Rate Equation Approach to Growing Networks

Sidney Redner, Boston University

Motivation: **Citation distribution**

Basic Model for Citations:

Barabási-Albert network

Rate Equation Analysis:

Degree and related distributions
Global properties
Finiteness & fluctuations
Who is the leader?

Protein Interaction Network:

Cluster size distribution
Lack of self averaging

Outlook

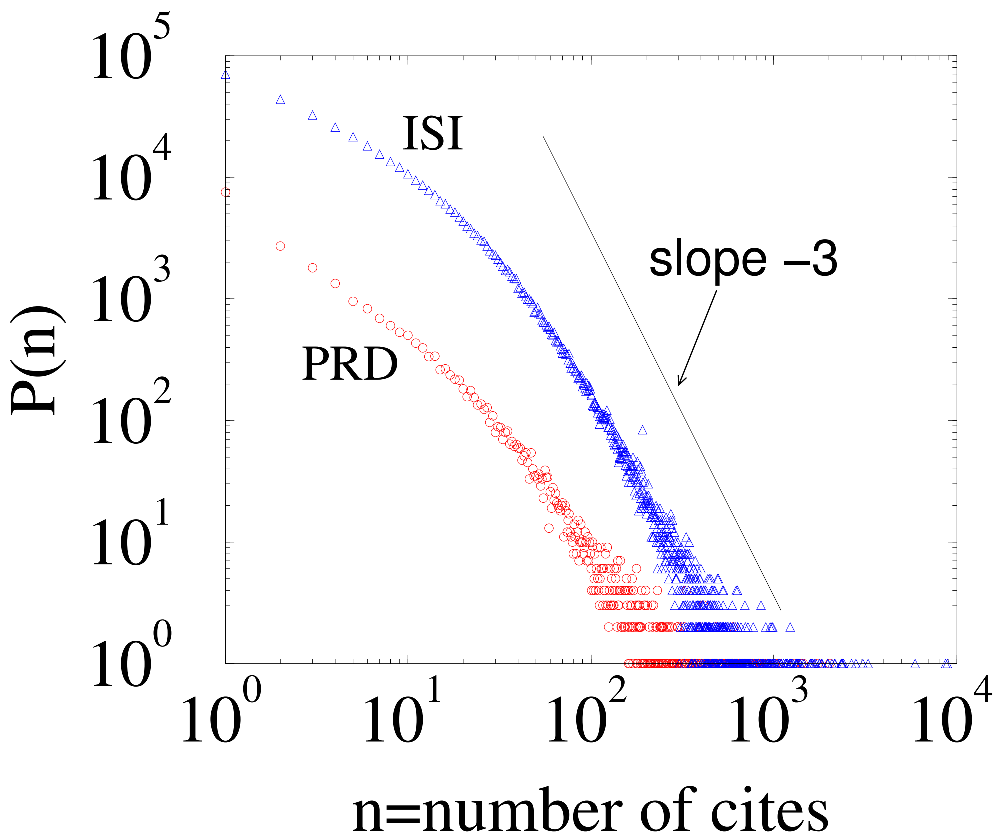
Paul Krapivsky (Boston University)
Francois Leyvraz (CIC, Mexico)
Geoff Rodgers (Brunel University, UK)
Jeenu Kim & Byungnam Kahng (SNU)

Citation Distribution

ISI: 783339 papers 6716198 cites, $\langle n \rangle = 8.6$.

1	paper	cited	8907	times
64	papers		>1000	times
282	papers		>500	times
2103	papers		>200	times
633391	papers		<10	times
368110	papers		0	times!

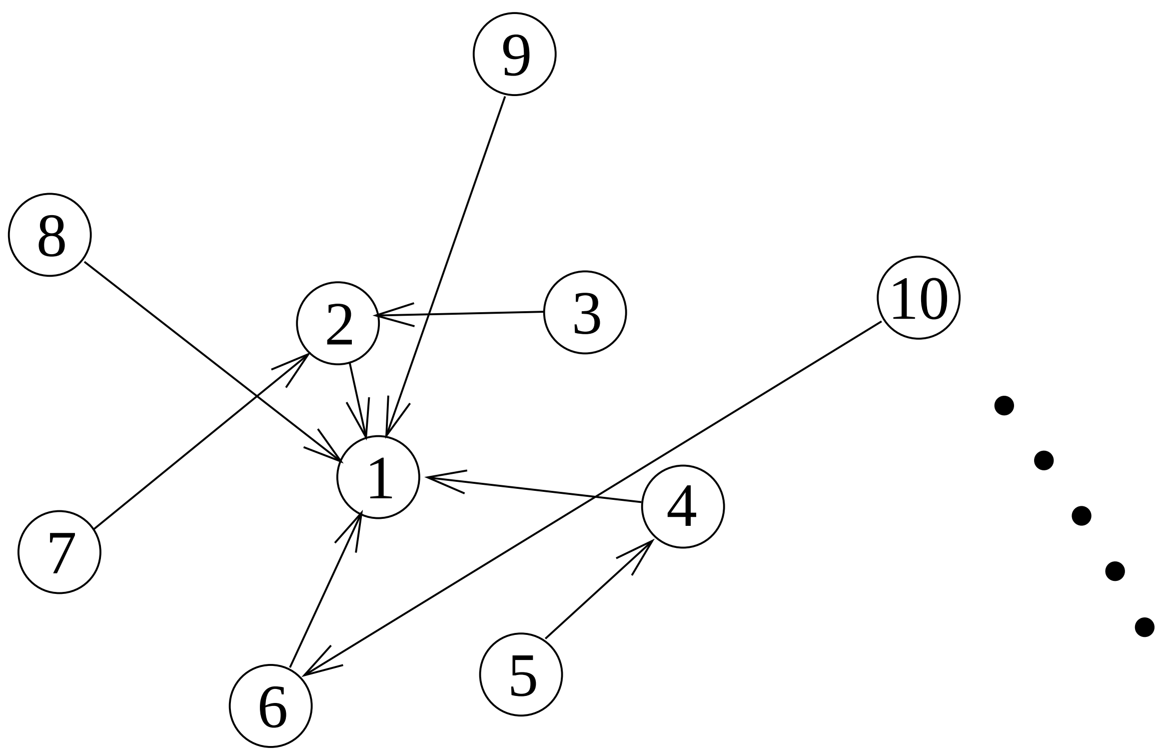
PRD: 24296 papers 351872 cites, $\langle n \rangle = 14.5$.



J. Laherrere and D. Sornette, EPJB (1998)
Redner, EPJB (1998)

Barabási-Albert Model

nodes $\longleftrightarrow$ publications
links $\longleftrightarrow$ citations



1. **Introduce nodes one at a time.**
2. **Attach to earlier node with k links at rate A_k.**

H. A. Simon, Biometrica (1955)
A. L. Barabási and R. Albert, Science (1999)

Rate Equation Approach

KRL, PRL (2000), KR, PRE (2001)
see also, Albert & Barabási, Rev. Mod. Phys. (2002)
Dorogovtsev & Mendes, Adv. Phys. (2002)

Basic Observable:

$N_k \equiv$ **Number of nodes with k links**
The degree distribution.

Rate Equation:

$$\frac{dN_k}{dt} = \frac{A_{k-1}N_{k-1} - A_k N_k}{A} + \delta_{k1}.$$

Attachment Rate:

$$A_k \sim k^\gamma$$

so that

$$A(t) = \sum_{j=1}^{\infty} A_j N_j = \sum_{j=1}^{\infty} j^\gamma N_j \equiv M_\gamma(t).$$

Moment equations:

$\dot{M}_0 \equiv \sum_k \dot{N}_k = 1; \quad \dot{M}_1 \equiv \sum_k k\dot{N}_k = 2$

These suggest: $A(t) = \sum j^\gamma N_j \propto \mu(\gamma)t$

$$N_k(t) \equiv t n_k.$$

Rate eqs. → Linear recursion relations

Formal Solution:

$$n_k = \frac{\mu}{A_k} \prod_{j=1}^{k} \left(1 + \frac{\mu}{A_j}\right)^{-1}$$

Asymptotics:

$$n_k \sim \begin{cases} k^{-\gamma} \exp\left[-\mu\left(\frac{k^{1-\gamma}-2^{1-\gamma}}{1-\gamma}\right)\right], & 0 \leq \gamma < 1; \\ k^{-\nu}, \quad \nu > 2, & \gamma = 1; \\ \text{best seller} & 1 < \gamma < 2; \\ \text{bible} & \gamma > 2. \end{cases}$$

Heterogeneity

Bianconi & Barabási, (2000); KR (2002).

Each node has intrinsic "attractiveness" η and attachment rate $A_k(\eta)$.

Rate equation:

$$\frac{dN_k(\eta)}{dt} = \frac{A_{k-1}(\eta)N_{k-1}(\eta) - A_k(\eta)N_k(\eta)}{A} + p_0(\eta)\delta_{k1}.$$

Solution for linear kernel $A_k(\eta) = \eta k$:

$$n_k(\eta) = \frac{\mu\, p_0(\eta)}{\eta}\, \frac{\Gamma(k)\, \Gamma\left(1 + \frac{\mu}{\eta}\right)}{\Gamma\left(k + 1 + \frac{\mu}{\eta}\right)}.$$

Asymptotics for total degree distribution:

Bounded support of $p_0(\eta)$:

$$n_k \sim k^{-(1+\mu/\eta_{\rm max})}\, (\ln k)^{-\omega},$$

with μ determined by $1 = \int d\eta\, p_0(\eta)\, \left(\frac{\mu}{\eta} - 1\right)^{-1}$.

Unbounded support: condensation!

Age Distribution: KR, PRE (2001).

$N_k(t,a)$: **# nodes of degree k, age a, time t.**

Rate Equation:

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)N_k=\frac{A_{k-1}N_{k-1}-A_kN_k}{A}+\delta_{k1}\delta(a).$$

Age-Degree Distribution:

$$A_k=k: \qquad N_k(t,a)=\sqrt{1-\frac{a}{t}}\left(1-\sqrt{1-\frac{a}{t}}\right)^{k-1}$$

$$A_k=1: \qquad N_k(t,a)=(1-a/t)\frac{|\ln(1-a/t)|^{k-1}}{(k-1)!}.$$

Average Age:

$$A_k=k: \qquad \langle a_k\rangle \sim t(1-12/k^2)$$

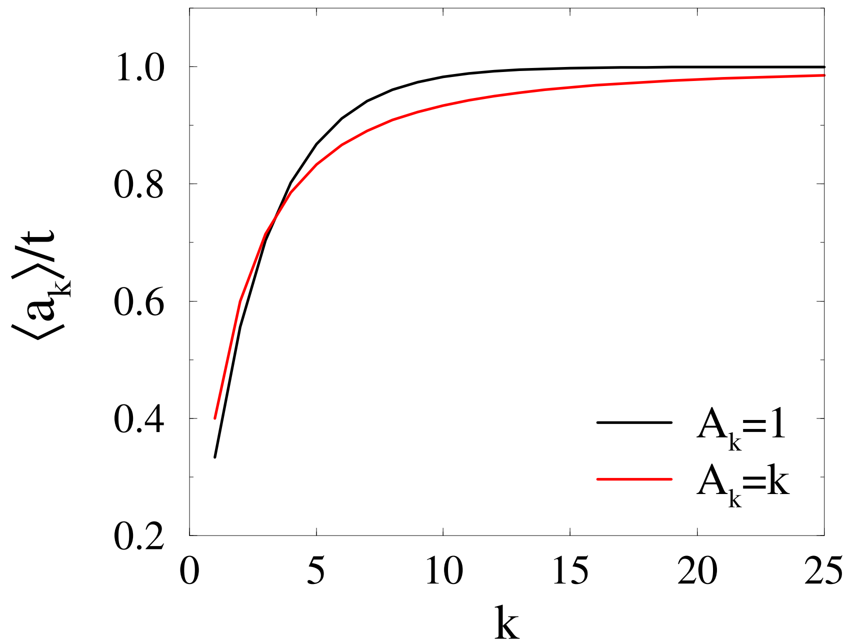
$$A_k=1: \qquad \langle a_k\rangle = t[1-(2/3)^k].$$

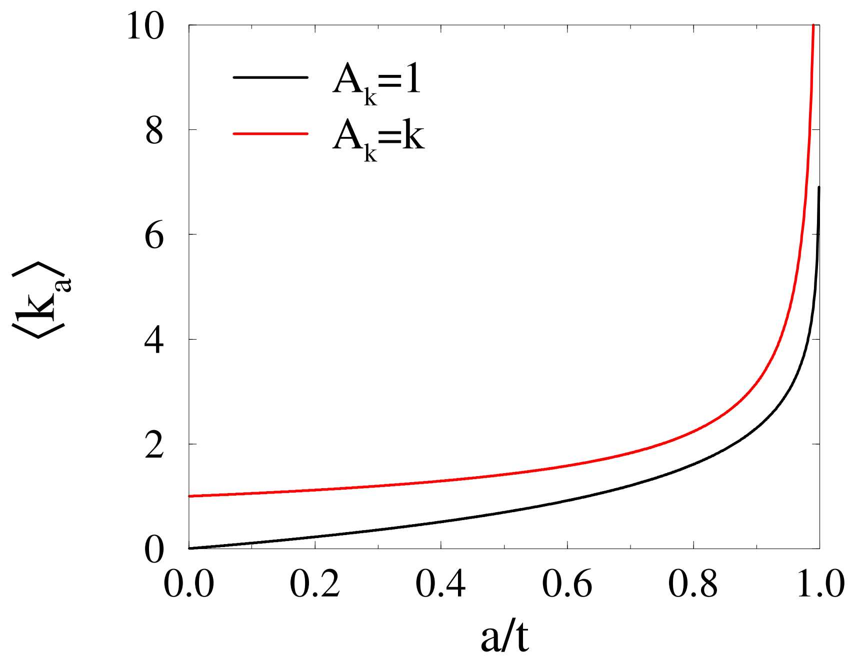
Average Degree:

$$A_k=k: \qquad \langle k_a\rangle \sim (1-a/t)^{-1/2}$$

$$A_k=1: \qquad \langle k_a\rangle = -\ln(1-a/t).$$

Age-degree statistics:



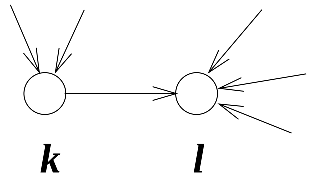


Message: $A_k = k$, **rich nodes must be old.**

$A_k = 1$, **rich nodes can be young.**

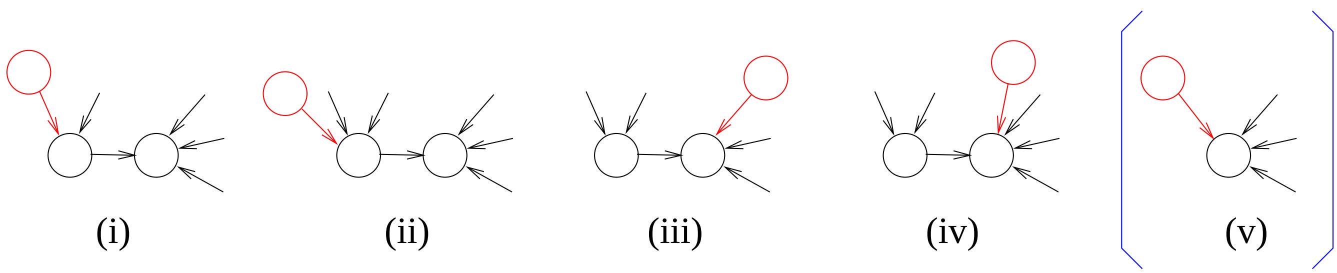
Degree Correlations:

$C_{kl}(t) \equiv$ number of nodes of degree k that attach to an ancestor node of degree l.



Rate equation (linear kernel):

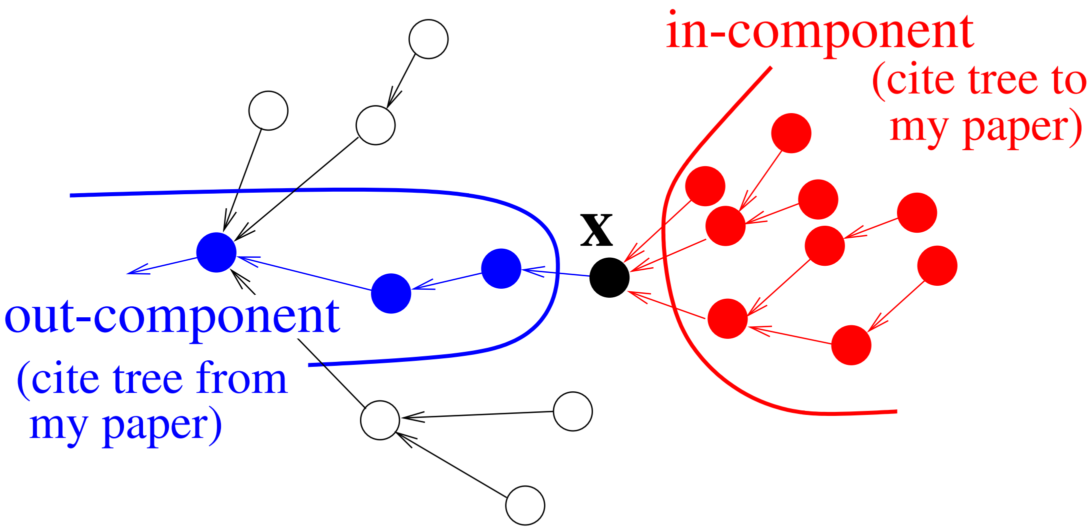
$$\frac{dC_{kl}}{dt} = \frac{1}{A}\Big\{ [(k-1)C_{k-1,l} - kC_{kl}] + [(l-1)C_{k,l-1} - lC_{kl}] \Big\} + (l-1)C_{l-1}\,\delta_{k1}.$$



Asymptotic solution: $(k, l \gg 1$ **and** $k/l \neq 1)$

$$c_{kl} \to \begin{cases} 16\,(l/k^5) & l \ll k, \\ 4/(k^2\, l^2) & l \gg k. \end{cases} \qquad c_{kl} \neq n_k n_l \propto (k\, l)^{-3}\ !$$

In- and Out-Components:



$I_s(t) \equiv$ **No. of in-components with** s **nodes**

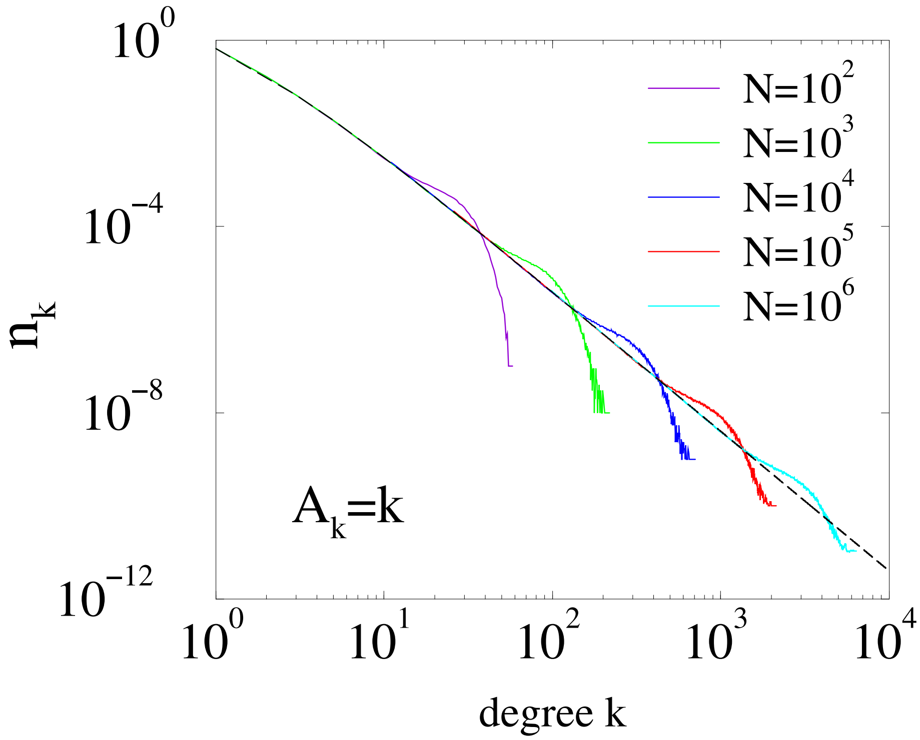
$$\sim \frac{t}{s^2}.$$

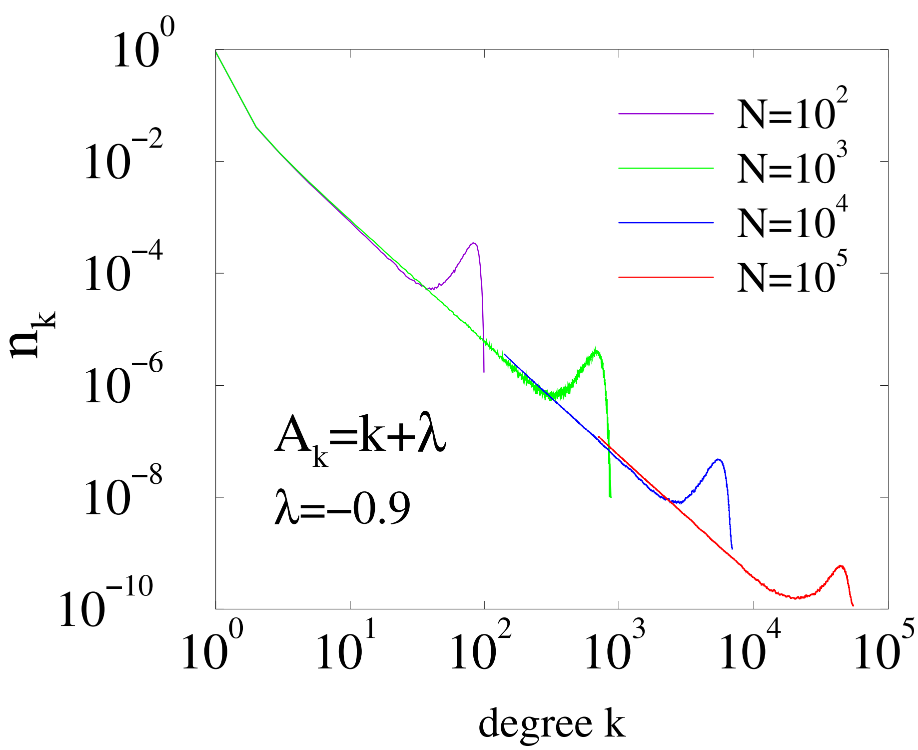
$O_s(t) \equiv$ **No. of out-components with** s **nodes**

$$\sim \frac{(\ln t)^s}{s!}.$$

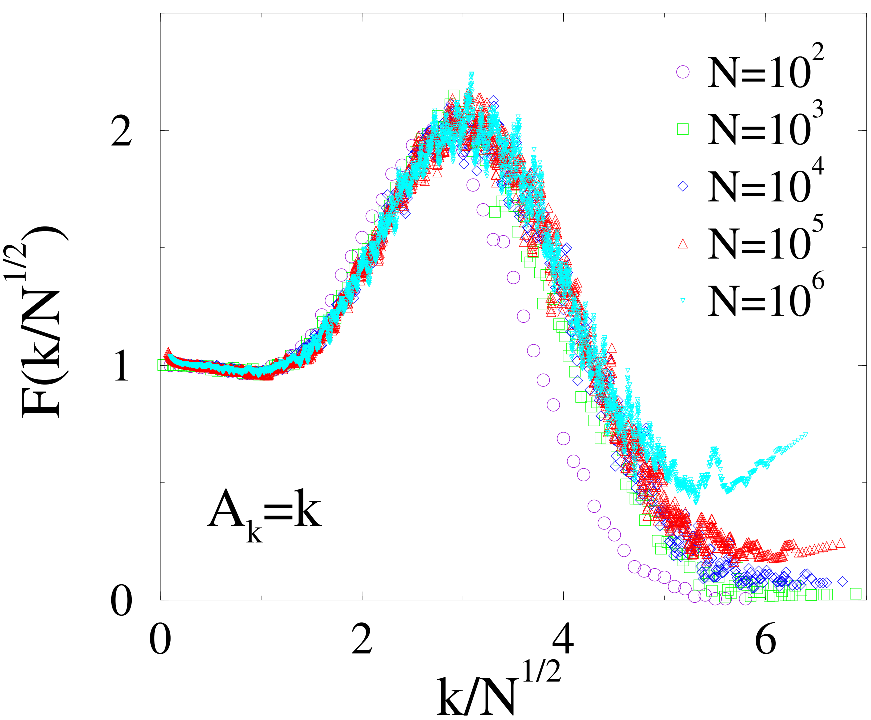
Degree Distribution of Finite Networks:

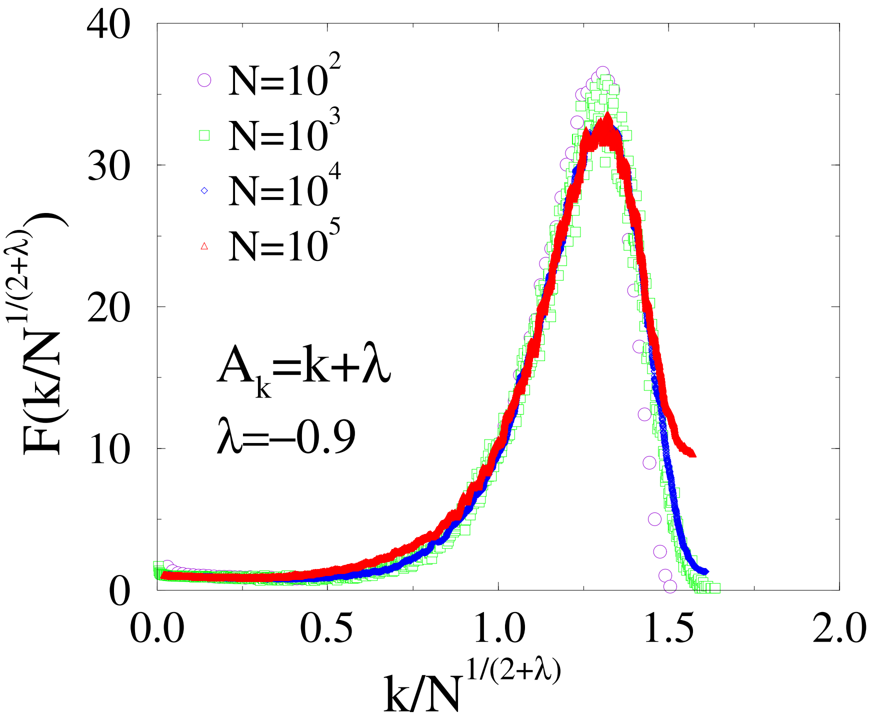
Dorogovtsev et al PRE ('01), Moreira et al cond-mat 0205411, KR ('02).



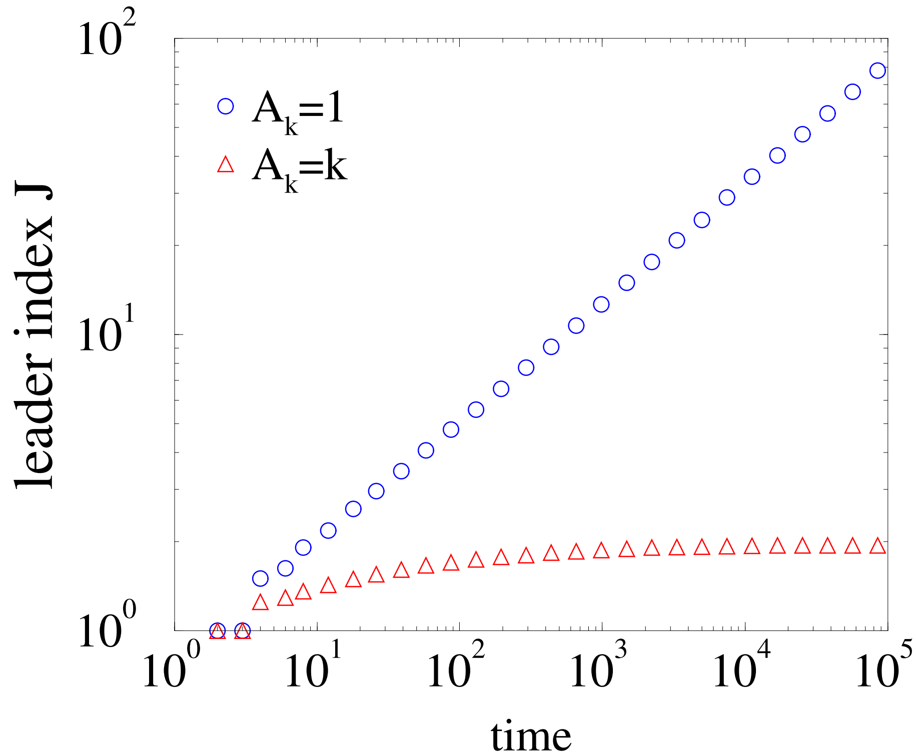


Scaling Near the Extreme:





Who is the most popular node?

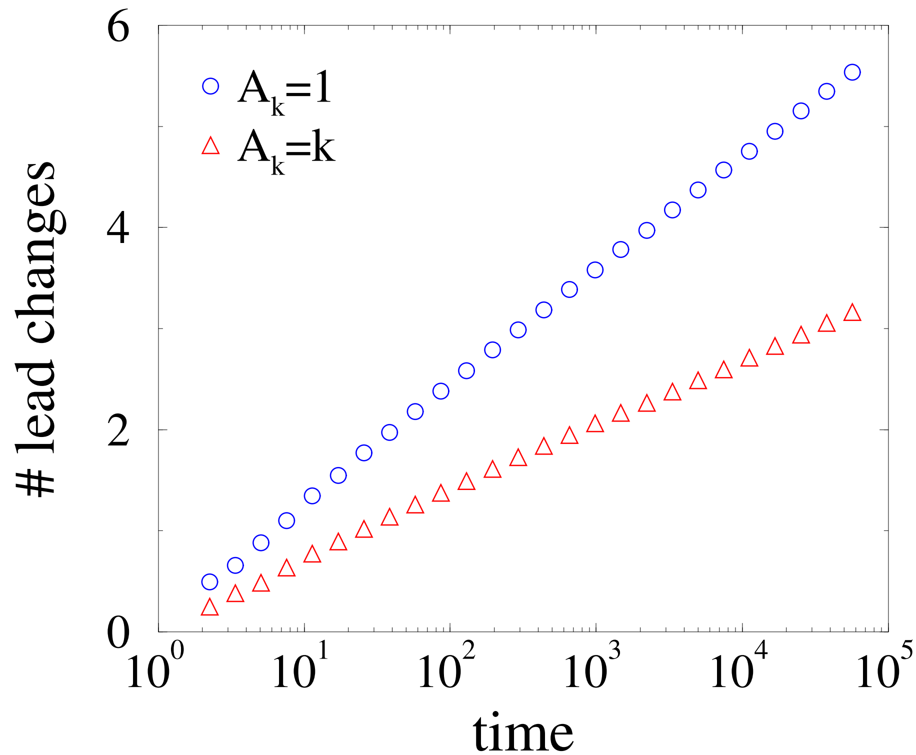


Using $J = t - a_{k_{\max}}$, results for $k_{\max}$ & $a_{k_{\max}}$:

- **For $A_k = k$, leader among the oldest!**
  $J \approx 1.9$. $P_1 \approx 0.44$, $P_2 \approx 0.21$, $P_3 \approx 0.10$,

- **For $A_k = 1$, leader index $J(t) \sim t^{\psi}$, with**
  $\psi = 1 + \ln(2/3)/\ln 2 \approx 0.415$.

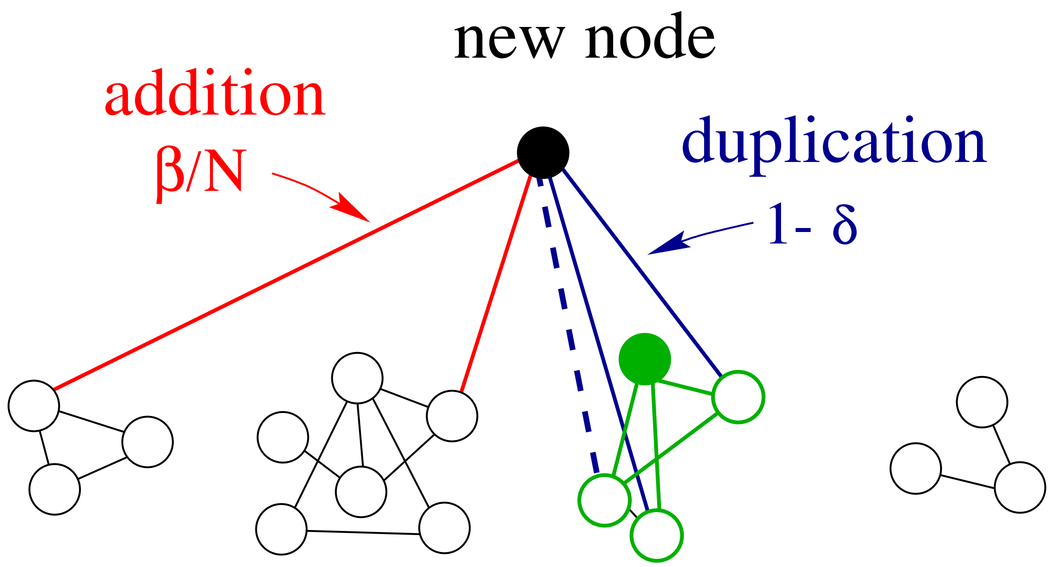
How many lead changes occur?



Basic feature:

Number of lead changes up to time t is proportional to $\ln t$.

Protein Interaction Network



Duplication: A new node duplicates a random pre-existing node by connecting to each of its neighbors with probability $1-\delta$.

Addition: A new node links to any previous node with probability β/N.

Uetz et al., Nature (2000).
Wagner, PNAS (1994).

Vazquez et al., cond-mat/0108043 (2001).
Solé et al., Adv. Complex Systems (2002).

Rate Equation (equiprob. node selection):

K^{3}R (2002)

$$\frac{dN_k}{dN} = \frac{A_{k-1}N_{k-1} - A_k N_k}{N} + G_k.$$

Attachment Rate:

$$A_k = \underbrace{(1-\delta)k}_{\text{duplication}} + \underbrace{\beta}_{\text{addition}}$$

"Source":

$$G_k = \sum_{a+b=k} \sum_{s=a}^{\infty} \underbrace{n_s \binom{s}{a} (1-\delta)^a \delta^{s-a}}_{\text{duplication: } a \text{ links}} \underbrace{\frac{\beta^b}{b!} e^{-\beta}}_{\text{addition: } b \text{ links}}$$

$$\to (1-\delta)^{\gamma-1} n_k.$$

$$n_k = N_k/N.$$

Average node degree:

In each event, number of links evolves as

$$\frac{dL}{dN} = \beta + (1-\delta)\,\frac{2L}{N},$$

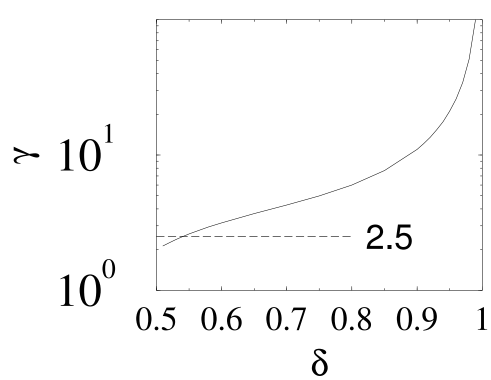
Combining with $\mathcal{D}(N) = 2L(N)/N$, the average node degree $\mathcal{D}$ is

$$\mathcal{D}(N) = \begin{cases} \text{finite} & \delta > 1/2, \\ \beta \ln N & \delta = 1/2, \\ \text{const.} \times N^{1-2\delta} & \delta < 1/2. \end{cases}$$

The degree distribution (for $\delta < 1/2$):

The rate equation is a recursion.
Substituting $n_k \sim k^{-\gamma}$ determines γ via:

$$\gamma = 1 + \frac{1}{1-\delta} - (1-\delta)^{\gamma-2}$$



Addition Only: A *non-local* percolation.

Rate equation for cluster size distr. C_s

$$\frac{dC_s}{dN} = - \underbrace{\beta \frac{sC_s}{N}}_{\text{loss by linking}} + \underbrace{\sum_{n=0}^{\infty} \frac{\beta^n}{n!} e^{-\beta} \sum_{s_1 \cdots s_n} \prod_{j=1}^{n} \frac{s_j C_{s_j}}{N}}_{\text{gain by } n-\text{body merging}},$$

sum $s_1 \geq 1, \ldots, s_n \geq 1$**, with** $s_1 + \cdots + s_n + 1 = s$**.**

Generating function:

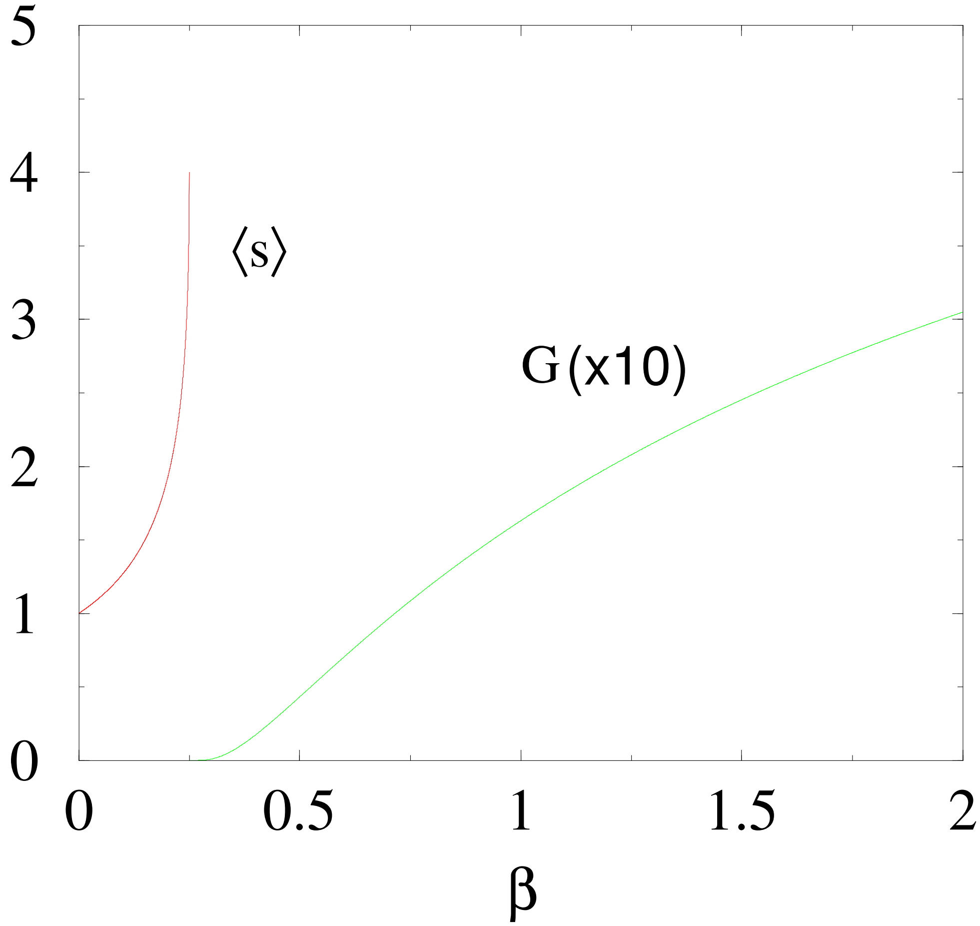
Define $g(z) = \sum_1^{\infty} sc_s e^{sz}$**, with** $c_s = C_s/N$

$$\rightarrow g = -\beta g' + (1 + \beta g')\, e^{z+\beta(g-1)}.$$

Basic results:

$$\begin{aligned} \langle s \rangle &= g'(0) = \frac{1 - 2\beta - \sqrt{1-4\beta}}{2\beta^2}; \\ c_s &\sim As^{-\tau}, \qquad \tau = 1 + \frac{2}{1 - \sqrt{1-4\beta}}; \\ G(\beta) &\sim e^{-\pi/\sqrt{4\beta - 1}}. \end{aligned}$$

Mean cluster size $\langle s \rangle$, percolation prob. G



Cluster size distribution:

$\beta < \frac{1}{4}$: **non-universal power law**

τ **rapidly decreasing in** β; $\tau \to 3$ **as** $\beta \to \frac{1}{4}$.

$\beta = \frac{1}{4}$: **logarithmic correction**

$c_s \sim 8 s^{-3} (\ln s)^{-2}$.

Duplication Only:

Rate equation for complete duplication

$$\frac{dN_k}{dN} = (k-1)\frac{[N_{k-1} - N_k]}{N}.$$

Solution: $N_k = 2\left(1 - \frac{2}{N}\right)^{k-1}$ – **irrelevant!**

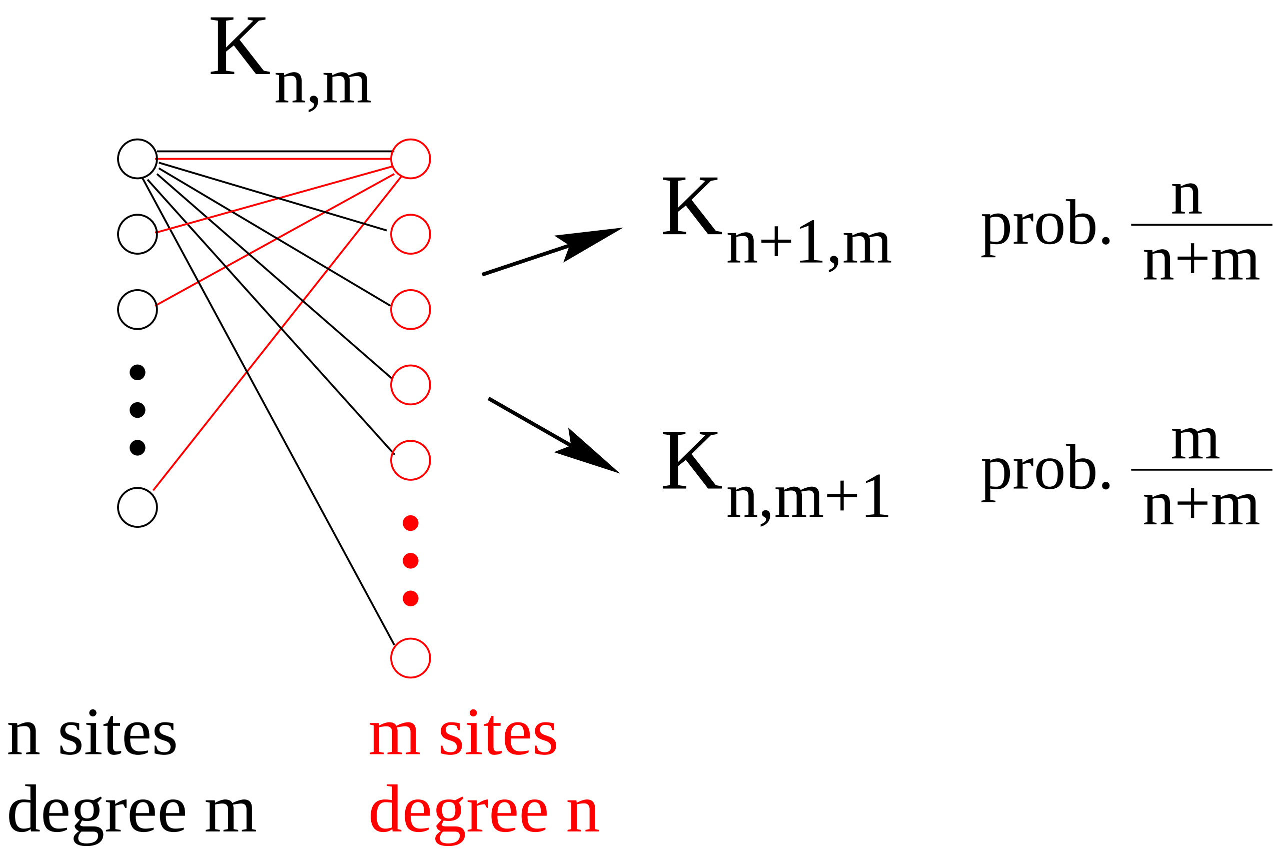
Instead: Strong sample-specific fluctuations

No self-averaging.

Asymptotic behavior:

$K_{1,1}$ **evolves to** $K_{n,m}$**.**

Each state $\{n, m\}$ **(with** $n + m = N$**) occurs with uniform probability.**



$K_{n,m}$
$K_{n+1,m}$
prob. $\frac{n}{n+m}$
$K_{n,m+1}$
prob. $\frac{m}{n+m}$
n sites
degree m
m sites
degree n

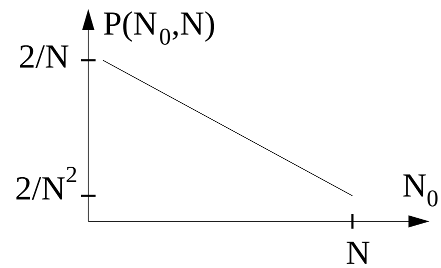
Failure of scaling:

If isolated sites created, they evolve independently and with strong sample-specific fluctuations.

Simple example: Start with $K_{1,1,0}$ ($\circ\!\!-\!\!\circ \quad \circ$).

After N sites, with complete duplication:

$$P(N_0, N) = 2\frac{N-1-N_0}{(N-1)(N-2)}$$



$$\to \langle N_0 \rangle = \frac{N}{3}, \qquad \text{while} \qquad (N_0)_{\rm mp} \approx 1.$$

Incomplete duplication: any isolated sites created will evolve independently of $N_{k>0}$!

Outlook

The Rate Equation!

- **Simple yet powerful tool.**

Basic Messages

- **Degree distribution easily computable:**
  - **Power law not generic or robust.**
  - **Stretched exponential is robust.**
- **Heterogeneity, age distribution, correlations, global features, extremes,** *etc.*
- **Fluctuations – unresolved.**

Other Growth Mechanisms:

- **Biological processes.**
  - **New percolation process.**
  - **Self averaging can fail.**
  - **“Errors” for robust behavior.**