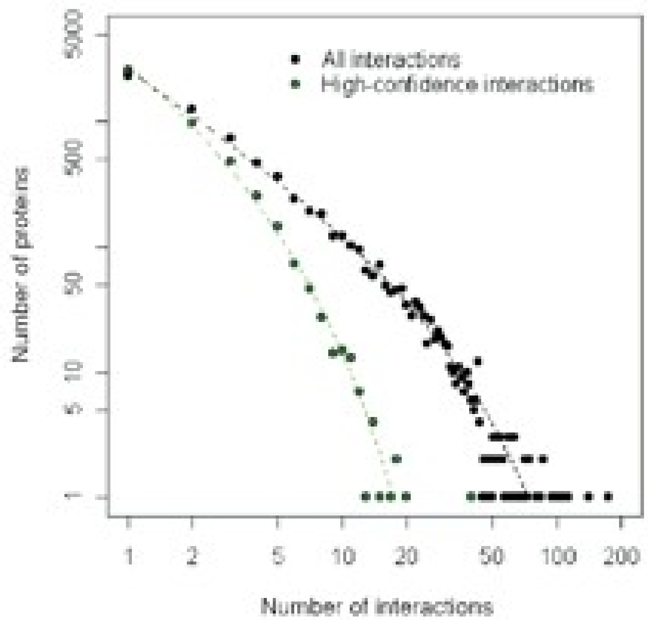
Protein interactions: Yeast two-hybrid method



(Courtesy of Eivind Almaas)

PIN for Drosophila
Giot, et al, Science 2003





GRNs (Courtesy of Julin Maloof)

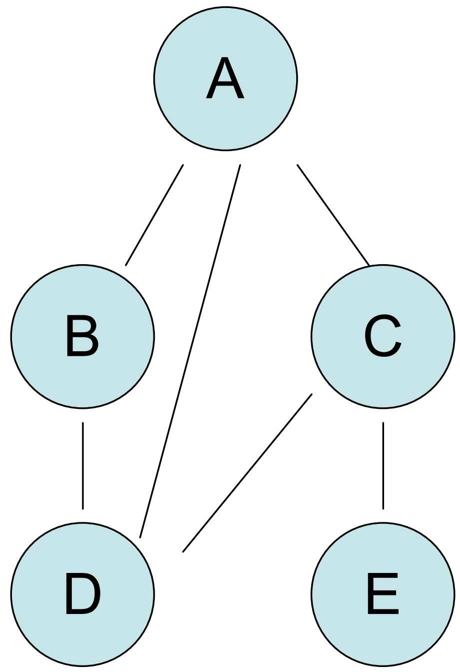
How can microarrays help us build GRNs?

- Co-expression or Relevance Network
  - measure gene expression across multiple samples
    - after perturbation
    - time course
    - different individuals
    - mutants
  - Create correlation matrix
  - Edges connect genes with correlation > threshold

Good Review: Markowetz, F. & Spang, R. Inferring cellular networks--a review. *BMC Bioinformatics* 8 Suppl 6, S5 (2007).

co-expression network

	A	B	C	D	E
A					
B	**.9**				
C	**.8**	.2			
D	**.7**	**.7**	**.6**		
E	.3	.3	**.7**	.1	



(Courtesy of Julin Maloof)