Statistical Physics of Popularity-Driven Networks

Sidney Redner, CNLS (on leave from Boston University)
collaborators: P. Krapivsky, F. Leyvraz

University of Arizona, January 14, 2005

Observations about scientific citations:

- amusing general facts/idle gossip
- analysis of citation data

Preferential attachment network

Master equation approach:

- degree distributions
- redirection & copying

Summary & Outlook

Phys. Rev Citation Data

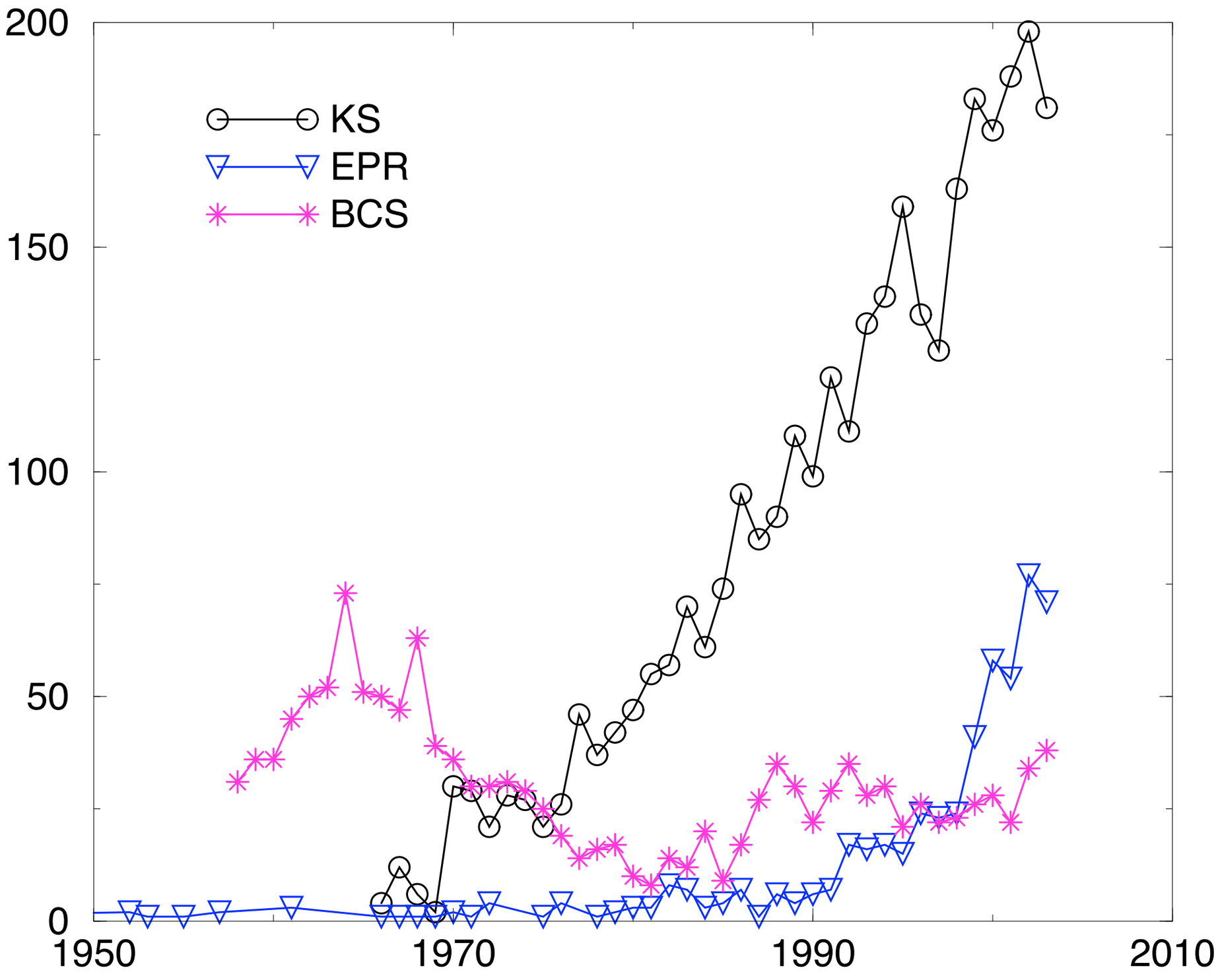
353,268 papers, 3,110,839 cites, av. = 8.81

11 papers with > 1000 citations
79 papers with > 500 citations
237 papers with > 300 citations
2340 papers with > 100 citations
8073 papers with > 50 citations
245459 papers with < 10 citations
84144 papers with 1 citation
23421 papers with 0 citations

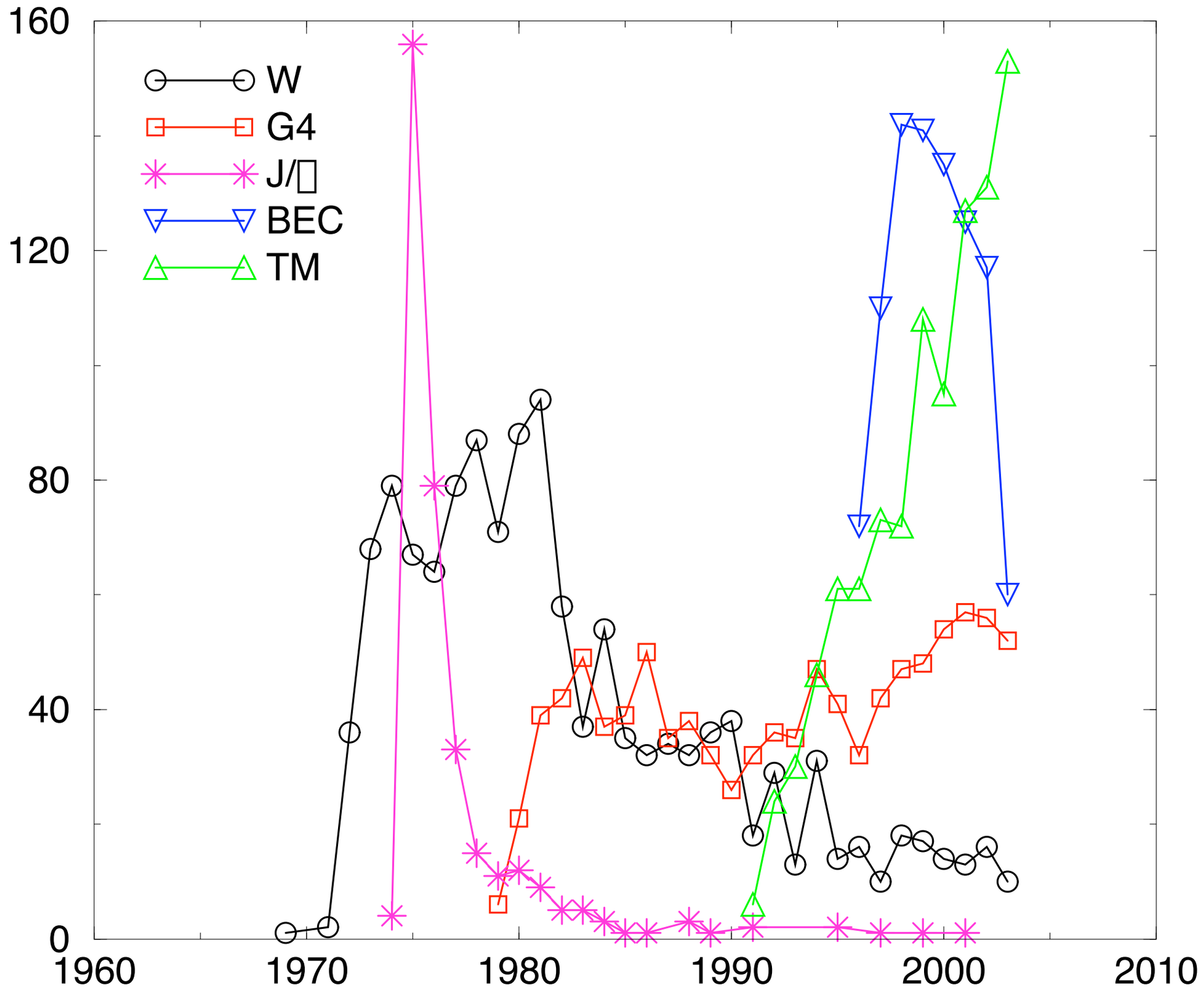
PR papers with >1000 cites

Cite Rank	Publication				# cites	Av. Age	Impact	Title	Author(s)
1	PR	140	A1133	1965	3227*	26.64	85972	Self-Consistent Equations...	W. Kohn & L. J. Sham
2	PR	136	B864	1964	2460*	28.70	70604	Inhomogeneous Electron Gas	P. Hohenberg & W. Kohn
3	PRB	23	5048	1981	2079	14.38	29896	Self-Interaction Correction to...	J. P. Perdew & A. Zunger
4	PRL	45	566	1980	1781	15.42	27463	Ground State of the Electron ...	D. M. Ceperley & B. J. Alder
5	PR	108	1175	1957	1364	20.18	27526	Theory of Superconductivity	J. Bardeen, L. N. Cooper, & J. R. Schrieffer
6	PRL	19	1264	1967	1306	15.46	20191	A Model of Leptons	S. Weinberg
7	PRB	12	3060	1975	1259	18.35	23103	Linear Methods in Band Theory	O. K. Andersen
8	PR	124	1866	1961	1178	27.97	32949	Effects of Configuration...	U. Fano
9	RMP	57	287	1985	1055	9.17	9674	Disordered Electronic Systems	P. A. Lee & T. V. Ramakrishnan
10	RMP	54	437	1982	1045	10.82	11307	Electronic Properties of...	T. Ando, A. B. Fowler, & F. Stern
11	PRB	13	5188	1976	1023	20.75	21227	Special Points for Brillouin-...	H. J. Monkhorst & J. D. Pack

Citation histories of classic PR papers

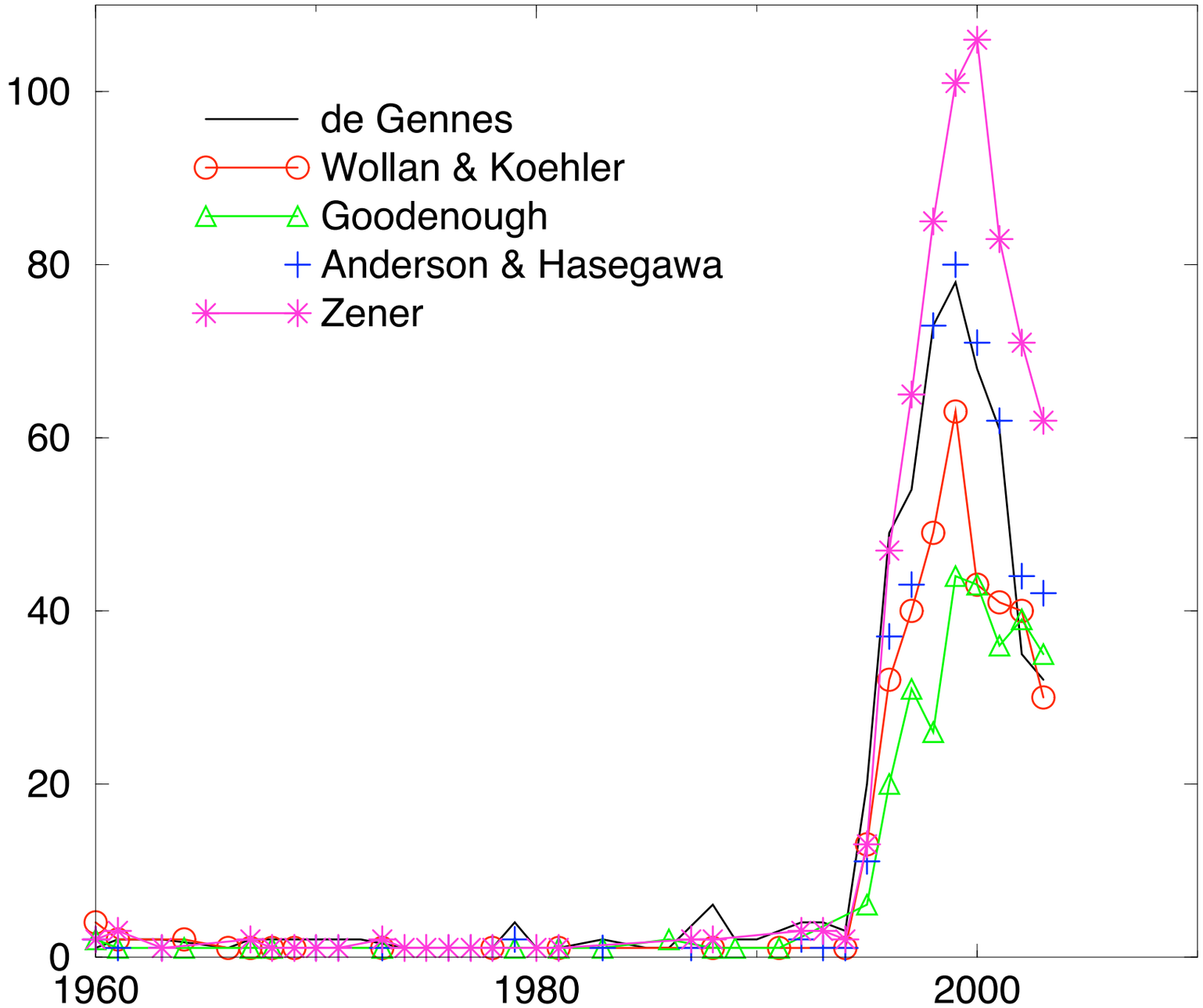


Citation histories of selected PR papers

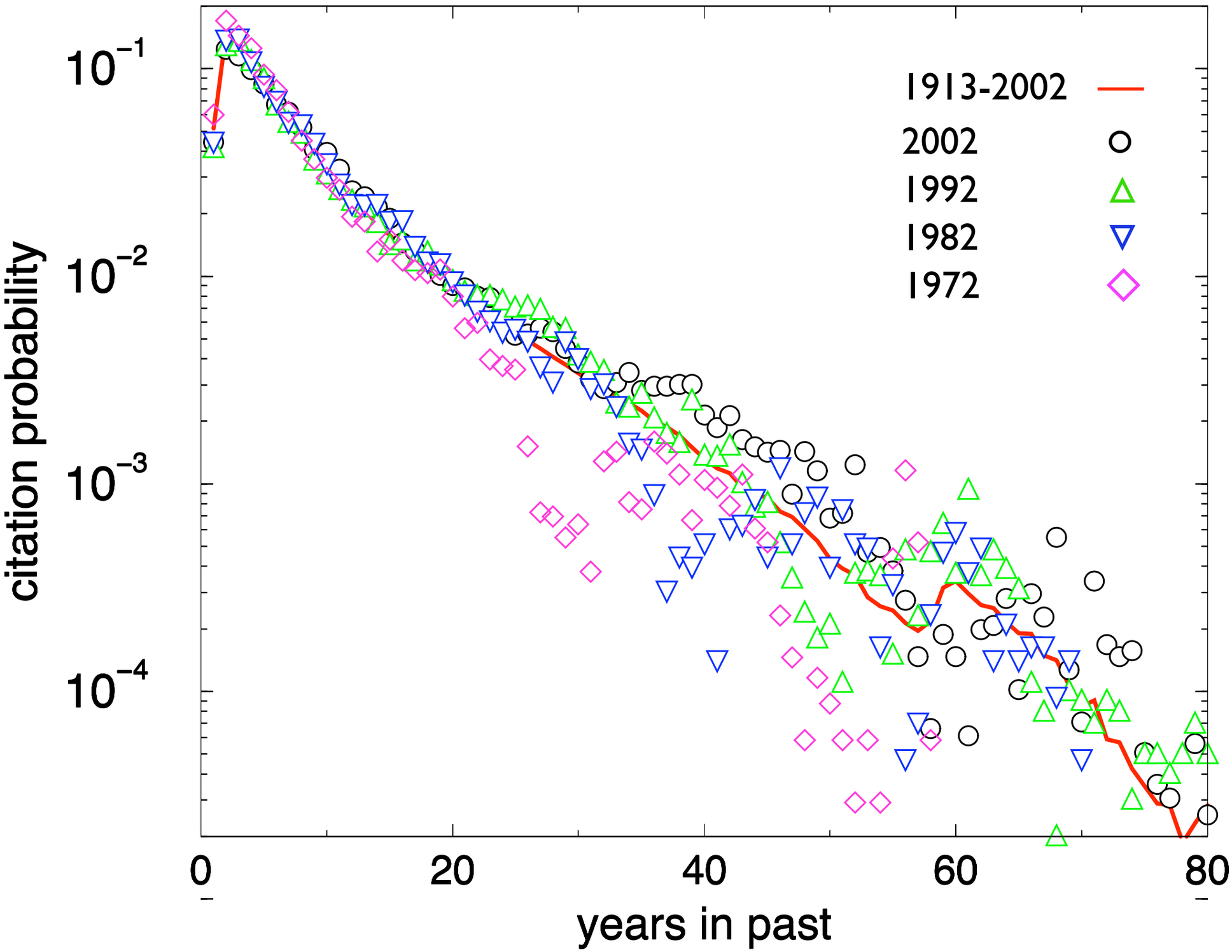


Citation histories of papers on double exchange

colossal magnetoresistance

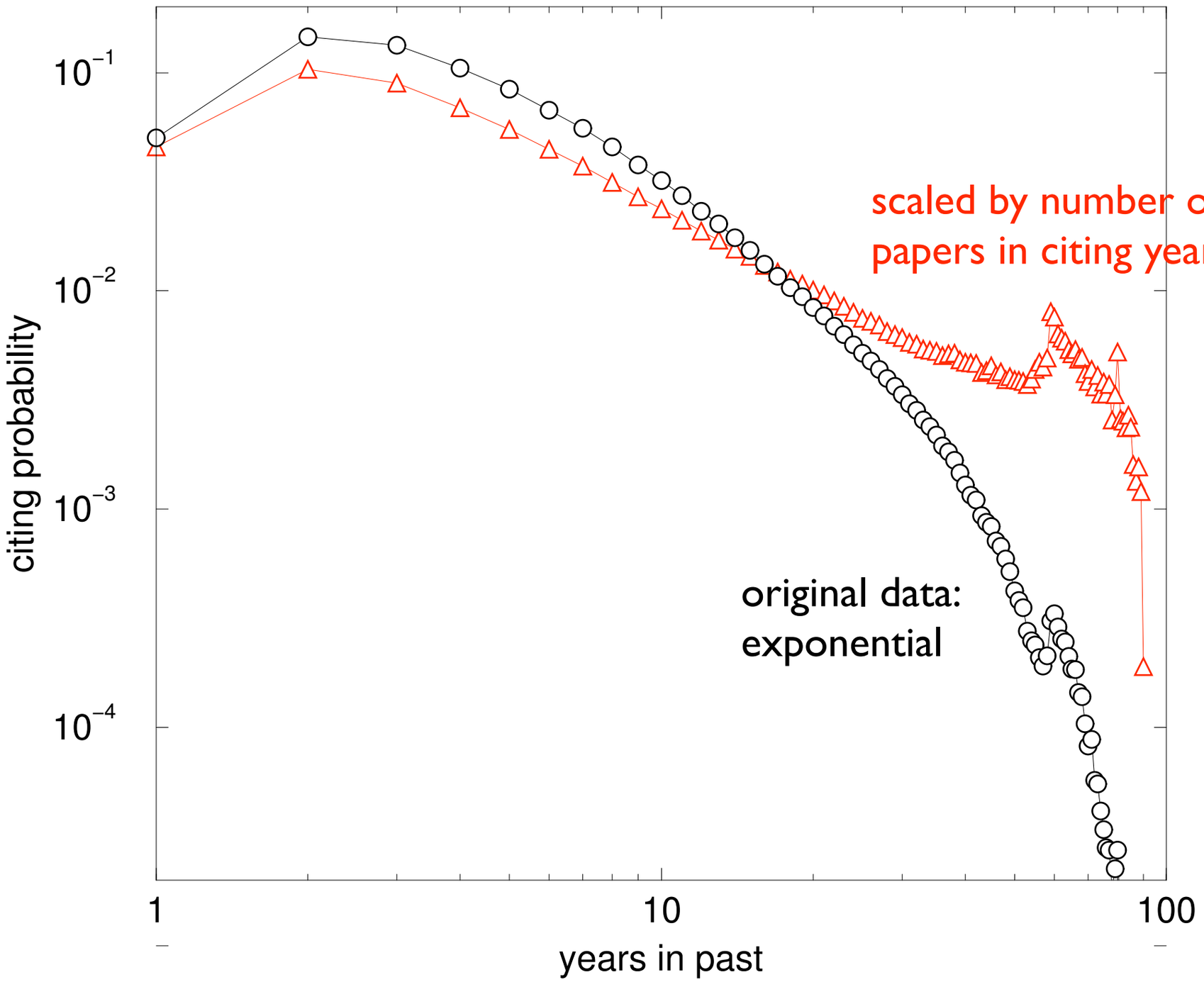


Citing History

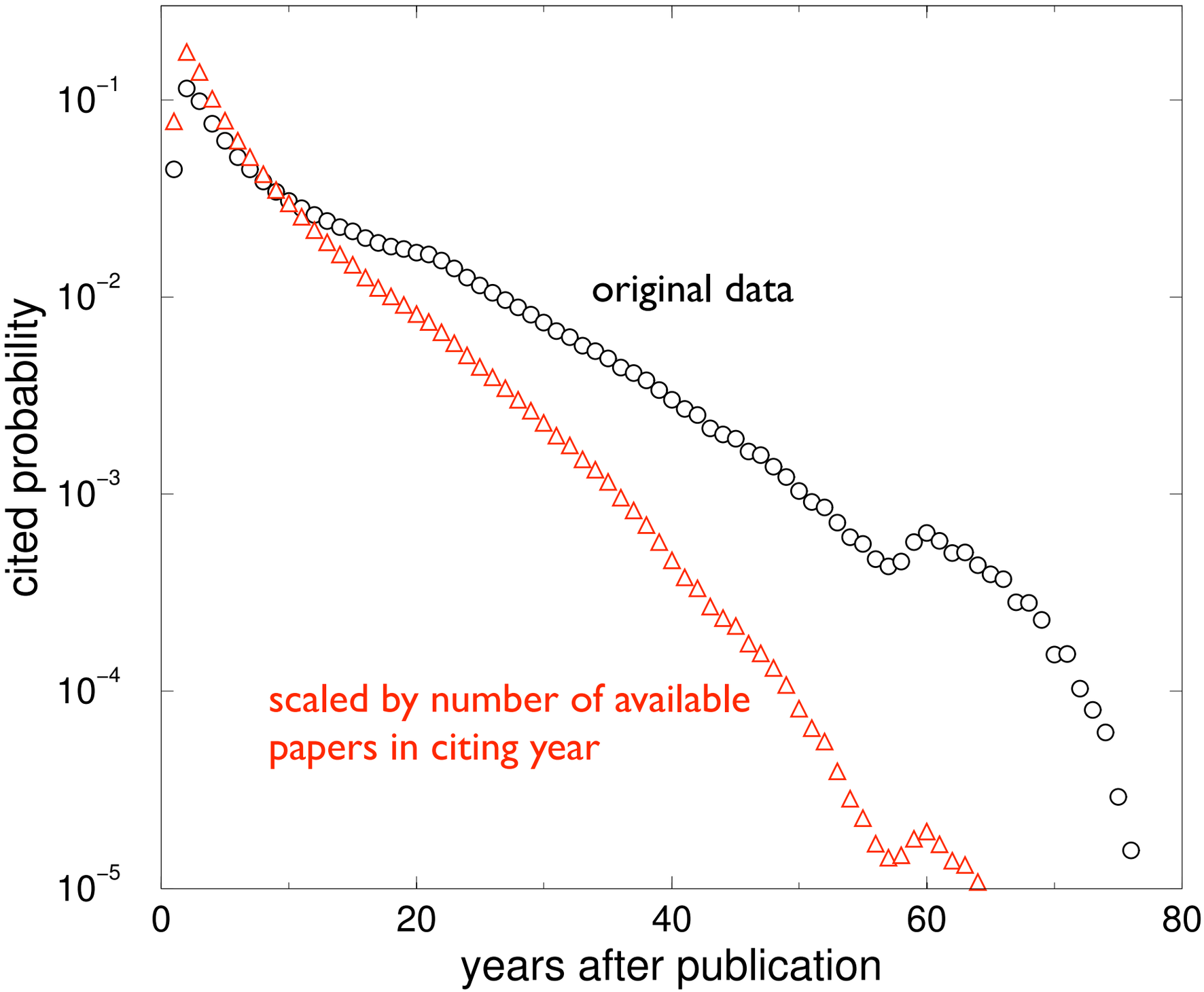


Scaled Citing History

Nakamoto (1988)



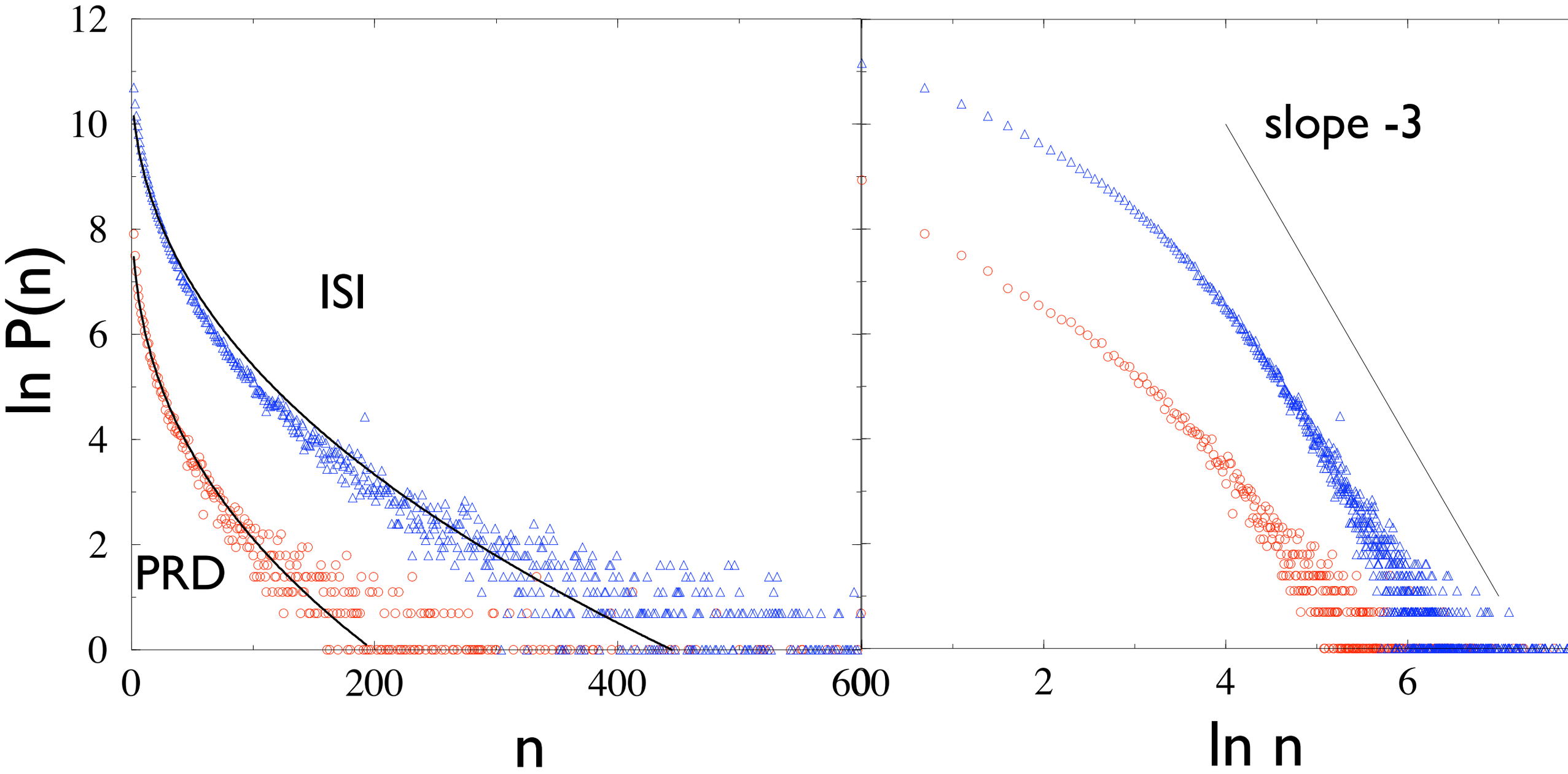
Scaled Cited History



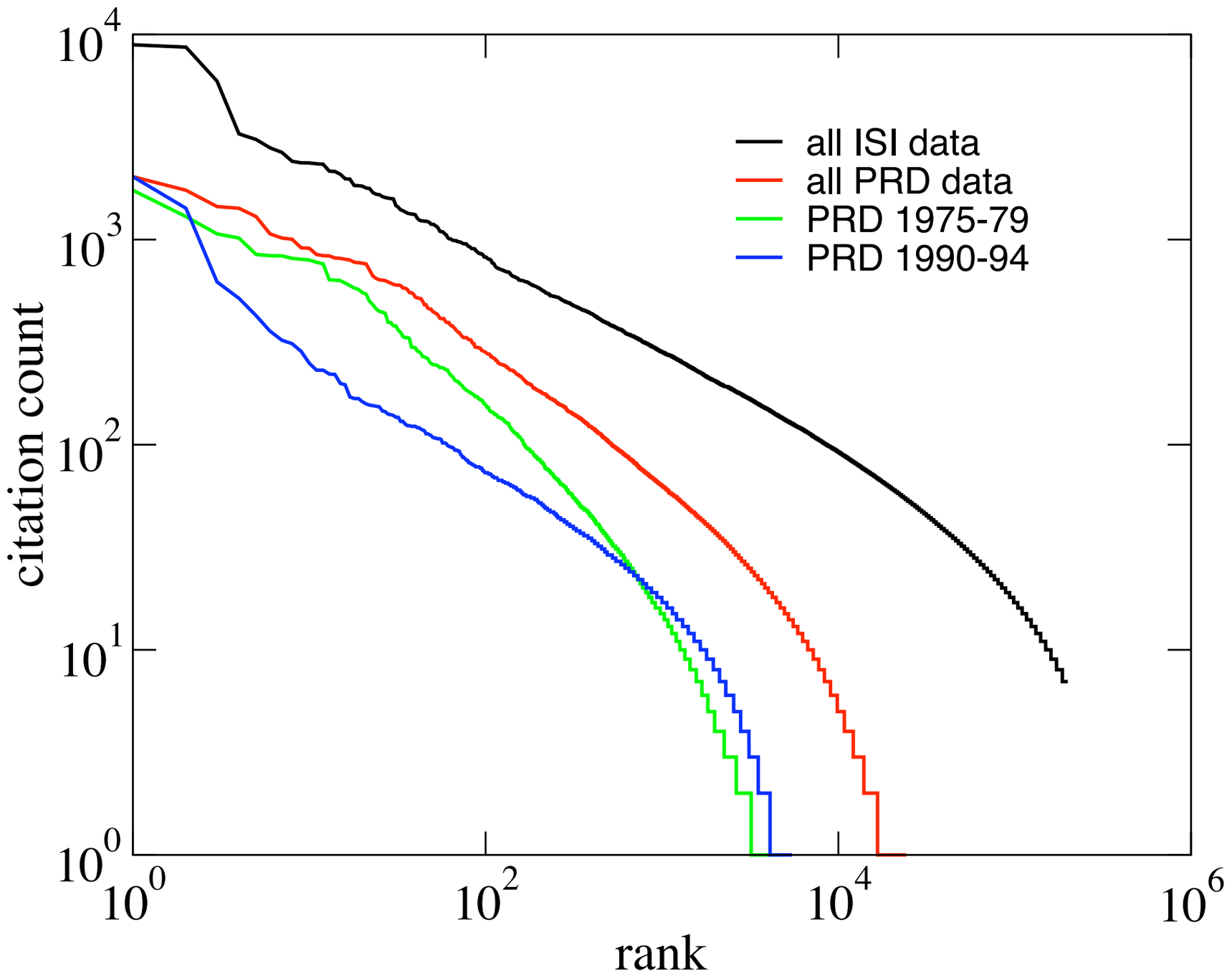
The Distribution of Citations

Lotka (1926)
Shockley (1957)
Price (1972)

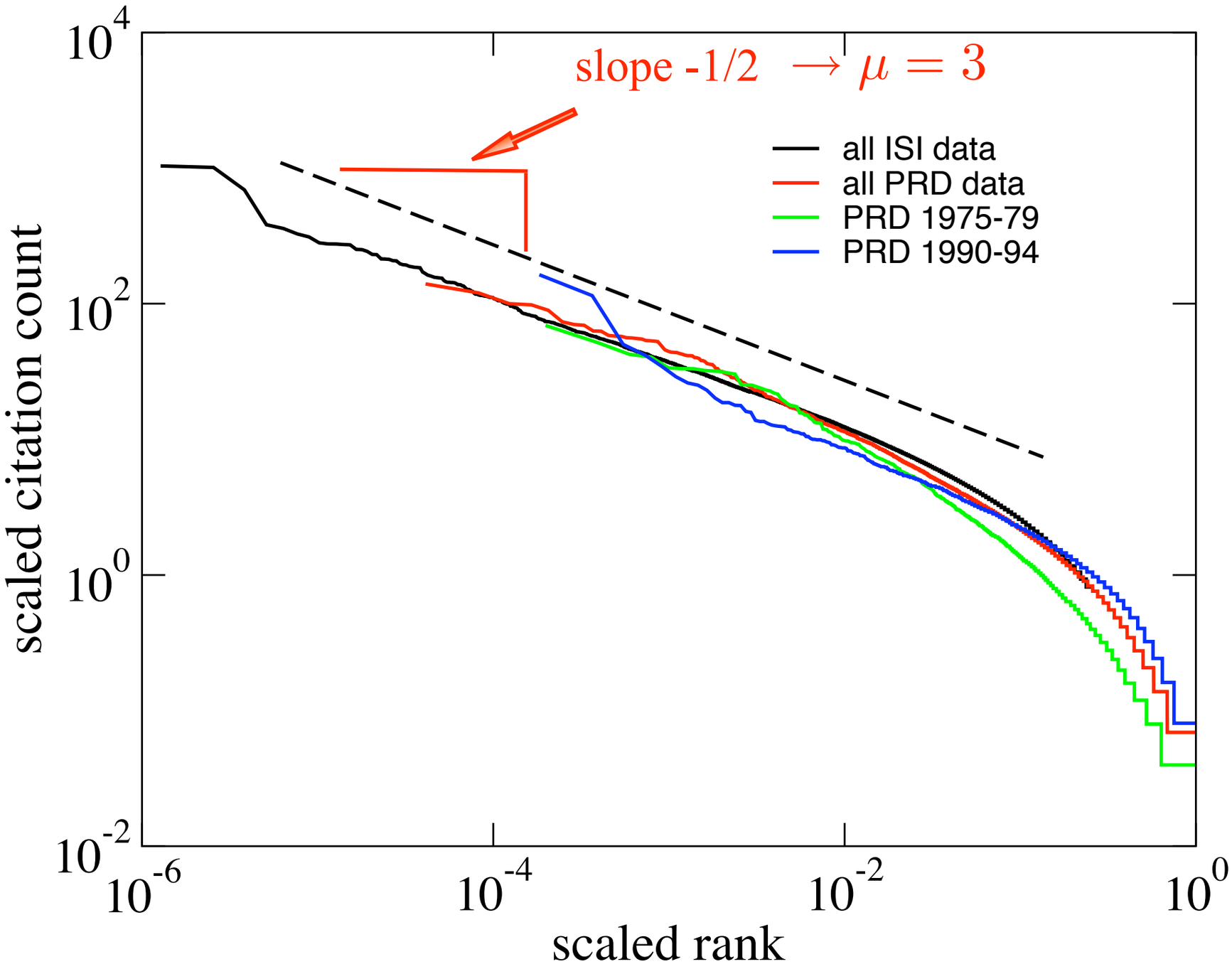
Laherrere & Sornette (1998)
SR (1998)



Zipf plot of citation data

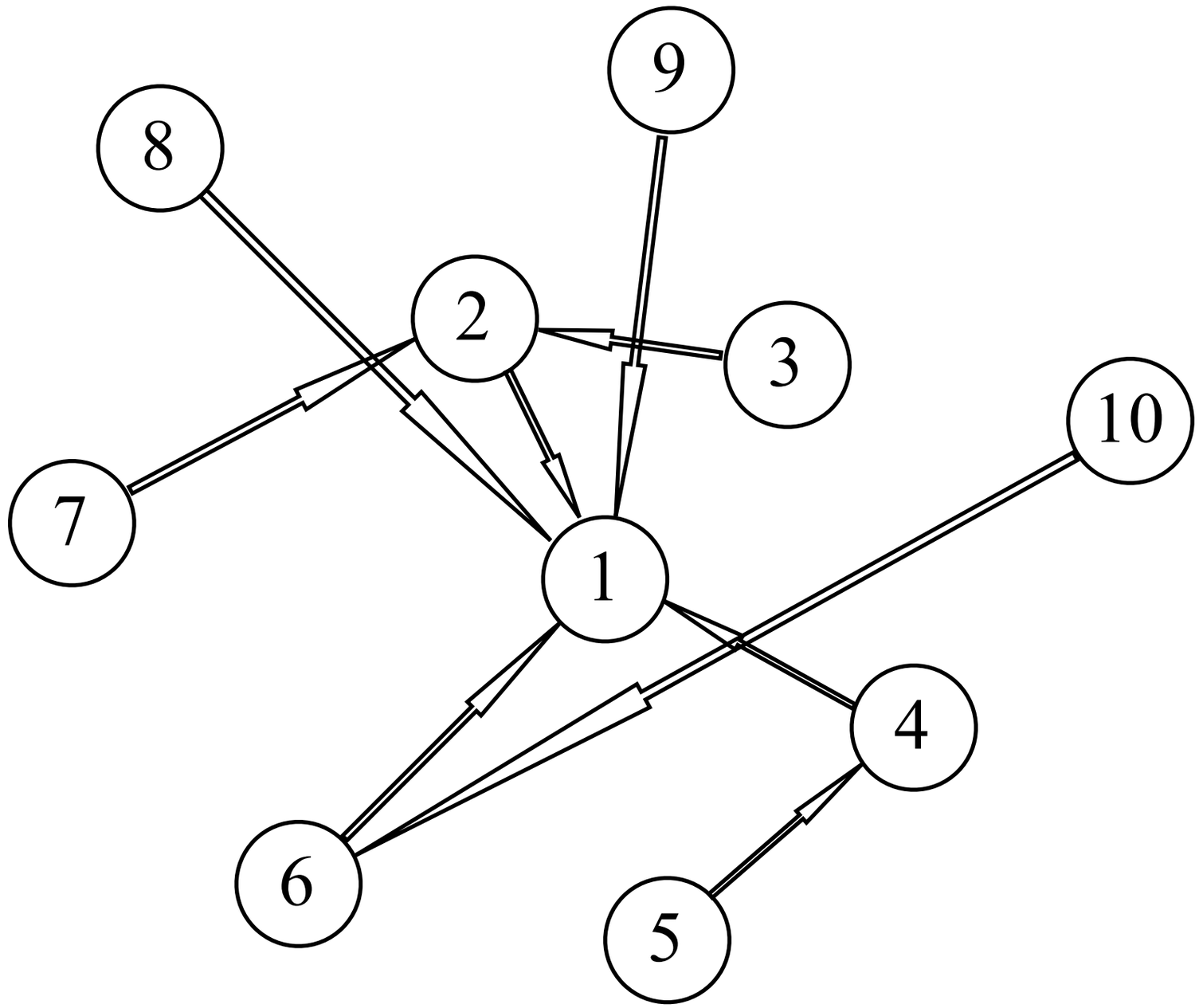


Scaled Zipf plot of citation data



Growing network model for citations

Simon (1955)
Barabasi & Albert (1999)



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

Master equation approach

KRL (2000)
Dorgovtsev & Mendes (2000)

Basic observable: N_k , the number of nodes with k links

the degree distribution

Master Equation:

$$\frac{dN_k}{dN} = \frac{A_{k-1}N_{k-1} - A_k N_k}{A} + \delta_{k,1} \qquad A = \text{total rate}$$

Attachment Rate: $A_k \sim k^\gamma$

Total Rate: $$A = A(N) = \sum_{j=1}^{\infty} A_j N_j = \sum_{j=1}^{\infty} j^\gamma N_j \equiv M_\gamma(N)$$

Basic Ansatz:

$$A(N) = \sum_j j^\gamma \, N_j \quad \propto \mu(\gamma) N \quad \text{for } 0 \leq \gamma \leq 1$$

$$N_k(N) \equiv N \; n_k$$

Converts the rate eqns. to linear recursions

$$\frac{dN_k}{dN} = \frac{A_{k-1} N_{k-1} - A_k N_k}{A} + \delta_{k,1}$$

$$\Rightarrow \quad n_k = \frac{A_{k-1} n_{k-1} - A_k n_k}{\mu} + \delta_{k,1}$$

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

Asymptotics for $A_k \sim k^\gamma$

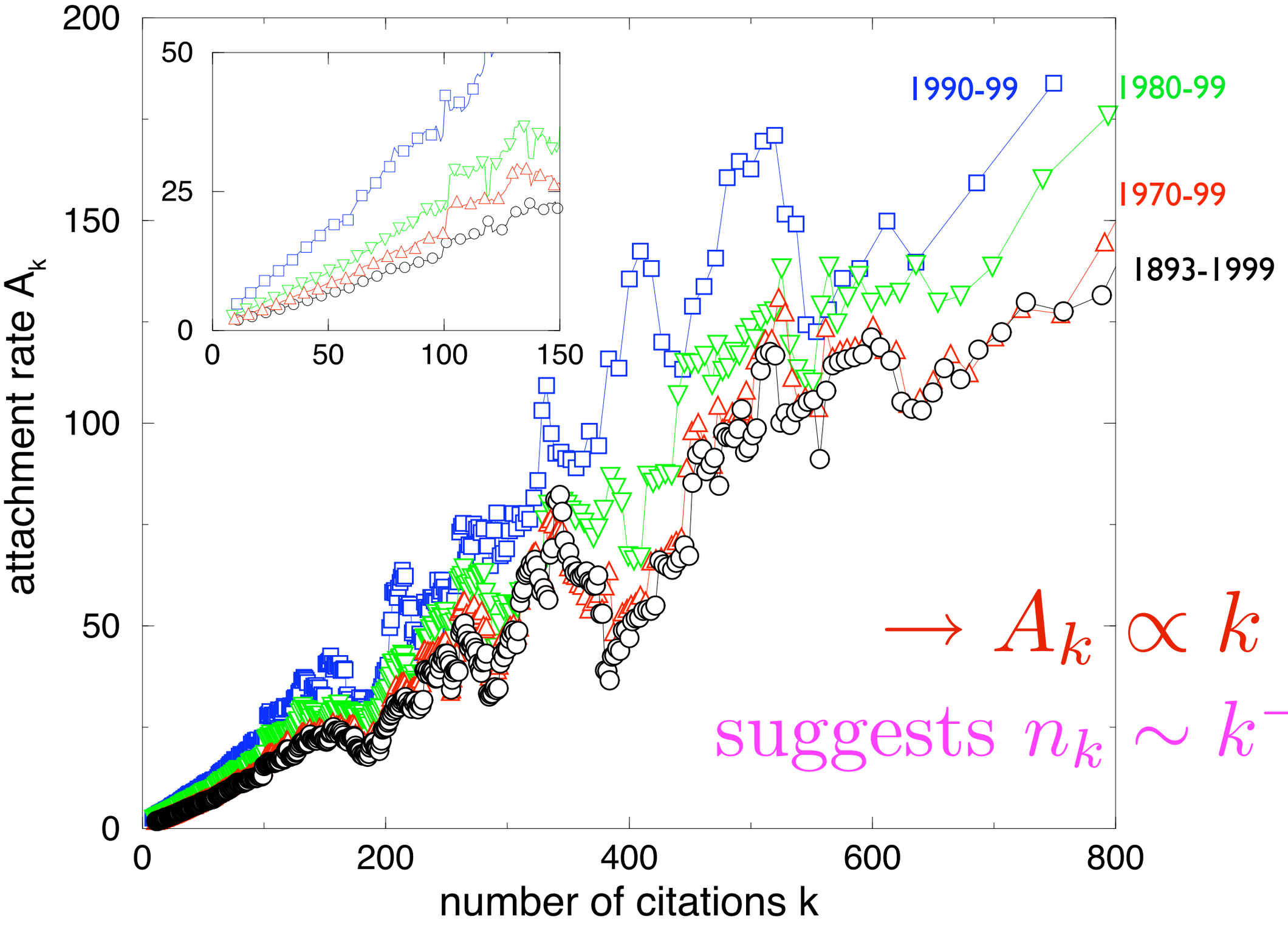
$$n_k \sim \begin{cases} k^{-\gamma} \exp\left[-\mu\left(\frac{k^{1-\gamma} - 2^{1-\gamma}}{1-\gamma}\right)\right] & 0 \le \gamma < 1 \\ k^{-\nu},\ \nu > 2 & \gamma = 1 \\ \text{best seller} & 1 < \gamma \le 2 \\ \text{bible} & \gamma > 2 \end{cases}$$

Important: $n_k \sim k^{-3}$ only for $A_k = k$

Example: If $A_k = k + \lambda$, then $n_k \sim k^{-(3+\lambda)}$ $(\lambda > -1)$

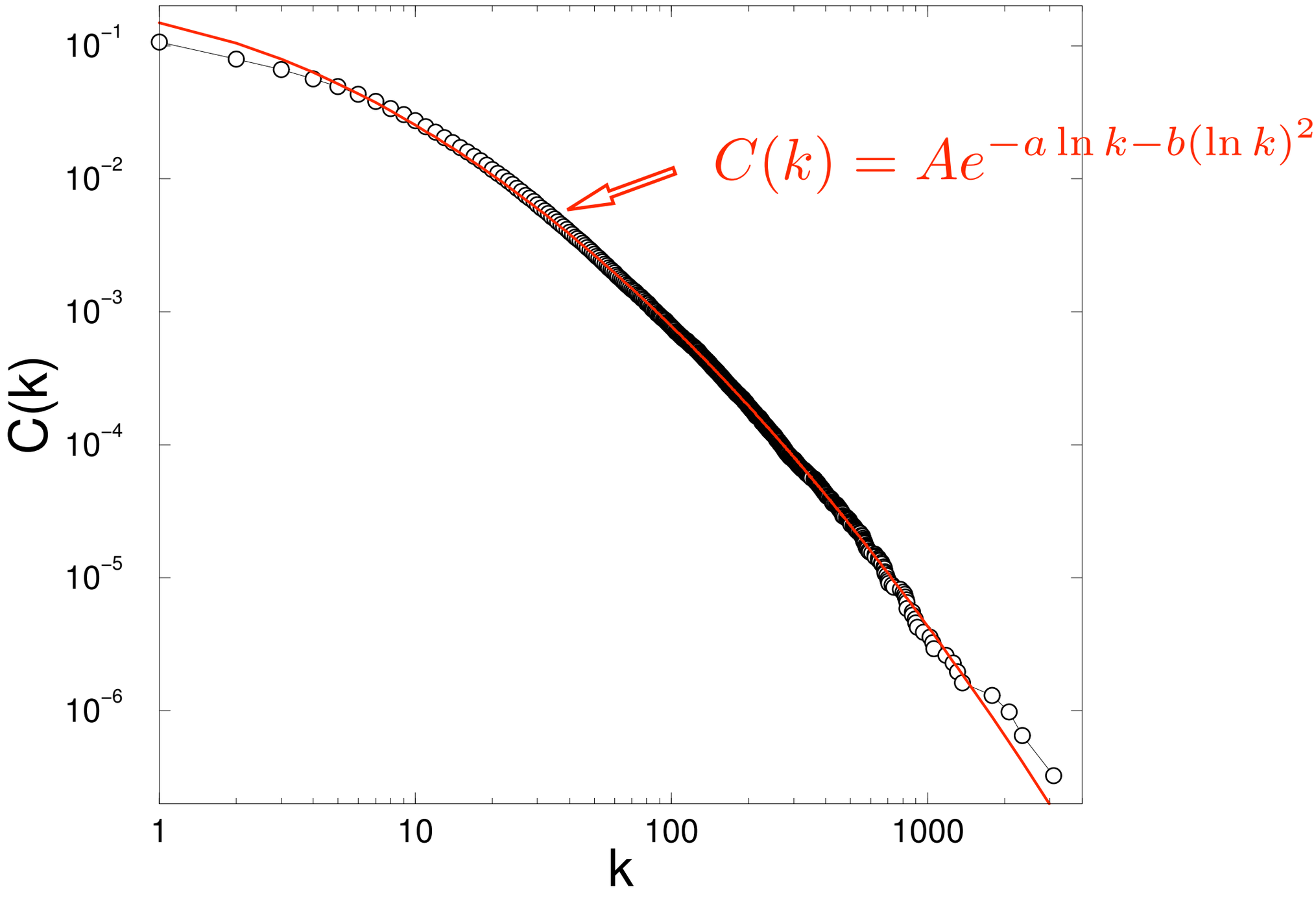
Attachment rate for PR publications

Jeong et al (2003)
SR (2004)



but... a power law is not the whole story

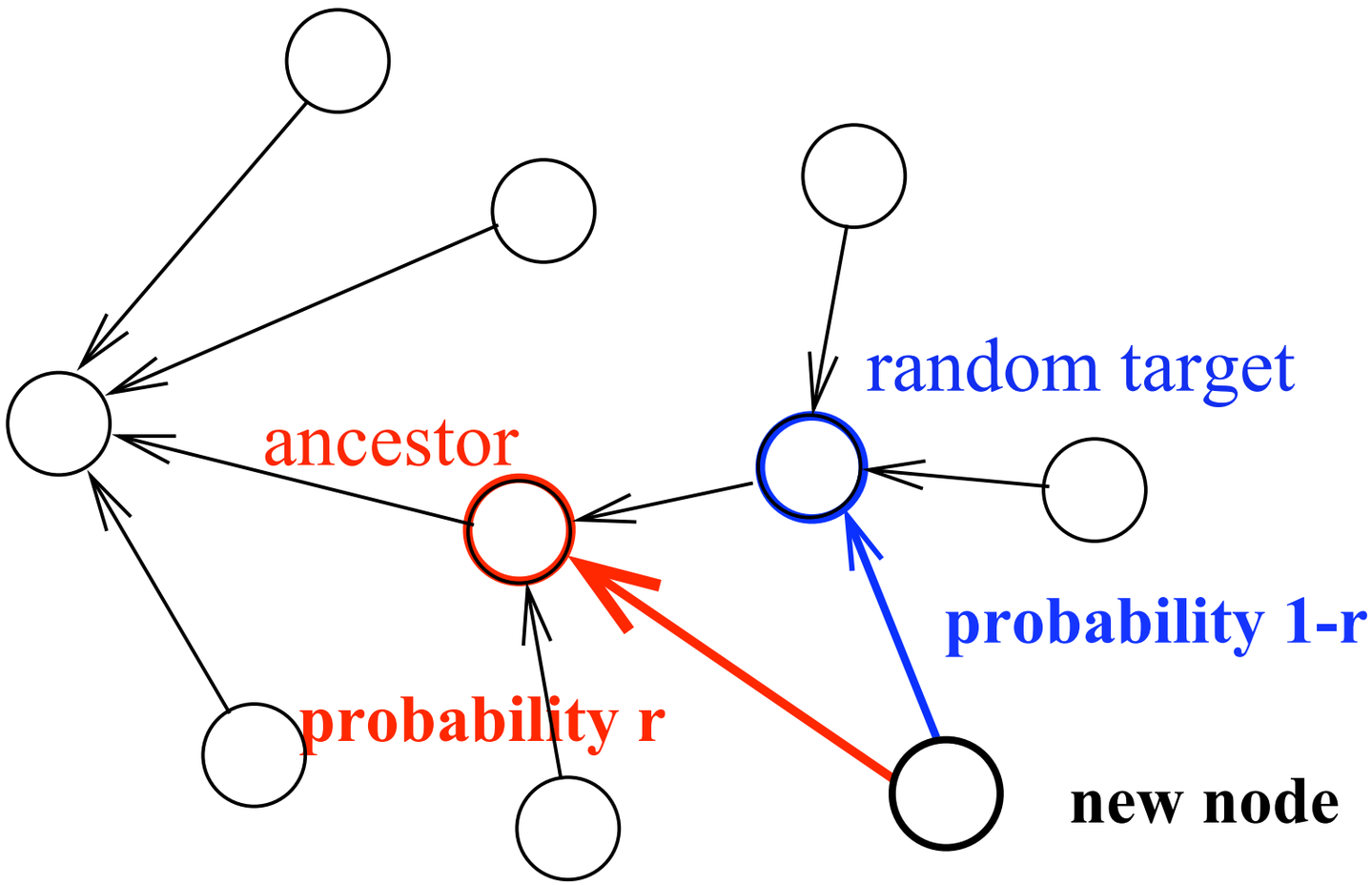
the cumulative citation distribution



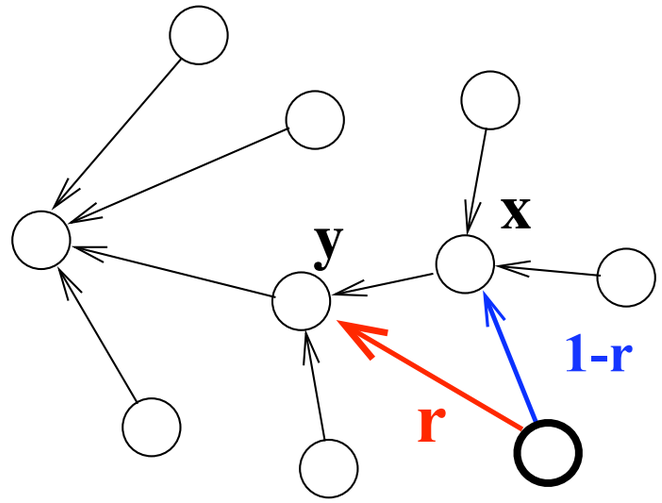
Random attachment + Redirection

Kleinberg et al (1999)
KR (2001)

initial
network



Master equations:



$$\frac{dN_k}{dN} = \frac{1-r}{M_0}\left[N_{k-1} - N_k\right]$$

$$+\frac{r}{M_0}\left[(k-2)N_{k-1} - (k-1)N_k\right] \quad + \delta_{k1}$$

$$= \frac{r}{M_0}\left\{\left[(k-1) + \frac{1}{r} - 2\right] N_{k-1} - \left[k + \frac{1}{r} - 2\right] N_k\right\} + \delta_{k1}$$

$\Rightarrow$ *shifted* linear attachment rate:

$$A_k = k + \left(\frac{1}{r} - 2\right)$$

$$\equiv k + \lambda$$

local rule produces preferential attachment!

$$\text{substitute into } n_k = \frac{\mu}{A_k}\prod_{j=1}^{k}\left(1 + \frac{\mu}{A_j}\right)^{-1} \sim k^{-(3+\lambda)} \quad (-1 < \lambda < \infty)$$

"Densifying" networks

Broder et al (2000)
Broido et al (2002)
Donato et al (2004)

Internet data:

year	1997	1998	1999	2000	2001
# AS	3060	4318	6107	9116	12155
AS links	5302	7874	12037	18196	25179
links/nodes	1.73	1.82	1.97	2.00	2.07

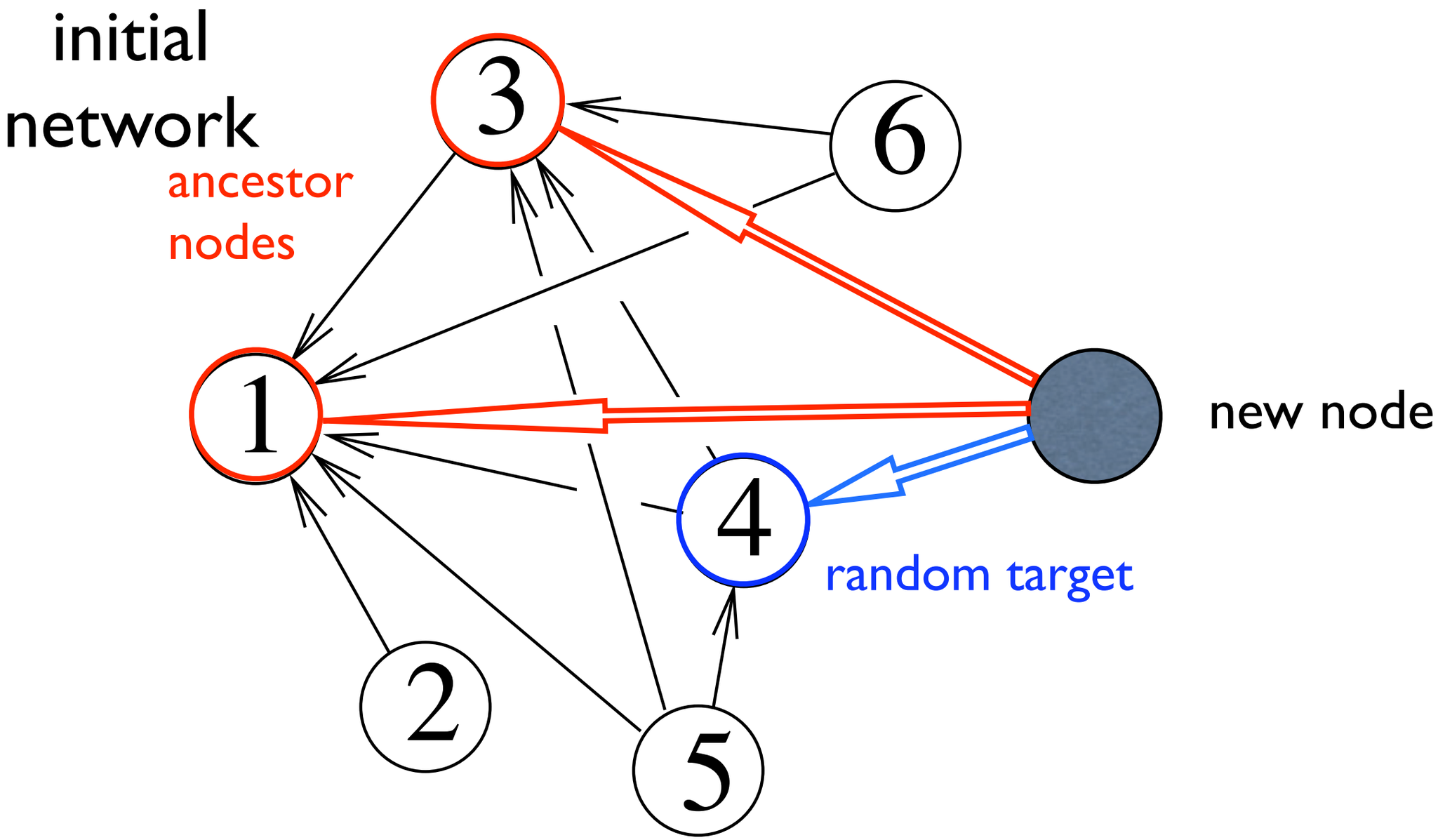
Broido et al (2002)

ratio of links to nodes is growing slowly with time

Random attachment + copying

KR (2004)

a lazy person's approach to references

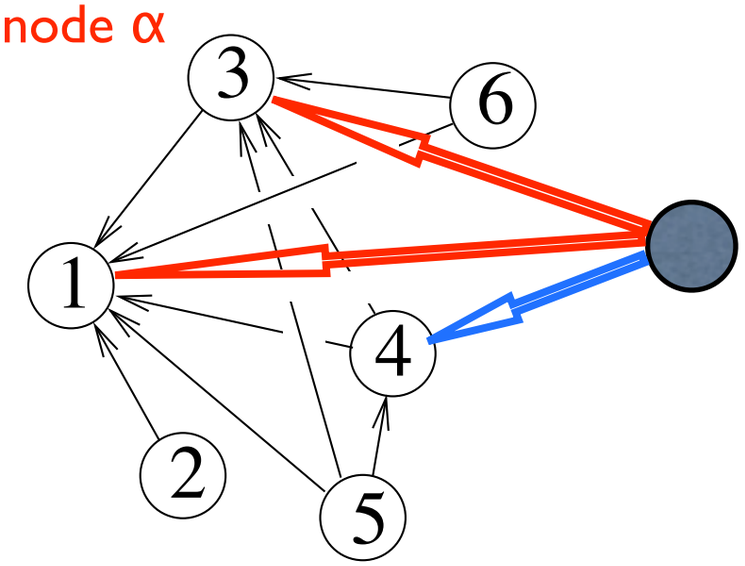


Mean Number of Links L(N)

Evolution equation:

$$L(N+1) = L(N) + \frac{1}{N}\sum_{\alpha}(1+j_\alpha)$$

$$= L(N) + 1 + \frac{L(N)}{N}$$

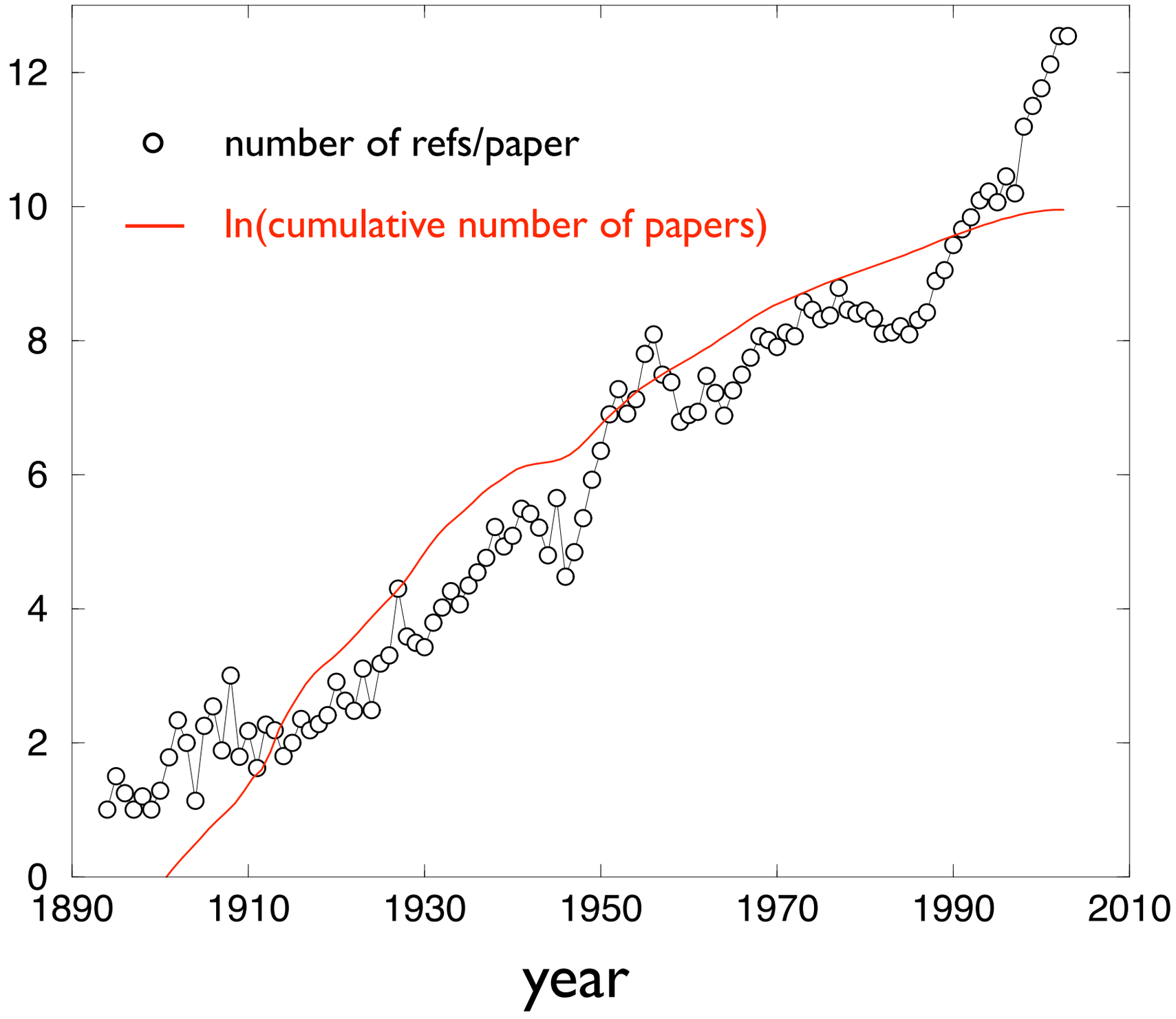


Solution:

$$L(N) = N(H_N - 1)$$

$$= N\ln N - N(1-\gamma) + \frac{1}{2} - \frac{1}{12N} + \cdots$$

Comparison with PR reference data



Summary & Outlook

Large-scale citation analysis helps motivate and test current theories of growing networks

Master equations -- a powerful way to understand many geometric properties of networks

For the future:

Deeper analysis of citation data:

contextual information, specialization

Biological networks:

giant fluctuations, infinite-order percolation

Social networks

frustration, cliques