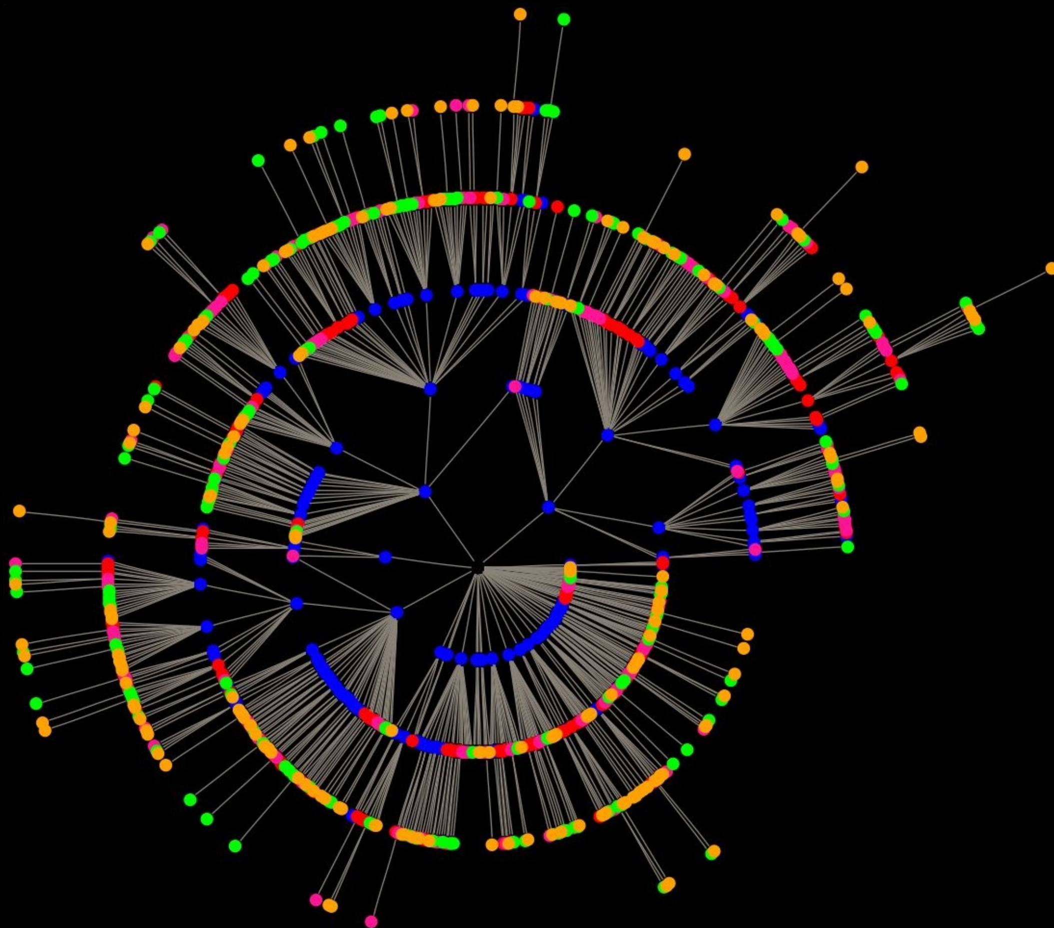
“Preferential Attachment from Optimization”



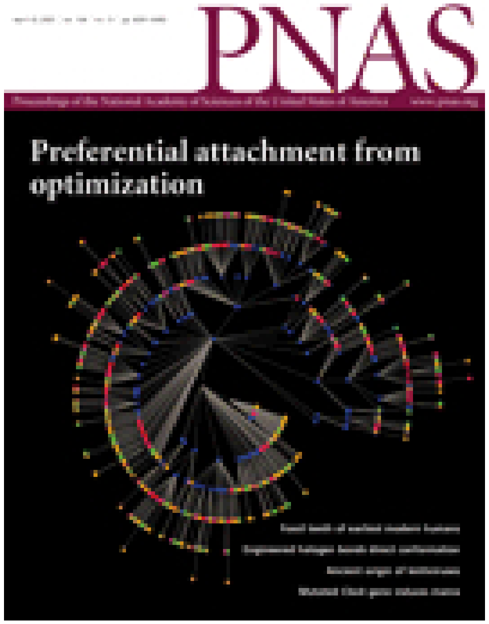
Raissa D’Souza **UC Davis**

Dept of Mechanical and Aeronautical Eng.

Center for Computational Science and Eng.

Santa Fe Institute

Part I: “Emergence of Tempered Preferential Attachment from Optimization”



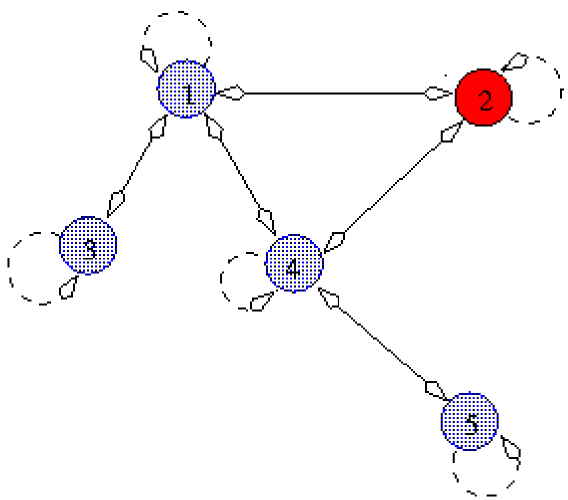
(April 10, 2007)

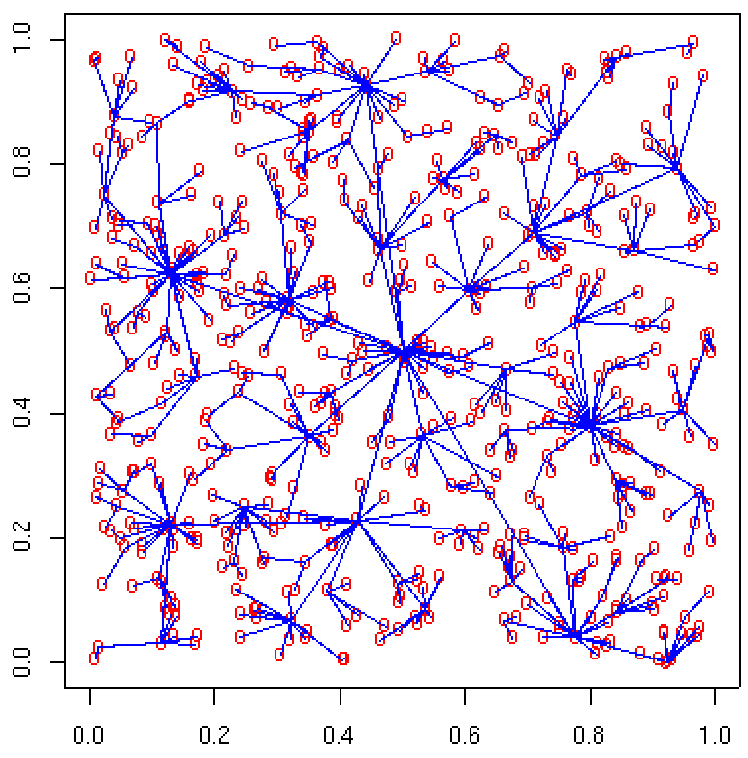
Jennifer Chayes *(MSR)*   Christian Borgs *(MSR)*

Noam Berger *(UCLA)*   Bobby Kleinberg *(Cornell)*

Part II:

Separation of timescales:
When can we consider dynamics on dynamic graphs?





"Power Laws" in data?

- in the WWW .... sure.
- in a social network ... possible.
- in earthquake magnitude ... yes, but to some cutoff.
- in the Internet???

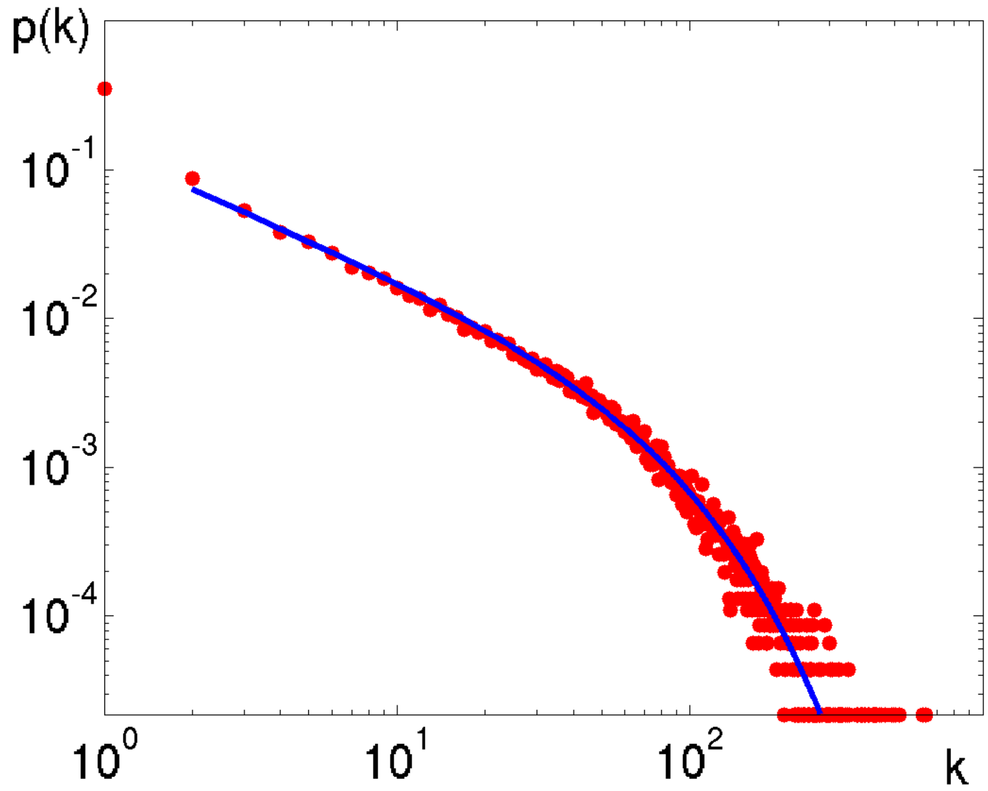
Why power laws cannot continue: Finite size effects, resource limitations, physical geometric (Internet) vs virtual geometry-free (WWW)....

Preferential Attachment?

- in the WWW .... sure.
- in a social network ... sure.
- in earthquake magnitude ... obviously not.
- in the Internet
  - Router level ... NO!
  - AS level ... probably not.

The "Who-is-Who" network in Budapest

(Analysis by Balázs Szendröi and Gábor Csányi)



Bayesian curve fitting $\longrightarrow p(k) = ck^{-\gamma}e^{-\alpha k}$

“Power law” $\to$ power law with exponential tail

Ubiquitous empirical measurements:

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

(Saturation and PA often put in apriori to explain)

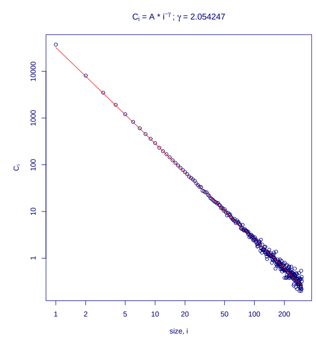
Power laws are observed

Social systems:

- Popularity of web pages: $N_k \sim k^{-1}$
- Rank of city sizes ("Zipf's Law"): $N_k \sim k^{-1}$

Connectivity of random graphs at critical point:

- Component sizes: $N_k \sim k^{-5/2}$



The AS-level Internet?

- Internet Autonomous Systems (AS), like ISP's; $N_k \sim k^{-2.2}$

Known Mechanisms for Power Laws

- Phase transitions (singularities)
- Random multiplicative processes (fragmentation)
- Combination of exponentials (e.g. word frequencies)
- **Preferential attachment / Proportional attachment**
  (Polya 1923, Yule 1925, Zipf 1949, Simon 1955, Price 1976, Barabási and Albert 1999)

  Attractiveness is proportional to size:

  $$\boxed{\frac{dP(s)}{dt} \propto s}$$

- Add in **saturation,** get PA with exponential decay.
  (Amaral, Scala, Barthélémy, Stanley *PNAS* 2000), (Börner, Maru, Goldstone *PNAS* 2004)

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 cost per word: $C = \sum p_j C_j$, where p_j is prob of jth word.
- Ave information per word: $H = -\sum p_j \log p_j$
- Minimize: $\frac{d}{dp_j}\left(\frac{C}{H}\right) \implies p_j \sim j^{-\alpha}$

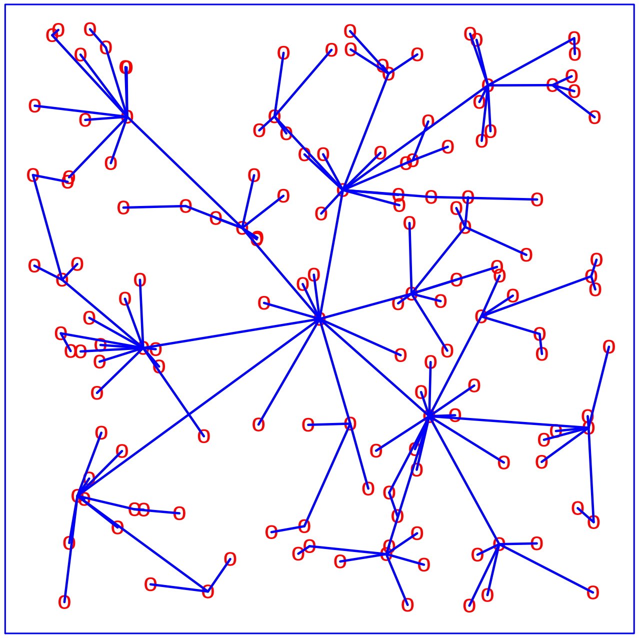
Optimization versus Preferential Attachment origin of power laws

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

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": $\alpha d_{ij} + h_j$



- "Bimodal": initial nodes hubs, remainder leaves. (CDF hides, PDF very clear).

Tempered Preferential Attachment

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

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

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

$\longrightarrow$ **PA**

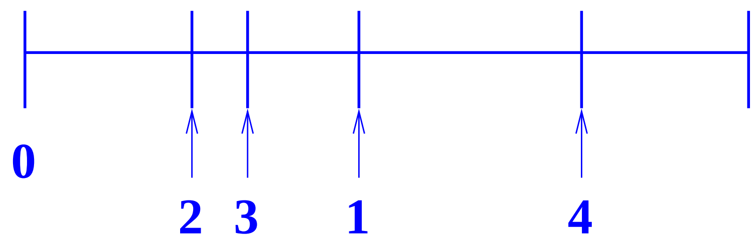
$\longrightarrow$ **Saturation**

$\longrightarrow$ **Viability**

(HETEROGENIETY: 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 $\quad \alpha_{tj} = \alpha \rho_{tj} = \alpha n_{tj} / d_{tj}.$

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

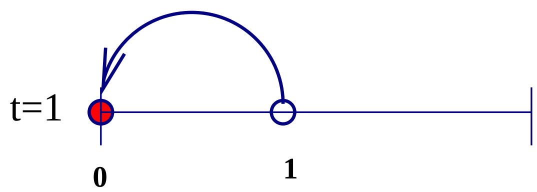
$$\boxed{F_{tj} = \min_j \left[\alpha n_{tj} + h_j\right]}$$

- $\alpha_{tj} = \alpha \rho_{tj}$ 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 $\alpha > 1$.

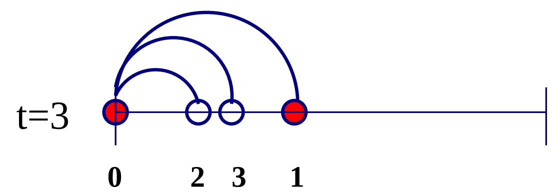
The process on the line (for $1/3 < \alpha < 1/2$)

"Border Toll Optimization Problem" (BTOP)

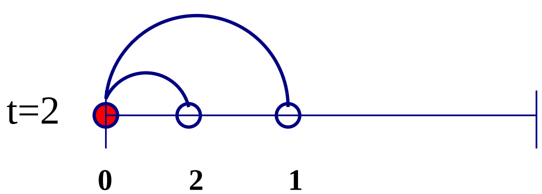
$$\boxed{F_{tj} = \min_j \left[\alpha n_{tj} + h_j\right]}$$



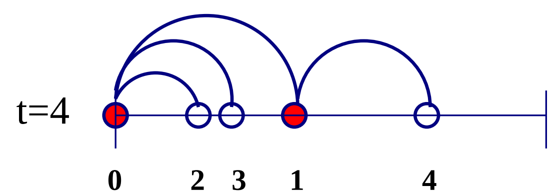
F(10) = 0



F(30) = α
F(32) = 1
F(31) = 1



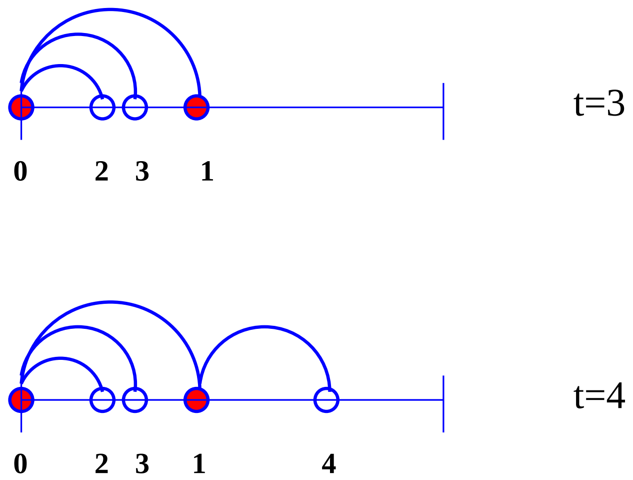
F(20) = 0
F(21) = 1



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.)

“Fertility”/Viability

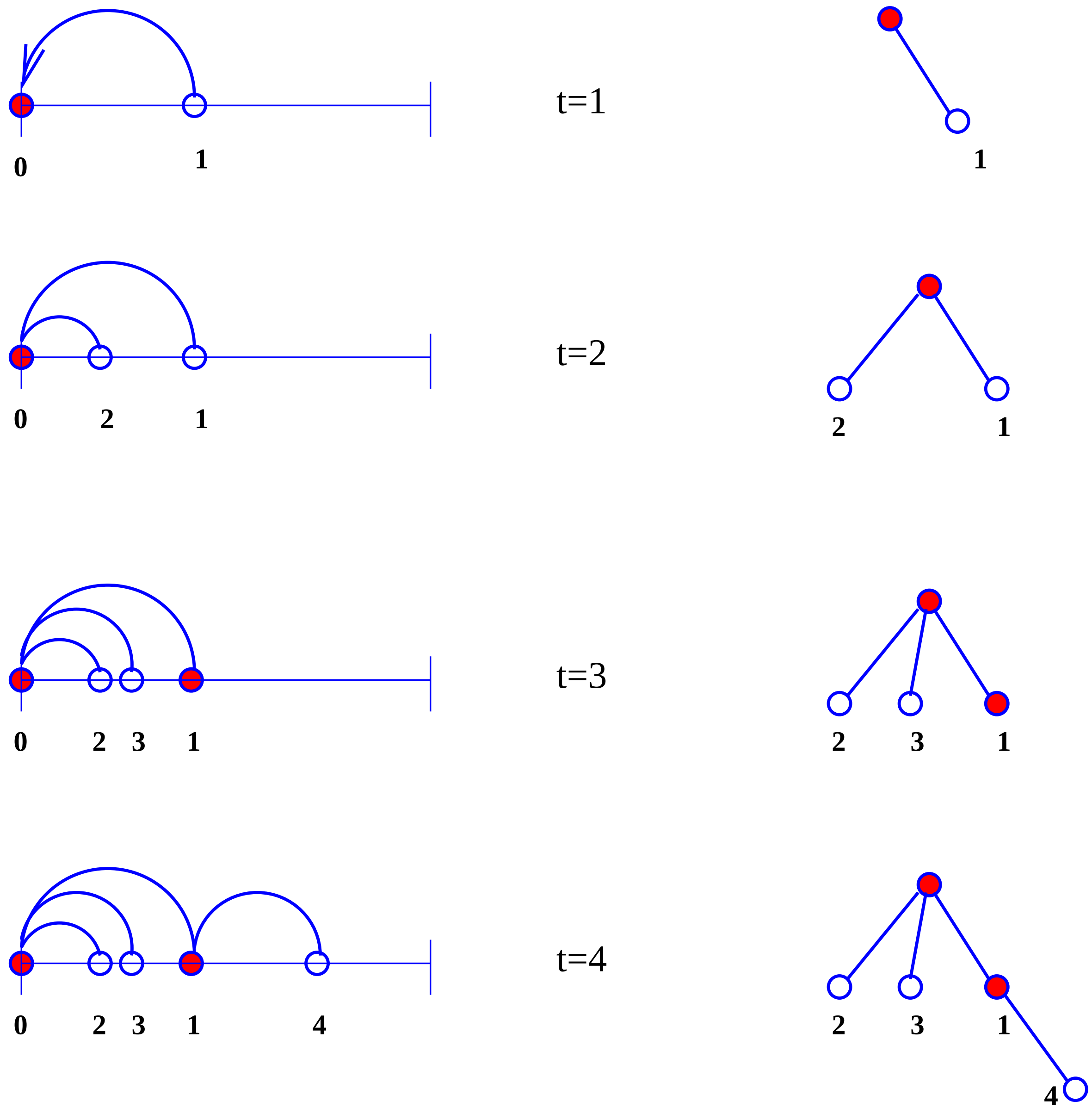


Node 1 becomes “fertile” at time $t = 3$.

- 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)



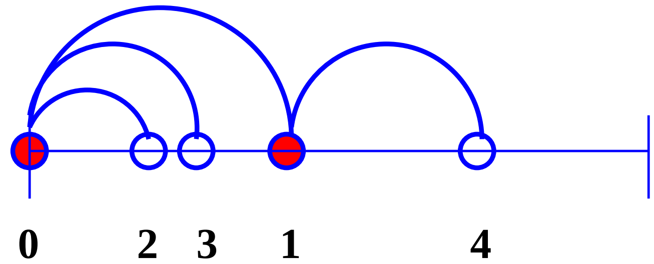
From line to tree

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

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

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

Preferential attachment with a cutoff



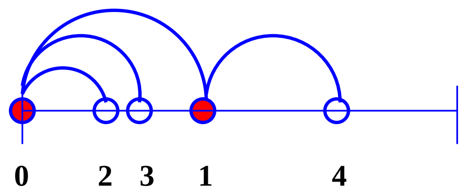
Let $d_j(t)$ equal the degree of fertile node j at time t.

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

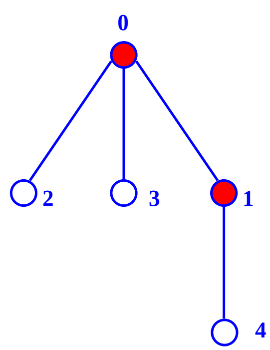
Probability node $(t+1)$ attaches to node j is:

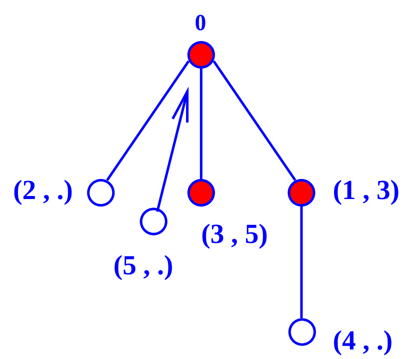
$$\boxed{Pr(t+1 \rightarrow j) = \max(d_j(t), A)/(t+1).}$$

The process on the tree



Equiv line:





Indistinguishable nodes with aging.

The process on degree sequence (The master equation)

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")

The process on degree sequence, cont.

Their expectations follow:

$$
\begin{aligned}
n_1(t+1) &= n_1(t) + \frac{A}{t} n_A(t) - \frac{1}{t} n_1(t) \\
n_k(t+1) &= n_k(t) + \frac{k-1}{t} n_{k-1}(t) - \frac{k}{t} n_k(t), \qquad 1 < k < A \\
n_A(t+1) &= n_A(t) + \frac{A-1}{t} n_{A-1}(t).
\end{aligned}
$$

Let: $p_k(t) = n_k(t)/t$.

In terms of $p_k(t)$:

$$
\begin{aligned}
p_1(t+1)(t+1) - p_1(t)(t) &= Ap_A(t) - p_1(t) \\
p_k(t+1)(t+1) - p_k(t)(t) &= (k-1)p_{k-1}(t) - kp_k(t), \quad 1 < k < A \\
p_A(t+1)(t+1) - p_A(t)(t) &= (A-1)p_{A-1}(t).
\end{aligned}
$$

Proposition 1 (Convergence of expectations to stationary distribution): $\quad p_k(t) \to p_k$.

$$
\begin{aligned}
p_1 &= Ap_A - p_1 \\
p_k &= (k-1)p_{k-1} - kp_k, \qquad 1 < k < A \\
p_A &= (A-1)p_{A-1}.
\end{aligned}
$$

Proposition (2): (Concentration) (i.e., How big are the fluctuations about $n_k(t)$?)

Recursion relation

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

Implies

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

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

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

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

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

Generalizing: from A to (A_1, A_2)

Let the “viability threshold” A_1 differ from the “attractiveness saturation” A_2.

Consider the Markov matrix describing the evolution of the degree sequence (A1 not equal A2)

Limits: $A_1 = 1, A_2 = \infty$, Preferential Attachment

$A_1 = 1, A_2$ finite, PA with a cutoff

$A_1 = A_2 = 1$, uniform.

Degree sequence (summary)

$$
\begin{aligned}
p_k &= c_1 k^{-\gamma} \quad \text{for} \quad k < A_2 \\
p_k &= c_2 \exp[-k/(A+1)] \quad \text{for} \quad k > A_2.
\end{aligned}
$$

- Power law for $d < A_2$, with $\gamma(A_1, A_2)$.
- Exponential decay for $d > A_2$.

MONOTONICITY

We can further show that the exponent γ is bounded, $1 < \gamma < 3$, and:

- γ is monotonically *decreasing* with A_1.
- γ is monotonically *increasing* with A_2.

This means we can get not only the technological exponents, $2 < \gamma < 3$, but also the "biological" exponents, $1 < \gamma < 2$.

“Power law” $\to$ power law with exponential tail

Ubiquitous empirical measurements:

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

Fitting the Internet AS-level topology data

- Three "standard" views (c.f. CAIDA, TR-2005-02, 2005, and *SIGCOMM* 2006.)
  - Traceroute
  - BGP tables
  - Whois

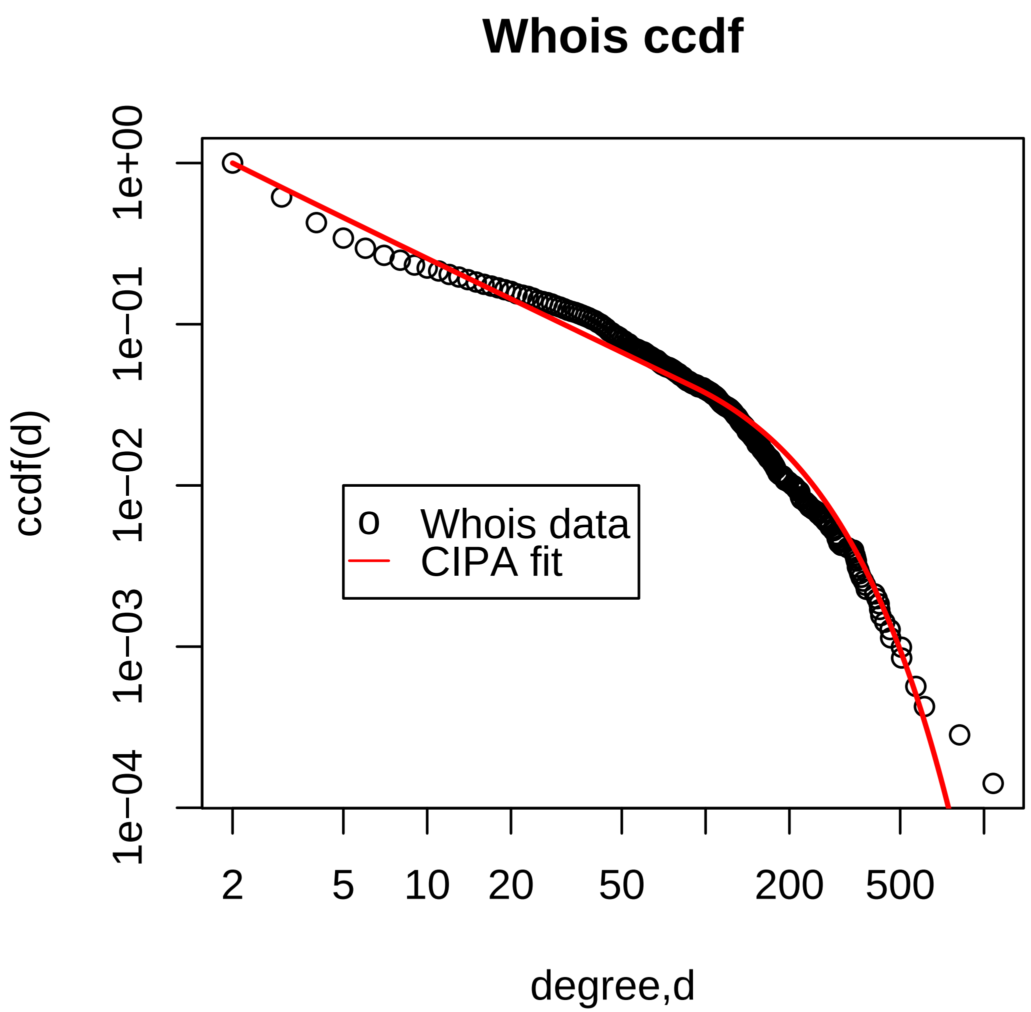
- First two shown to have power-law degree distributions, in accords with the standard view first put forth by (Faloutsos)^3 *SIGCOMM*, 1999. (And cited over 1100 times).

- Both views established via traceroute sampling.

Traceroute Sampling

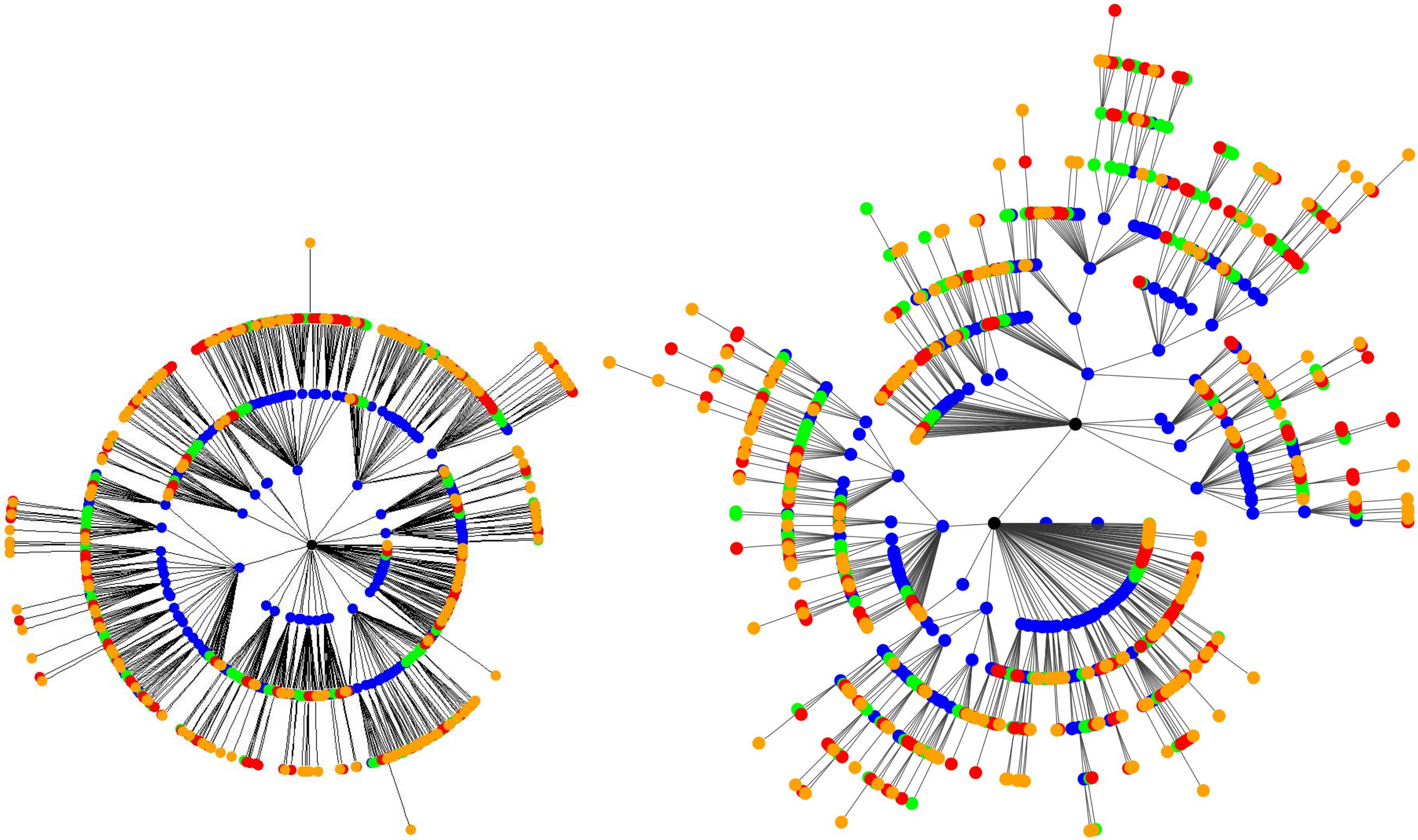
- Recently shown that traceroute methods can bias results, making underlying graph appear to have power-laws:
  - Lakhina, Byers, Crovella, Xie *INFOCOM*, 2003;
  - Achlioptas, Clauset, Kempe, Moore *STOC*, 2005.
- Counter-arguments– can distinguish *heavy-tailed* from simple exponential, but not precise form. Also depends on "betweeness", how many probes.....
  - Dall'Asta, Alvarez-Hamelin, Barrat, Vazquez, Vespignani *Theo. Comput. Sci.* 2006
- Third view (Whois), not power law and no explanation.

TPA and Whois



- TPA fit, $R^2 = 0.97$ with $A_1 = 187$, and $A_2 = 90$.

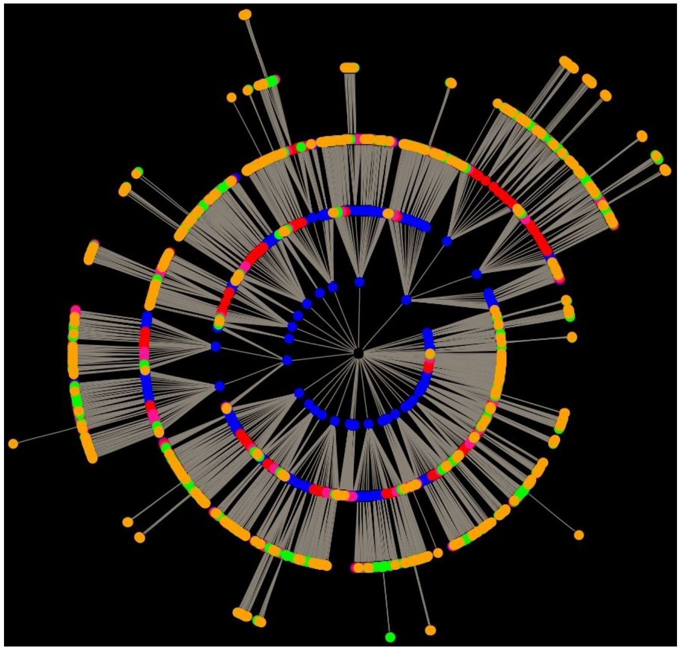
Comparing TPA and PA graphs



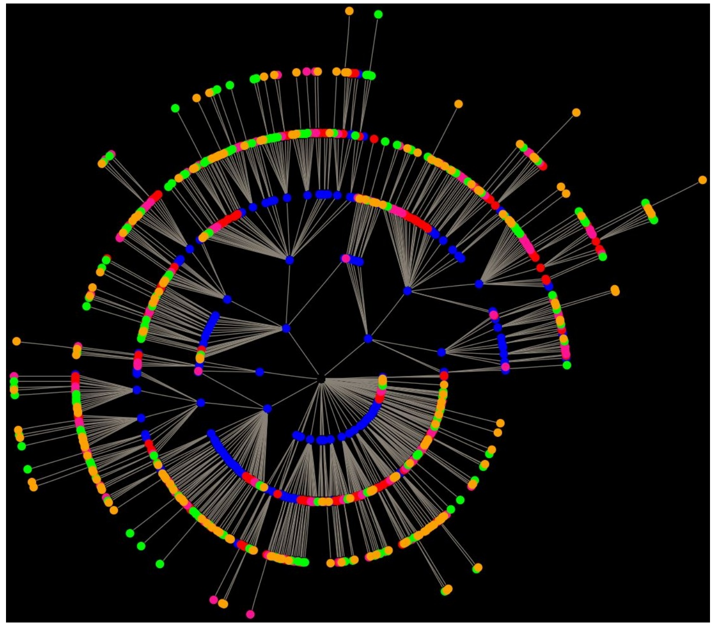
TPA

PA

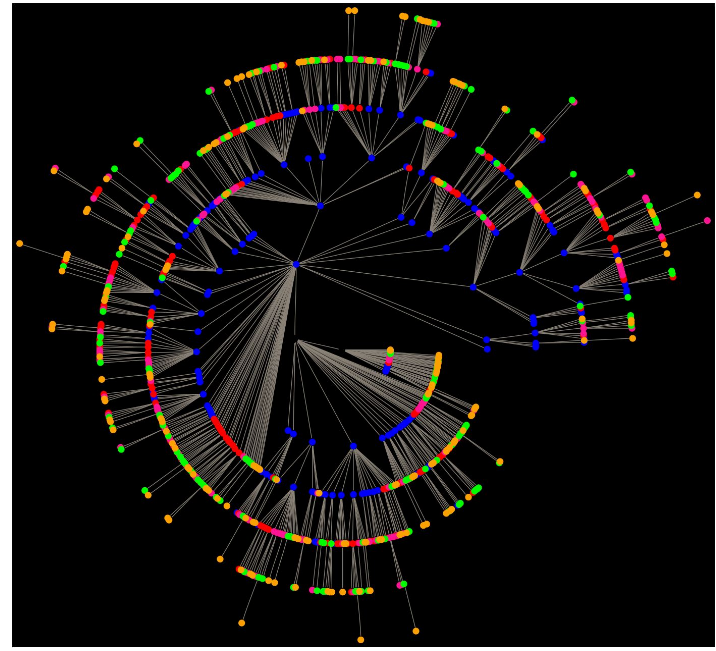
Example range of TPA graphs



$A_1 = 10, A_2 = 25$



$A_1 = 3, A_2 = 20$



$A_1 = 3, A_2 = 20$, (3-roots)

Numerical efficiency

Simulation of a TPA graph of size N requires:

- $cN \ln N$ space.
- Time linear in N.

(c is slightly larger than for standard PA, since need to keep list of "viable".)

Note, most optimization models require time $\sim N^2$ (exploration of all possible alternatives).

Extensions

- Different cost functions and geometries:
  – Biological choices? (modularity versus efficiency)
  – Open-source software ("systems' motifs")
  – Economics (whom should you trade with)

- What is the fine-structure? (joint deg dist)

- Validation! — historical data on Internet growth.

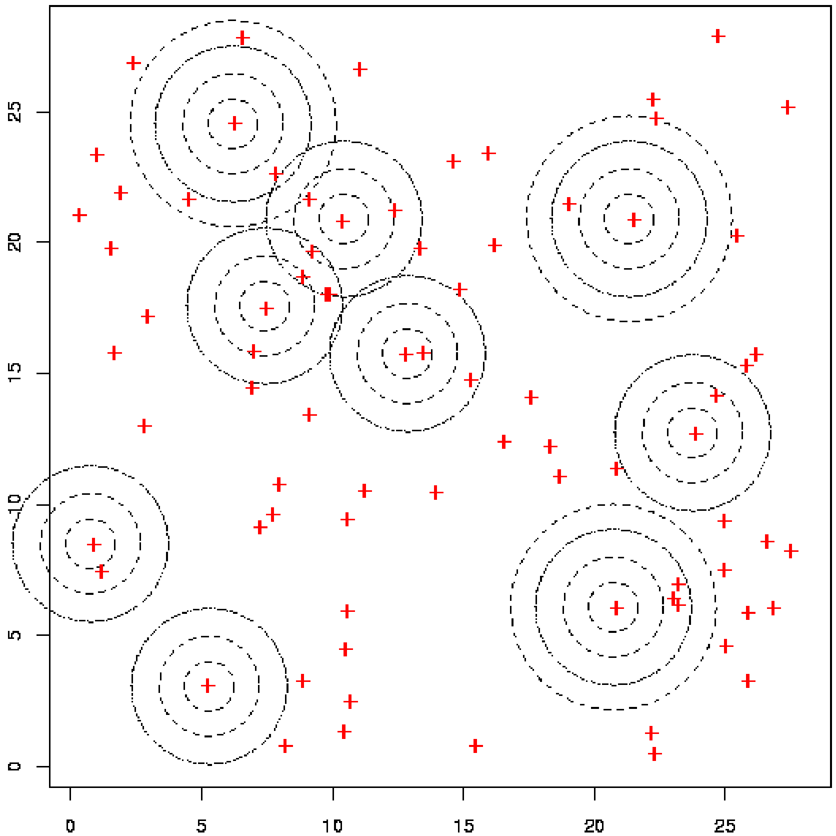
- Hierarchy and feedback – emergent for performance/robustness reasons

Part II: Timescales/Separation of Timescales

Relaxation time as the "characteristic" time for information flow on a network

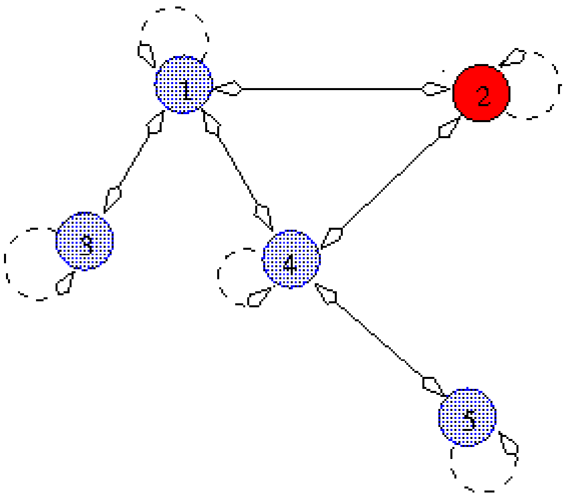
R.D, S. Ramanathan, and D. Temple Lang. "Measuring performance of ad hoc networks using timescales for information flow", *INFOCOM* 2003.

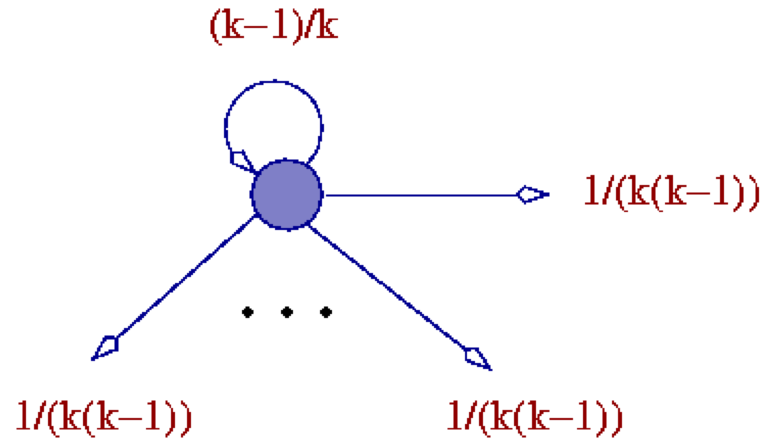
Build up a connected network of mobile nodes



- Can we ever do any routing?
- (i.e., do the nodes move too fast for us to make use of instantaneous topology?)

Consider a random walk on the instantaneous topology





Communication model:

- Only one conversation at a time ("unicast").
- So penalize for degree: k neighbors, get to transmit only $1/k$ of the time.
- Can choose any other model, what follows holds.

Relaxation time

- Let M be the resulting state-transition matrix (column normalized).
- Principle eigenvalue $\lambda_1 = 1$. (All other $\lambda_i < 1$.)
- $M^t \vec{u}_i = (\lambda_i)^t \vec{u}_i = \frac{1}{e} \vec{u}_i \implies t_i = -1/\ln(\lambda_i)$
- The relaxation time:

$$\boxed{\tau_{max} = -1/\ln(\lambda_2)}$$

where λ_2 is penultimate eigenvector.

Separation of timescales

(When can you study dynamics on dynamic graphs?)

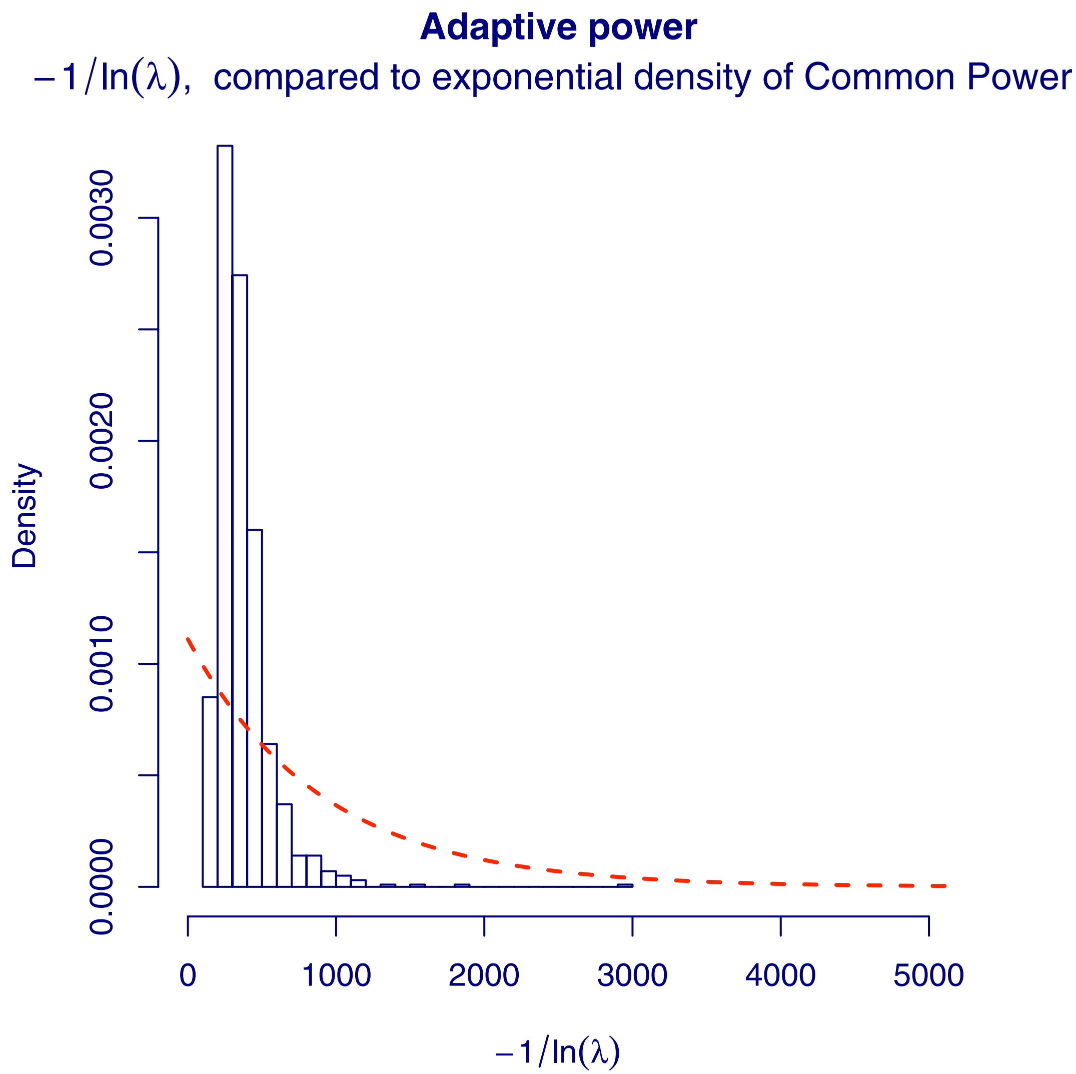
- τ_{max} estimates characteristic time for information flow on network.

- Let T be time it takes for sensor to move far enough that its connectivity changes (i.e., timescale for network topology to change).

- If $\tau_{max} \ll T$ network topology static during packet delivery $\Longrightarrow$ routing table useful.

- if $\tau_{max} > T$ network topology changes faster than packet can be delivered $\Longrightarrow$ no clever routing possible (just flooding).

Relaxation time as performance metric



R.D, S. Ramanathan, and D. Temple Lang. “Measuring performance of ad hoc networks using timescales for information flow”, *INFOCOM* 2003.

Network growth with feedback

- $\lambda(t)$: **Rate** of node arrival

- **Range**, $k(t) = \lceil cN(t)/\lambda(t) \rceil$: Only get a finite set of candidate parent nodes. (Some fraction, $c \leq 1$, of the network allocated to growth, the rest busy doing other things).

- **Optimization function**, F_{ij}: How attachment node j is chosen from k candidates.

- **Fitness**, $\mathcal{F}(G(t))$: The characteristic timescale for information flow, $\tau_{\max}$

- **Feedback**: $\lambda(t)$ evolves in time in response to feedback on the changes in the fitness of the system as follows:

Feedback

- **Feedback**: $\lambda(t)$ evolves in time in response to feedback on the changes in the fitness of the system as follows:

$$\lambda(t+1) = \begin{cases} \lambda(t), & \text{if } \mathcal{F}(G(t)) = \mathcal{F}(G(t-\delta)) \\ \lambda(t)+1, & \text{if } \mathcal{F}(G(t)) > \mathcal{F}(G(t-\delta)) \\ \max[\lambda(t)-1, 1], & \text{if } \mathcal{F}(G(t)) < \mathcal{F}(G(t-\delta)) \end{cases}$$

Preliminary results

With feedback, can grow larger, more fit networks in less time.

	$\delta \to \infty$	$\delta = 10$, $N_{\mathrm{stop}} = 501$	$\delta = 10$, $t_{\mathrm{stop}} = 167$
$\langle N(t) \rangle$	501	501	705.3 ± 10.3
$\langle \mathrm{Time} \rangle$	167	130.7 ± 1.8	167
$\langle h_j \rangle$	1.661 ± 0.002	1.753 ± 0.004	1.753 ± 0.004
$\langle \max h_j \rangle$	2.47 ± 0.05	2.58 ± 0.05	2.50 ± 0.05
$\langle \max d_j \rangle$	169.1 ± 0.9	132.1 ± 1.8	169.8 ± 1.0
$< \tau_c / \ln(N)^{2.35} >$	3.0 ± 0.1	3.2 ± 0.1	3.5 ± 0.1

Summary

- Mixing time, cover time, relaxation time as performance metrics.

- Establish separation of timescales. (When can we treat a dynamic graph as static).

- Optimization and network growth

Tempered Preferential Attachment

- Start from **optimization** framework.
- Gives an **underlying mechanism for how PA can arise.**
- Gives an **underlying mechanism for aging and saturation.**
- Introduces **Viability.**

TPA: Fitting data

- Degree distribution, power law with exponential tail

- $A_1 \neq A_2$
  – Get all the exponents, from $1 < \gamma < 3$.
  – Continuously interpolate from uniform attachment, to PA with cutoff, to PA.

- Numerical efficiency – even though an optimization model. Requires minimal additional numerical overhead when compared with standard PA.

TPA: A new class of models for analysis and simulation of networks.
