Bibliometrics
Research metrics
Rankings

Michel Beaudouin-Lafon
Université Paris-Sud

What is the problem?

- How to evaluate research?
- What is a good paper? a good journal?
- What is a good project? a good lab?
- What does “good” mean?

- Research breakthrough are hard to predict

Bibliometrics

- Measuring the impact of publications
- Main metric = number of citations
- Derived metrics
  - h-index
  - impact factor

J. Hirsch

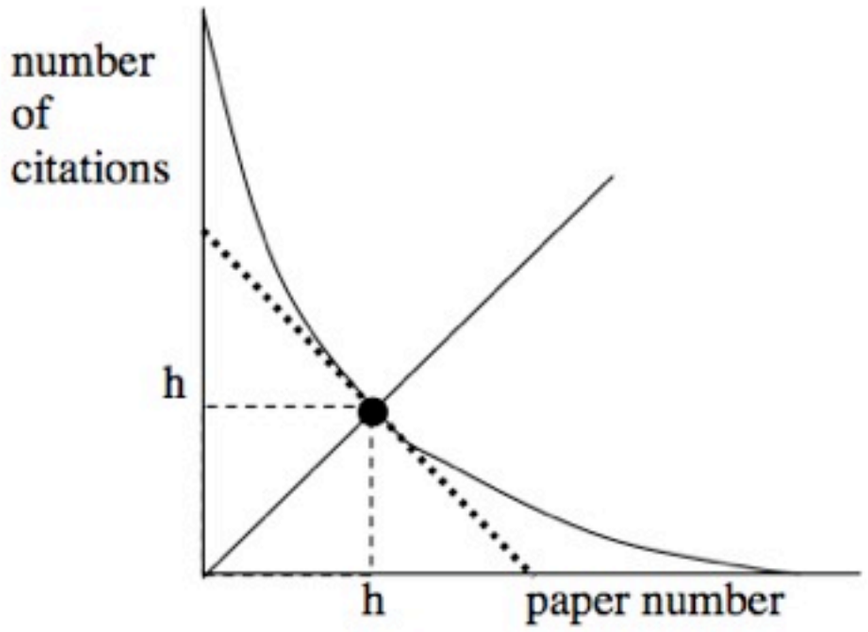


FIG. 1: The intersection of the 45 degree line with the curve giving the number of citations versus the paper number gives h. The total number of citations is the area under the curve. Assuming the second derivative is non-negative everywhere, the minimum area is given by the distribution indicated by the dotted line, yielding a=2 in Eq. 1.

From inspection of the citation records of many physicists I conclude:

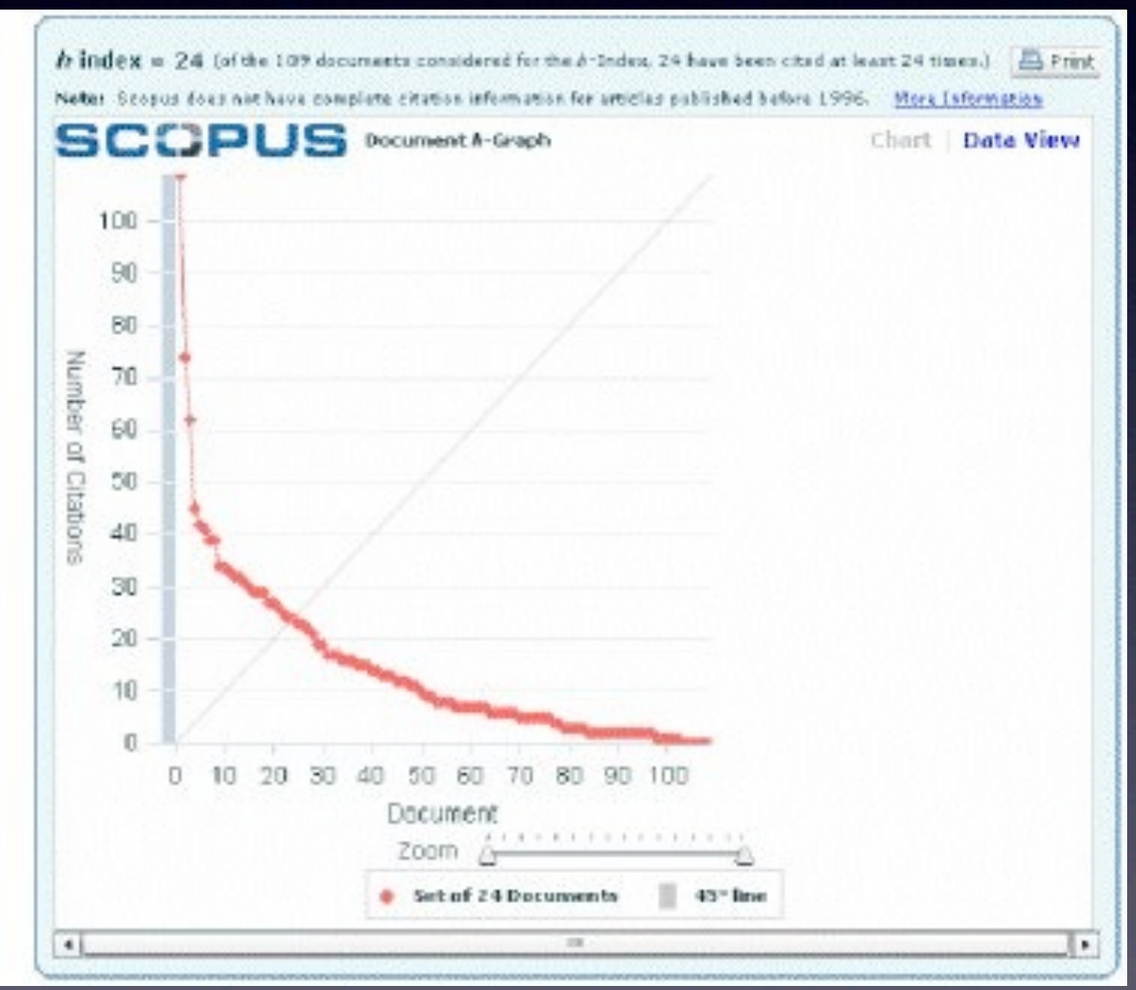
(1) A value $m \sim 1$, i.e. an h index of 20 after 20 years of scientific activity, characterizes a successful scientist.

(2) A value $m \sim 2$, i.e. an h-index of 40 after 20 years of scientific activity, characterizes outstanding scientists, likely to be found only at the top universities or major research laboratories.

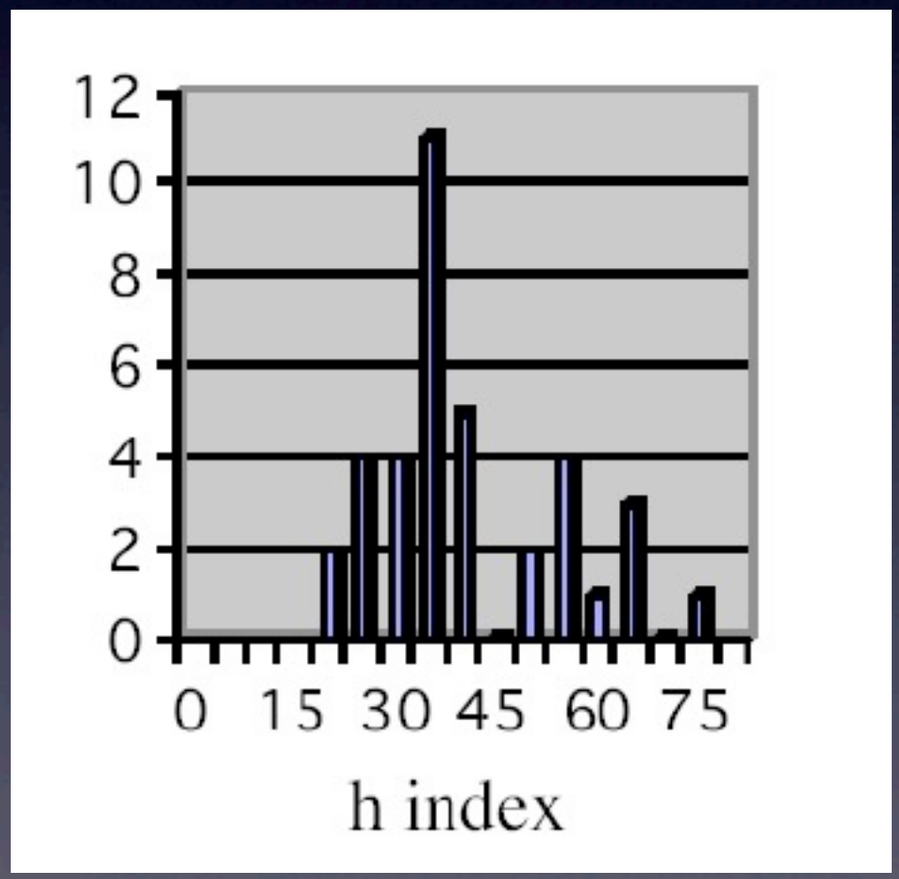
(3) A value $m \sim 3$ or higher, i.e. an h-index of 60 after 20 years, or 90 after 30 years, characterizes truly unique individuals.

In summary, I have proposed an easily computable index, h, which gives an estimate of the importance, significance and broad impact of a scientist's cumulative research contributions. I suggest that this index may provide a useful yardstick to compare different individuals competing for the same resource when an important evaluation criterion is scientific achievement, in an unbiased way.

Power law
vs Normal law

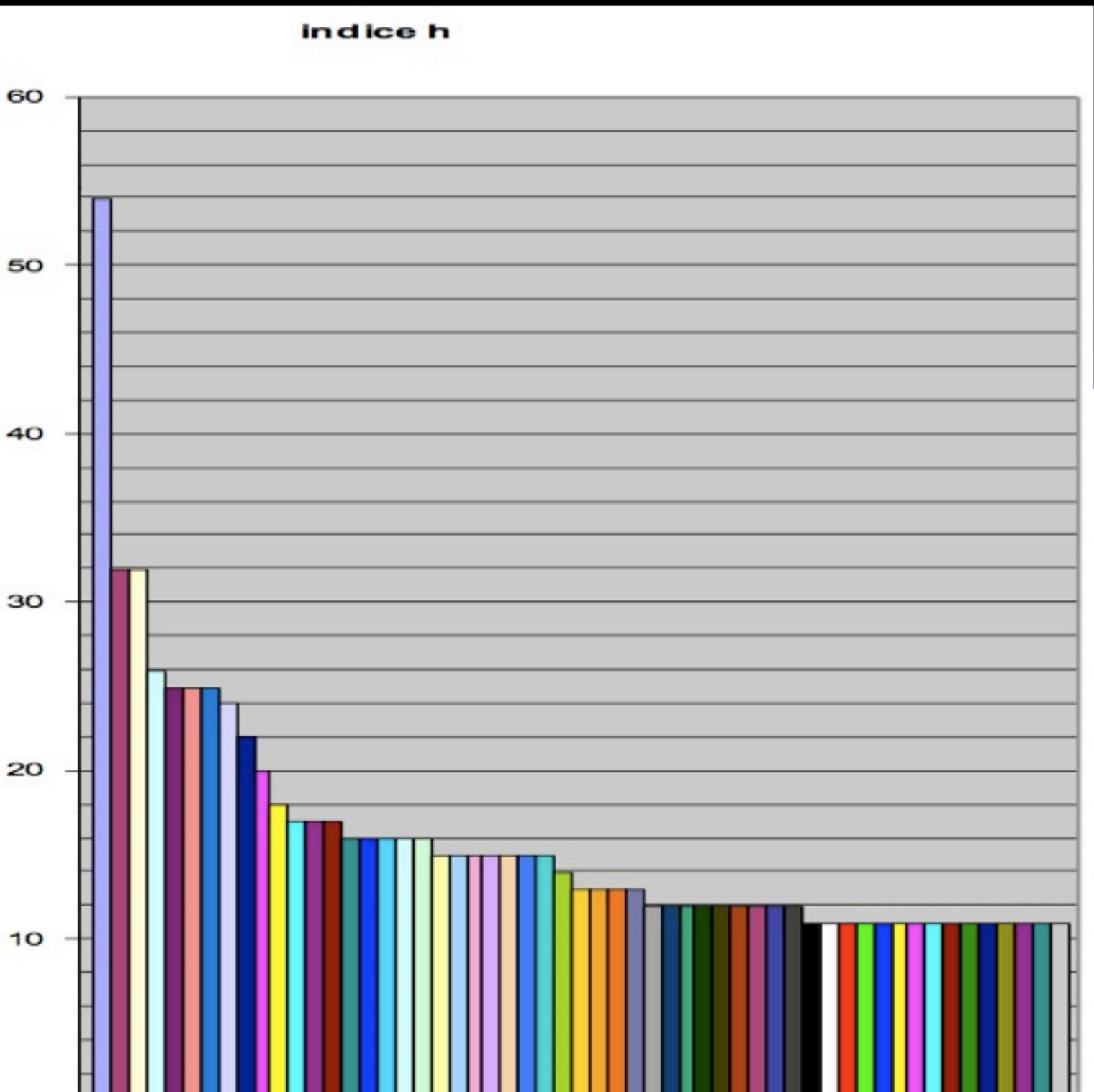


Scopus

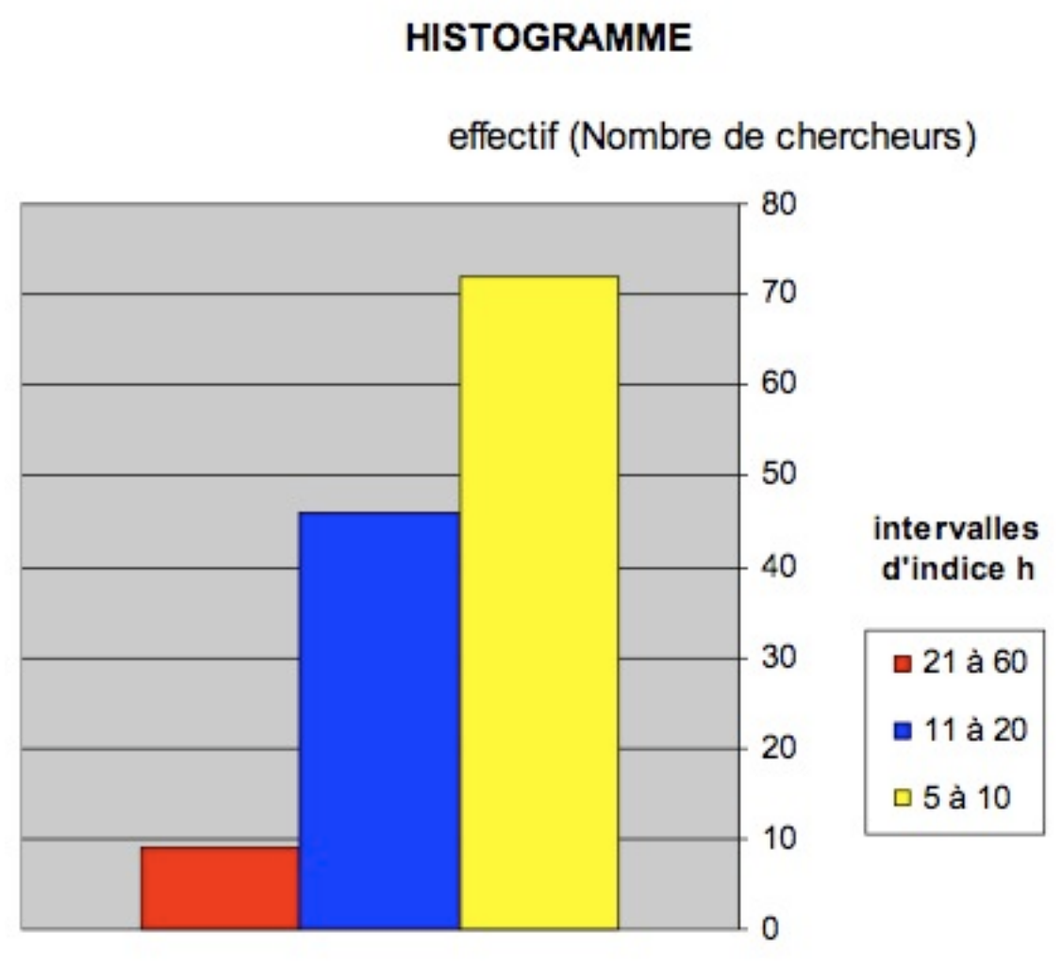


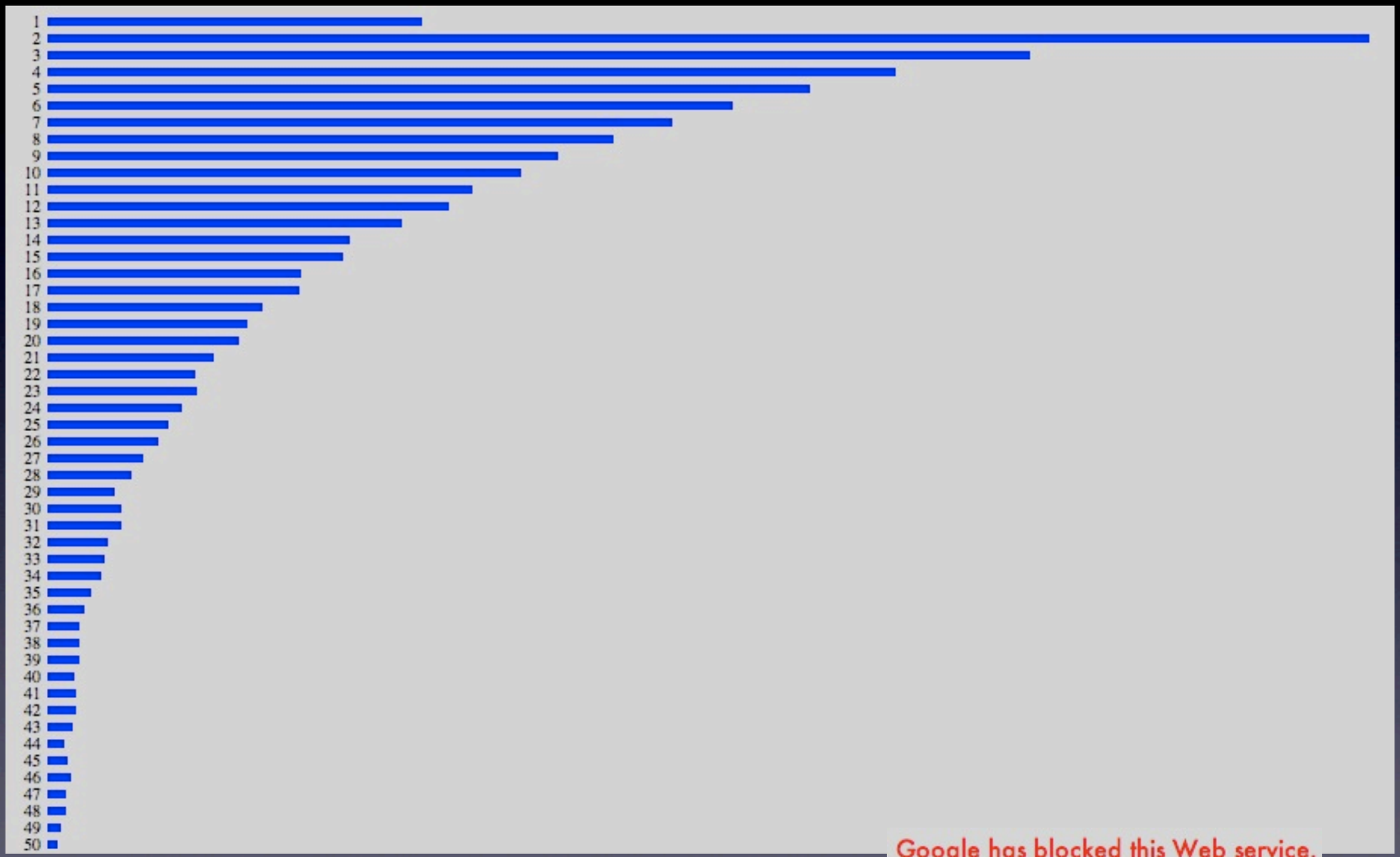
J. Hirsch - h-index des Prix Nobel

Distribution of h-index



Digiteo, 2008





J. Palsberg - distribution of h-index in Computer Science (2009)

Year	Researcher	H Index[2]
2001	Stuart K. Card	44
2001	James D. Foley	28
2001	Morten Kyng	23
2001	Thomas P. Moran	33
2001	Donald A. Norman	49
2001	Judith S. Olson	28
2001	Ben Shneiderman	64
2002	William A. S. Buxton	15
2002	John M. Carroll	52
2002	Douglas C. Engelbart	11
2002	Sara Kiesler	39
2002	Thomas K. Landauer	38
2002	Lucy A. Suchman	10
2003	Thomas Green	39
2003	James D. Hollan	18
2003	Robert E. Kraut	34
2003	Gary M. Olson	29
2003	Peter G. Polson	21

S. Mackenzie - h-index of ACM SIGCHI Academy (2005)

h-index

- Publication practices are very different from one (sub)-discipline to the next
- Always increases (even after the author dies!)
- Sensitive to the quality of the data

=> Can be used *at best* to compare researchers with similar seniority in the same (sub-)discipline

More indices

- Normalize by number of co-authors
- m-index = h-index / n (n=scientific age)
- g-index uses average number of citations
- c-index accounts for “quality” of citations
- e-index accounts for ignored citations
- create yours!

Power laws (again)

- 20% of journals cover 80% of "interesting" articles (Bradford law)
- 20% of articles represent 80% of citations
- ...

=> *simple statistics (mean)*
DO NOT MEAN ANYTHING
with such distributions

Impact Factor

Number of citations of articles published
in years N-1 & N-2 and cited in year N

Number of articles published in years N-1 & N-2

(only takes into account citations
in known journals)

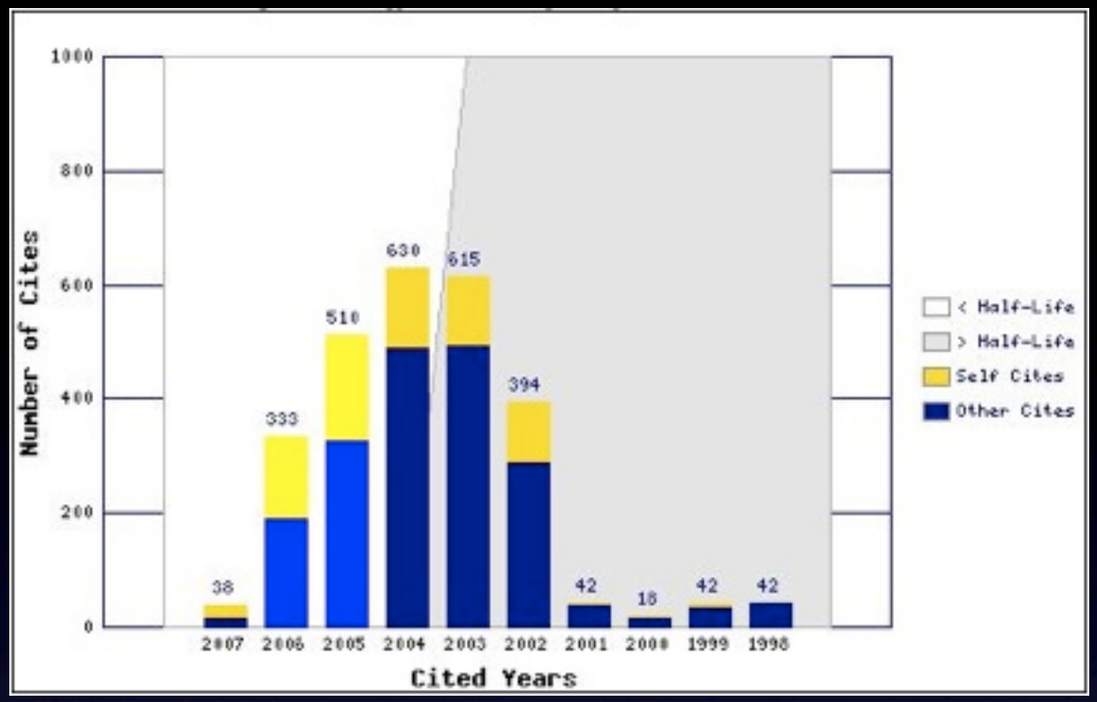
ISI Journal Citation Report

Journals 1 - 20 (of 84) [1 | 2 | 3 | 4 | 5] **Page 1 of 5**

MARK ALL | UPDATE MARKED LIST

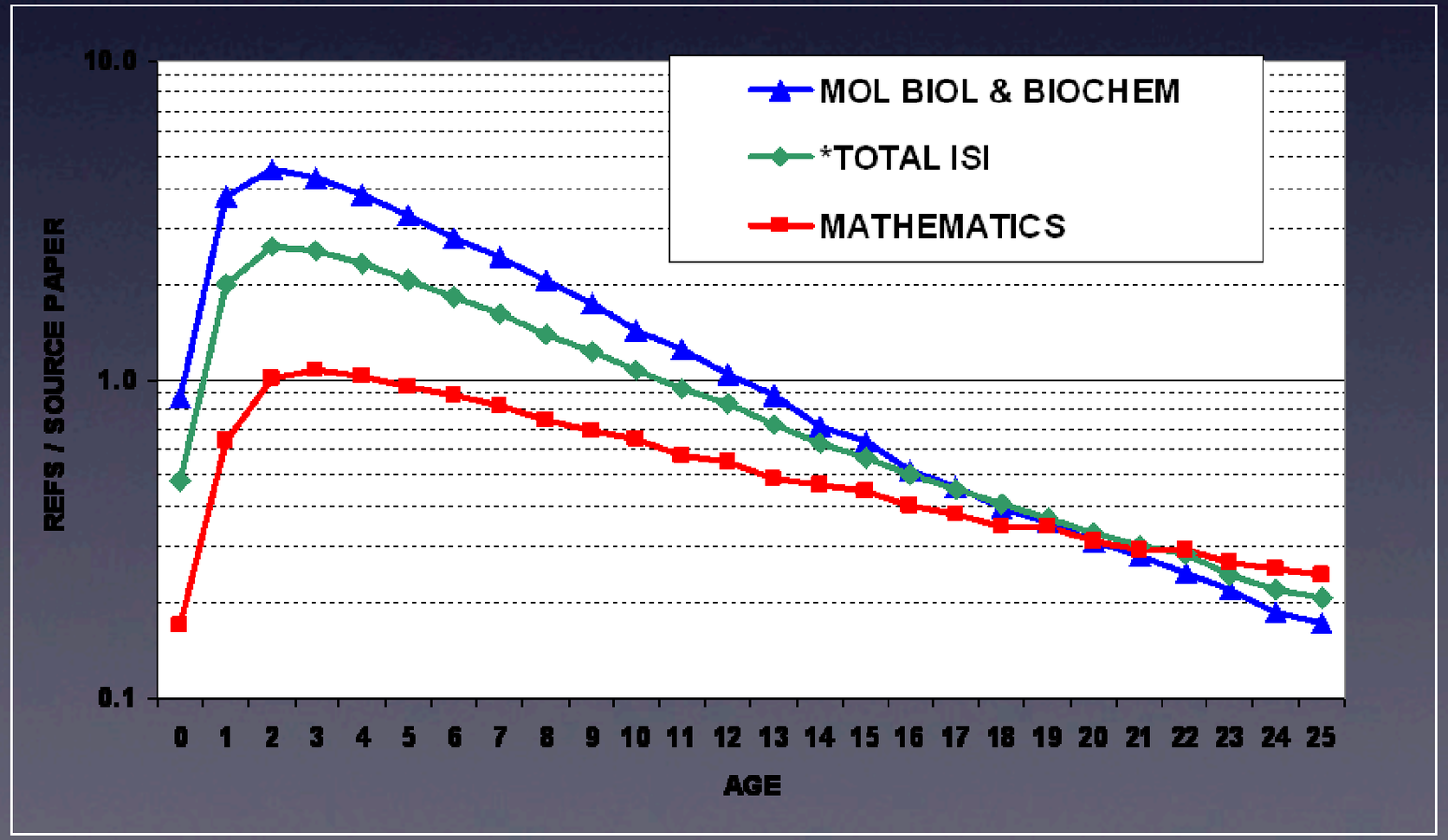
Ranking is based on your journal and sort selections.

Mark	Rank	Abbreviated Journal Title (linked to journal information)	ISSN	JCR Data						*Eigenfactor*™ Metrics	
				Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Articles	Cited Half-life	*Eigenfactor*™ Score	*Article Influence*™ Score
☐	1	ACM T GRAPHIC	0730-0301	3102	3.413	4.683	0.297	128	4.1	0.02259	1.833
☐	2	J WEB SEMANT	1570-8268	328	3.410		0.500	22	3.2	0.00147	
☐	3	J ACM	0004-5411	4894	3.136	3.925	0.312	32	>10.0	0.00962	2.587
☐	4	ACM T SOFTW ENG METH	1049-331X	534	2.792	3.667	0.417	12	7.3	0.00150	1.077
☐	5	REAL-TIME IMAGING	1077-2014	284	2.270	1.287			4.6	0.00187	0.537



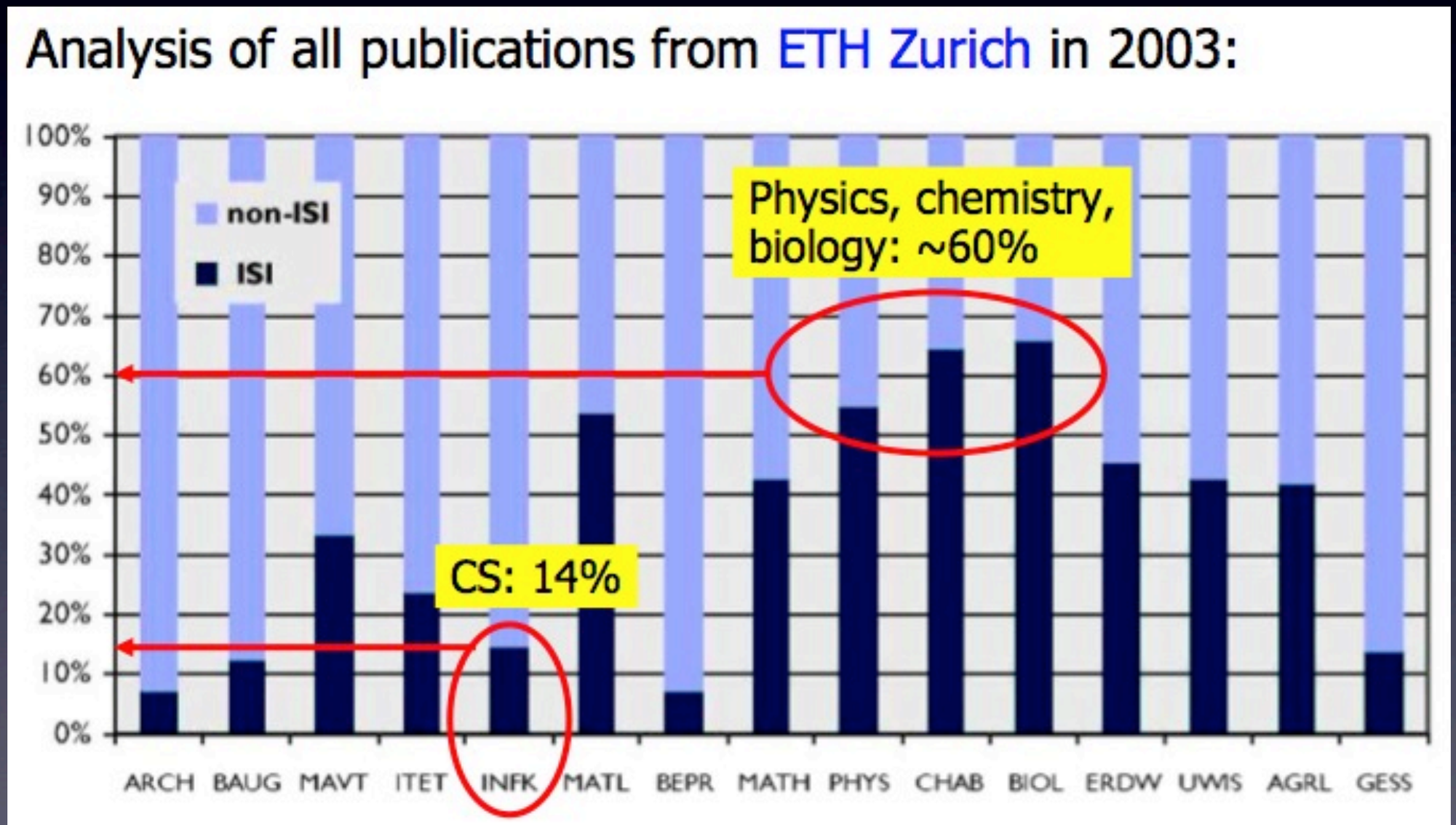
ISI JCR

"Half of the journal's cited items were published more recently than the half-life"



Age des citations (Elsevier)

ISI : coverage problem for computer science



J. Mattern

Why?

- ISI does not index conferences, which are the primary means of publication in many areas of computer science, with a process as rigorous and selective as journals
- ISI is not very welcoming to index a number of journals, especially those published by non-profit organizations

Commercial publishers

- Scientific publishing has become a big business, which uses a lot of free labor: the researchers themselves (editorial boards, reviewers, editors, authors)
- Impact factor is a marketing argument for commercial publishers
- The customers are the research institutions that subscribe to the journals. They are a captive audience, so prices rise.

Evolution of publication business models

- Open Access: freely available to the world
- Who pays the cost of publishing?
  Charge authors instead of readers!
- 1500€ to 8000€ per article!!
- Commercial editors are embracing this model ...
  while keeping the subscription model:
  They charge BOTH readers and authors!

Ranking of Universities

- Shanghai ranking
- Times ranking
- Mines ParisTech ranking
- Webometrics ranking
- ...

Shanghai ranking

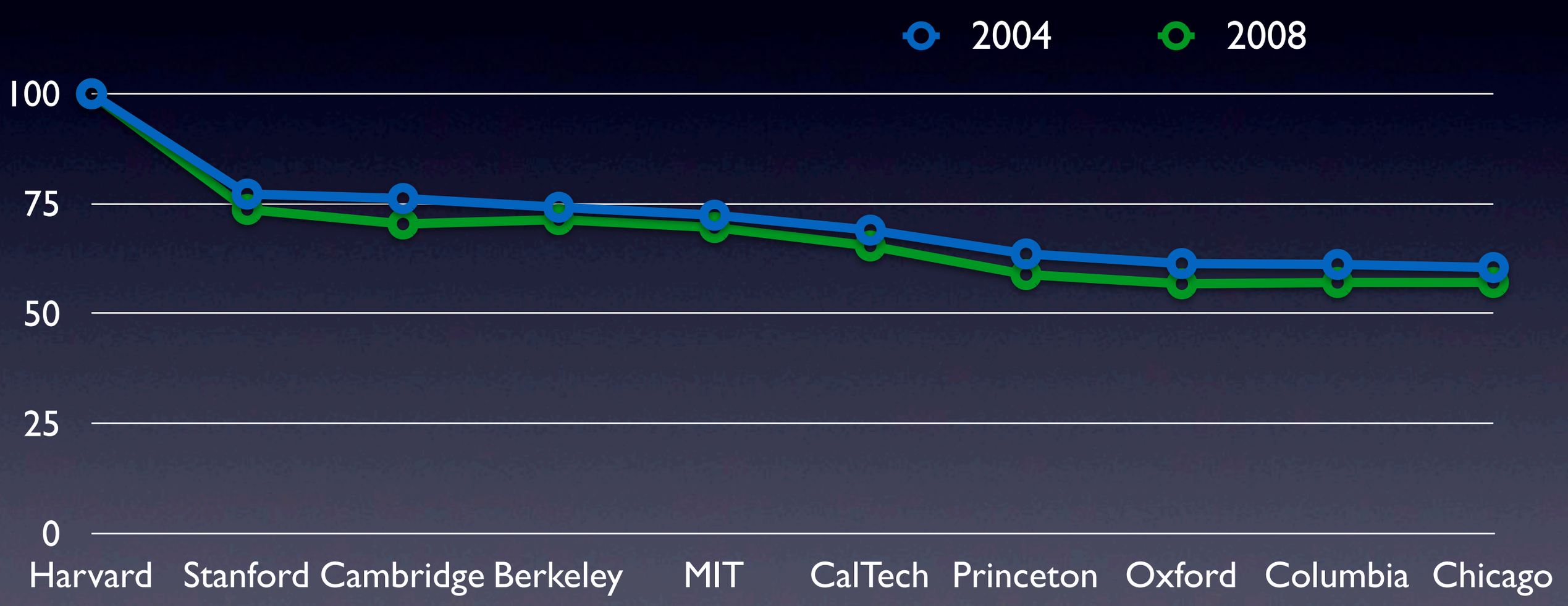
Criteria	Indicator	Code	Weight
Quality of Education	Alumni of an institution winning Nobel Prizes and Fields Medals	*Alumni*	10%
Quality of Faculty	Staff of an institution winning Nobel Prizes and Fields Medals	*Award*	20%
	Highly cited researchers in 21 broad subject categories	*HiCi*	20%
Research Output	Articles published in *Nature* and *Science*	*N&S**	20%
	Articles Indexed in Science Citation Index-Expanded and Social Science Citation Index	*SCI*	20%
Size of Institution	Academic performance with respect to the size of an institution	*Size*	10%
Total			100%

Shanghai : 2004 vs 2008

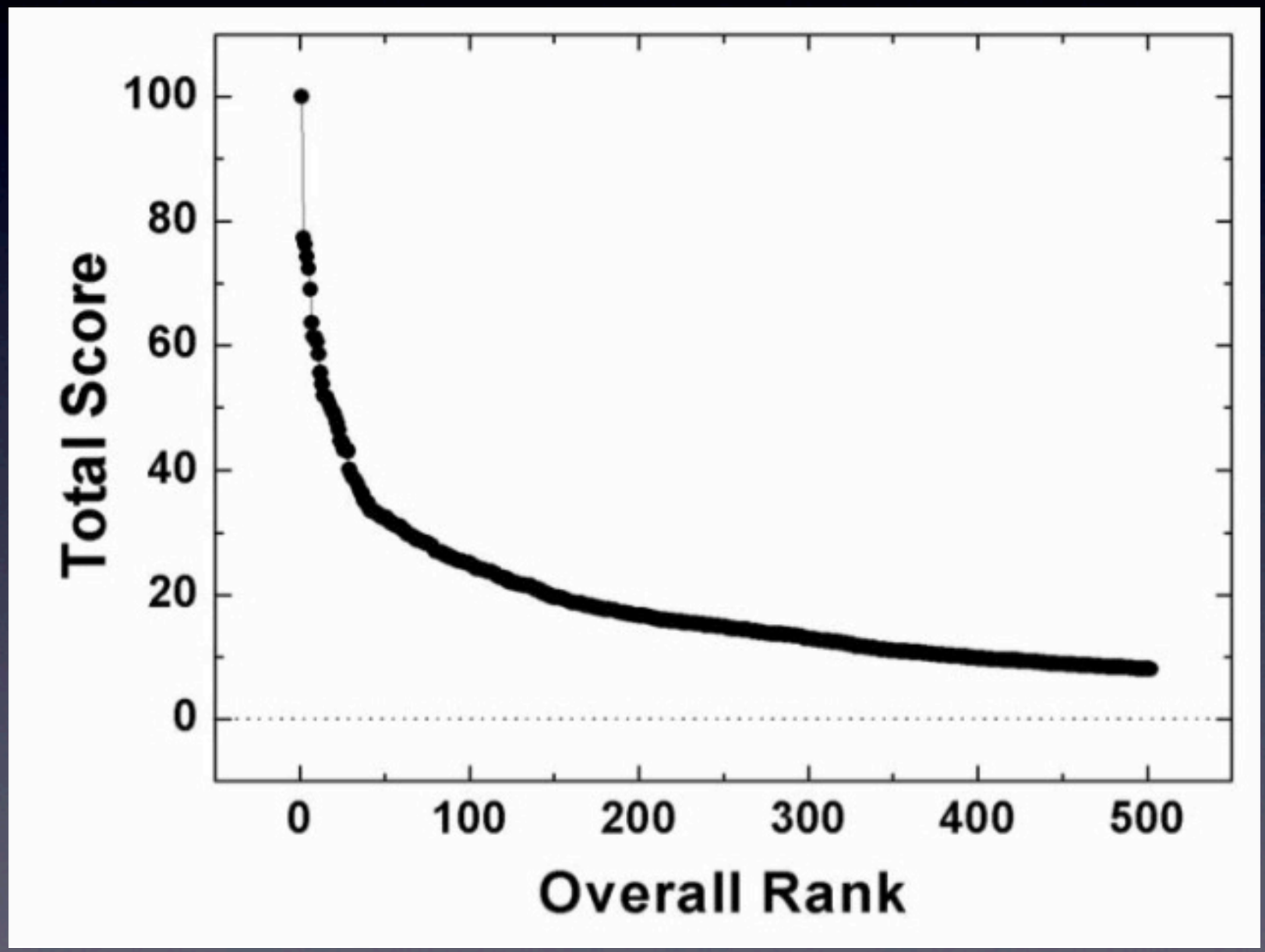
World Rank	Institution	Country
1	Harvard Univ	USA
2	Stanford Univ	USA
3	Univ Cambridge	UK
4	Univ California - Berkeley	USA
5	Massachusetts Inst Tech (MIT)	USA
6	California Inst Tech	USA
7	Princeton Univ	USA
8	Univ Oxford	UK
9	Columbia Univ	USA
10	Univ Chicago	USA

World Rank	Institution*	Region
1	Harvard Univ	Americas
2	Stanford Univ	Americas
3	Univ California - Berkeley	Americas
4	Univ Cambridge	Europe
5	Massachusetts Inst Tech (MIT)	Americas
6	California Inst Tech	Americas
7	Columbia Univ	Americas
8	Princeton Univ	Americas
9	Univ Chicago	Americas
10	Univ Oxford	Europe

Shanghai : 2004 vs 2008



Distribution of scores



France in the Shanghai ranking

Institution	Rang 2004 - 2008	Score 2004 - 2008
Paris-6	41 - 42	32.6 - 33.1
Paris-11	48 - 49	31.2 - 32.1
Strasbourg	82 - ~125	25.7 - 22.3
ENS Ulm	85 - 73	25.4 - 27.7

Times ranking

World University Rankings 2008

THE THE TOP 200 WORLD UNIVERSITIES

2008 RANK	2007 RANK	INSTITUTION	COUNTRY	PEER REVIEW SCORE	EMPLOYER REVIEW SCORE	STAFF/STUDENT SCORE	CITATIONS/STAFF SCORE	INTERNATIONAL STAFF SCORE	INTERNATIONAL STUDENTS SCORE	OVERALL SCORE
1	1	Harvard University	US	100	100	96	100	87	81	100
2	2=	Yale University	US	100	100	100	98	89	71	99.8
3	2=	University of Cambridge	UK	100	100	99	89	98	95	99.5
4	2=	University of Oxford	UK	100	100	100	85	96	96	98.9
5	7=	California Institute of Technology	US	100	74	98	100	100	93	98.6
6	5	Imperial College London	UK	99	100	100	83	98	100	98.4
7	9	University College London	UK	96	99	100	89	96	100	98.1
8	7=	University of Chicago	US	100	99	98	91	78	83	98.0
9	10	Massachusetts Institute of Technology	US	100	100	90	100	33	94	96.7
10	11	Columbia University	US	100	99	98	94	29	89	96.3

INDICATORS TO EVALUATE THE OVERALL POSITION OF A UNIVERSITY

Academic Peer Review
A global survey of academics asking respondents to identify universities they consider excellent in their own broad field of knowledge.

Employer Review
A global survey of employers, with experience of recruiting from universities, asking respondents to identify universities they consider to be best at preparing their graduates for the workplace.

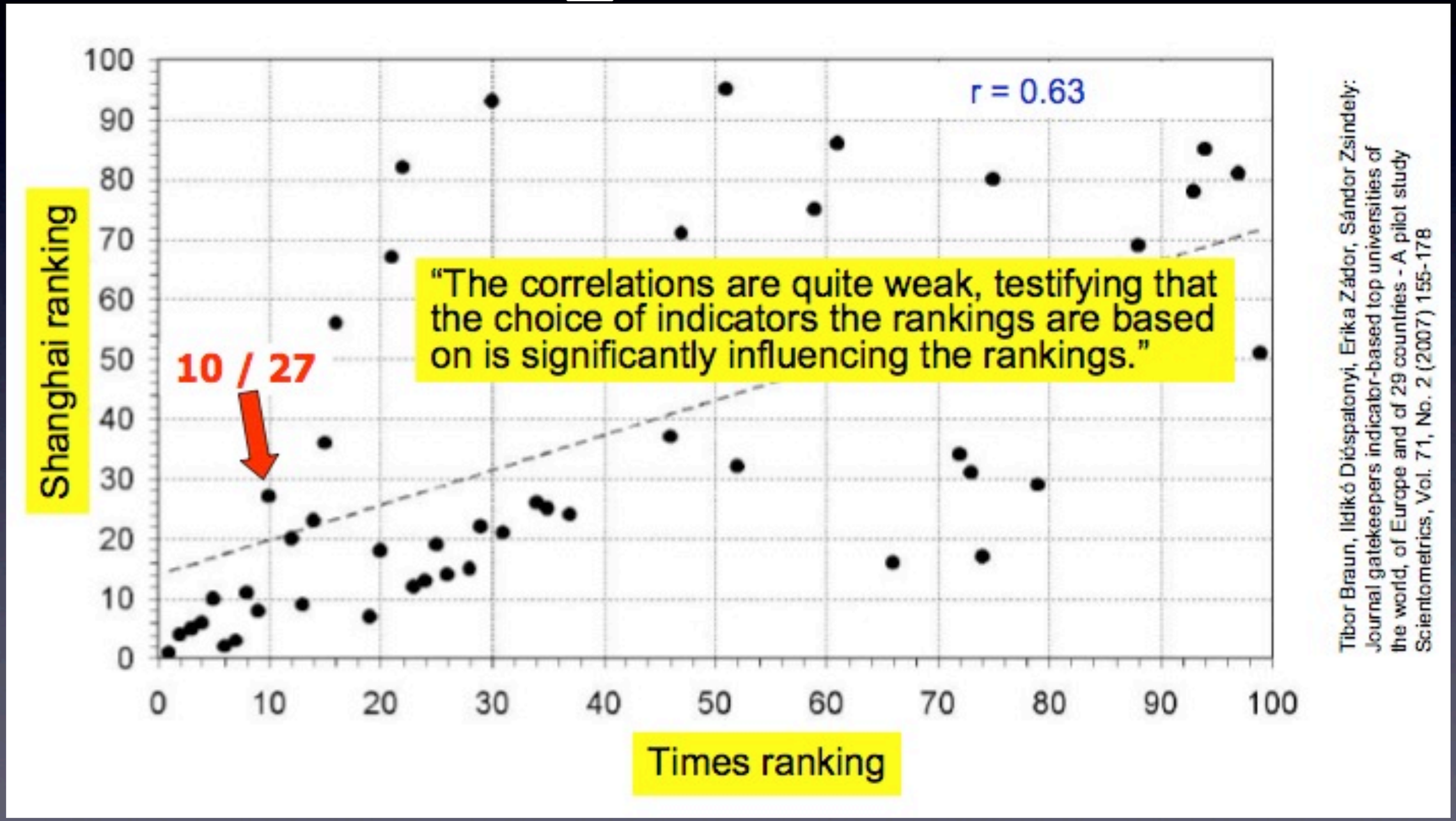
Citations per Faculty Member
Using exported information from Scopus, this measure combines research productivity and quality, taking into account the scale of an institution.

Student Faculty Ratio
In lieu of a globally available evaluation of teaching quality, this indicator is designed to serve as a widely available proxy for commitment to teaching.

International Students
The proportion of international students is used to evaluate not only a university's broad approach to internationalization but also to give prospective students feeling for an institution's commitment and facilities
for students from overseas.

International Faculty
This indicator supports the previous one in assessing a university's commitment to providing a global education for an increasingly global post-education environment.

Correlation Shanghai-Times



J. Mattern

France in the Time ranking (2008)

- 28 : ENS Ulm
- 34 : Ecole Polytechnique
- 140 : ENS Lyon
- 149 : Paris 6

How to improve your rank?

- Create your own ranking!

Rang	Institution	Score
1	Tokyo Univ	100,0
2	Harvard Univ	89,0
3	Stanford Univ	57,0
4	Keio Univ	48,0
5	Univ Pennsylvania	39,0
6	Waseda Univ	36,0
7	HEC	35,0
8	Kyoto Univ	30,0
8	Univ Oxford	30,0
10	ENA	29,5
11	Inst for Study of Politics - Paris	26,5
12	Yale Univ	26,0
13	Massachusetts Inst Tech (MIT)	24,0
13	Osaka Univ	24,0
15	Ecole Polytechnique	23,5
16	Seoul Natl Univ	23,0
17	Univ Muenster	22,0
18	Fordham Univ	21,0
18	Pennsylvania State Univ - Univ Park	21,0
20	Duke Univ	20,0
20	Ecole Natl Super Mines - Paris	20,0
20	Columbia Univ	20,0
23	Chalmers Univ Tech	18,0
23	Chuo Univ	18,0
23	Univ Chicago	18,0
23	Univ Iowa	18,0
27	INSEAD	15,5

Mines ParisTech ranking

Nous avons donc choisi de retenir un critère simple, non déclaratif et vérifiable : **le nombre d'anciens élèves occupant le poste de n°1 exécutif (Chief Executive Officer ou équivalent) dans une des 500 plus grandes entreprises internationales**, à la date et avec les éléments (entreprises et dirigeants) du classement « **Fortune Global 500** » établi par le magazine Fortune en 2007 à partir du chiffre d'affaires publié par les entreprises mondiales.

Ce critère se veut être l'équivalent au niveau des entreprises du critère « anciens élèves ayant obtenu le Prix Nobel ou la médaille Fields » utilisé par le classement de l'Université Jiaotong à Shanghaï, les nombres d'anciens concernés étant du même ordre de grandeur. Mais à la différence du classement de Shanghaï, ce critère permet de mettre l'accent sur les performances des formations délivrées dans les établissements d'enseignement supérieur, plutôt que sur les performances de recherche de ces établissements.

Webometrics ranking

90	ROYAL NETHERLANDS METEOROLOGICAL INSTITUTE		175	125	45	240
91	URBAN INSTITUTE		205	62	451	41
92	THOMAS JEFFERSON LAB NATIONAL ACCELERATOR FACILITY		98	198	28	170
93	ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES		117	133	243	183
94	JAPAN AEROESPACE EXPLORATION AGENCY		67	157	116	315
95	RUSSIAN ACADEMY OF SCIENCES		103	171	94	218
96	NYU MEDICAL CENTER NYU SCHOOL OF MEDICINE		104	119	169	364
97	UMR8623 LABORATOIRE DE RECHERCHE EN INFORMATIQUE		61	221	77	147
98	ENTE PER LE NUOVE TECNOLOGIE L'ENERGIA EL'AMBIENTE		96	201	60	191
99	ISTITUTO NAZIONALE DI STATISTICA		105	96	215	445
100	INSTITUT PASTEUR		126	132	213	262

1 | 101 | 201 | 301 | 401 | Institutes 1 to 100 of 500

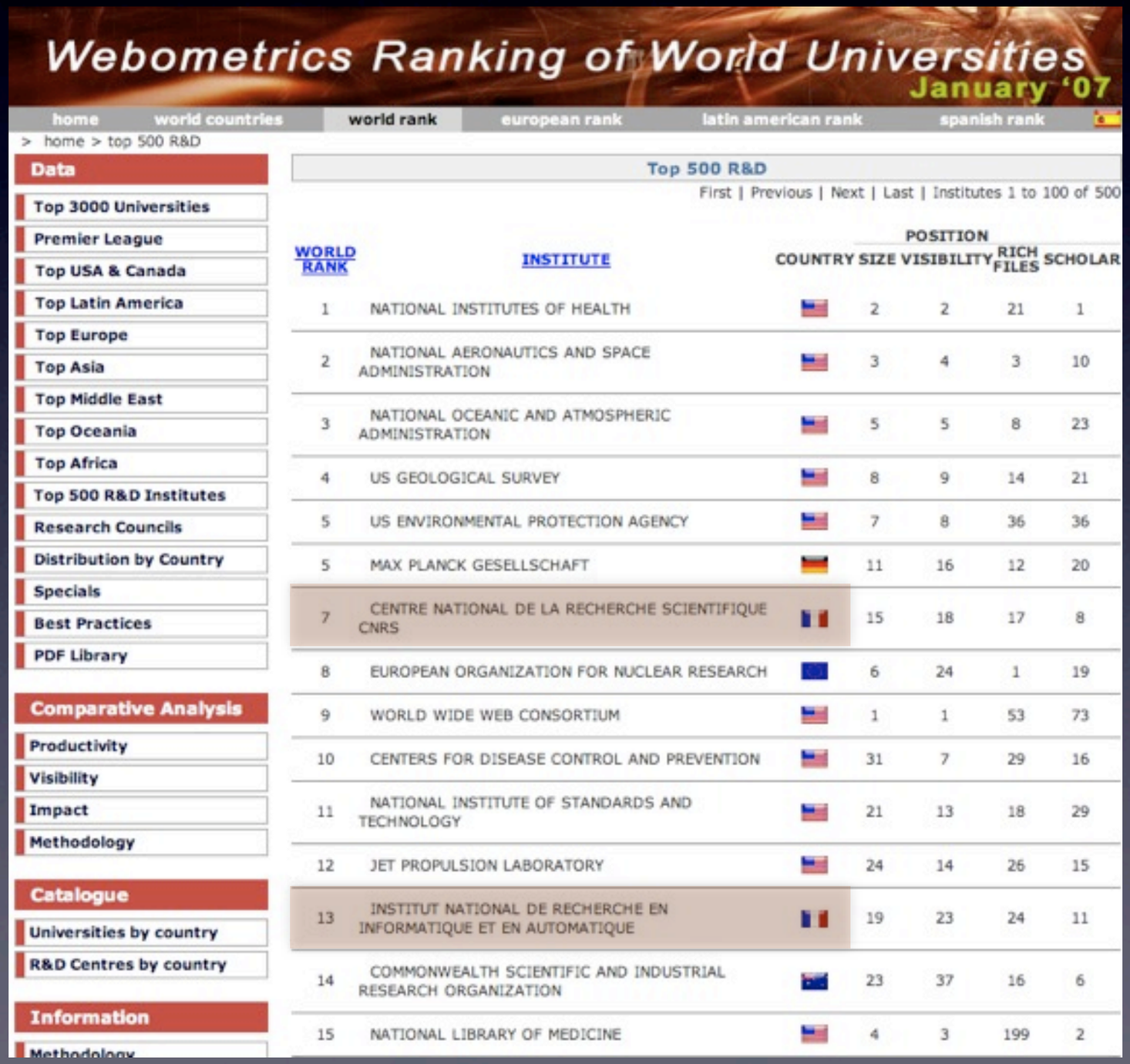
2007

Webometrics

WEBOMETRICS RANK

VISIBILITY (external inlinks) 50%	SIZE (web pages)	20%
	RICH FILES	15%
	SCHOLAR	15%

2007



Webometrics 2009

Top 4000 Universities

First | Previous | Next | Last | Universities 1 to 50 of 4000

WORLD RANK	UNIVERSITY	COUNTRY	POSITION SIZE	VISIBILITY	RICH FILES	SCHOLAR
1	Massachusetts Institute of Technology		1	3	2	6
2	Stanford University		2	2	3	12
3	Harvard University ***		3	1	17	1
4	University of California Berkeley		6	4	5	24
5	Cornell University		4	5	8	37
6	University of Michigan		10	6	15	22
7	California Institute of Technology ***		8	8	21	17
8	University of Minnesota		9	16	4	19
9	University of Illinois Urbana Champaign *		14	10	6	38
10	University of Texas Austin		11	9	10	45

- What do these ranking mean?
- How useful are they? Who uses them?

=> Mostly the media,
but also at the political level

In France, AERES evaluates research labs and grades them from A+ to C

Affects their future funding...

- Reliability issues
  - Quality of the sources
  - Precision of the measures
- Validity issues
  - Relevance of the measures
  - Measuring affects the system being observed

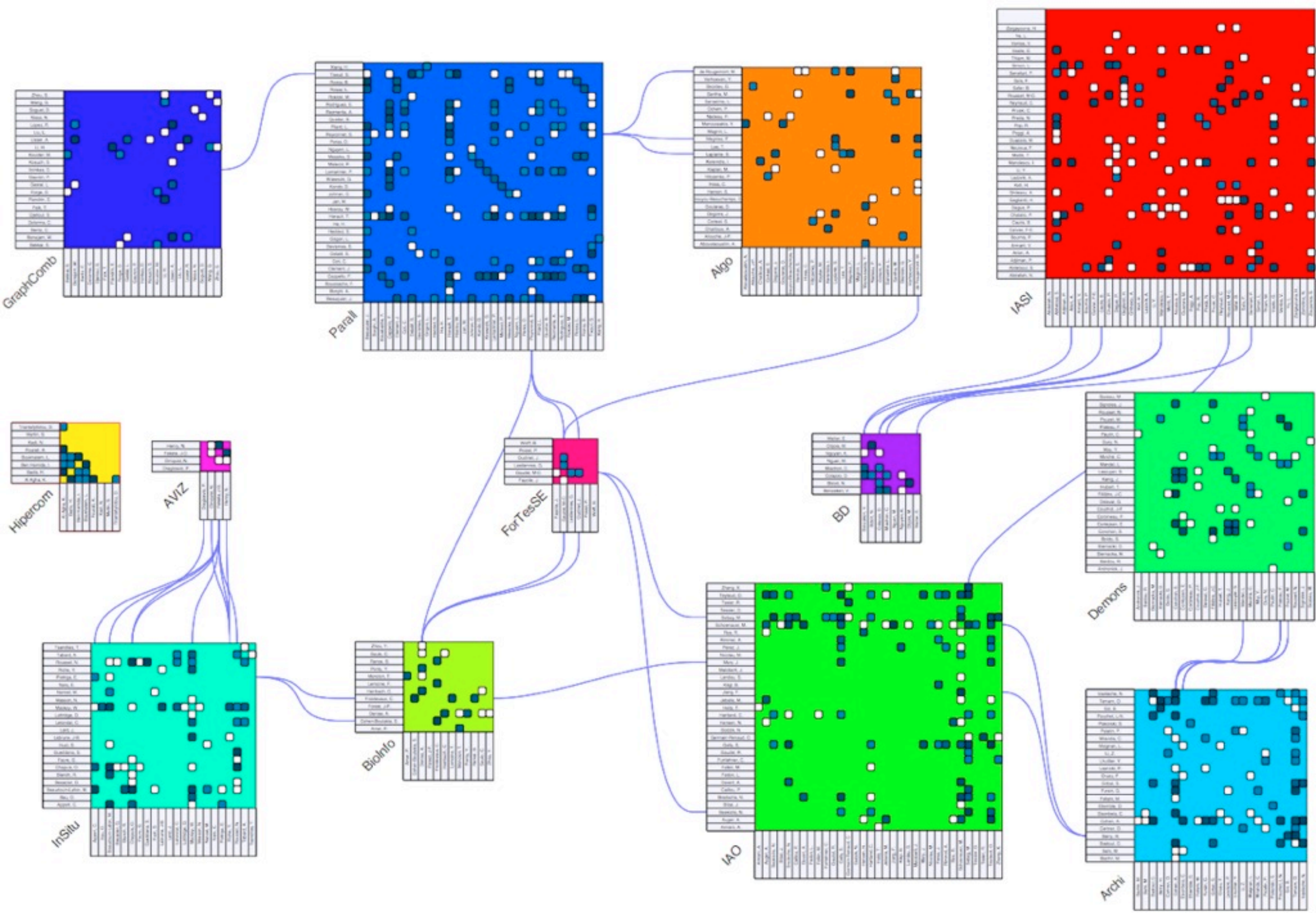
In conclusion...

Correlation between rankings and indices

- Rank correlations of 0.22 between the peer evaluation based quality rating of Netherlands computer science groups and citation impact indicators of their papers
  - Peer rating of 42 academic computer science groups in the Netherlands in 2003 (QANU)
  - ISI database plus conference proceedings from ACM, LNCS, IEEE

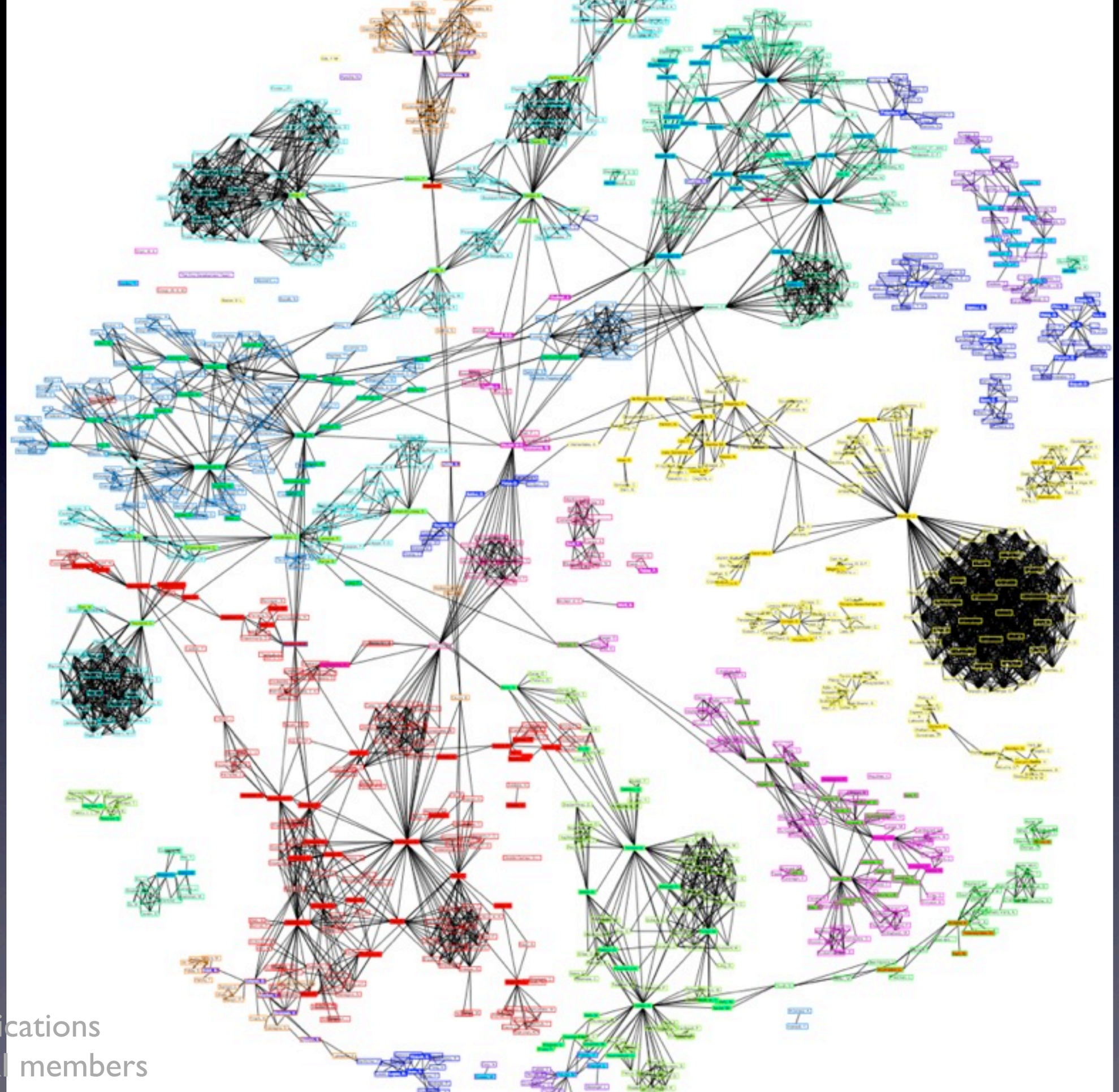
J. Mattern

- "The Mismeasurement of Science" Peter A. Lawrence
  - Measuring a system affects it
  - It is possible to manipulate the measures
    - Publishing practices evolve quickly
- Power law = law of rarity
  - Prediction inherently difficult
- Take rankings and indices for what they are



Co-publications within LRI (2008)

J.D. Fekete - N. Henry



Co-publications
between LRI members

J.D. Fekete
N. Henry