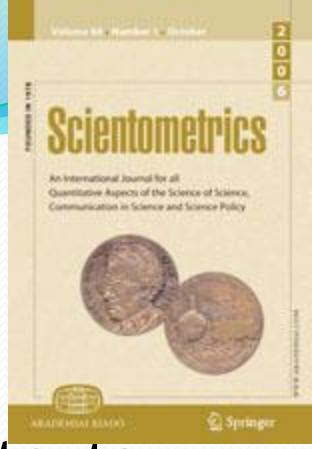
Scientific indices and research evaluation

Reedik Mägi



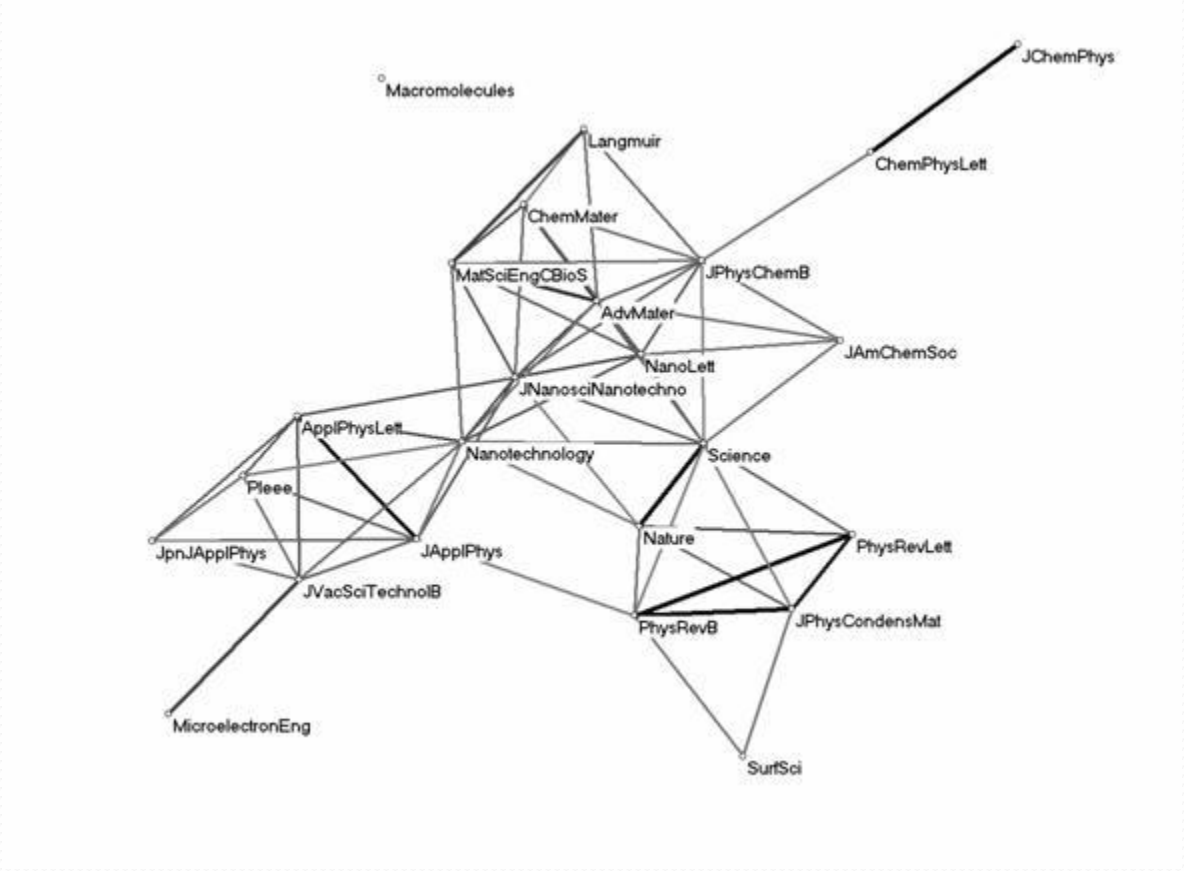
Definition

- **Bibliometrics** is the study, or measurement, of texts and information. Content analysis is a type of bibliometrics. While it is most often used in the field of library and information science, it has wide applications in other areas.

- **Scientometrics** is the science of measuring and analysing science. In practise, scientometrics is often done using bibliometrics that is measurement of (scientific) publications.

http://en.wikipedia.org/

The citation environment of the journal *Nanotechnology* in 2003



How to identify research groups

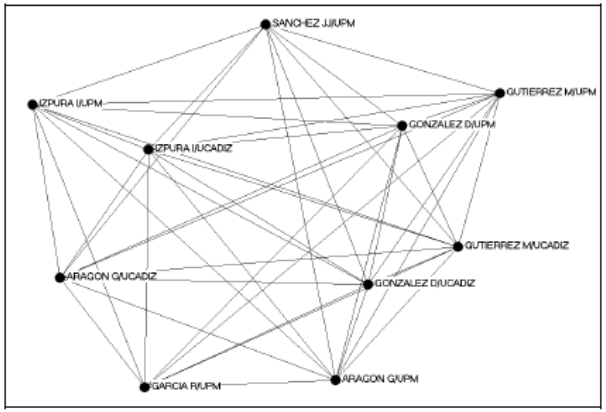


Figure 3. 8-core subgraph based on activity similarity relations
Each pair of AOCS connected represents a similar research profile.

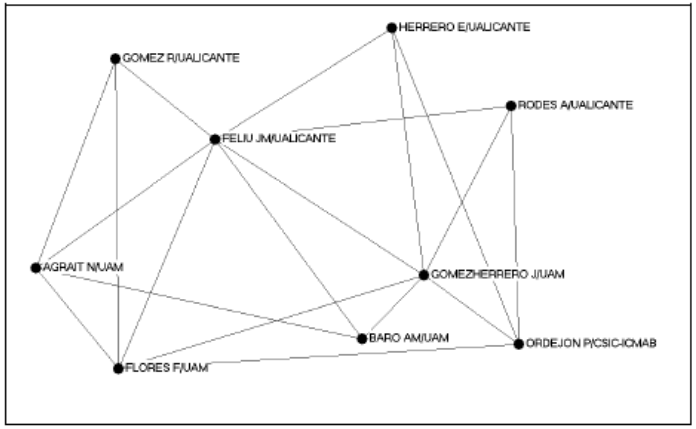


Figure 5. 3-core subgraph based on activity similarity relations
Each pair of AOCS connected represents a similar research profile.

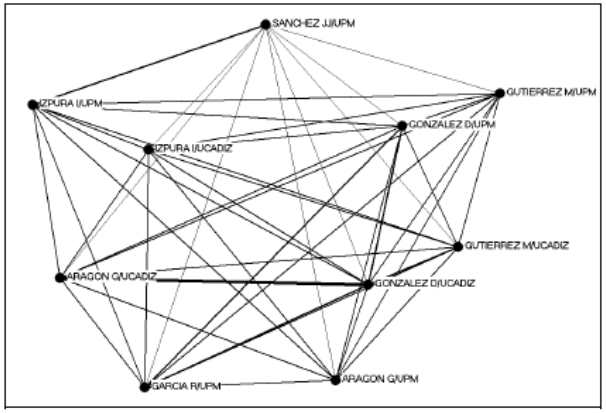


Figure 4. Subgraph based on co-authorship relations
Each pair of AOCS connected shows a co-publishing activity.

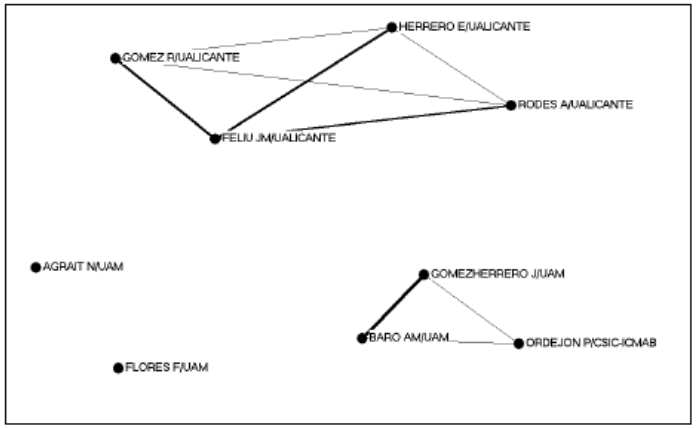


Figure 6. Subgraph based on co-authorship relations
Each pair of AOCS connected shows a co-publishing activity.

C. CALERO et al. (2006) Scientometrics

Citation index

- A **citation index** is an index of citations between publications, allowing the user to easily establish which later documents cite which earlier documents.

- The first citation indices were legal citators such as Shepard's Citations (1873)

  Given a reference of a legal decision, a citator allows the researcher to find newer documents which cite the original document and thus to reconstruct the judicial history of cases and statutes. Using a citator in this way is colloquially referred to as "Shepardizing".

http://en.wikipedia.org/

Different citation indices

- Thompson Scientific – ISI – Web of Science
- Elsevier – Scopus
- CiteSeer – computers and informatics
- RePec – economics
- Google Scholar – new articles only (2006)

http://en.wikipedia.org

ISI

- In 1960, Eugene Garfield's Institute for Scientific Information (ISI) introduced the first citation index for papers published in academic journals, starting with the *Science Citation Index* SCI.

- Accessible through TÜ Library page (utlib.ee):

ISI Web of Knowledge

- Since 2006 http://scholar.google.com/

Contains **Bibliography Manager**

Links to import citations into:

- Endnote
- Bibtex
- Refman
- …

Journal impact factor

- A = the number of times articles published in 2001-2 were cited in indexed journals during 2003
- B = the number of articles, reviews, proceedings or notes published in 2001-2 2003
- impact factor = A/B

Journal Top 20

Rank	Abbreviated Journal Title (linked to journal information)	ISSN	Total Cites	Impact Factor	Immediacy Index	Articles
1	CA-CANCER J CLIN	0007-9235	4218	49.794	21.300	20
2	ANNU REV IMMUNOL	0732-0582	14745	47.400	10.828	29
3	NEW ENGL J MED	0028-4793	167894	44.016	13.422	308
4	ANNU REV BIOCHEM	0066-4154	16313	33.456	4.857	28
5	NAT REV CANCER	1474-175X	9823	31.694	3.935	77
6	SCIENCE	0036-8075	345991	30.927	6.398	827
7	NAT REV IMMUNOL	1474-1733	8686	30.458	3.792	72
8	REV MOD PHYS	0034-6861	19446	30.254	5.633	30
9	NAT REV MOL CELL BIO	1471-0072	11438	29.852	6.225	80
10	CELL	0092-8674	132371	29.431	6.238	319
11	NATURE	0028-0836	372784	29.273	5.825	1065
12	NAT MED	1078-8956	40386	28.878	6.600	155
13	PHYSIOL REV	0031-9333	14943	28.721	4.788	33
14	NAT IMMUNOL	1529-2908	16989	27.011	5.362	130
15	NAT GENET	1061-4036	52387	25.797	5.921	190
16	ANNU REV NEUROSCI	0147-006X	8563	24.184	2.263	19
17	LANCET	0140-6736	131616	23.878	7.347	360
18	ANNU REV CELL DEV BI	1081-0706	7097	23.690	0.857	28
19	JAMA-J AM MED ASSOC	0098-7484	95715	23.494	5.082	380
20	NAT BIOTECHNOL	1087-0156	20914	22.738	5.210	124

Journal Citation Reports Science Edition 2005

Acknowledgment index

- An **acknowledgment index** is an experimental method for analyzing the scientific literature; it quantifies the acknowledgements in scientific journals.

Quantification of scientific output of individual scientific authors

- Total number of papers (N_p).

  Advantage: measures productivity.

  Disadvantage: does not measure importance or impact of papers.

- Total number of citations ($N_{c,tot}$).

  Advantage: measures total impact.

  Disadvantage: hard to find and may be inflated by a small number of "big hits," which may not be representative of the individual if he or she is a coauthor with many others on those papers. Another disadvantage is that $N_{c,tot}$ gives undue weight to highly cited review articles versus original research contributions.

Hirsch 2005 PNAS

Quantification of scientific output of individual scientific authors

- Citations per paper (i.e., ratio of $N_{c,tot}$ to N_p). Advantage: allows comparison of scientists of different ages.

  Disadvantage: hard to find, rewards low productivity, and penalizes high productivity

- Number of "significant papers," defined as the number of papers with >*y* citations (for example, *y* = 50).

  Advantage: eliminates the disadvantages of criteria *i*, *ii*, and *iii* and gives an idea of broad and sustained impact.
  Disadvantage: *y* is arbitrary and will randomly favor or disfavor individuals, and *y* needs to be adjusted for different levels of seniority.

Quantification of scientific output of individual scientific authors

- Number of citations to each of the q most-cited papers (for example, $q = 5$).

  Advantage: overcomes many of the disadvantages of the criteria above.

  Disadvantage: It is not a single number, making it more difficult to obtain and compare. Also, q is arbitrary and will randomly favor and disfavor individuals

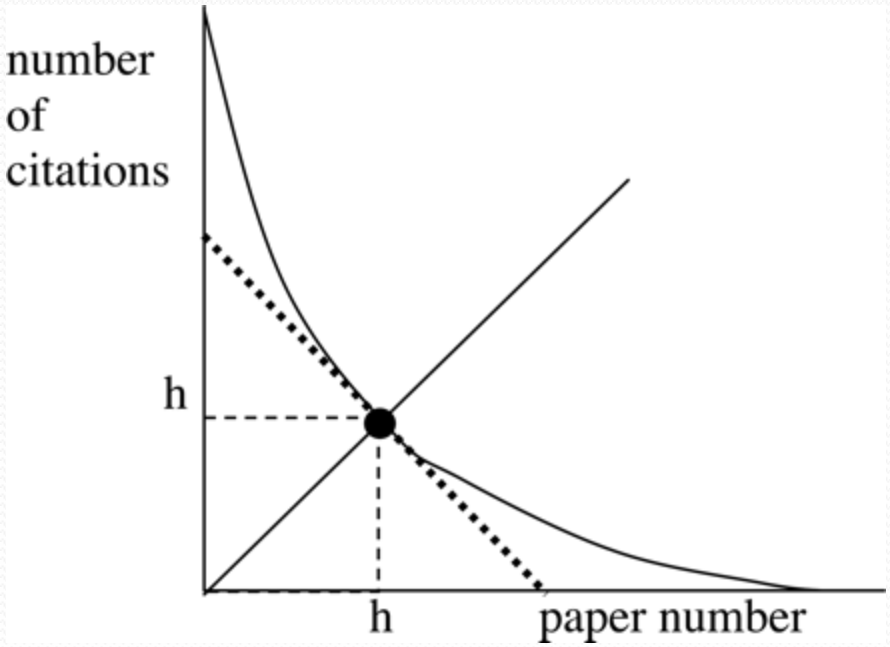
H-index

- The *h*-index, also known as the Hirsch number is a number suggested by Jorge E. Hirsch in 2005 for the quantification of scientific output of individual scientific authors

- A scientist has index h if h of his/her N_p papers have at least h citations each, and the other (N_p - h) papers have fewer than h citations each.

  In other words, a scholar with an index of *h* has published *h* papers with at least *h* citations each.

H-index



Records 1 to 10 PRINT E-MAIL SAVE

This report reflects citations to source items indexed within Web of Science. Perform a Cited Reference Search to include citations to items not indexed within Web of Science.

14 results found
Records 1 -- 10

Go to Page: 1 of 2 GO
[1 | 2 ...]

Times Cited SORT

Use the checkboxes to remove individual items from Citation Report or restrict to items processed between 1980 and 2007 GO

		2003	2004	2005	2006	2007	Total	Average Citations per Year
		18	21	41	30	5	125	11.36
1.	Zobel M, Otsus M, Liira J, et al. Is small-scale species richness limited by seed availability or microsite availability? ECOLOGY 81 (12): 3274-3282 DEC 2000	12	12	22	18	2	73	10.43
2.	Eichelmann H, Oja V, Rasulov B, et al. Photosynthetic parameters of birch (Betula pendula Roth) leaves growing in normal and in CO2- and O-3-enriched atmospheres PLANT CELL AND ENVIRONMENT 27 (4): 479-495 APR 2004		1	7	5	0	13	3.25
3.	Salumets A, Suikkari AM, **Mols T**, et al. Influence of oocytes and spermatozoa on early embryonic development FERTILITY AND STERILITY 78 (5): 1082-1087 NOV 2002	5	1	2	3	0	11	1.83
4.	Kangur K, **Mols T**, Milius A, et al. Phytoplankton response to changed nutrient level in Lake Peipsi (Estonia) in 1992-2001 HYDROBIOLOGIA 506 (1-3): 265-272 NOV 15 2003	0	5	2	2	0	9	2.25
5.	Paal J, Fremstad E, **Mols T** Responses of the Norwegian alpine Betula nana community to nitrogen fertilization CANADIAN JOURNAL OF BOTANY-REVUE CANADIENNE DE BOTANIQUE 75 (1): 108-120 JAN 1997	0	1	1	0	0	4	0.36
6.	Timm H, Ivask M, **Mols T** Response of macroinvertebrates and water quality to long-term decrease in organic pollution in some Estonian streams during 1990-1998 HYDROBIOLOGIA 464 (1-3): 153-164 NOV 2001	1	0	1	1	0	3	0.50
7.	**Mols T**, Starast H, Milius A, et al. The hydrochemical state of Lake Peipsi-Pihkva HYDROBIOLOGIA 338 (1-3): 37-47 NOV 8 1996	0	1	1	0	0	3	0.27
8.	Timm H, **Mols T** Macrozoobenthos of lake verevi HYDROBIOLOGIA 547: 185-195 SEP 15 2005			2	0	0	2	0.67
9.	Sammul M, Kull K, Niitla T, et al. A comparison of plant communities on the basis of their clonal growth patterns EVOLUTIONARY ECOLOGY 18 (5-6): 443-467 SEP 2004		0	1	0	1	2	0.67
10.	Fremstad E, Paal J, **Mols T** Impacts of increased nitrogen supply on Norwegian lichen-rich alpine communities: a 10-year experiment JOURNAL OF ECOLOGY 93 (3): 471-481 JUN 2005			0	0	2	2	0.67

14 results found
Records 1 -- 10

Go to Page: 1 of 2 GO
[1 | 2 ...]

H-index

- Physics:

  h=10 – scientist,

  h=20 – prof. of good university,

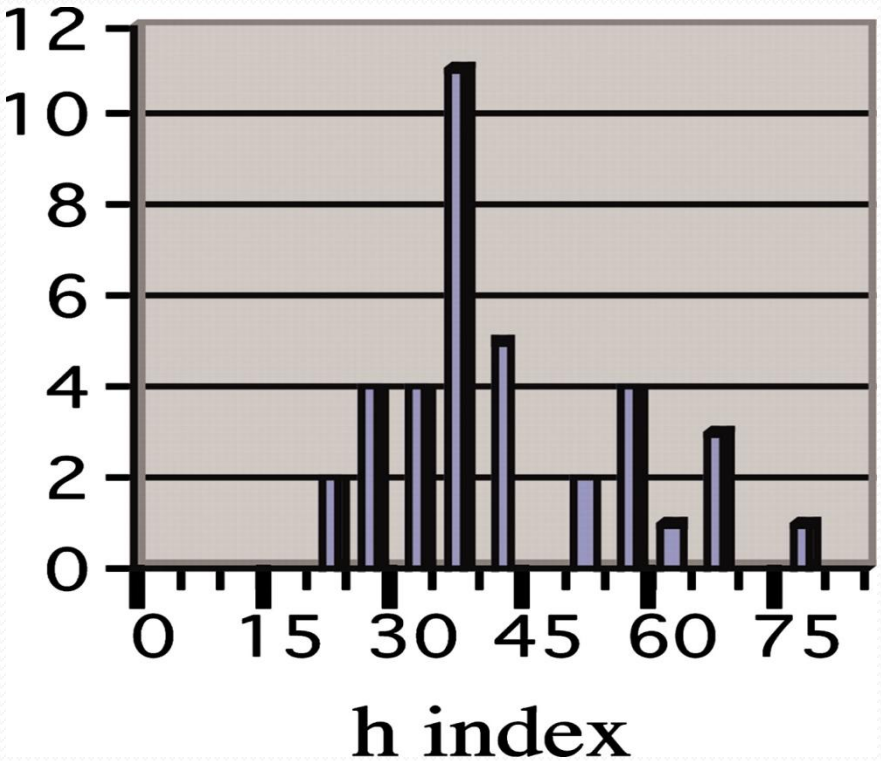
  h=40 – top scientist

  h=60 – Nobel winner

(after 20 years of scientific activity)

Aaviksoo PDF

Histogram giving the number of Nobel prize recipients in physics



… in the last 20 years versus their *h* index.
The peak is at the *h* index between 35 and 39

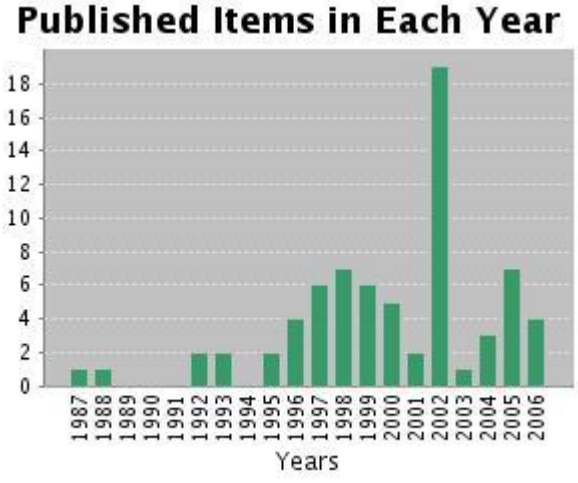
g-index

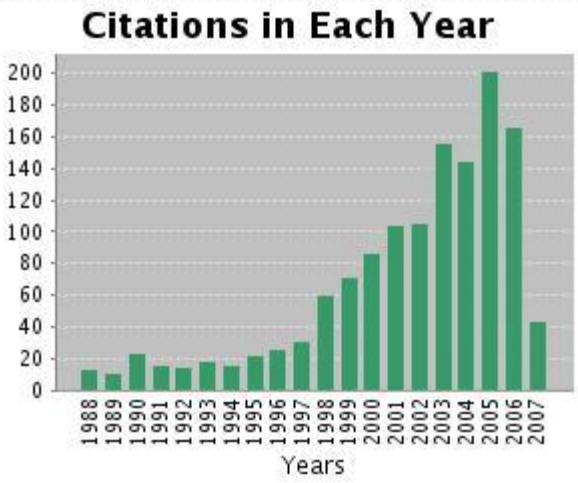
- was suggested in 2006 by Leo Egghe.
- The index is calculated based on the distribution of citations received by a given researcher's publications.
- *Given a set of articles ranked in decreasing order of the number of citations that they received, the g-index is the (unique) largest number such that the top g articles received (together) at least g^2 citations.* This index is very similar to the h-index, and attempts to address its shortcomings

Google Pagerank

- The Google Pagerank system has also been proposed to assess citation impact.

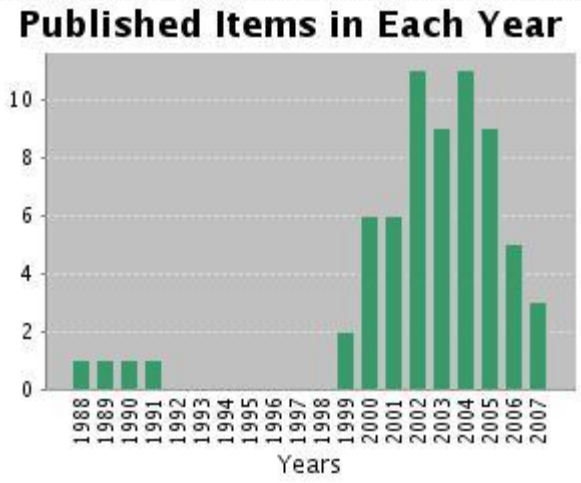
Andres Metspalu

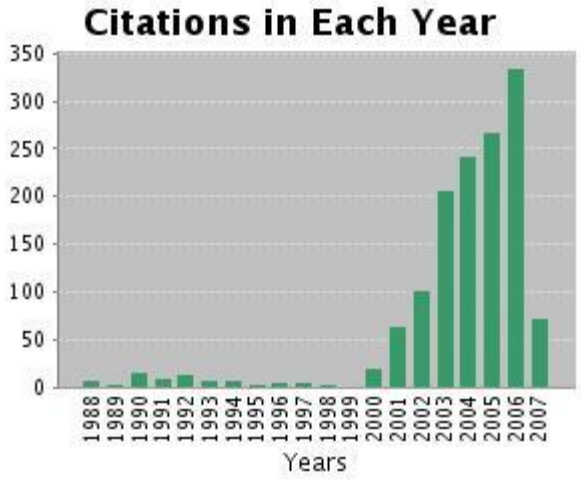




Results found:	76
Sum of the Times Cited :	1,375
Average Citations per Item :	18.09
h-index :	16

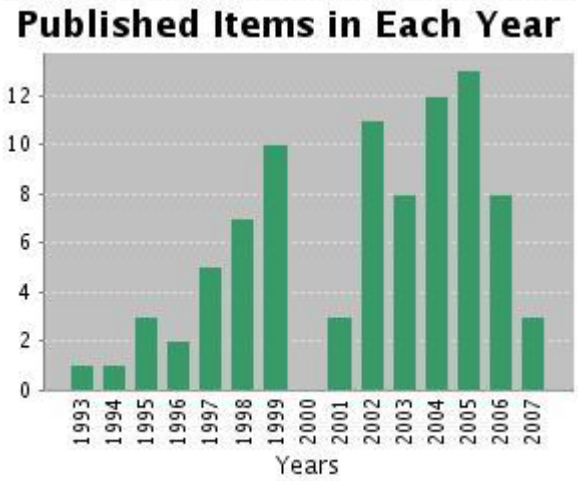
Richard Villems

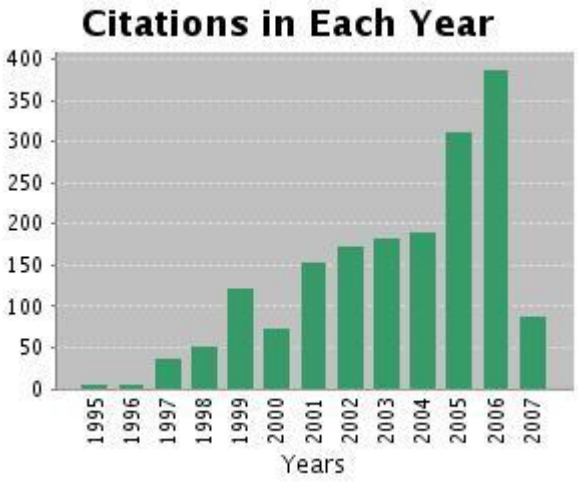




Results found:	89
Sum of the Times Cited :	1,502
Average Citations per Item :	16.88
h-index :	19

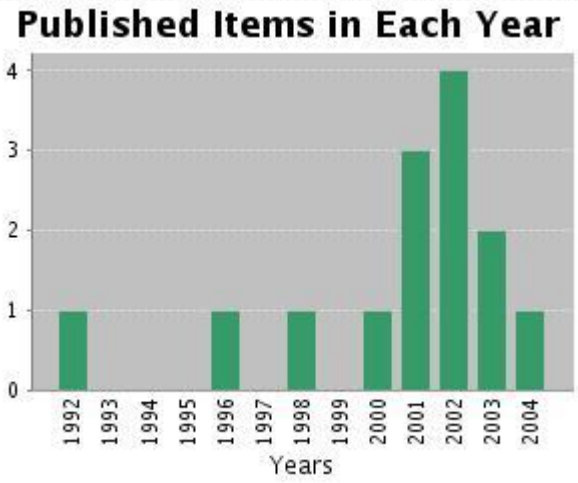
Ülo Niinemets

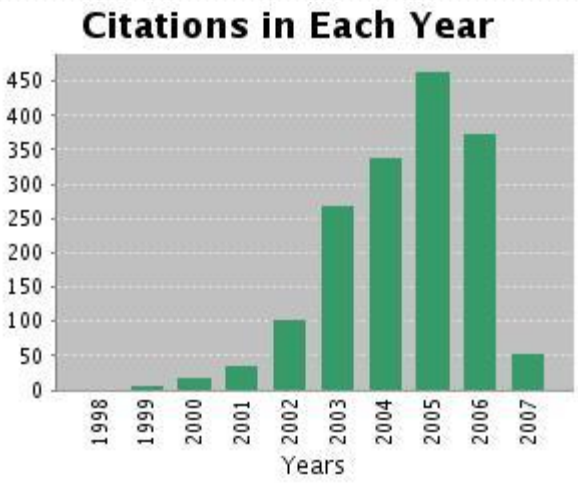




Results found:	87
Sum of the Times Cited :	1,789
Average Citations per Item :	20.56
h-index :	23

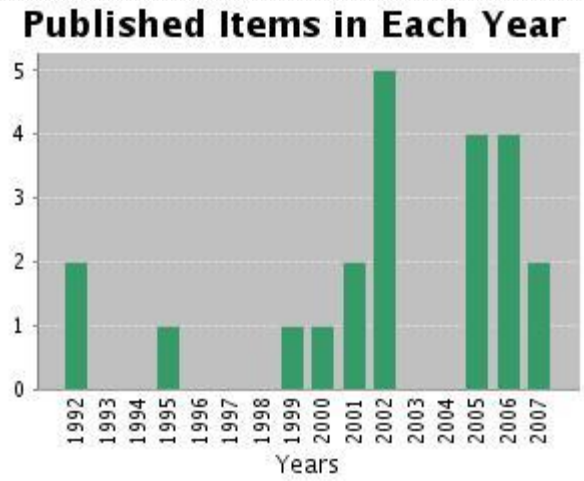
Jaak Vilo

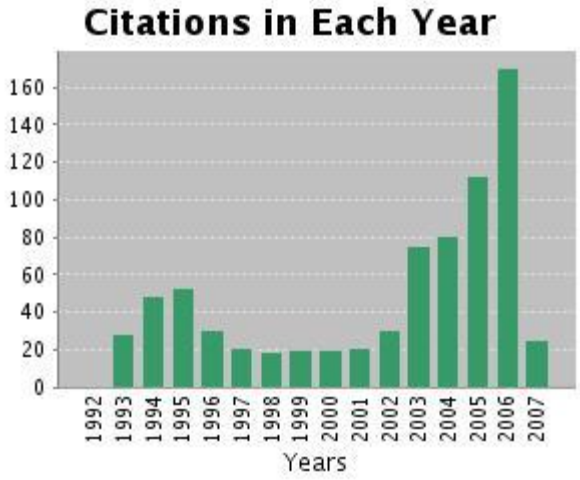




Results found:	14
Sum of the Times Cited :	1,672
Average Citations per Item :	119.43
h-index :	10

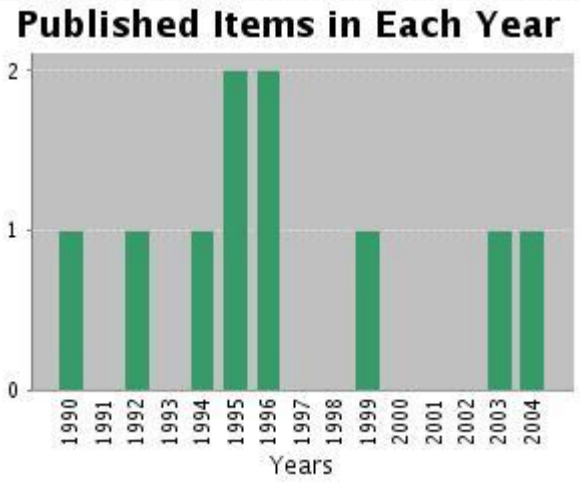
Maido Remm

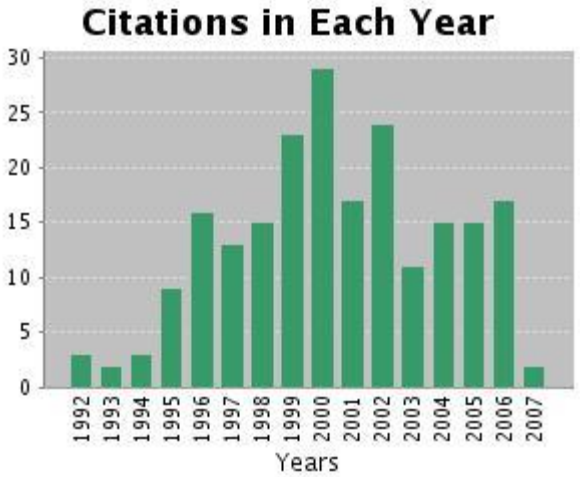




Results found:	22
Sum of the Times Cited :	757
Average Citations per Item :	34.41
h-index :	8

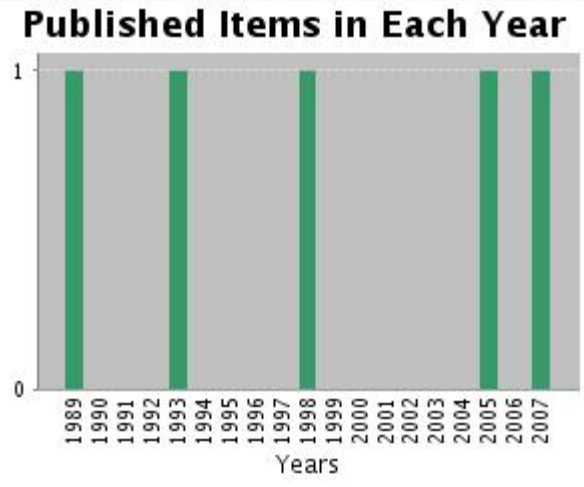
Ülo Puurand

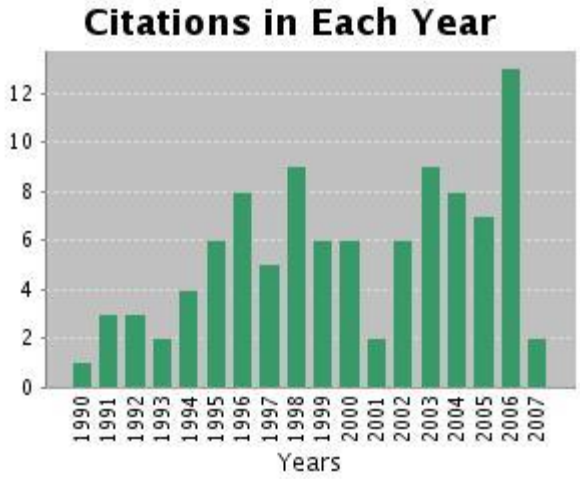




Results found:	10
Sum of the Times Cited :	214
Average Citations per Item :	21.40
h-index :	8

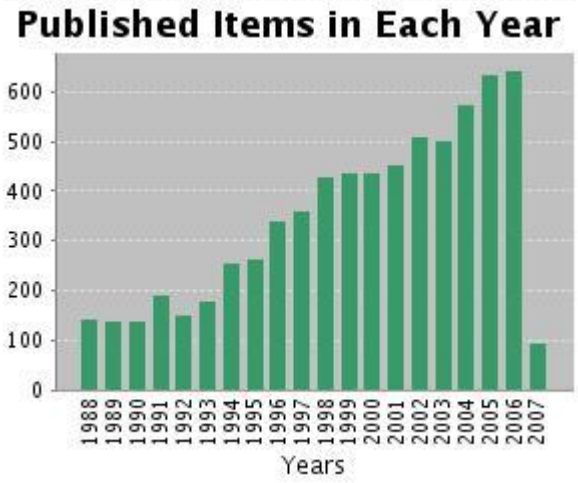
Tõnu Margus

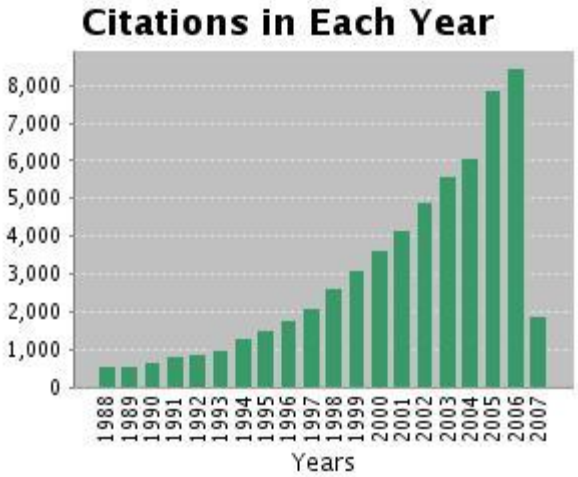




Results found:	5
Sum of the Times Cited :	100
Average Citations per Item :	20.00
h-index :	4

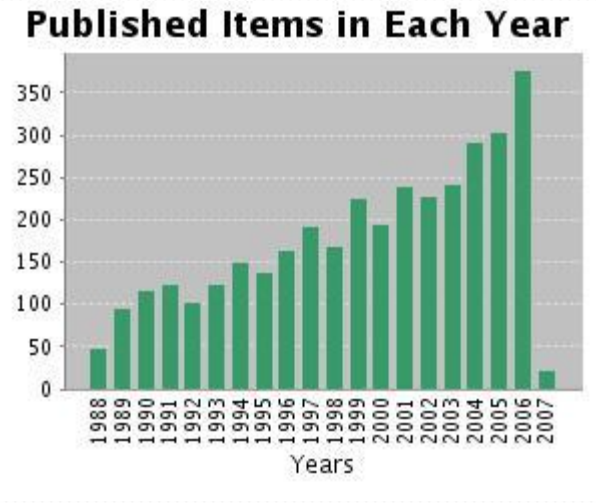
Tartu

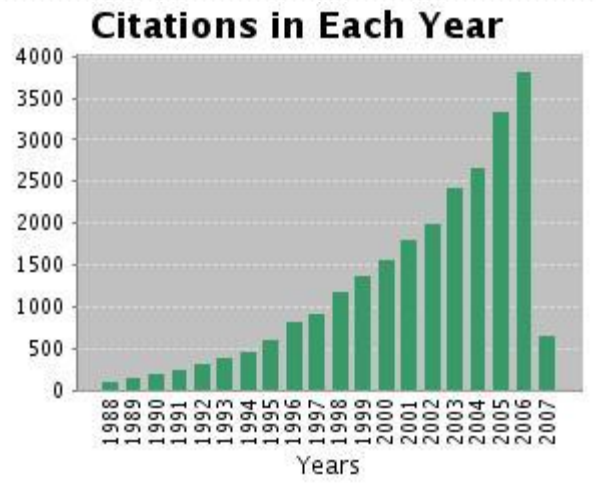




Results found:	7,945
Sum of the Times Cited :	61,504
Average Citations per Item :	7.74
h-index :	81

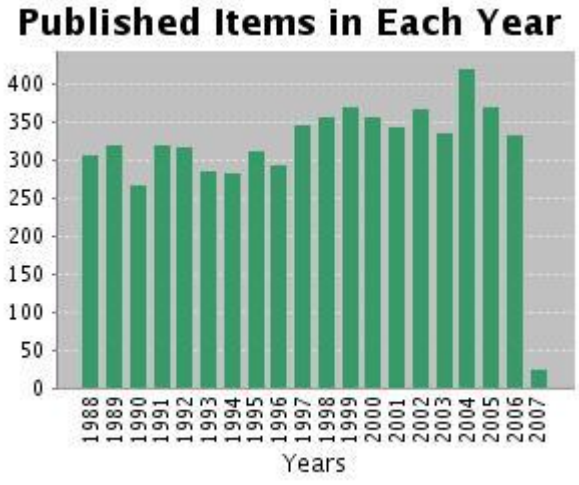
Tallinn

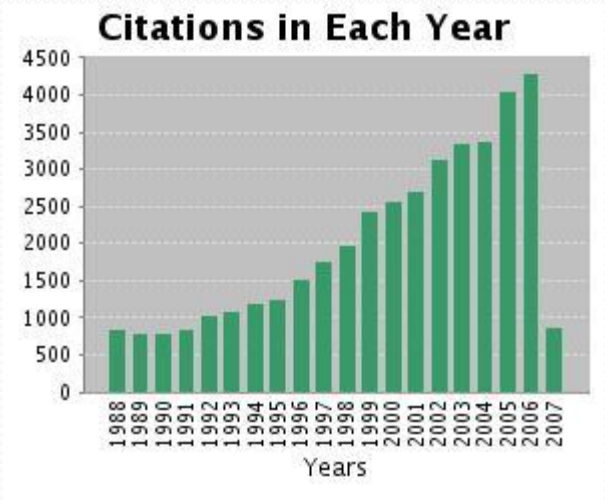




Results found:	3,634
Sum of the Times Cited :	25,316
Average Citations per Item :	6.97
h-index :	57

Riia





Results found:	8,565
Sum of the Times Cited :	43,404
Average Citations per Item :	5.07
h-index :	65

Helsinki

91,258 results found

Citation Report:

The Citation Report feature is not available from a set containing more than 10,000 articles.

Country rankings by publication performance 1994–2003

Country	Population (000,000s)	Papers	Citations	Citations per paper	Papers per 1,000,000 population	GERD in 000,000/per 1000 papers
Bulgaria	7.8	17003	70774	4.16	2180	5.2
Cyprus	0.8	1827	10195	5.58	2284	21.1
Czech	10.3	47200	237462	5.03	4583	21.7
Estonia	1.4	5882	37008	6.29	4201	10.5
Latvia	2.4	3587	16569	4.62	1495	10.8
Lithuania	3.6	5221	23940	4.59	1450	21.2
Malta	0.04	492	3249	6.6	12300	n.a.
Poland	38.7	105530	503306	4.77	2734	10.3
Romania	22.4	19310	63482	3.29	862	10.6
Slovakia	5.4	21843	92685	4.24	4045	7.5
Slovenia	1.9	14252	66074	4.64	7501	26.2
Turkey	65.7	64307	187746	2.92	979	19.9
Sweden	9.0	162201	1804859	11.13	18022	64.5
US	293.0	2832621	36297842	12.81	9668	94.8
Japan	127.3	759449	5775093	7.6	5966	173.5

Must_scientometrics_2006

The number of publications by fields 1994–2003

Country / Field	BG	CZ	CY	EE	HU	LV	MT	LT	PL	RO	SK	SI	TR	SE	US	Total
Agriculture	236	1222	0	0	1294	0	0	0	1122	63	410	212	1306	1867	38744	46480
Biology, Biochemistry	1650	3175	46	328	3169	174	0	324	6694	410	2237	843	2158	13852	210809	245887
Chemistry	3756	8559	0	646	8447	854	0	799	23064	5877	3937	2122	6261	13530	216632	294526
Clinical Medicine	884	3520	0	661	4544	0	0	339	7088	530	1156	1710	19691	43754	650999	734889
Computer Science	236	610	117	46	856	59	0	56	1380	318	249	348	786	2205	56853	64121
Ecology	219	909	0	302	522	0	0	0	1706	114	636	278	1422	5892	70783	82786
Economics	53	794	132	0	244	0	0	0	244	16	536	100	464	1762	64096	68443
Engineering	1481	2531	200	293	2768	364	0	453	7873	2153	973	2085	5430	8486	190301	225406
Geosciences	350	1158	0	507	654	34	0	67	1409	289	489	117	1359	3748	77393	87579
Immunology	0	367	0	71	478	0	0	37	676	76	112	85	309	4712	53304	60228
Materials Science	1420	2294	0	161	1319	461	0	421	4713	1686	1320	1034	1831	5420	67375	89468

The number of publications by fields 1994–2003

Country / Field	BG	CZ	CY	EE	HU	LV	MT	LT	PL	RO	SK	SI	TR	SE	US	Total
Mathematics	790	1636	113	125	2414	0	0	270	4170	1296	644	607	824	2240	62660	77798
Microbiology	203	991	0	107	514	67	0	64	729	32	694	171	321	2764	49634	56294
Molecular Biology, Genetics	363	888	62	144	942	55	0	104	1588	75	342	208	543	4893	113090	123301
Multi-disciplinary	370	25	3	10	66	6	0	0	70	0	0	0	0	206	8454	9212
Neuroscience	251	591	0	174	1949	0	0	41	2199	58	252	149	1141	7469	115264	129540
Pharmacology, Toxicology	443	524	0	88	1026	22	0	0	1935	146	417	209	1579	3358	47122	56873
Physics	3057	6269	386	783	5266	840	0	1107	21979	4459	2919	1931	3526	14240	210162	276957
Plant & Animal Science	916	3890	0	502	3043	121	0	188	6673	161	1510	539	3197	9070	148440	178264
Psychiatry, Psychology	58	476	0	108	257	11	8	20	276	28	431	88	506	2731	113174	118173
Social Sciences	135	662	58	139	550	0	64	52	488	0	306	254	617	4179	184536	192043
Space Science	383	785	0	194	624	0	0	102	2079	158	247	45	288	1758	49768	56435

How much science costs?

- Mean expence per scientist (with infrastructure and machines) is ~1.5 mil kr. per year (US 1 mil $, EU 0.5 MEUR)

- Critical minimus size of workgroup is 4-5 persons

- In-house research minimum investment starts with 10MEEK * 5 years

Aaviksoo PDF