

A Physicist's Perspective on Citation Analysis

*Impact and Productivity Measurements in a Changing Research Environment
Washington, D.C., October 27, 2010*

P. Chen (BU), S. Maslov (BNL), H. Xie (BNL)

A Physicist's Perspective on Citation Analysis

*Impact and Productivity Measurements in a Changing Research Environment
Washington, D.C., October 27, 2010*

P. Chen (BU), S. Maslov (BNL), H. Xie (BNL)

Observations about scientific citations:

amusing facts/idle gossip

evolution is governed by linear preferential attachment

Google page rank analysis

hidden “gems”

h-index analysis

danger of reliance on h

predictive value of outliers

Summary & Outlook

Phys Rev Citation Data (as of July 07)

Phys Rev Citation Data (as of July 07)

362,886 papers, 4,370,893 cites

$\langle \# \text{ cites} \rangle = 12.04$, $\langle \text{cite age} \rangle = 7.82$

Phys Rev Citation Data (as of July 07)

362,886 papers, 4,370,893 cites

$\langle \# \text{ cites} \rangle = 12.04$, $\langle \text{cite age} \rangle = 7.82$

N.B.: *Internal* citations only; undercount by factor of 3-5.
(for highly cited HEP papers; SPIRES)

Phys Rev Citation Data (as of July 07)

362,886 papers, 4,370,893 cites

$\langle \# \text{ cites} \rangle = 12.04$, $\langle \text{cite age} \rangle = 7.82$

N.B.: *Internal* citations only; undercount by factor of 3-5.
(for highly cited HEP papers; SPIRES)

32 papers with > 1000 citations

Phys Rev Citation Data (as of July 07)

362,886 papers, 4,370,893 cites

$\langle \# \text{ cites} \rangle = 12.04$, $\langle \text{cite age} \rangle = 7.82$

N.B.: *Internal* citations only; undercount by factor of 3-5.
(for highly cited HEP papers; SPIRES)

32 papers with > 1000 citations

158 papers with > 500 citations

Phys Rev Citation Data (as of July 07)

362,886 papers, 4,370,893 cites

$\langle \# \text{ cites} \rangle = 12.04$, $\langle \text{cite age} \rangle = 7.82$

N.B.: *Internal* citations only; undercount by factor of 3-5.
(for highly cited HEP papers; SPIRES)

32 papers with > 1000 citations

158 papers with > 500 citations

426 papers with > 300 citations

Phys Rev Citation Data (as of July 07)

362,886 papers, 4,370,893 cites

$\langle \# \text{ cites} \rangle = 12.04$, $\langle \text{cite age} \rangle = 7.82$

N.B.: *Internal* citations only; undercount by factor of 3-5.
(for highly cited HEP papers; SPIRES)

32 papers with > 1000 citations

158 papers with > 500 citations

426 papers with > 300 citations

3693 papers with > 100 citations

Phys Rev Citation Data (as of July 07)

362,886 papers, 4,370,893 cites

$\langle \# \text{ cites} \rangle = 12.04$, $\langle \text{cite age} \rangle = 7.82$

N.B.: *Internal* citations only; undercount by factor of 3-5.
(for highly cited HEP papers; SPIRES)

32 papers with > 1000 citations

158 papers with > 500 citations

426 papers with > 300 citations

3693 papers with > 100 citations

12519 papers with > 50 citations

Phys Rev Citation Data (as of July 07)

362,886 papers, 4,370,893 cites

$\langle \# \text{ cites} \rangle = 12.04$, $\langle \text{cite age} \rangle = 7.82$

N.B.: *Internal* citations only; undercount by factor of 3-5.
(for highly cited HEP papers; SPIRES)

32 papers with > 1000 citations

158 papers with > 500 citations

426 papers with > 300 citations

3693 papers with > 100 citations

12519 papers with > 50 citations

244281 papers with < 10 citations

Phys Rev Citation Data (as of July 07)

362,886 papers, 4,370,893 cites

$\langle \# \text{ cites} \rangle = 12.04$, $\langle \text{cite age} \rangle = 7.82$

N.B.: *Internal* citations only; undercount by factor of 3-5.
(for highly cited HEP papers; SPIRES)

32 papers with > 1000 citations

158 papers with > 500 citations

426 papers with > 300 citations

3693 papers with > 100 citations

12519 papers with > 50 citations

244281 papers with < 10 citations

51487 papers with 1 citation

Phys Rev papers with > 1000 cites (July 2003)

Phys Rev papers with >1000 cites (July 2003)

		article		cites	⟨age⟩	title	author(s)
1	PR	140	A1133 (1965)	3227	26.7	Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136	B864 (1965)	2460	28.7	Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn

Phys Rev papers with >1000 cites (July 2003)

	article		cites		⟨age⟩	title	author(s)
1	PR	140 A1133 (1965)	3227	26.7		Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136 B864 (1965)	2460	28.7		Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23 5048 (1981)	2079	14.4		Self-Interaction Correction to..	J. P. Perdew & A. Zunger

Phys Rev papers with >1000 cites (July 2003)

	article			cites <age>		title	author(s)
1	PR	140	A1133 (1965)	3227	26.7	Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136	B864 (1965)	2460	28.7	Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23	5048 (1981)	2079	14.4	Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL	45	566 (1980)	1781	15.4	Ground State of the Electron..	D. M. Ceperley & B. J. Alder

Phys Rev papers with >1000 cites (July 2003)

	article			cites \langle age \rangle		title	author(s)
1	PR	140	A1133 (1965)	3227	26.7	Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136	B864 (1965)	2460	28.7	Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23	5048 (1981)	2079	14.4	Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL	45	566 (1980)	1781	15.4	Ground State of the Electron..	D. M. Ceperley & B. J. Alder
5	PR	108	1175 (1957)	1364	20.2	Theory of Superconductivity	Bardeen, Cooper, Schrieffer

Phys Rev papers with >1000 cites (July 2003)

	article			cites <age>		title	author(s)
1	PR	140	A1133 (1965)	3227	26.7	Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136	B864 (1965)	2460	28.7	Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23	5048 (1981)	2079	14.4	Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL	45	566 (1980)	1781	15.4	Ground State of the Electron..	D. M. Ceperley & B. J. Alder
5	PR	108	1175 (1957)	1364	20.2	Theory of Superconductivity	Bardeen, Cooper, Schrieffer
6	PRL	19	1264 (1967)	1306	15.5	A Model of Leptons	S. Weinberg

Phys Rev papers with >1000 cites (July 2003)

	article		cites		⟨age⟩		title	author(s)
1	PR	140 A1133 (1965)	3227	26.7	Self Consistent Equations..	W. Kohn & L. J. Sham		
2	PR	136 B864 (1965)	2460	28.7	Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn		
3	PRB	23 5048 (1981)	2079	14.4	Self-Interaction Correction to..	J. P. Perdew & A. Zunger		
4	PRL	45 566 (1980)	1781	15.4	Ground State of the Electron..	D. M. Ceperley & B. J. Alder		
5	PR	108 1175 (1957)	1364	20.2	Theory of Superconductivity	Bardeen, Cooper, Schrieffer		
6	PRL	19 1264 (1967)	1306	15.5	A Model of Leptons	S. Weinberg		
7	PRB	12 3060 (1975)	1259	18.4	Linear Methods in Band Theory	O. K. Andersen		
8	PR	124 1866 (1961)	1178	28.0	Effects of Configuration..	U. Fano		
9	RMP	57 287 (1985)	1055	9.2	Disordered Electronic Systems	P. A. Lee & T. V. Ramakrishnan		
10	RMP	54 437 (1982)	1045	10.8	Electronic Properties of..	T. Ando, A. B. Fowler, & F. Stern		
11	PRB	13 5188 (1976)	1023	20.8	Special Points for Brillouin..	H. J. Monkhorst & J. D. Pack		

Phys Rev papers with >1000 cites (July 2003)

	article		cites		⟨age⟩	title	author(s)
1	PR	140 A1133 (1965)	3227	26.7		Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136 B864 (1965)	2460	28.7		Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23 5048 (1981)	2079	14.4		Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL	45 566 (1980)	1781	15.4		Ground State of the Electron..	D. M. Ceperley & B. J. Alder
5	PR	108 1175 (1957)	1364	20.2		Theory of Superconductivity	Bardeen, Cooper, Schrieffer
6	PRL	19 1264 (1967)	1306	15.5		A Model of Leptons	S. Weinberg
7	PRB	12 3060 (1975)	1259	18.4		Linear Methods in Band Theory	O. K. Andersen
8	PR	124 1866 (1961)	1178	28.0		Effects of Configuration..	U. Fano
9	RMP	57 287 (1985)	1055	9.2		Disordered Electronic Systems	P. A. Lee & T. V. Ramakrishnan
10	RMP	54 437 (1982)	1045	10.8		Electronic Properties of..	T. Ando, A. B. Fowler, & F. Stern
11	PRB	13 5188 (1976)	1023	20.8		Special Points for Brillouin..	H. J. Monkhorst & J. D. Pack

top 10 cited Phys Rev papers (July 2007)

Phys Rev papers with >1000 cites (July 2003)

	article			cites <age>		title	author(s)
1	PR	140	A1133 (1965)	3227	26.7	Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136	B864 (1965)	2460	28.7	Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23	5048 (1981)	2079	14.4	Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL	45	566 (1980)	1781	15.4	Ground State of the Electron..	D. M. Ceperley & B. J. Alder
5	PR	108	1175 (1957)	1364	20.2	Theory of Superconductivity	Bardeen, Cooper, Schrieffer
6	PRL	19	1264 (1967)	1306	15.5	A Model of Leptons	S. Weinberg
7	PRB	12	3060 (1975)	1259	18.4	Linear Methods in Band Theory	O. K. Andersen
8	PR	124	1866 (1961)	1178	28.0	Effects of Configuration..	U. Fano
9	RMP	57	287 (1985)	1055	9.2	Disordered Electronic Systems	P. A. Lee & T. V. Ramakrishnan
10	RMP	54	437 (1982)	1045	10.8	Electronic Properties of..	T. Ando, A. B. Fowler, & F. Stern
11	PRB	13	5188 (1976)	1023	20.8	Special Points for Brillouin..	H. J. Monkhorst & J. D. Pack

top 10 cited Phys Rev papers (July 2007)

	article			cites	title	author(s)
1	PR	140	A1133 (1965)	4930	Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136	B864 (1965)	3564	Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23	5048 (1981)	3007	Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL	45	566 (1980)	2514	Ground State of the Electron..	D. M. Ceperley & B. J. Alder

Phys Rev papers with >1000 cites (July 2003)

	article			cites	<age>	title	author(s)
1	PR	140	A1133 (1965)	3227	26.7	Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136	B864 (1965)	2460	28.7	Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23	5048 (1981)	2079	14.4	Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL	45	566 (1980)	1781	15.4	Ground State of the Electron..	D. M. Ceperley & B. J. Alder
5	PR	108	1175 (1957)	1364	20.2	Theory of Superconductivity	Bardeen, Cooper, Schrieffer
6	PRL	19	1264 (1967)	1306	15.5	A Model of Leptons	S. Weinberg
7	PRB	12	3060 (1975)	1259	18.4	Linear Methods in Band Theory	O. K. Andersen
8	PR	124	1866 (1961)	1178	28.0	Effects of Configuration..	U. Fano
9	RMP	57	287 (1985)	1055	9.2	Disordered Electronic Systems	P. A. Lee & T. V. Ramakrishnan
10	RMP	54	437 (1982)	1045	10.8	Electronic Properties of..	T. Ando, A. B. Fowler, & F. Stern
11	PRB	13	5188 (1976)	1023	20.8	Special Points for Brillouin..	H. J. Monkhorst & J. D. Pack

top 10 cited Phys Rev papers (July 2007)

	article			cites	title	author(s)
1	PR	140	A1133 (1965)	4930	Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136	B864 (1965)	3564	Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23	5048 (1981)	3007	Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL	45	566 (1980)	2514	Ground State of the Electron..	D. M. Ceperley & B. J. Alder
5	PRL	77	3865 (1996)	2478	Generalized Gradient Approx...	Perdew, Burke, Ernzerhof
6	PRB	13	5188 (1976)	2277	Special Points for Brillouin..	H. J. Monkhorst & J. D. Pack
7	PRB	54	11169 (1996)	1933	Efficient Iterative Schemes....	G. Kresse & J. Furthmuller
8	PRB	43	1993 (1991)	1776	Efficient Pseudopotentials for	N. Troullier & J.L. Martins
9	PRB	41	7892 (1990)	1749	Soft Self-Consistent Pseudopotentials...	D. Vanderbilt

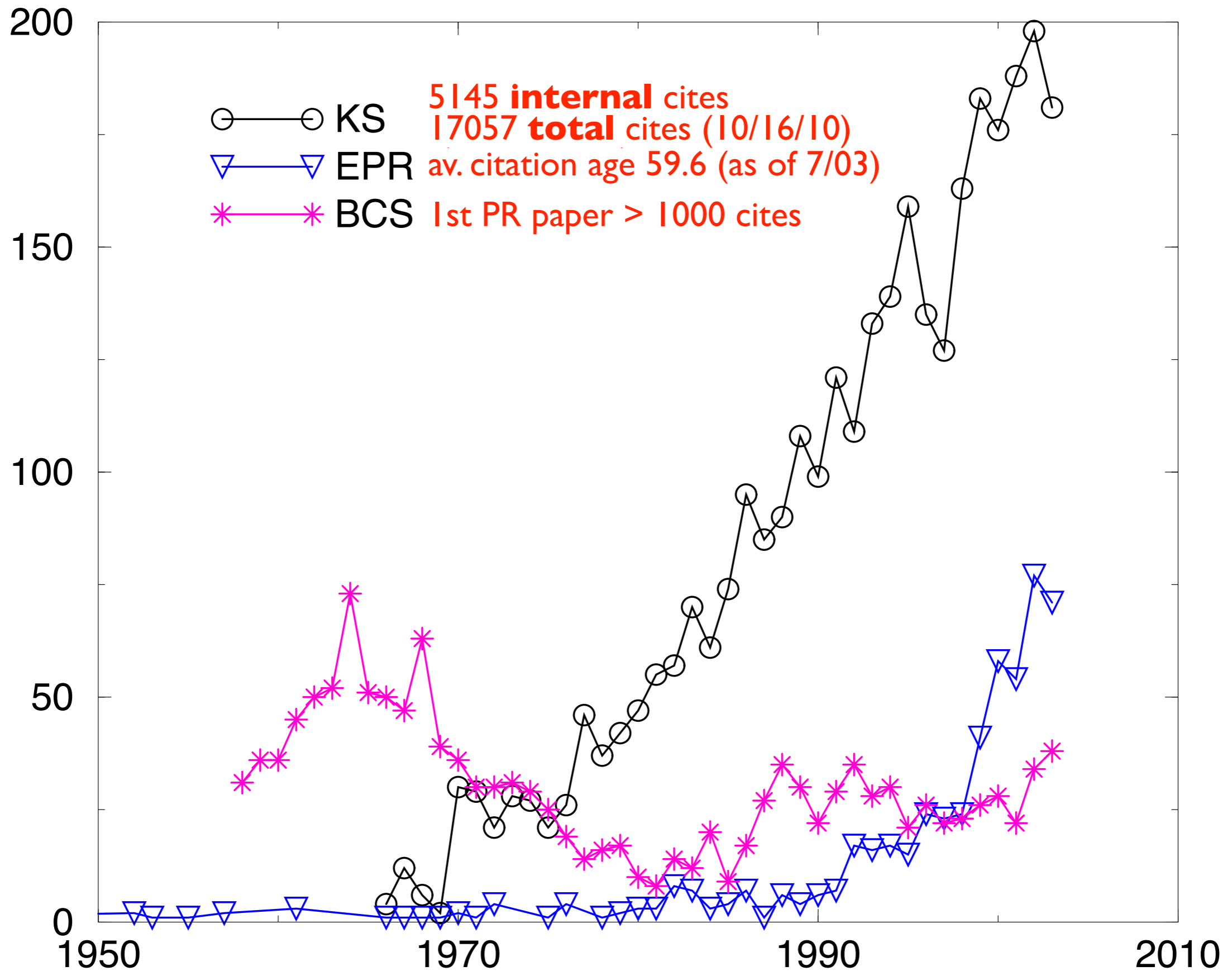
Phys Rev papers with >1000 cites (July 2003)

	article		cites		⟨age⟩	title	author(s)
1	PR	140 A1133 (1965)	3227	26.7		Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136 B864 (1965)	2460	28.7		Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23 5048 (1981)	2079	14.4		Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL	45 566 (1980)	1781	15.4		Ground State of the Electron..	D. M. Ceperley & B. J. Alder
5	PR	108 1175 (1957)	1364	20.2		Theory of Superconductivity	Bardeen, Cooper, Schrieffer
6	PRL	19 1264 (1967)	1306	15.5		A Model of Leptons	S. Weinberg
7	PRB	12 3060 (1975)	1259	18.4		Linear Methods in Band Theory	O. K. Andersen
8	PR	124 1866 (1961)	1178	28.0		Effects of Configuration..	U. Fano
9	RMP	57 287 (1985)	1055	9.2		Disordered Electronic Systems	P. A. Lee & T. V. Ramakrishnan
10	RMP	54 437 (1982)	1045	10.8		Electronic Properties of..	T. Ando, A. B. Fowler, & F. Stern
11	PRB	13 5188 (1976)	1023	20.8		Special Points for Brillouin..	H. J. Monkhorst & J. D. Pack

top 10 cited Phys Rev papers (July 2007)

	article		cites		title	author(s)
1	PR	140 A1133 (1965)	4930		Self Consistent Equations..	W. Kohn & L. J. Sham
2	PR	136 B864 (1965)	3564		Inhomogeneous Electron Gas..	P. Hohenberg & W. Kohn
3	PRB	23 5048 (1981)	3007		Self-Interaction Correction to..	J. P. Perdew & A. Zunger
4	PRL	45 566 (1980)	2514		Ground State of the Electron..	D. M. Ceperley & B. J. Alder
5	PRL	77 3865 (1996)	2478		Generalized Gradient Approx...	Perdew, Burke, Ernzerhof
6	PRB	13 5188 (1976)	2277		Special Points for Brillouin..	H. J. Monkhorst & J. D. Pack
7	PRB	54 11169 (1996)	1933		Efficient Iterative Schemes....	G. Kresse & J. Furthmuller
8	PRB	43 1993 (1991)	1776		Efficient Pseudopotentials for	N. Troullier & J.L. Martins
9	PRB	41 7892 (1990)	1749		Soft Self-Consistent Pseudopotentials...	D. Vanderbilt
10	PR	108 1175 (1957)	1650		Theory of Superconductivity	Bardeen, Cooper, Schrieffer

Citation histories of 3 classic PR papers



“Sleeping Beauties”

cites > 300
<cite age>/paper age > 3/4

“Sleeping Beauties” $\# \text{ cites} > 300$
 $\langle \text{cite age} \rangle / \text{paper age} > 3/4$
8 papers total, 5 on double exchange
 \Rightarrow colossal magnetoresistance

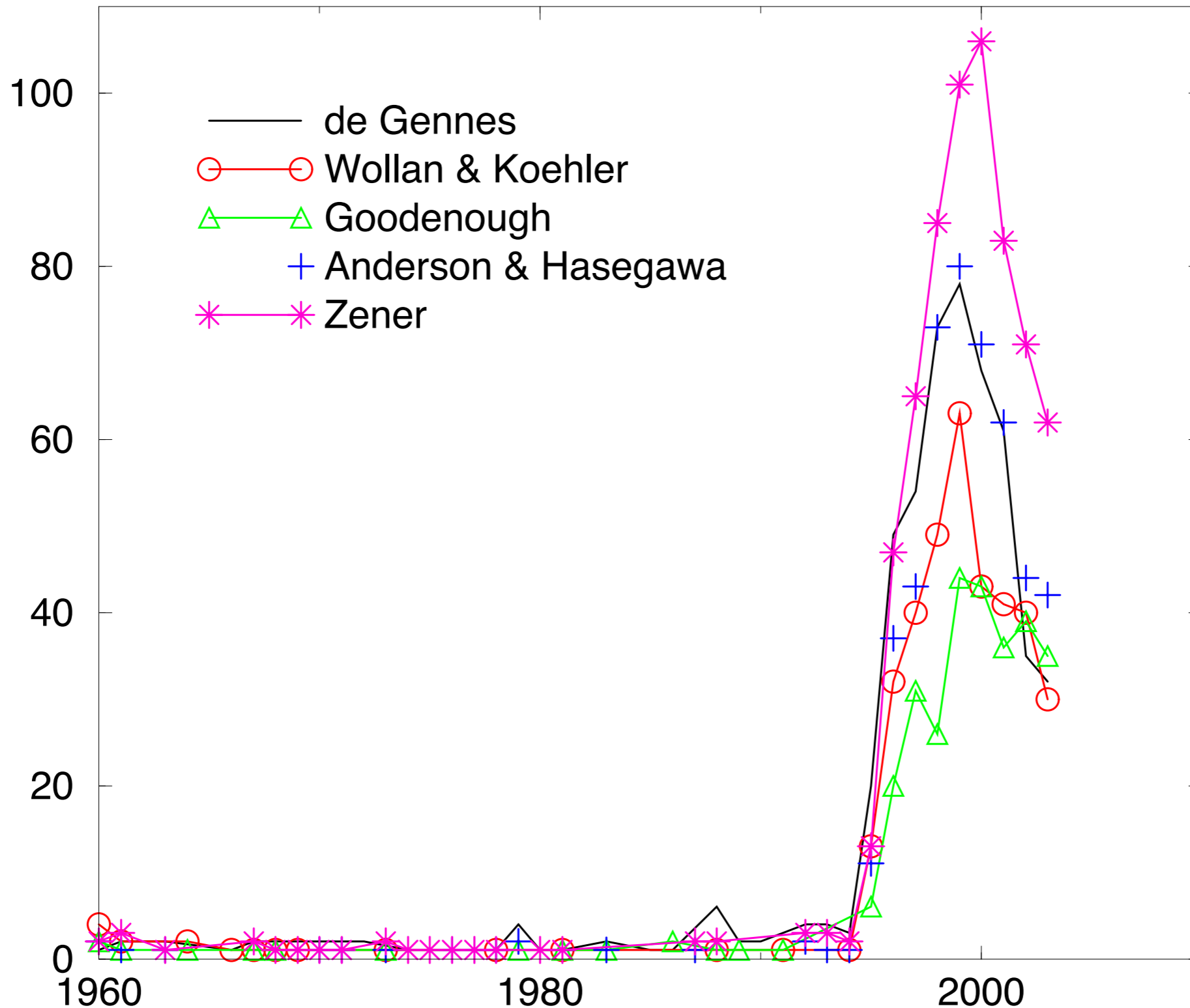
“Sleeping Beauties”

cites > 300

$\langle \text{cite age} \rangle / \text{paper age} > 3/4$

8 papers total, 5 on double exchange

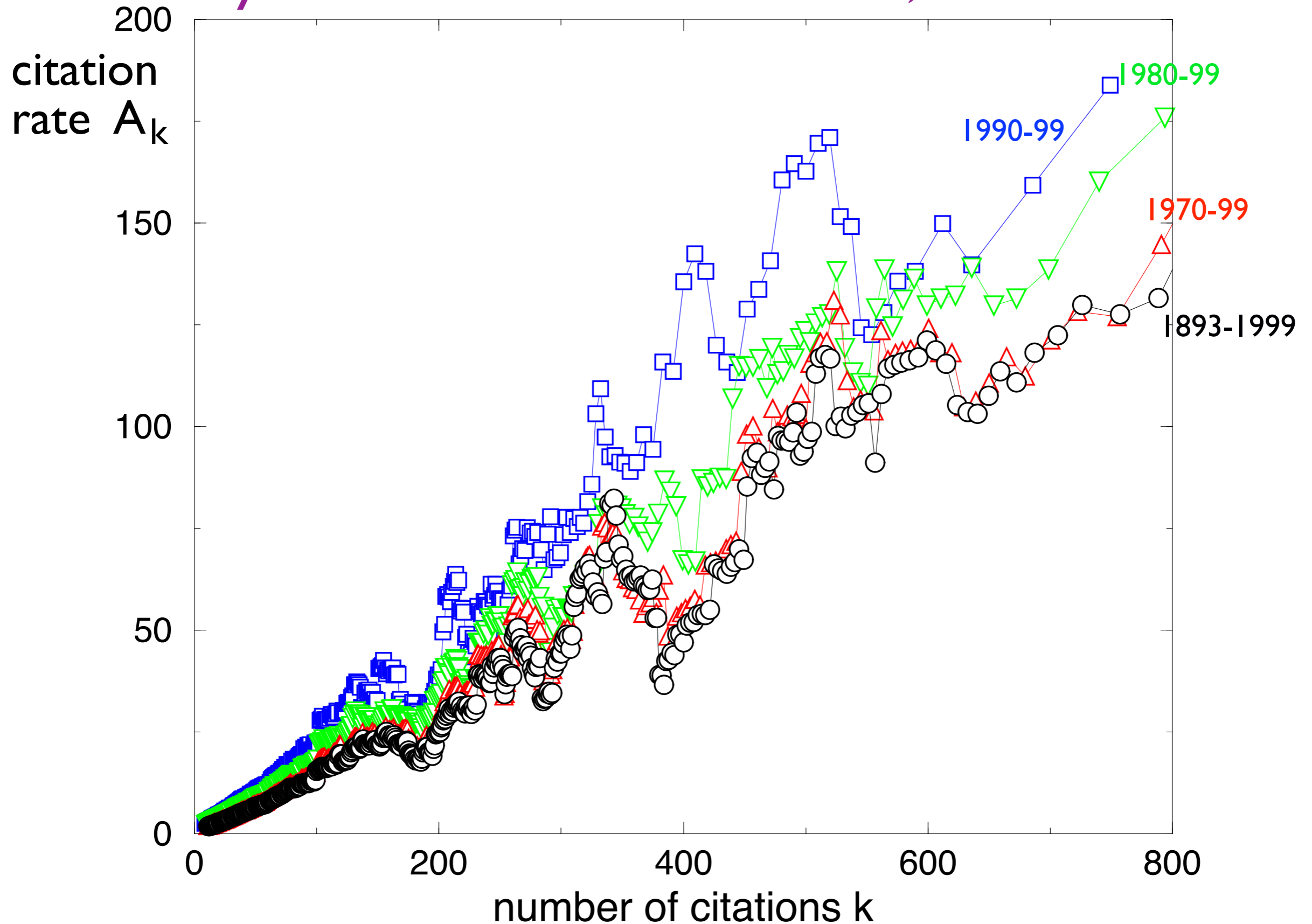
\Rightarrow colossal magnetoresistance



What is the Citation Rate?

Physics Today **58**, 49 (2005)

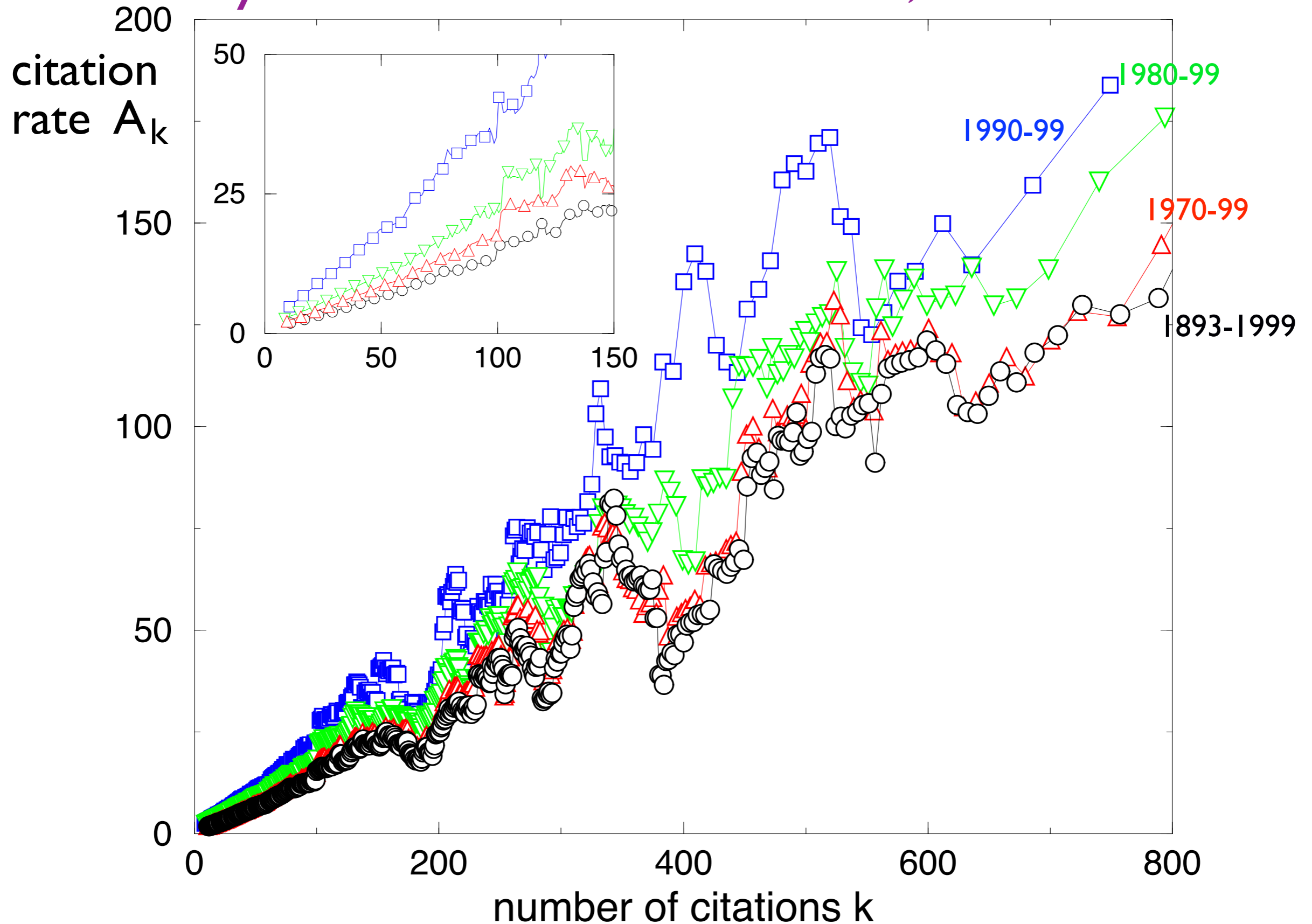
Physical Review Citation Data, 1893-2003



What is the Citation Rate?

Physics Today **58**, 49 (2005)

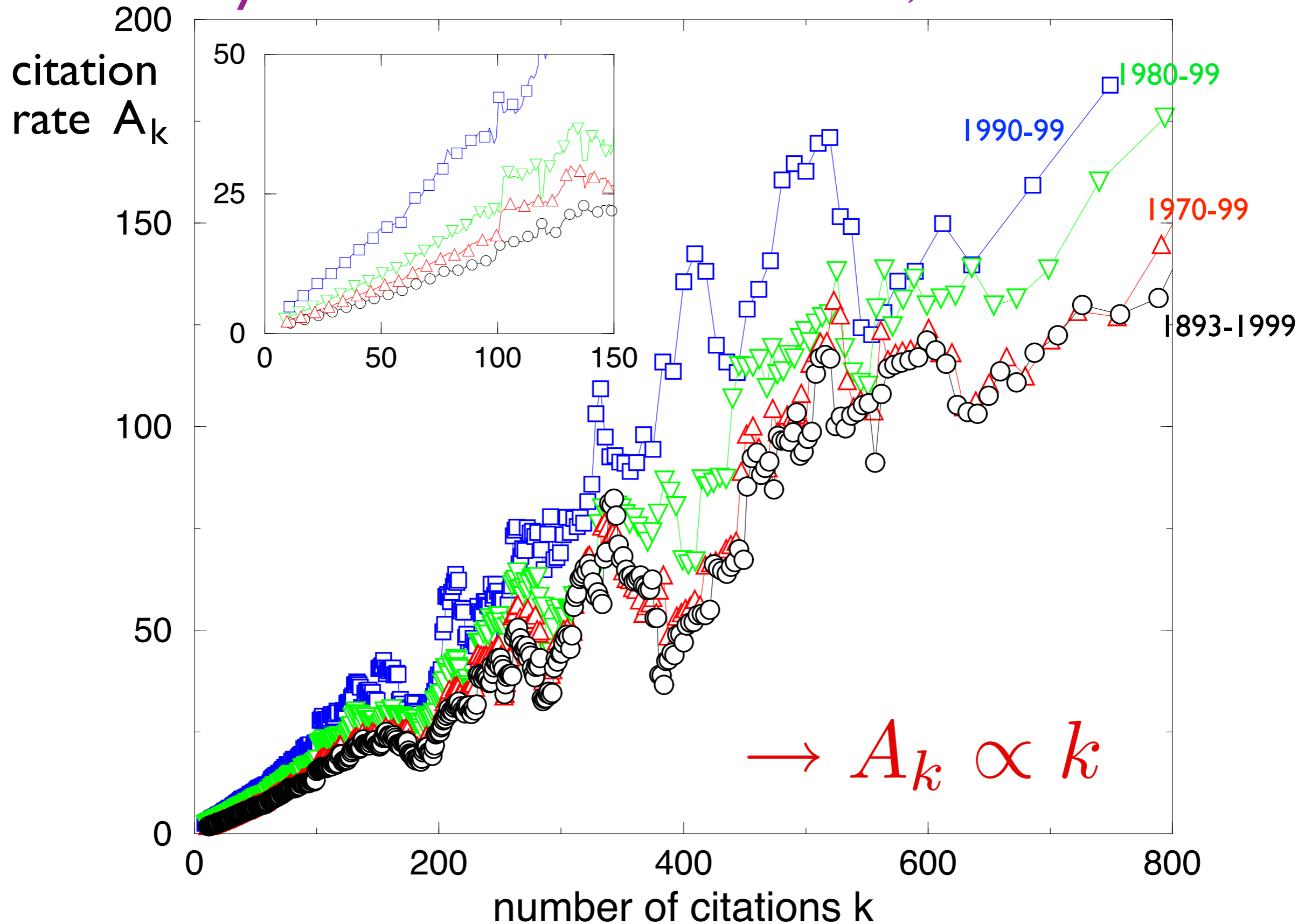
Physical Review Citation Data, 1893-2003



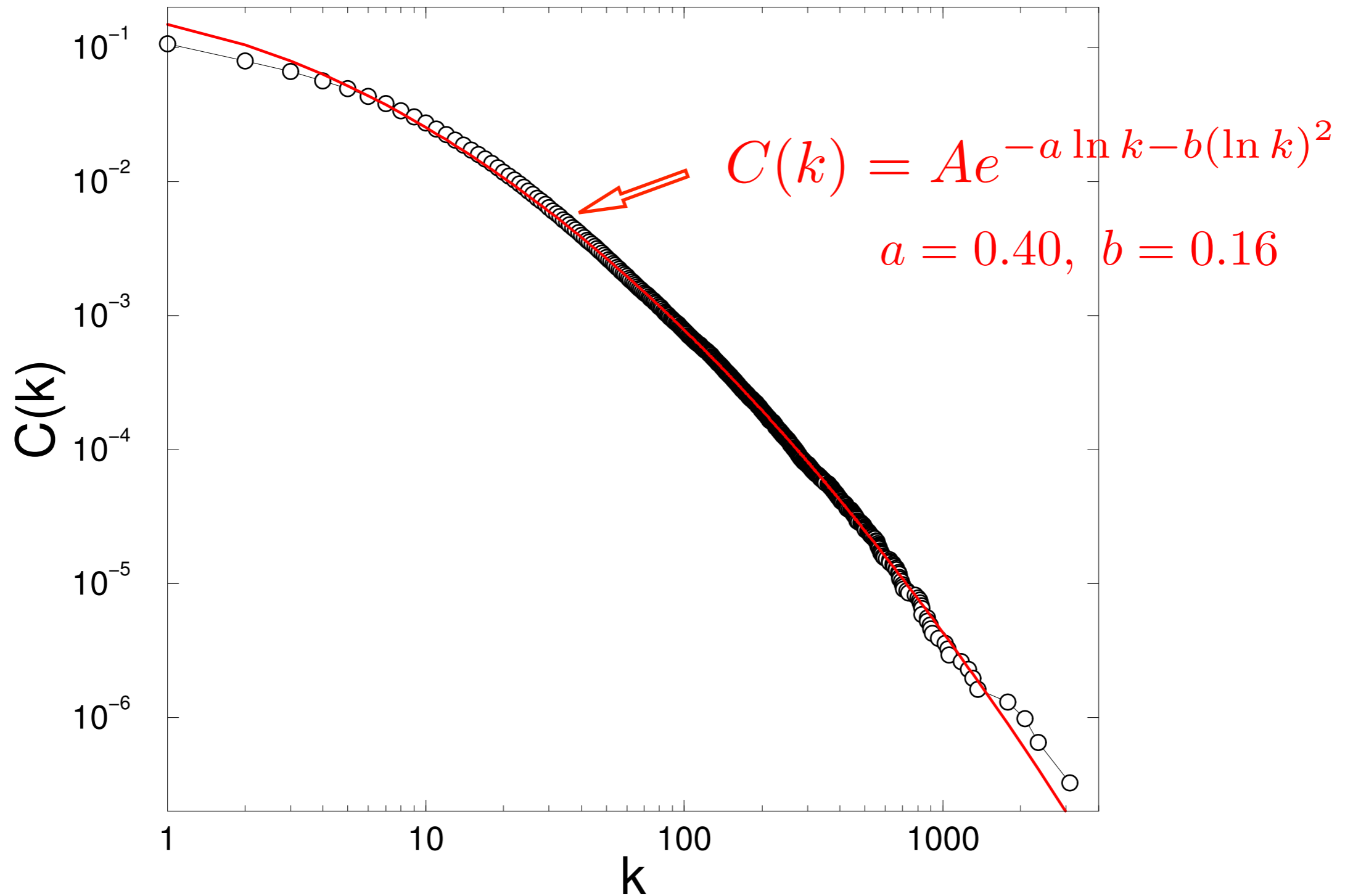
What is the Citation Rate?

Physics Today **58**, 49 (2005)

Physical Review Citation Data, 1893-2003



Cumulative Citation Distribution



Google Page Rank for Citations Brin & Page (1999)

Evolution equation for websurfers on a network:

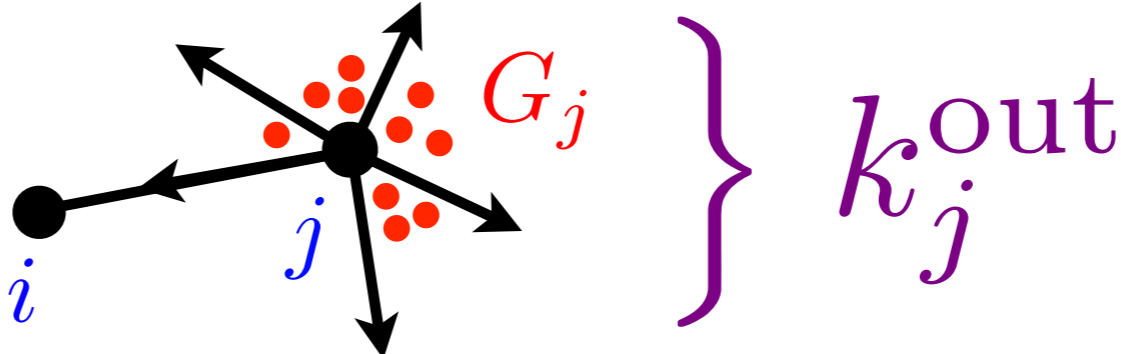
G_i = number of websurfers on node i

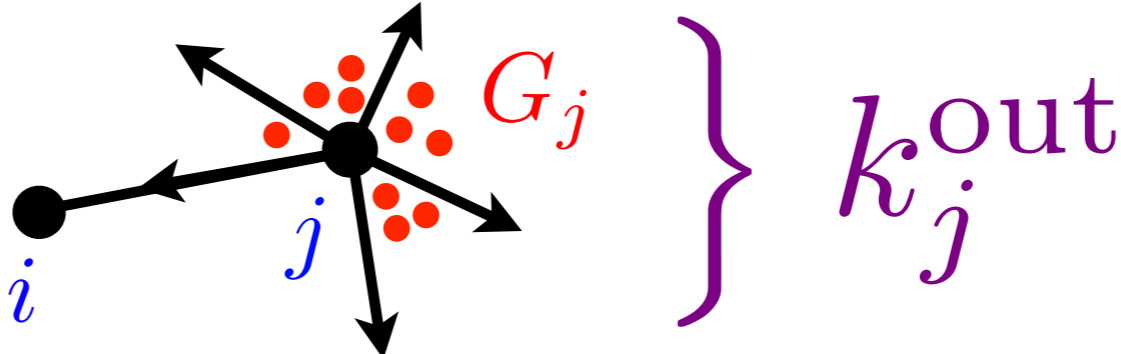
Google Page Rank for Citations Brin & Page (1999)

Evolution equation for websurfers on a network:

G_i = number of websurfers on node i

$$G_i = (1 - d) \sum_{j \text{ nn } i} \frac{G_j}{k_j^{\text{out}}} + \frac{d}{N}$$

random walk propagation  manna from heaven



The diagram shows a central node j with several outgoing edges. A cluster of red dots representing websurfers is located at node j , with the label G_j next to it. A purple bracket on the right side of the diagram is labeled k_j^{out} , indicating the number of outgoing edges from node j . A red arrow points from the text 'random walk propagation' to the summation term in the equation, and another red arrow points from 'manna from heaven' to the $\frac{d}{N}$ term.

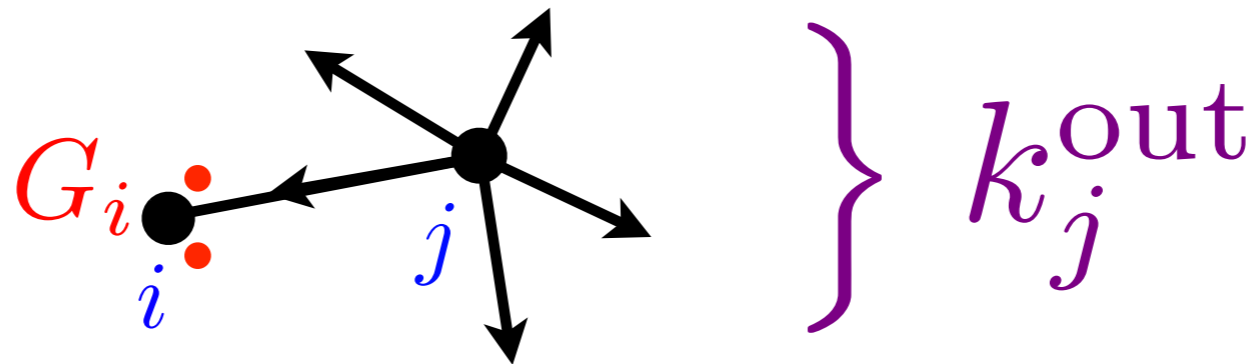
Google Page Rank for Citations Brin & Page (1999)

Evolution equation for websurfers on a network:

G_i = number of websurfers on node i

$$G_i = (1 - d) \sum_{j \text{ nn } i} \frac{G_j}{k_j^{\text{out}}} + \frac{d}{N}$$

random walk propagation \rightarrow $\frac{G_j}{k_j^{\text{out}}}$ \leftarrow manna from heaven $\frac{d}{N}$



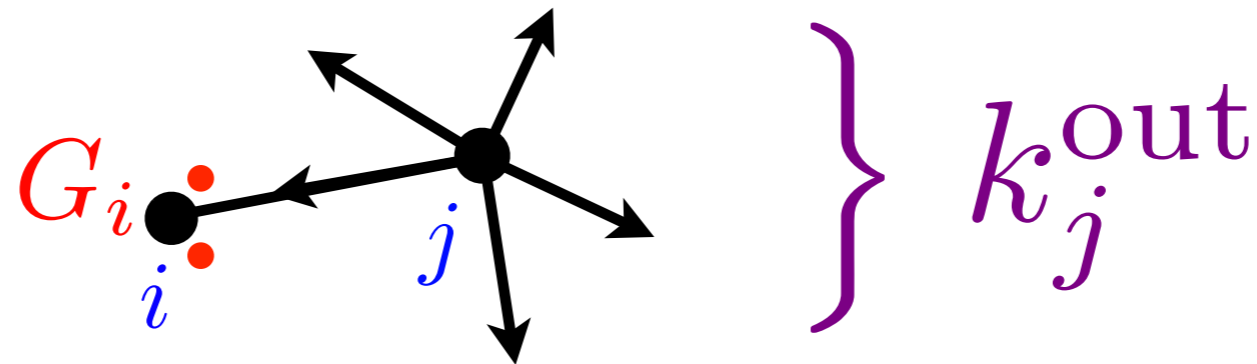
Google Page Rank for Citations Brin & Page (1999)

Evolution equation for websurfers on a network:

G_i = number of websurfers on node i

$$G_i = (1 - d) \sum_{j \text{ nn } i} \frac{G_j}{k_j^{\text{out}}} + \frac{d}{N}$$

random walk propagation \rightarrow $\frac{G_j}{k_j^{\text{out}}}$ \leftarrow manna from heaven $\frac{d}{N}$



Brin/Page: $d=0.15$ (bored after 6 clicks)

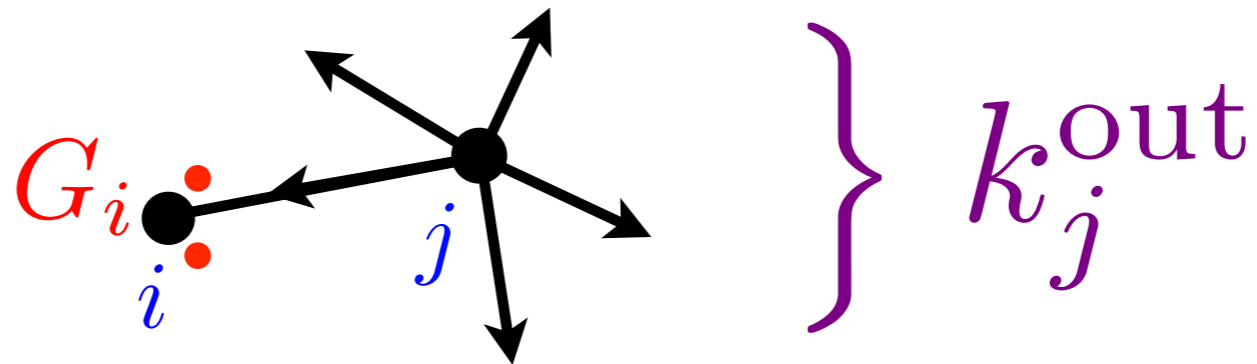
Google Page Rank for Citations Brin & Page (1999)

Evolution equation for websurfers on a network:

G_i = number of websurfers on node i

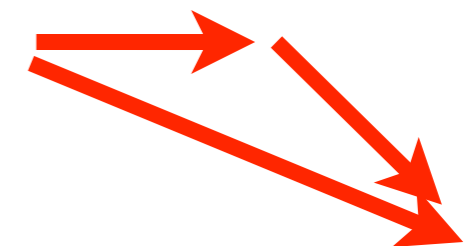
$$G_i = (1 - d) \sum_{j \text{ nn } i} \frac{G_j}{k_j^{\text{out}}} + \frac{d}{N}$$

random walk propagation → ← manna from heaven



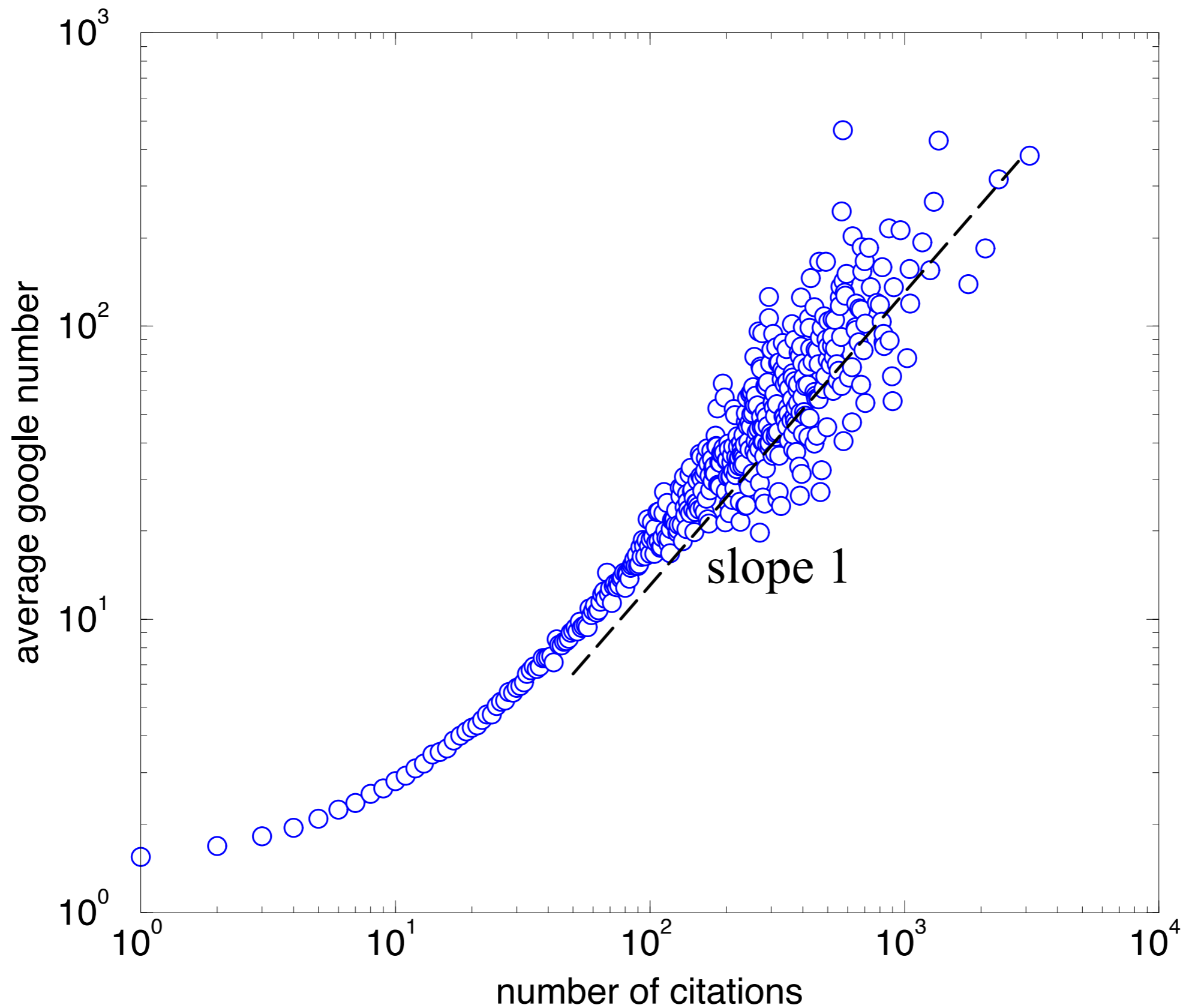
Brin/Page: $d=0.15$ (bored after 6 clicks)

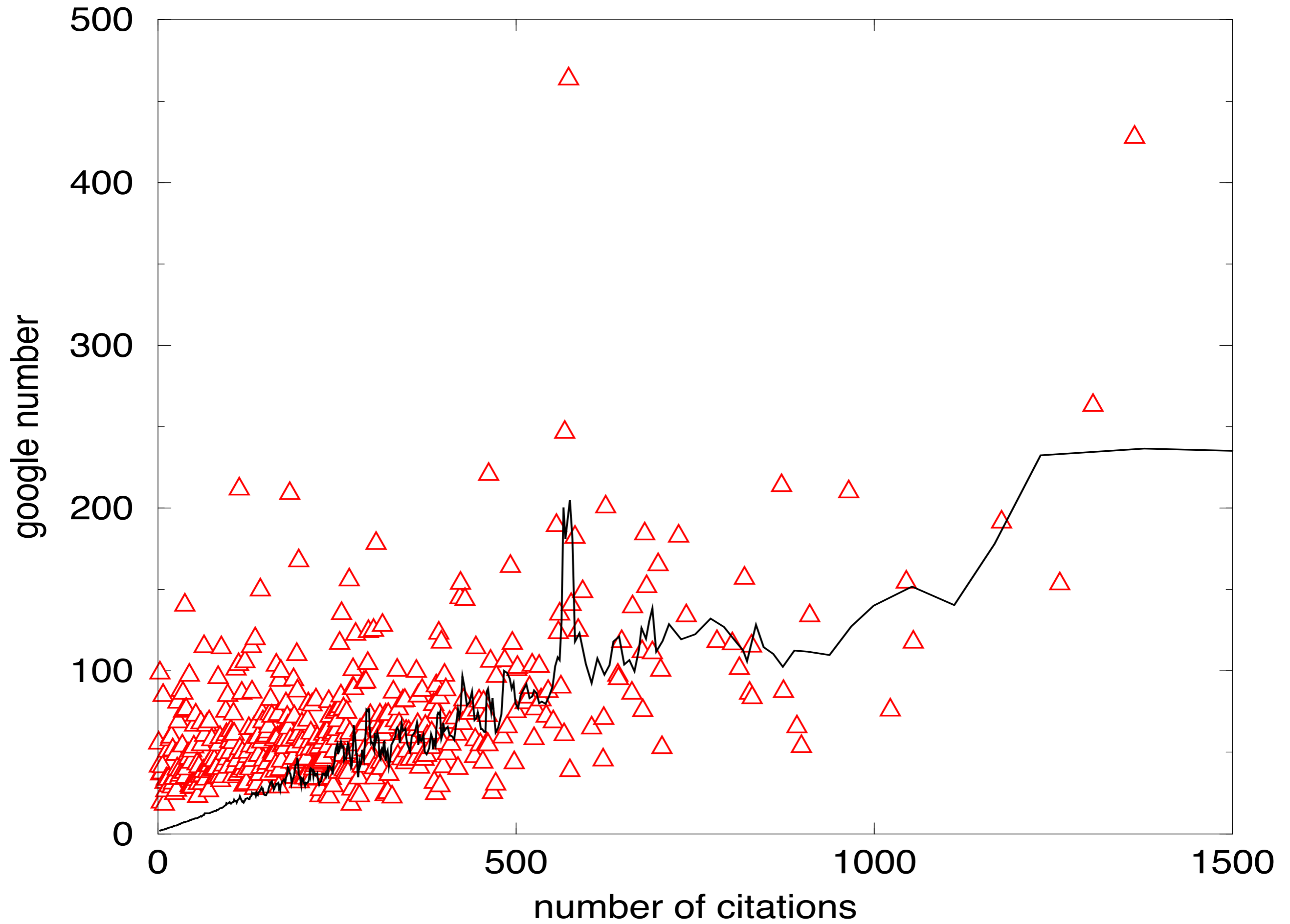
Here: $d=0.50$ (don't cite beyond 2 generations)

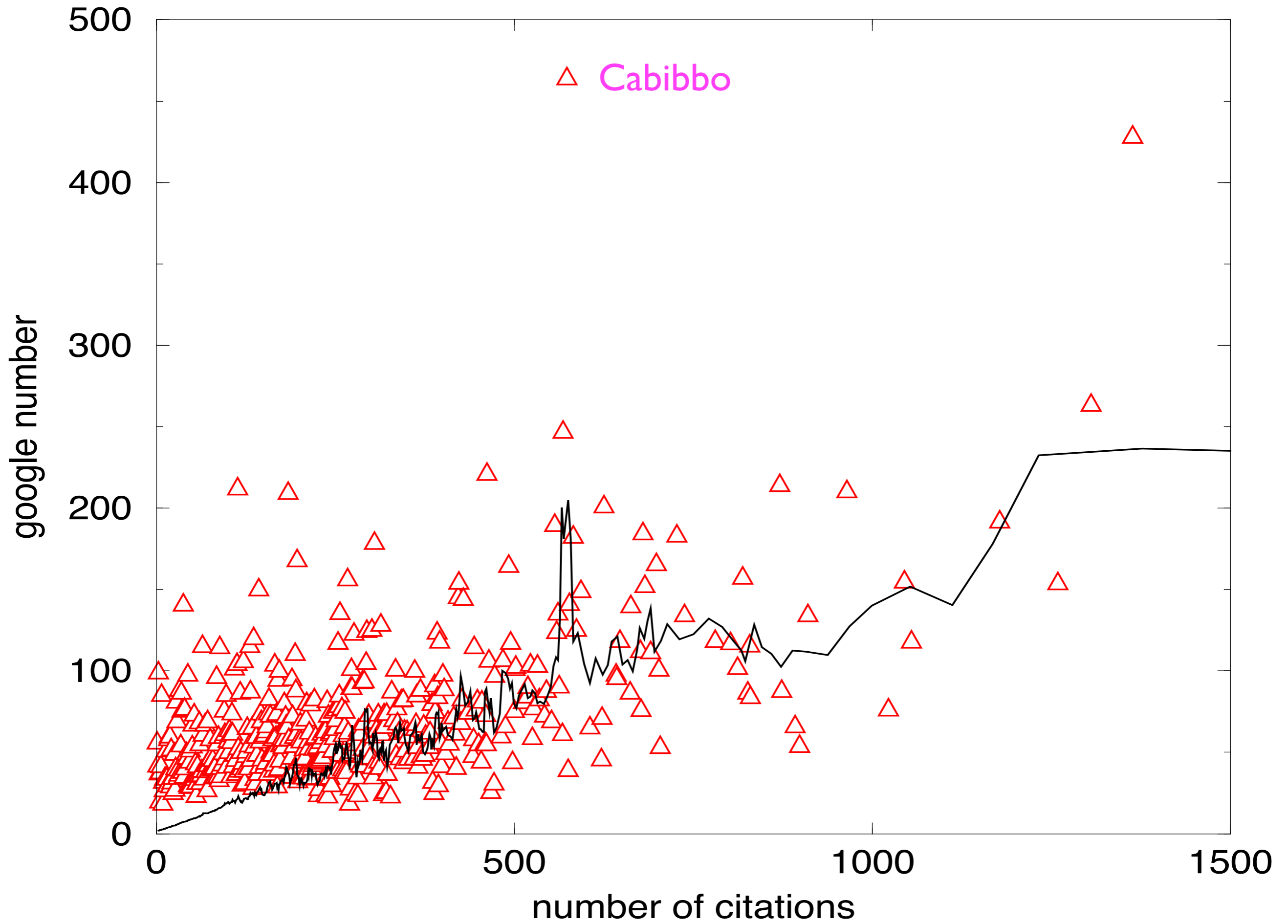


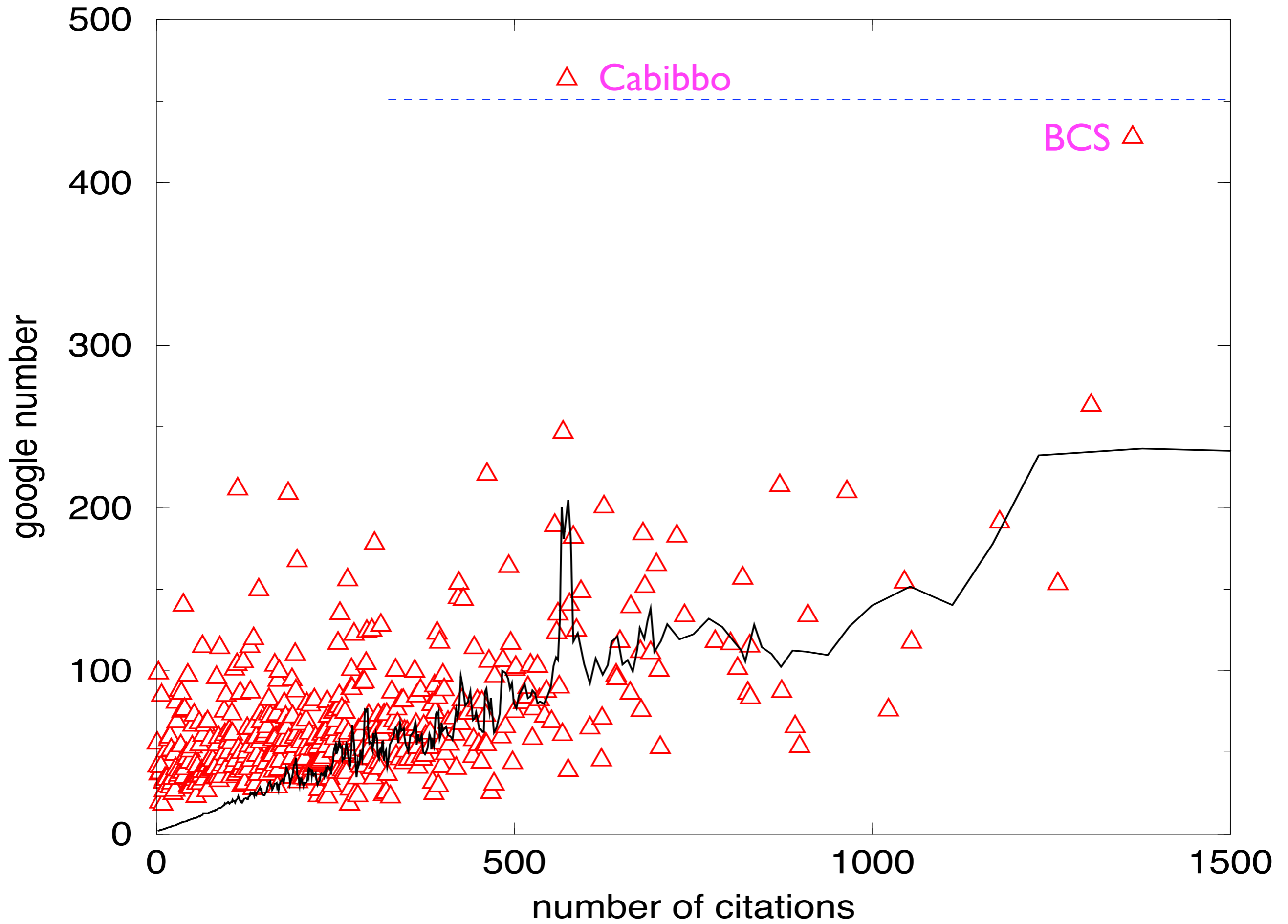
Google Number & Citation Counts

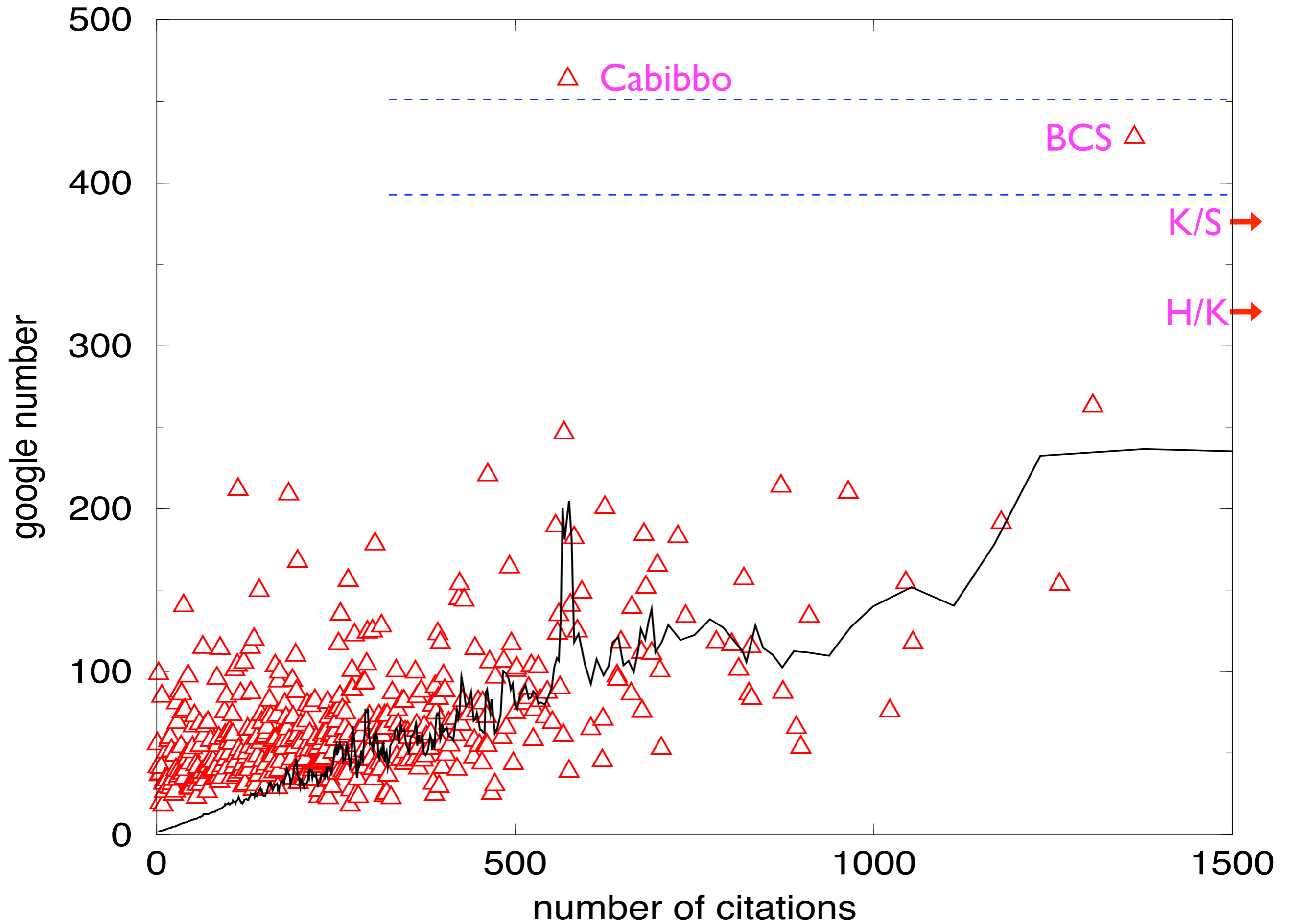
P. Chen, H. Xie, S. Maslov, and S. Redner, *J. Informetrics* **1**, 8 (2007)

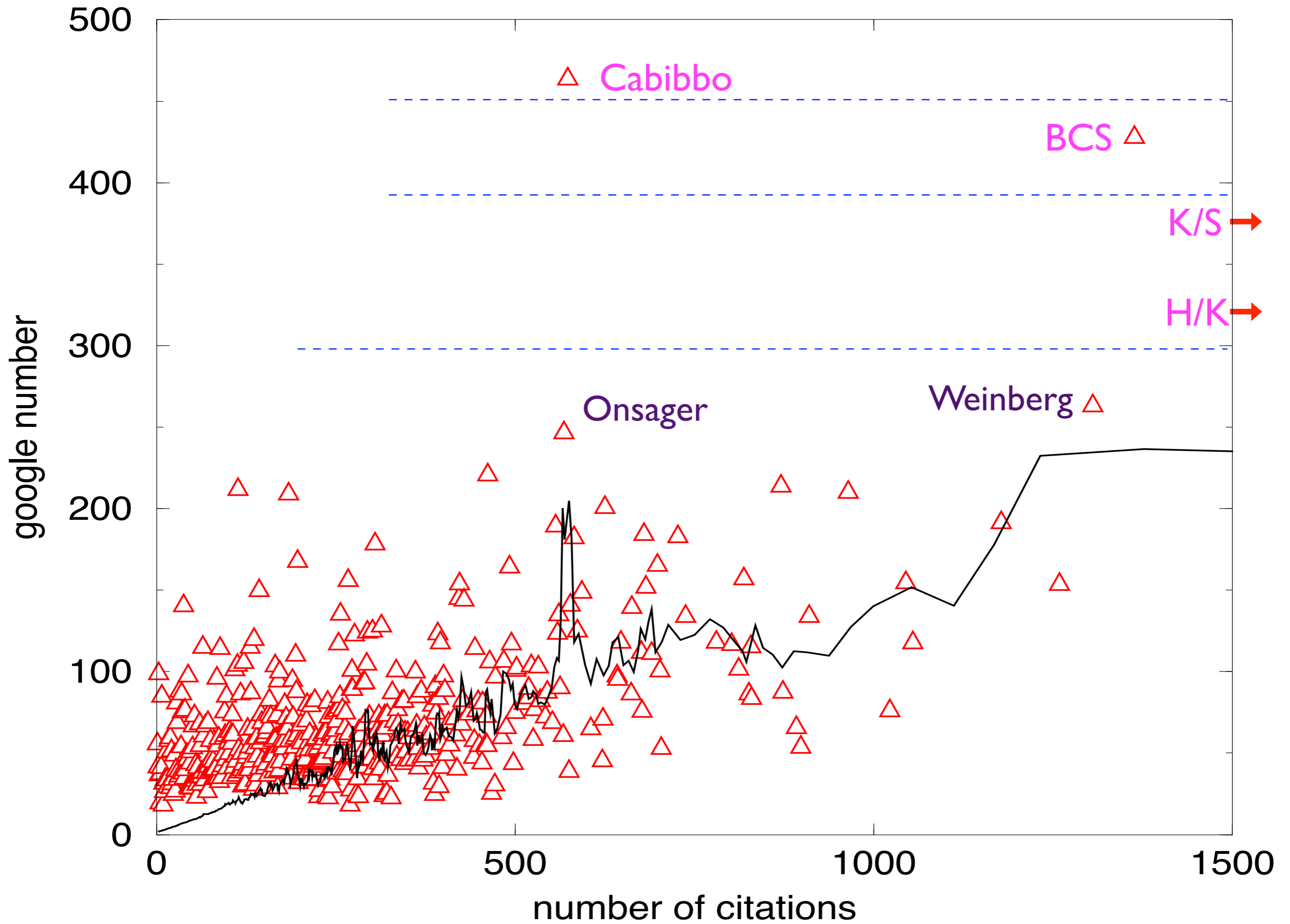


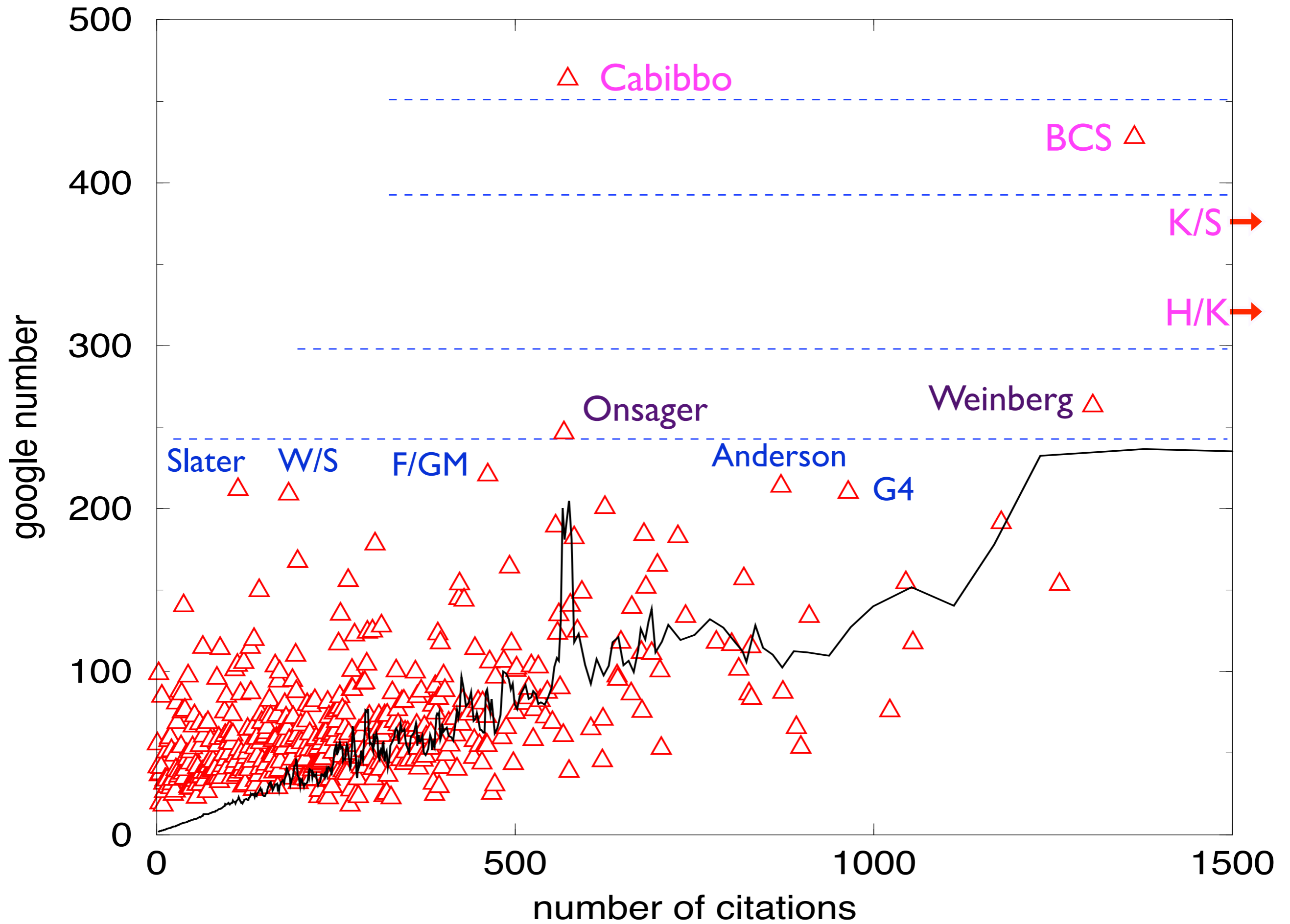


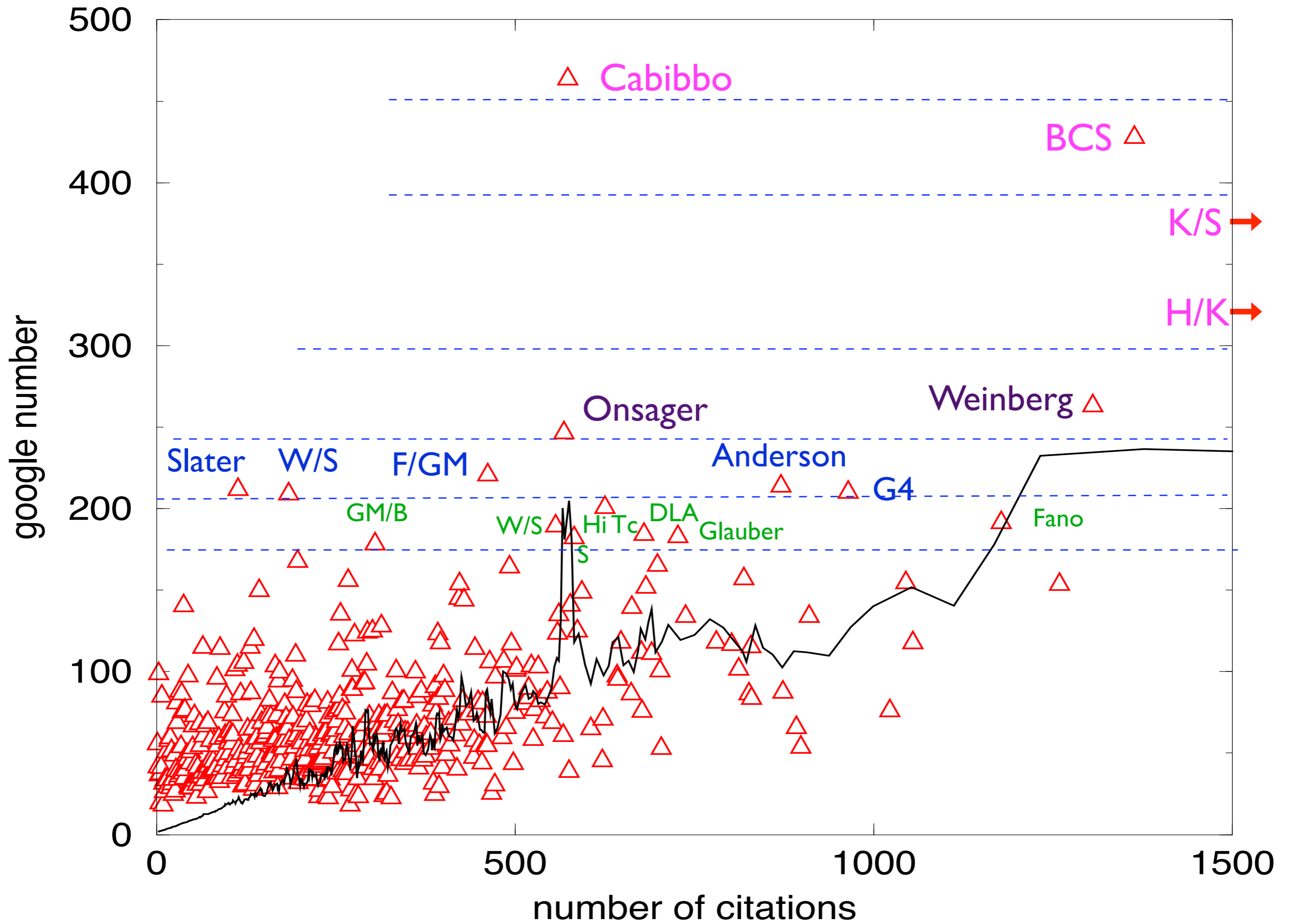










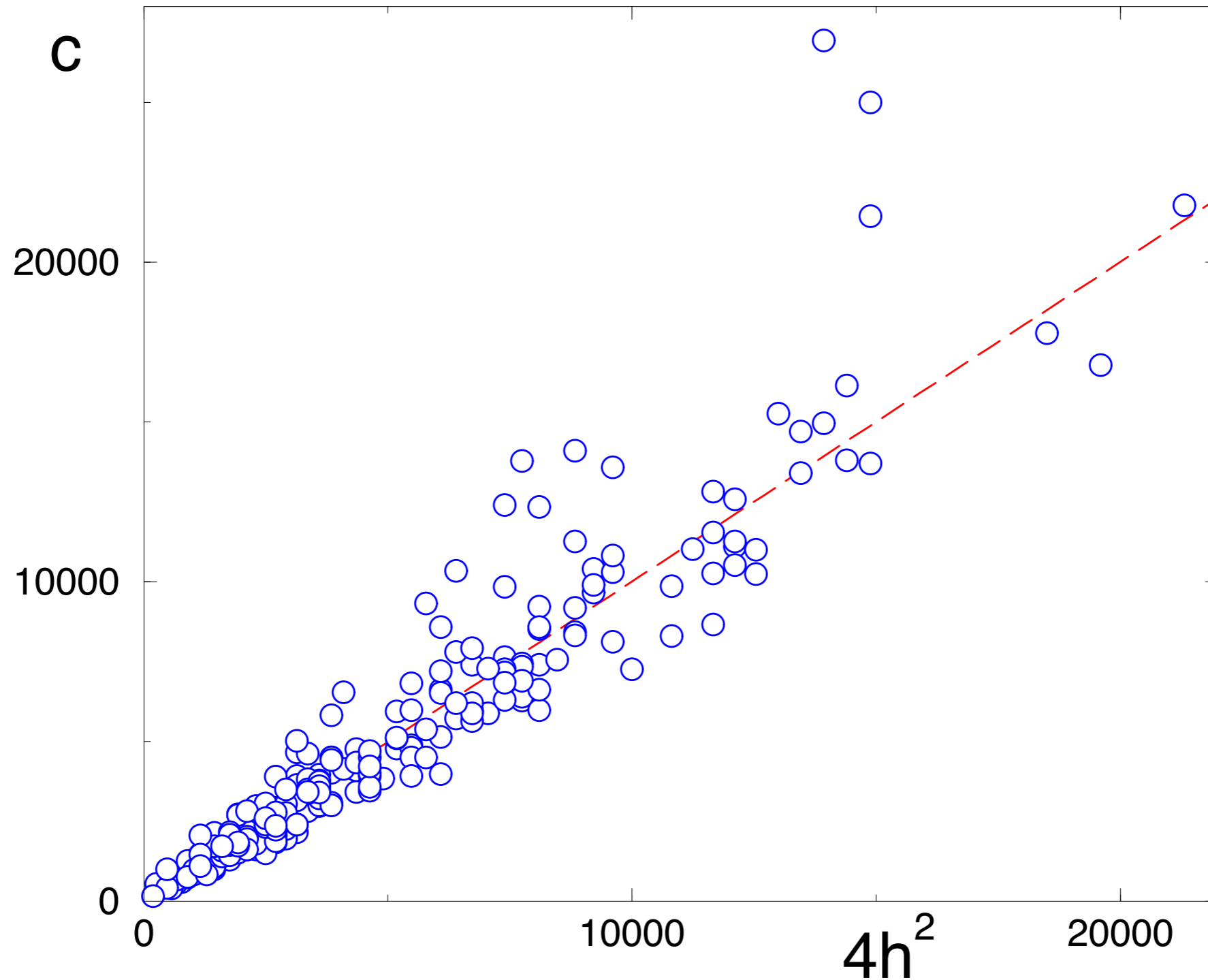


The h-Index

h definition: \cong h publication with \cong h citations

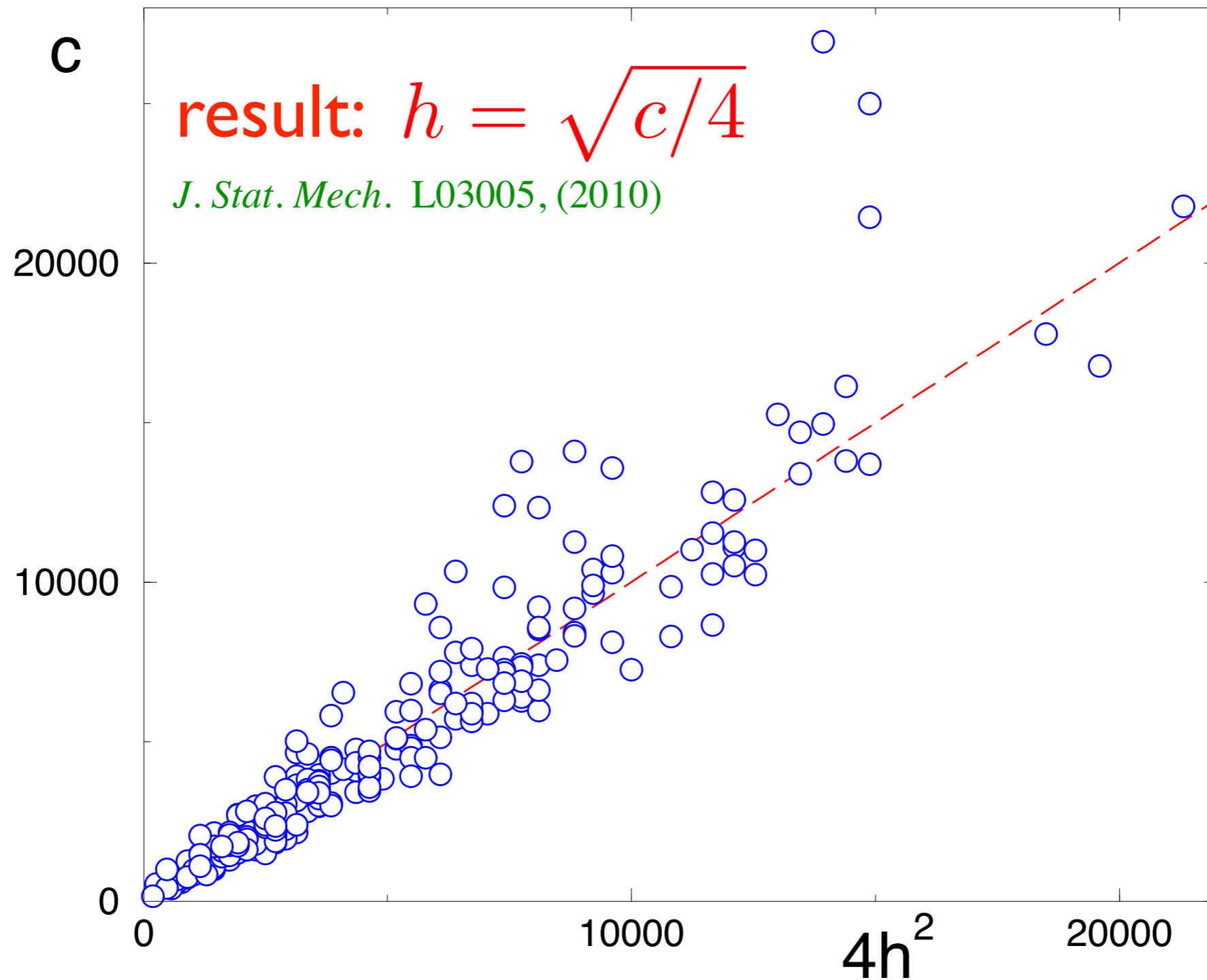
The h-Index

h definition: \cong h publication with \cong h citations

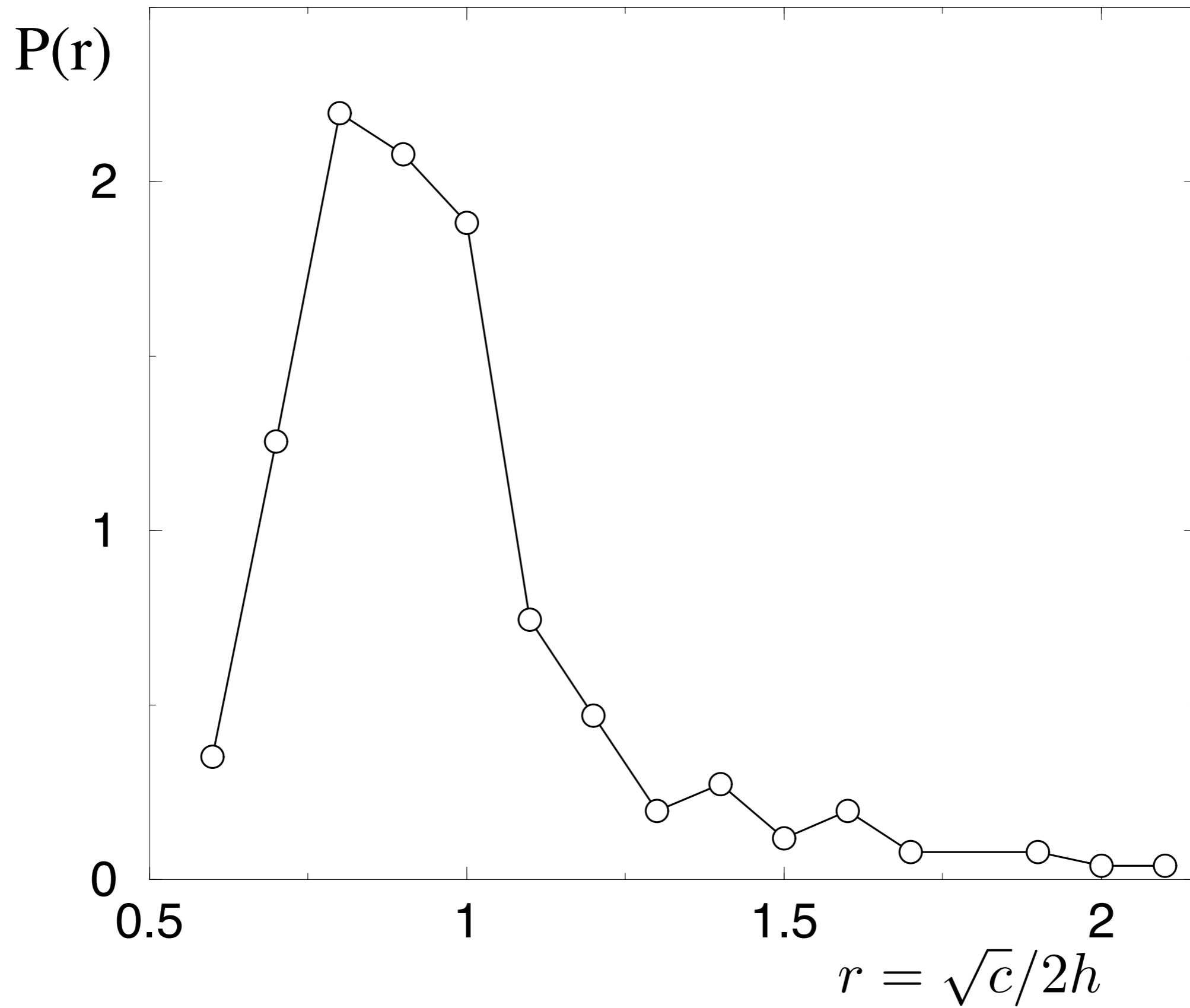


The h-Index

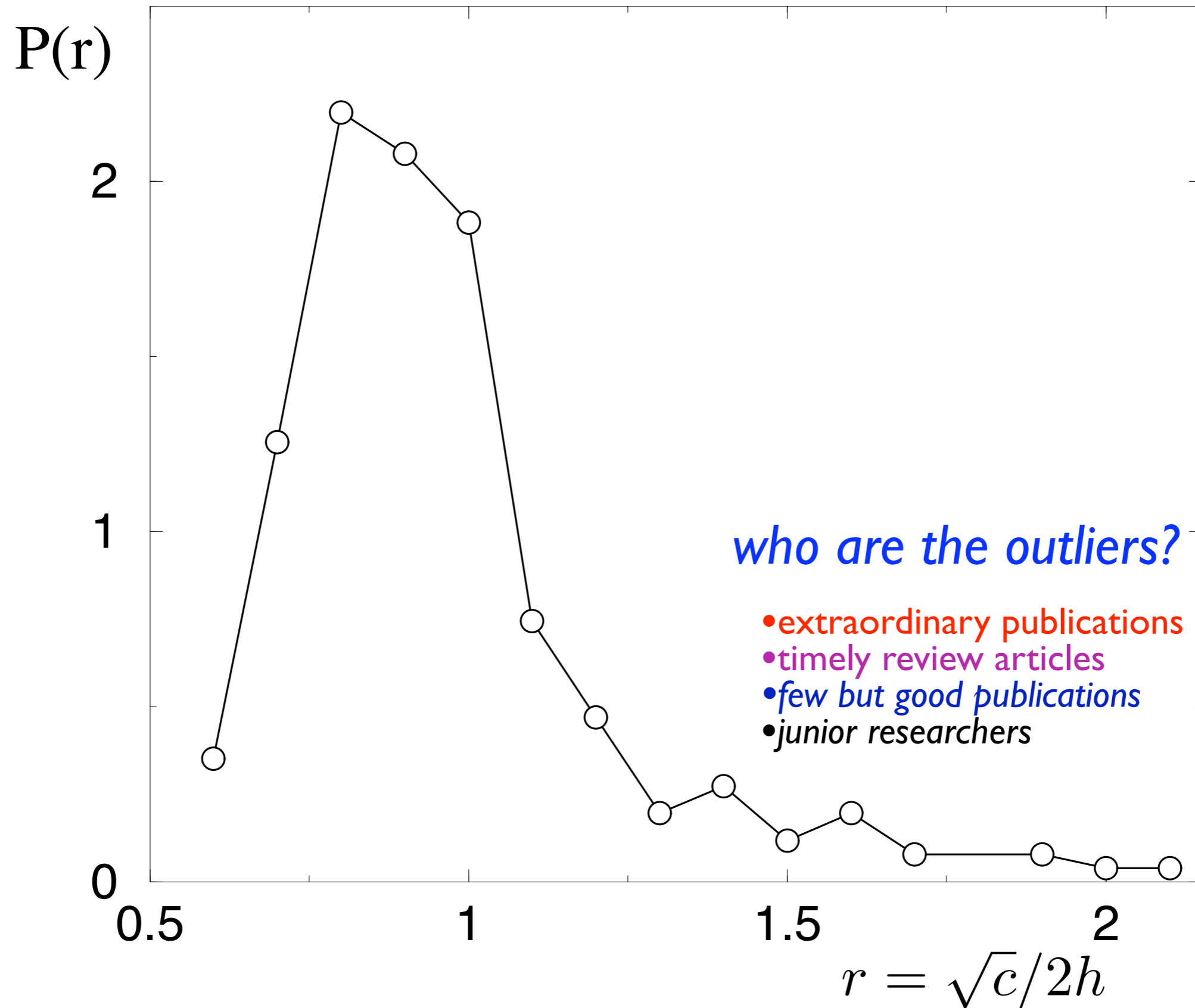
h definition: \cong h publication with \cong h citations



Distribution of $\sqrt{c}/2h$



Distribution of $\sqrt{c}/2h$



Summary & Outlook

Summary & Outlook

Citation analysis is fun! (but a Pandora's box)

Regularities and revealing irregularities

Data motivate and test theories of growing citation networks

Page rank analysis: *helps uncover hidden “gems”*

h-index analysis: *h is equivalent to \sqrt{c}
but outliers, properly interpreted, are revealing*

Summary & Outlook

Citation analysis is fun! (but a Pandora's box)

Regularities and revealing irregularities

Data motivate and test theories of growing citation networks

Page rank analysis: *helps uncover hidden “gems”*

h-index analysis: *h is equivalent to \sqrt{c}
but outliers, properly interpreted, are revealing*

Future: Deeper analysis of citation data:
contextual information, specialization

Larger data sources:
test universality of citation statistics