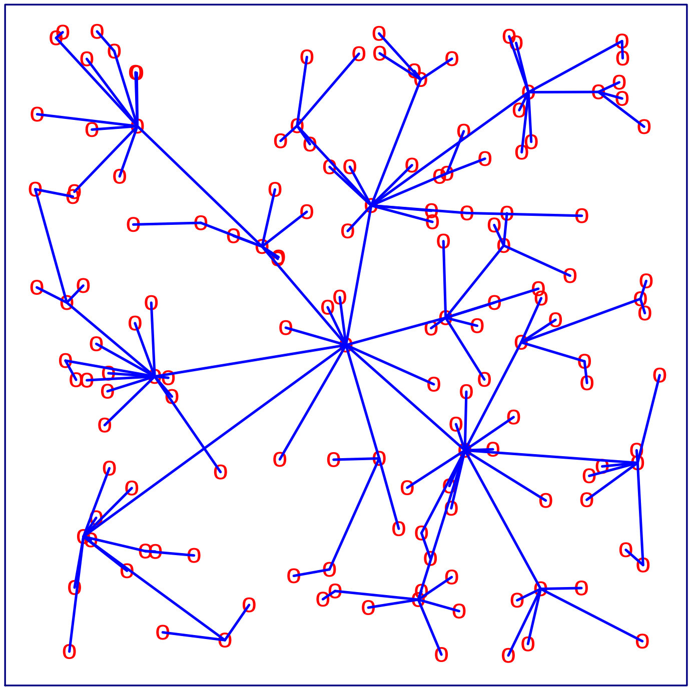
MAE 298, Lecture 5

April 13, 2006



"Optimization and network growth"

Recall: Preferential Attachment process: Origins of preferential attachment

- 1923 — Polya, urn models.
- 1925 — Yule, explain genetic diversity.
- 1949 — Zipf, distribution of city sizes (*1/f*).
- 1955 — Simon, distribution of wealth in economies. ("The rich get richer").
- [Interesting note, in sociology this is referred to as the *Matthew effect* after the biblical edict, "For to every one that hath shall be given ... " (Matthew 25:29)]

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, 1999
- Fabrikant, Koutsoupias, and Papadimitriou, 2002
- Solé, 2002

From Barabási and Albert to FKP

- Barabási and Albert, "Emergence of Scaling in Random Networks", *Science* **286**, 1999.
  A "preferential attachment" model of network growth.

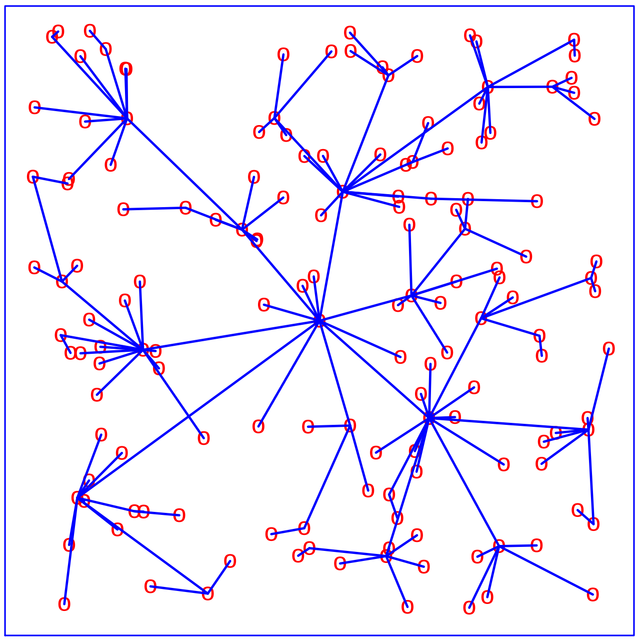
- A. Fabrikant, E. Koutsoupias, and C. H.Papadimitriou, "Heuristically Optimized Trade-offs" *Lecture Notes In Computer Science (ICALP 2002)* **2380**, 2002.
  FKP extend the ideas of Carlson and Doyle to network context:

- J.M. Carlson and J. Doyle, "Highly optimized tolerance: A mechanism for power laws in designed systems", *Physical Review E*, 1999.
  (See also: J.M. Carlson and J. Doyle, "Complexity and Robustness" *PNAS* 2002.

- For a recent press account of optimization versus power laws see: Sara Robinson, "Recent Research Provides New Picture of Router-Level Internet," *Computing in Science & Engineering*, **8** (2) March/April 2006, pp. 3-6.

FKP (Fabrikant, Koutsoupias, and Papadimitriou, 2002)

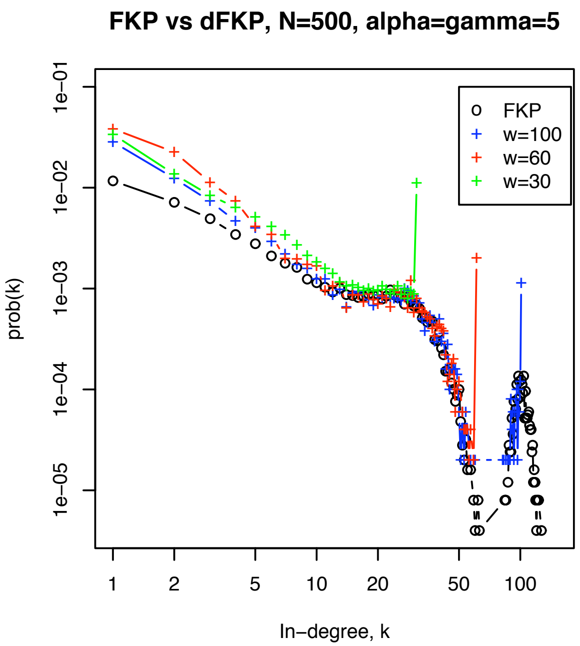
- Nodes arriving sequentially at random in a unit square.
- Upon arrival, each node connects to an already existing node that minimizes "cost": $\alpha d_{ij} + h_j$

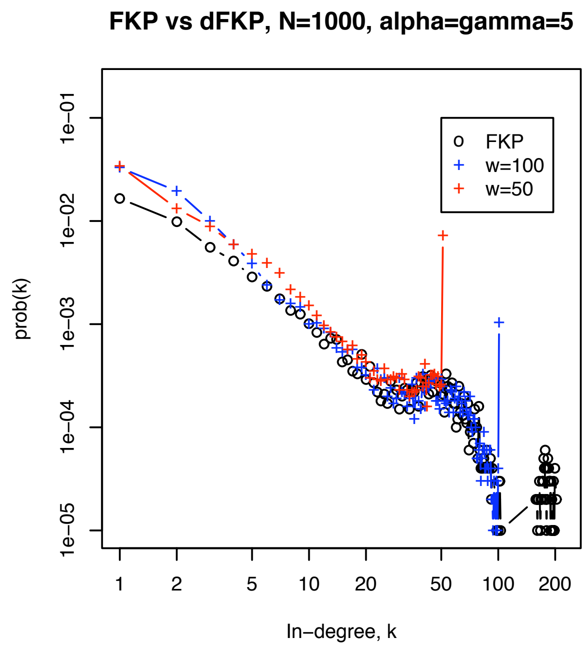


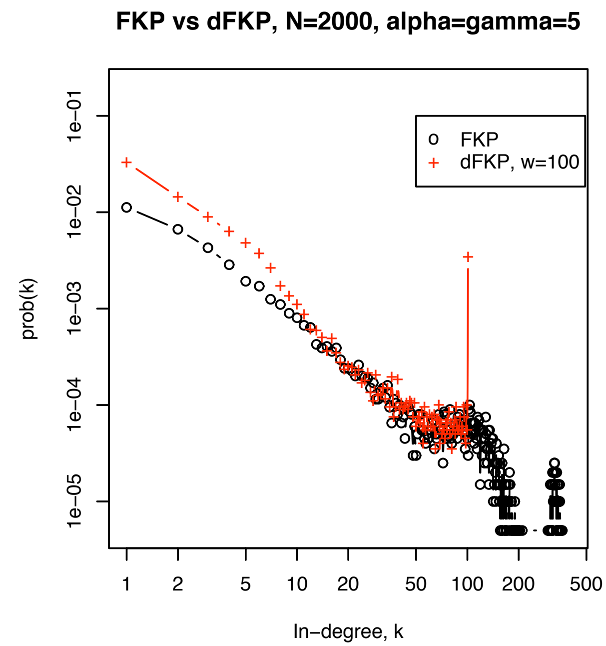
Using R to explore the FKP model

- $\alpha = 0$ limit
- $\alpha \to \infty$ limit
- in-between

FKP degree distribution







A bimodal distribution (hubs and leaves, but almost all nodes are leaves). For details see:

N. Berger and B. Bollobas and C. Borgs and J. Chayes and O. Riordan, “Degree distribution of the FKP network model”, *Lecture Notes In Computer Science (ICALP 2003)*, 2003.

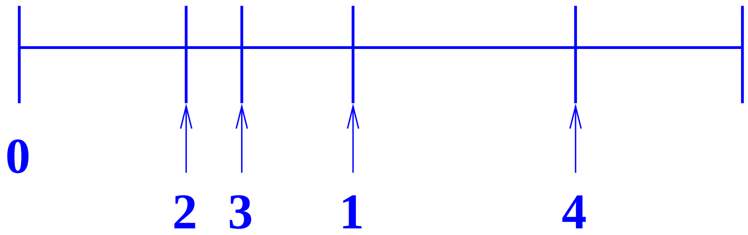
Competition-Induced Preferential Attachment

[N. Berger, C. Borgs, J. Chayes, R.D., R. Kleinberg, ICALP 2004]

- Like FKP, start with linear tradeoffs, but consider a scale-free metric. (Plus will result in local model.)
- Show how mechanism of PA arises, but with eventual saturation.
- PA w/ Saturation in turn gives with to Power Laws with eventual Exponential Decay.
- Such distributions observed ubiquitously in nature.
- Saturation, like PA, has previously been used as axiom to explain data.

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

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

- $\alpha_{tj} = \alpha \rho_{tj}$ local density, e.g. real estate in Manhattan.

- Reduces to n_{tj} — number of points in the interval between t and j

- "Transit domains" — captures realistic aspects of Internet costs (i.e. AS/ISP-transit requires BGP and peering).

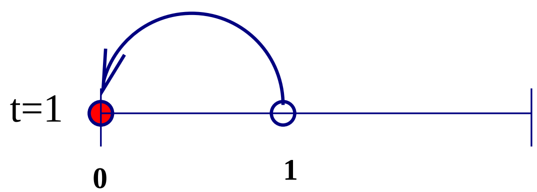
- Like FKP, tradeoff intial connection cost versus usage cost.

- Note cases $\alpha = 0$ and $\alpha > 1$.

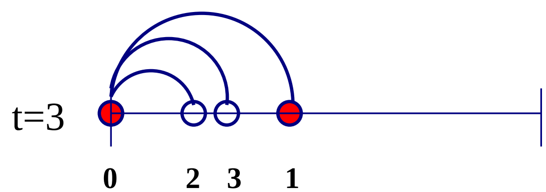
The process on the line (for $1/3 < \alpha < 1/2$)

"Border Toll Optimization Problem" (BTOP)

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



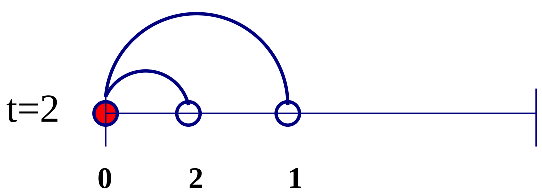
F(10) = 0



F(30) = α

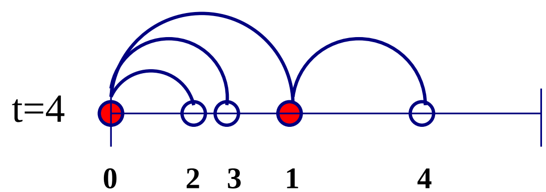
F(32) = 1

F(31) = 1



F(20) = 0

F(21) = 1



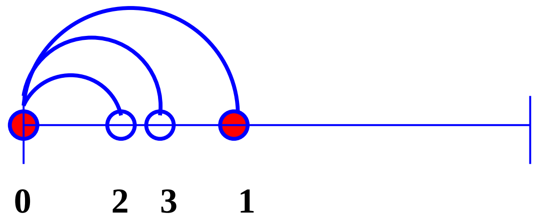
F(40) = 3 α

F(42) = 1 + 2 α

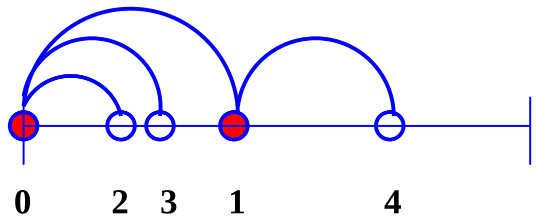
F(43) = 1 + α

F(41) = 1

“Fertility”



t=3



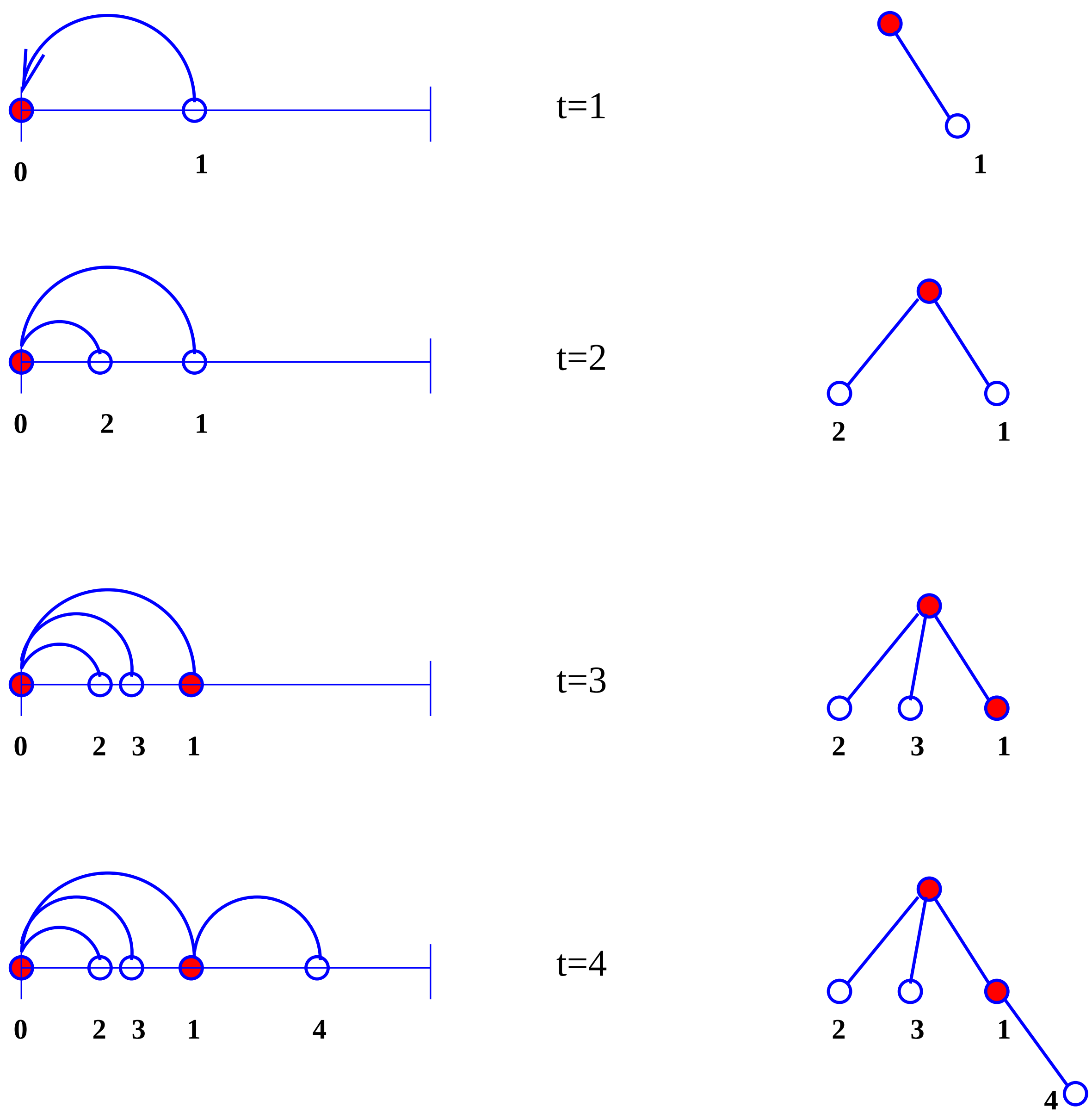
t=4

Node 1 becomes “fertile” at time $t = 3$.

- Define $A = \lceil 1/\alpha \rceil$
- A node must have $A - 1$ “infertile” children before giving birth to a “fertile” child.

Mapping onto a tree

(equal in distribution to the line)



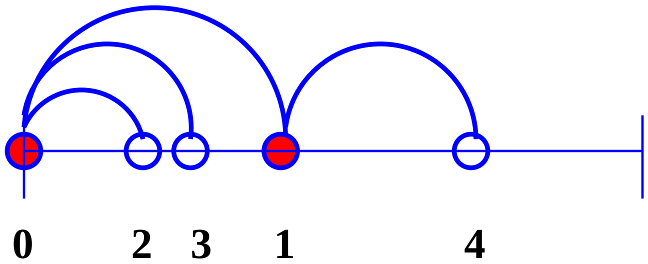
From line to tree

Integrating out the dependence on interval length from the conditional probability:

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

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 $\max(d_j(t), A)$.

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

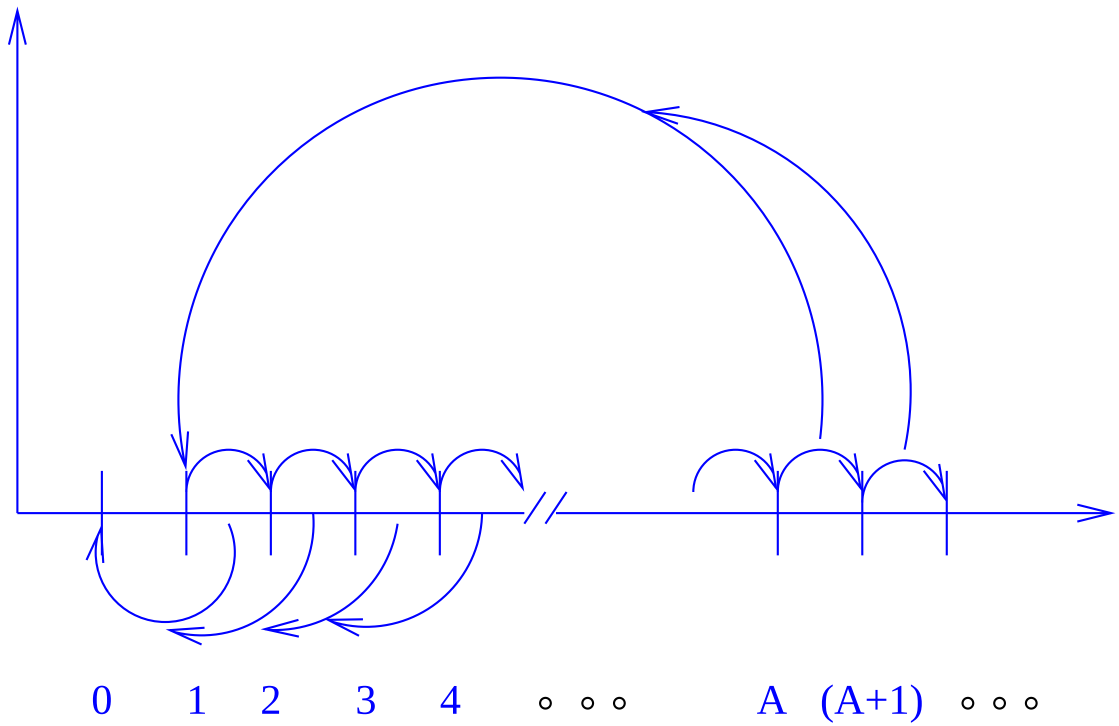
$$Pr(t+1 \rightarrow j) = \max(d_j(t), A)/(t+1).$$

The process on degree sequence
(Can go through a similar heuristic derivation as with PA)

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")



0
1
2
3
4
A
(A+1)

Recursion relation

$$p_k = (k-1)p_{k-1}(t) - kp_k(t), \qquad 1 < k < A.$$

and

$$\left(p_{k-1} - p_k\right), \qquad k \geq A.$$

Implies:

$$\boxed{p_k = \prod_{i=2}^{k} \left(\frac{i-1}{i+1}\right) p_1, \qquad 1 < k < A.}$$

and

$$\boxed{p_k = \left(\frac{A}{A+1}\right)^{k-A} p_A, \qquad k \geq A.}$$

Power law for $1 < k < A$

$$\frac{p_k}{p_1} = \prod_{i=2}^{k} \left(\frac{i-1}{i+1} \right) = \frac{2}{k(k+1)}$$

$$\sim c\, k^{-2}$$

Exponential decay for $k > A$

Recursion relation: $p_k = A\left(p_{k-1} - p_k\right), \qquad k \geq A.$

Implies

$$p_k = \left(\frac{A}{A+1}\right)^{k-A} p_A, \qquad k \geq A.$$

$$p_k = \left(1 - \frac{1}{A+1}\right)^{k-A} p_A = \left[\left(1 - \frac{1}{A+1}\right)^{A+1}\right]^{(k-A)/(A+1)} p_A$$

$$\sim \exp\left[-(k-A)/(A+1)\right] p_A.$$

Degree sequence (summary)

$$p_k = c_1 k^{-\gamma} \quad \text{for} \quad k < A$$

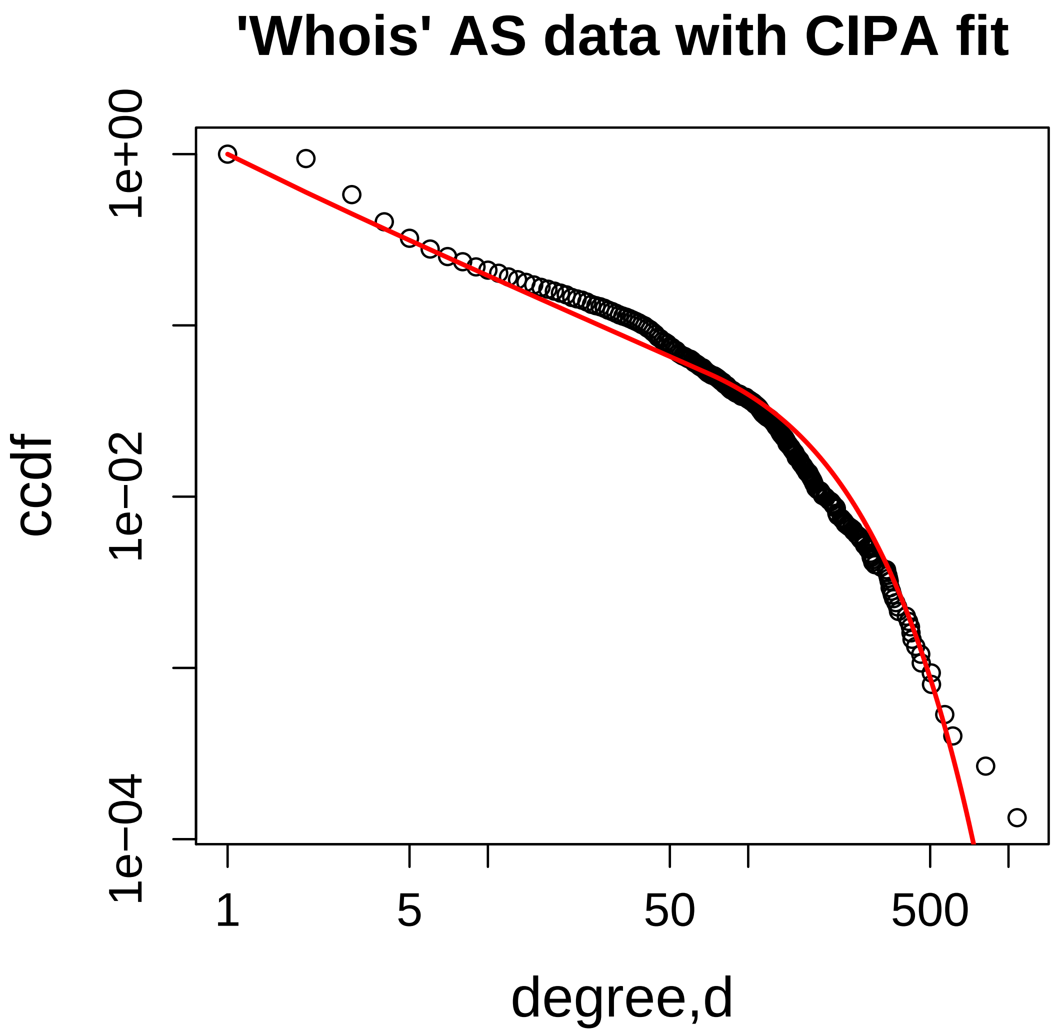
$$p_k = c_2 \exp[-k/(A+1)] \quad \text{for} \quad k > A.$$

"Power law" $\rightarrow$ power law with exponential tail

Ubiquitous empirical measurements:
(Saturation and PA often put in apriori to explain)

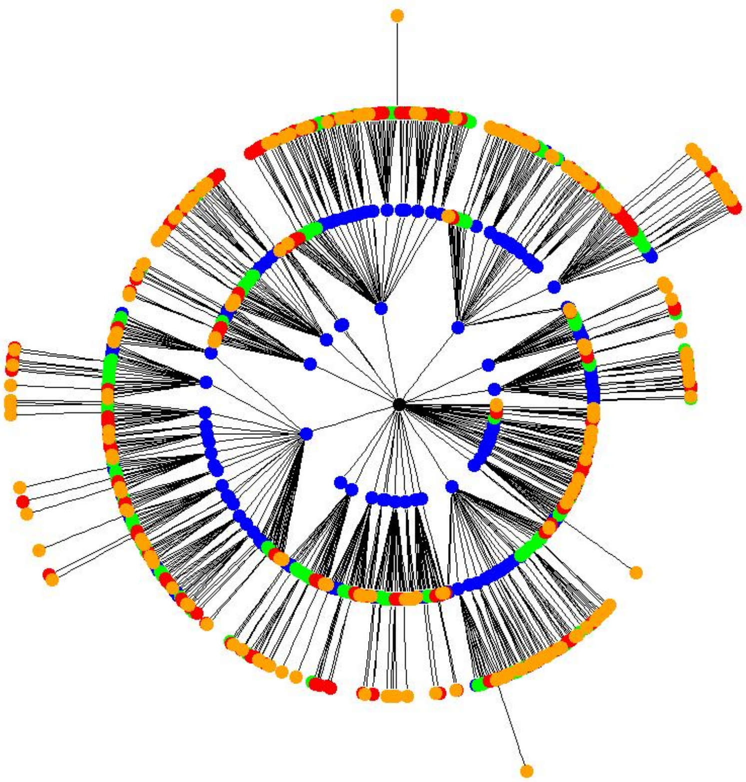
System with: $p(x) \sim x^{-B} \exp(-x/C)$	B	C
Full protein-interaction map of *Drosophila*	1.20	0.038
High-confidence protein-interaction map of *Drosophila*	1.26	0.27
Gene-flow/hydridization network of plants as function of spatial distance	0.75	10^5 m
Earthquake magnitude	1.35 - 1.7	$\sim 10^{21}$ Nm
Avalanche size of ferromagnetic materials	1.2 - 1.4	$L^{1.4}$
ArXiv co-author network	1.3	53
MEDLINE co-author network	2.1	~ 5800
PNAS paper citation network	0.49	4.21
WHOIS AS Internet data	0.59	178

Fitting the “WHOIS” AS level Internet data

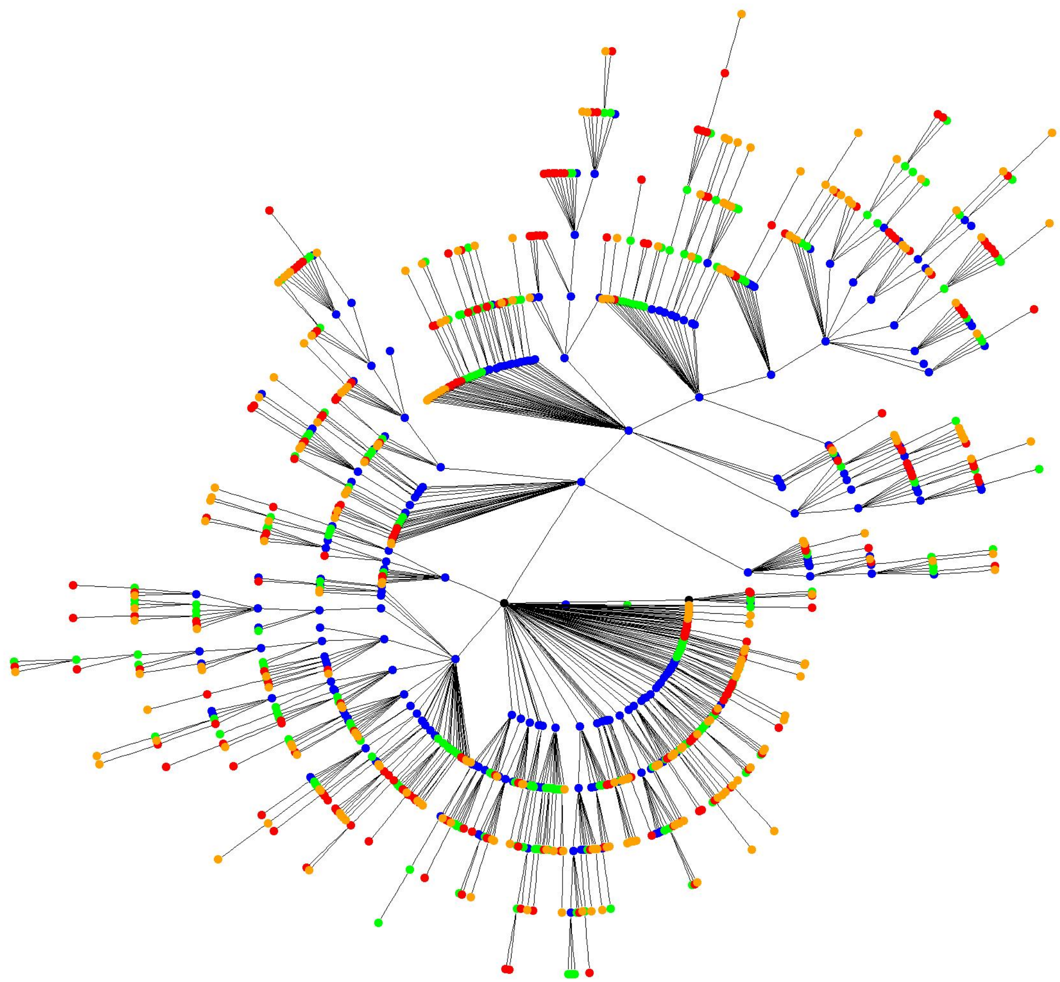


Definitely not a power law, so previously no model to explain distribution.

Comparing CIPA and PA graphs



CIPA



PA

Network growth models

- Preferential attachment.
- Heuristically Optimized Tolerance
- Competition-Induced Preferential Attachment
- Validation! — (but what does this really mean, can only validate aspects).

Further reading

- A. Fabrikant, E. Koutsoupias, and C. H.Papadimitriou, "Heuristically Optimized Trade-offs" *Lecture Notes In Computer Science (ICALP 2002)* **2380**, 2002.
  FKP extend the ideas of Carlson and Doyle to network context:
- J.M. Carlson and J. Doyle, "Highly optimized tolerance: A mechanism for power laws in designed systems", *Physical Review E*, 1999.
  (See also: J.M. Carlson and J. Doyle, "Complexity and Robustness" *PNAS* 2002.
- N. Berger, C. Borgs, J. Chayes, R. D'Souza, R. Kleinberg, "Competition-Induced Preferential Attachment", *Lecture Notes in Computer Science (ICALP 2004)* **3142** 208-221, 2004.