Which h-index? – A comparison of WoS, Scopus and Google Scholar

JUDIT BAR-ILAN

Department of Information Science, Bar-Ilan University, Ramat Gan (Israel)

This paper compares the h-indices of a list of highly-cited Israeli researchers based on citations counts retrieved from the Web of Science, Scopus and Google Scholar respectively. In several case the results obtained through Google Scholar are considerably different from the results based on the Web of Science and Scopus. Data cleansing is discussed extensively.

Introduction

Until just a few years ago, when citation information was needed the single most comprehensive source was the ISI Citation Indexes. Although the Citation Indexes were often criticized for various reasons, there was no other source to rely on. Data from the ISI Citation Indexes and the Journal Citation Reports are routinely used by promotion committees at universities all over the world. In this paper we refer to the Web version of Citation Indexes, i.e., to the Web of Science (WOS) (http://portal.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame).

Recently two alternatives to the ISI Citation Indexes have become available. One of them is Scopus (http://www.scopus.com/) developed by Elsevier and the other is the freely available Google Scholar (http://scholar.google.com/). Each of these has a

different collection policy which affects both the publications covered and the number of citations to the publications. How different are these citation databases? In this paper we tried to provide a partial answer by considering the h-indexes [Hirsch, 2005a, b] of a group of highly cited researchers based on each of the three citation databases.

Literature review

Comparing the databases

The Science Citation Index was first published in print in 1963 with citation data from 1961 [Garfield, 1963]. "The Web of Science provides seamless access to current and retrospective multidisciplinary information from approximately 8,700 of the most prestigious, high impact research journals in the world." [Thomson Scientific, no date]. The references from all the indexed items are extracted and the cited reference interface lists all citation to works of an author regardless of whether the cited items are indexed by WOS or not.

Until very recently, the Web of Science was the only comprehensive database to provide citation data. However, in November 2004 the citation database scenery changed considerably at once by the launching of Scopus on November 3, 2004 [Reed Elsevier, 2004] and Google Scholar on November 18, 2004 [Payne, 2004]. Scopus provides full citation coverage from 1996 and onwards, and claims to be "the largest abstract and citation database of research literature and select results from the web. Scopus covers 27 million abstracts, 230 million references and 200 million web pages" [Scopus, no date]. Scopus provides citation data only for the items indexed by it.

Google, probably on purpose, does not provide any explicit information either about the number of records or about its time coverage. Google Scholar, unlike WOS and Scopus, is freely accessibly. They index data from publishers only if the publisher is willing to provide at least the abstract of the paper freely [Google Scholar, 2005] – viewing the full text may be fee or subscription based. The data comes from other sources as well, like freely available full text from preprint servers or personal websites as well, thus in many cases the full text is freely available for all users. References are automatically extracted from the full text of the indexed items. In case the reference itself is not indexed by Google Scholar, only the number of citation to that item appear in the search results.

Google Scholar was and is received with mixed feelings. For example, Giles [2005] reports: "Although there are no detailed studies, many librarians report that faculty members and students are beginning to use the search engine; some suspect that Scholar will replace more established, and more costly, search tools. It already directs more online traffic to Nature websites than any other multidisciplinary science search engine." [Giles, 2005 : 554]. Librarians seem to be less enthusiastic than their clients:

in the summer of 2005 only a minority of the 113 ARL library sites linked to Google Scholar [MULLEN & HARTMAN, 2006].

Evaluation studies, mainly comparing Google Scholar (GS) with the Web of Science (WOS) had mixed results as well. In an early study, BAUER & BAKKALBASI [2005] analyzed the citation counts of JASIST articles published in 1985 and 2000. The results for 1985 were inconclusive, but for 2000, the citation counts in GS were considerably higher than either in WOS or in Scopus. They conclude: "Based on our preliminary examination and discovery of higher citation counts, we recommend that researchers should consult Google Scholar in addition to Web of Science or Scopus, especially for a relatively recent article, author or subject area." Indeed when we carried out the literature search on the h-index for this paper, Google Scholar retrieved a considerably larger number of items than either from WOS or Scopus. However, one must note that a large number of these references were to preprint repositories. Preprint repositories allow researchers to become updated on recent developments, but they should be evaluated with care, since they are not peer-reviewed publications. JACSO [2006] criticizes the attention the BAUER & BAKKALBASI [2005] article received by the media (news and blogs), where only the results for 2000, based on 105 articles were highlighted.

Jacso emphasizes shortcomings of Google Scholar in several articles; he shows inconsistencies in Google Scholar. For example [JACSO, 2006] one of the most prolific authors according to Google Scholar is "I Introduction" with 40,100 reported documents authored by him/her. We rerun the test on November 4, 2006 – the number of items authored by "I Introduction" (the number of results for the query *author:"I Introduction"*) increased to 689,000 (!). Another example, more related to informetrics and webometrics was located by us, when we looked for the Almind & Ingwersen article "Informetric analyses on the World Wide Web". Google Scholar indexed this article which was published in the Journal of Documentation in 1997, and reports that it was cited 197 times (as of November 4, 2006), however Google Scholar is sure that the paper was authored by D. Copenhagen (see Figure 1). Citation counts are deflated both in WOS and in GS, when citations are not grouped together. This is especially emphasized for WOS, which counts exact matches only, as can be seen when carrying out a cited reference search. On the other hand, citations are sometimes inflated in Google Scholar, since [JACSO, 2006] it indexes non-scholarly sources as well. Google Scholar often indexes both the preprint and the journal version of a paper (see Figure 2 for an interesting example we found). What should be the true citation count in this case? The sum of the two counts? A closer examination of the citing papers show that often both sources are cited side-by-side, and it is hard to imagine that the citing authors meant to count both citations separately. A possible exception is the current paper we cited both the journal and the preprint version of the HIRSCH paper [2005A, B] on purpose. The issue of multiple manifestations of a work was discussed extensively in [BAR-ILAN, 2006].
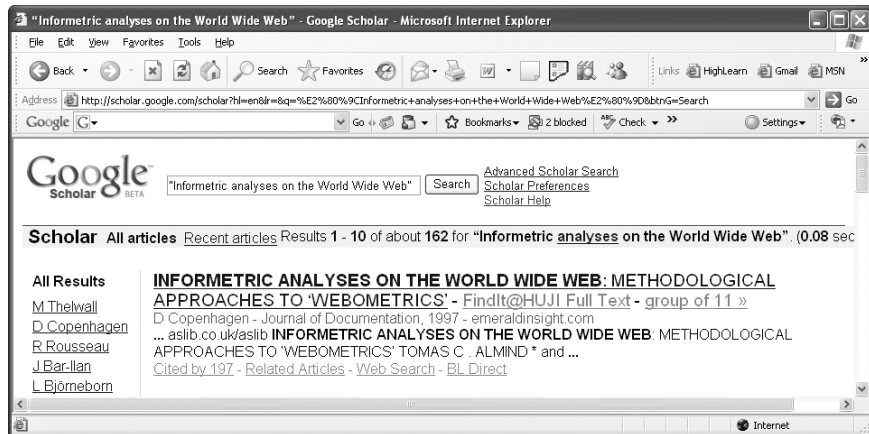
Figure 1. A highly cited paper by "D Copenhagen"

JACSO [2006] also notes that neither the Boolean operators nor the range operator (for limiting the date of publication) work properly in GS. These problems are probably directly "inherited" from Google [BAR-ILAN, 2005]. Google Scholar is not always able to correctly identify the publication year of the item, and citations are not always attributed to the correct publication [JACSO, 2006, 2005A]. JACSO [2006] concludes that Google Scholar cannot be a substitute for WOS, unlike the conclusion of PAULY & STERGIOU [2005] based on testing the citation counts of 114 papers from different scientific disciplines. RAHM & THOR [2005] point out the usefulness of Google Scholar in evaluations in the area of computer science. Note that in computer science a major publication venue are proceedings that are only very partially indexed by the Web of Science [BAR-ILAN, 2006; VISSER & MOED, 2005, 2006]. GARDNER & ENG [2006] compared Google Scholar with well-known Social Science databases, and although aware of its current shortcomings, they conclude that "Google Scholar is still in beta testing, so it has the potential to improve significantly before it becomes fully operational".

JACSO [2005A] examined citations received between 1996–2005 for Garfield's 1955 paper in Science: WOS listed 83 citations, Scopus 76 citations and Google Scholar 82; however only 33 of the citing items appeared in all three databases – showing that citation counts are not everything. KOUSHA & THELWALL [2006B] also found that the overlap of citing documents between WOS and Google Scholar is rather low in some cases (only 33% for chemistry). In terms of the number of indexed items (unknown for GS), JACSO [2005A] found that Scopus overtakes WOS by 2000. It seems that Google Scholar does not fully index items on its partner sites, as demonstrated by JACSO [2005B] and by the search interface specially developed by Peter Jacso to illustrate the

spotty coverage (available at http://www2.hawaii.edu/~jacso/scholarly/side-by-side2.htm). Rather interestingly, NEUHAUS & AL. [2006] found excellent coverage in the sciences and the life science – 100% for PubMed, and only complained about the low coverage in the Social Science. On the other hand, KOUSHA & THELWALL [2006A] found good coverage of the Social Sciences. NORUZI [2005] studied the citation counts reported by WOS and Google Scholar on a set of webometrics papers: in most cases GS had higher citation counts than WOS. These findings are supported by the results of VAUGHAN & SHAW [2006] for Information Science in general. BELEW [2005] tested the citation counts of 203 publications reported by WOS and Google Scholar respectively, and found "surprisingly good agreement between data citation counts provided by the two services". SMITH'S [2006] results show that there is high correlation between the Google Scholar citation counts and New Zealand's Performance Based Research Funding research assessment exercise.

BAKKALBASI & AL. [2006] compared citation counts reported by WOS, Scopus and GS, to publications in twenty two journals in oncology and in condensed matter physics published in 1993 and 2003. The results varied with publication year and discipline, and the findings could not "identify any of these three resources as the answer to all citation tracking needs". BAR-ILAN & AL. [2007] compared the rankings of the publications of highly-cited Israeli researchers induced by the citations counts reported by WOS, Scopus and Scholar. The computed measures show high similarity between Scopus and WOS and lower similarities between Google Scholar and the other tools, indicating that Google Scholar's coverage is considerably different from that of WOS and Scopus.

A few studies compared WOS with Scopus – they emphasize the wider coverage (in terms of the number of indexed publications), the user friendliness and the shorter time span of Scopus when compared with WOS [JACSO, 2004; LAGUARDIA, 2005; DEIS & GOODMAN, 2005; BURNHAM, 2006].

The h-index

The new bibliometric measure, the *h-index* was introduced by Jorge Hirsch in August 2005 [HIRSCH, 2005A, B], and it is defined as follows

A scientist has index h if h of his/her Np papers have at least h citations each, and the other (Np − h) papers have no more than h citations each [HIRSCH, 2005B : 16569].

The new measure raised considerable interest in informetric circles and in a short period of time, a considerable number of publications discussed and further developed the ideas introduced by Hirsch.

The h-index was applied to compare scientists. HIRSCH [2005A, B] calculated the h-indices of prominent physicists. GLÄNZEL & PERSSON [2005] computed the h-index

for Price medalists based on data from the Web of Science, Bar-Ilan [2006b] recomputed the values for the same list of people based on data from Google Scholar. Cronin & Meho [2006] computed the h-indices of prominent American information scientists, Butler & McAllister [2006] studied the applicability of the h-index for researchers in the Social Sciences, Saad [2006] compared data obtained from WOS and GS for consumer scholars, Bornmann & Daniel [2005] studied the relation between the h-index and the acceptance of post-doctoral grants.

Van Raan [2005] and Moed [2005] prefer to consider the research group as the basic unit for computing the h-index. A number of researchers noted that one should take into account the scientific ages of the authors when calculating the h-index, since researchers who have been around for a longer time, have better chances of having high h-indexes (see [Egghe, 2006a; Kelly & Jennions, 2006; Liang, 2006]). H-index for journals was introduced by Braun & al. [2005 & 2006]. Rousseau [2006a] calculated the h-index of JASIS and also studied the effect of time on the h-index. Schubert & Glänzel [2006] found through regression analysis a definite relationship between the IF and the h-index for journal. In contrast, Miller [2006] concluded that for physics periodicals there is no correlation between the IF and the h-index. Costas & Bordons [2006] found good correlations between the h-index and number of publications and the number of citations of individuals. Banks [2006] applied the concept to compound names and scientific topics in publications.

Models for the h-index have also been proposed (see [Egghe & Rousseau, 2006; Glänzel, 2006; Rousseau, 2006b; Burrell, 2006]). Some researchers recommend different types of improvements to the h-index [Batista & al., 2006; Egghe, 2006b; Sidiropoulus & al., 2006; Iglesias & Pechorroman, 2006]. Finally, as usual, there is also criticism regarding the new measure [Purvis, 2006].

Methods

For this study, we used the ISI HighlyCited database (http://isihighlycited.com/) as our starting point, which lists 47 Israeli researchers as of November 2006. Even though this list can be criticized – for example it does not include any of the three Israeli Nobel Prize winners in the last two years – it is a list published by a highly respected institution.

A few researchers from the list for whom it was extremely difficult to disambiguate their work from works of others with the same or similar name, were excluded from further analysis, and the names of the three recent Nobel Prize winners were added. The final list was comprised of 40 names. The name of each researcher was searched in ISI's Web of Science (time span: 1996–2006), Elsevier's Scopus (1996 to present) and Google Scholar (queries of the type *author: "J Doe"*), with publication dates limited to 1996–2006. We had to limit the publication years from 1996 and onwards since Scopus

has complete citation data from 1996 and onwards, and we wanted to compare the three citation databases on a "fair" basis. In a few cases there were no or almost no publications during this period, since the ISI Highly-cited databases is based on citation data for publications of the listed researchers between 1981 and 1999 [ISI Highlycited.com, no date_a; ISI Highlycited.com, no date_b].

We searched WOS only through its "General search" interface where the citation counts are only for items indexed by WOS, and citations to items that do not exactly match the indexed citation are ignored. More complete (but much more time consuming) citation counts could have been achieved had we consulted the "Cited reference search" as well (like in [Cronin & Meho, 2006]). Most other h-index studies to this day relied on the results of the "General search" interface only.

The result sets were sorted by times cited, and the bibliographic details of all the publications that received more citations than the h-index of the author as defined by the specific database were downloaded.

The next step was data cleansing, especially for Google Scholar. When providing the initial and the family name of an author, it retrieves publications by authors whose initials are included in the specified initial, for example when searching for *author: "L Gillis"*, papers published by HL Gillis and LL Gillis are also retrieved (see Figure 3).
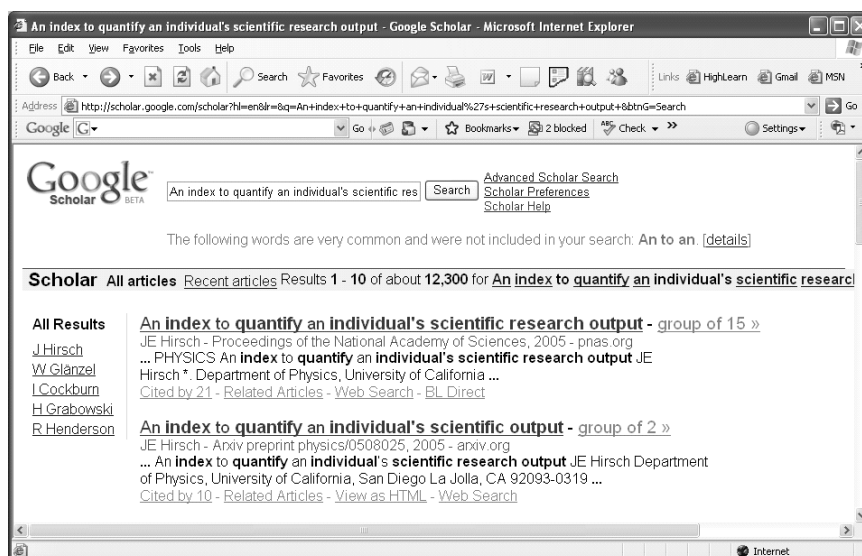

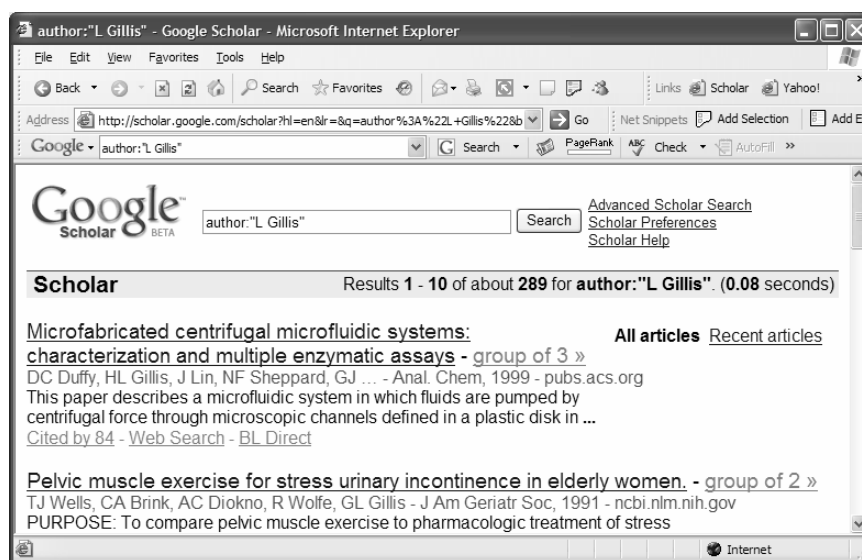Figure 2. Multiple manifestations of a work not grouped together

Figure 3. Google Scholar allows inexact matches for the authors' initials

Google Scholar is not clean of mistakes either, as can be seen from the literauter review above. Data for the publications of Avi Wigderson was collected from Google Scholar on January 16, 2006 – at them time the top cited item was "Probabilistic encryption" cited 915 times. This item was incorrectly attributed to Avi Wigderson (a mistake that has been corrected since). There are also examples where highly cited publications do not appear when searching for an author. Consider, for example the previously mentioned case of *author:"P Ingwersen"* (see Figure 1). In some cases the same publication is listed more than once, in these cases the citation counts were combined.

In addition, Google Scholar often incorrectly identifies the publication date of the item. This happened especially frequently with publications of the American Physical Society – thus we had to check each item on the Google Scholar lists, to make sure that the item was actually published after 1996, and it was indeed authored by the specific scientist. We had to double-check the title of the item as well – in some cases it was incorrect, while in other cases it was shortened, and for a few authors (especially in high-energy physics) the initial part of two or more papers are identical, for example (1) Ackerstaff et al. QCD studies with e(+)e(–) annihilation data at 161 GeV and (2) Abbiendi et al. QCD studies with e(+)e(–) annihilation data at 172–189 GeV.

Searching Scopus was not entirely straightforward either. Although it has an "Author search" interface, where an attempt is made to group together items published

by the same author (i.e., to differentiate between two or more authors with the same name), the system is not always successful. Thus we preferred to search through the "Basic search" interface and to search using only the initials of the authors. However, unlike WOS, searching for an author with a single letter initial retrieves publications of all authors with the same surname and the specific letter appearing as one of the initials. Thus data cleansing had to be done for Scopus as well. Note that Scopus does not index all the authors of a publication (only the first one hundred). A maximum of one hundred authors is sufficient under regular circumstances, but not for high-energy physics publications (e.g., the OPAL or ATLAS groups).

Results and discussion

Table 1 displays the selected scientists, their research field(s) as defined by ISI and their h-indices for the period 1996-to present (2006) computed based on WOS, Scopus and Google Scholar respectively. The table contains 40 names – 37 from the 47 Israeli highly cited researchers, as defined by ISI (ten names were excluded because of the difficulty in disambiguating their publications from publications of other researchers with identical names or because they had no publications in the 1996–2006 period) and the three recent Israeli Nobel prize winners – Robert Aumann (economics, 2005), Aaron Ciechanover and Avram Hershko (chemistry, 2004).

Recall that ISI created the list of highly cited researchers based on citations for publications between 1981 and 1999, and we computed the h-indices based on publications from 1996 and onwards. During this period some of the previously highly cited researchers were inactive or relatively inactive and had much lower publication and citation counts. We had to base our computations on the 1996–2006 period, because Scopus provides full citation information only for items published after 1995.

Except for a few cases the differences in the h-indices between WOS and Scopus are not significant, except for Mikenberg and Wigderson. Avi Wigderson is a theoretical computer scientist and Scopus, unlike WOS, indexes the two major theoretical computer science conference series, STOC and FOCS which explains the difference in the h indices. The case of Giora Mikenberg is rather interesting: in the list of highly cited Israeli researchers, there are three high-energy physicists, Alexander, Duchovni and Mikenberg all three of them are members of both the OPAL and ATLAS (http://atlas.web.cern.ch/Atlas/index.html) groups (each with hundreds of members) where all members of these groups author all (or most) their publications. Alexander and Duchovni are among the first one hundred authors indexed by Scopus, while Mikenberg is not. The authors are usually listed in alphabetical order, WOS indexes all the authors, and this is the major reason for the huge difference between the number of items indexed by WOS and Scopus for Mikenberg.

Table 1. H-index according to WoS, Scopus and Google Scholar for highly cited Israeli researchers for 1996–2006

Researcher	Category	WoS	Scopus	Google Scholar
Alexander, Gideon	Physics	32	30	20
Alon, Noga	Mathematics, Computer Science	14	17	27
Aumann, Robert J.	Mathematics	8	6	11
Aurbach, Doron	Materials Science	29	29	19
Beeri, Catriel	Computer Science	3	3	8
Chet, Ilan	Plant & Animal Science	21	21	20
Ciechanover, Aaron	Biology & Biochemistry	33	34	30
Cohen, Irun R.	Immunology	29	32	26
Dagan, Gedeon	Engineering, Ecology/Environment	13	14	12
Dekel, Avishai	Space Sciences	25	25	24
Dolev, Daniel	Computer Science	5	7	18
Duchovni, Ehud	Physics	32	29	15
Geiger, Benjamin	Molecular Biology & Genetics	34	33	31
Gohberg, Israel	Mathematics	8	8	11
Goldreich, Oded	Computer Science	12	14	32
Harel, David	Computer Science	9	10	22
Hershko, Avram	Biology & Biochemistry	21	21	21
Hochberg, Yosef	Mathematics	4	4	7
Jortner, Joshua	Chemistry	26	26	21
Kanner, Joseph	Agricultural Sciences	9	9	6
Kerem, Batsheva	Molecular Biology & Genetics	20	19	18
Kotler, Burt P.	Ecology/Environment	10	10	9
Leviatan, Yehuda	Computer Science	6	5	5
Lubotzky, Alex	Mathematics	5	5	10
Mechoulam, Raphael	Pharmacology	28	29	26
Mikenberg, Giora	Physics	31	10	4
Moran, Shlomo	Computer Science	7	6	11
Netzer, Hagai	Space Sciences	28	28	18
Oren, Moshe	Molecular Biology & Genetics	47	49	38
Peleg, David	Computer Science	8	11	21
Piran, Tsvi	Space Sciences	32	32	33
Pnueli, Amir	Computer Science	11	8	23
Procaccia, Itamar	Physics	18	19	18
Shainberg, Isaac	Ecology/Environment	8	10	9
Shamai, Shlomo	Computer Science	16	17	20
Sharir, Micha	Engineering, Computer Science	13	15	17
Shelah, Saharon	Mathematics	10	10	15
Sklan, David	Agricultural Sciences	18	20	15
Turkel, Eli	Mathematics	9	9	11
Wigderson, Avi	Computer Science	8	13	23

The differences between GS and the two other databases are much more considerable. We partitioned the researchers into three groups:

1. The GS h-index is at least 30% lower than the average of the h indices based on WOS and Scopus.
2. The GS h-index is at least 30% higher than the average of the h indices based on WOS and Scopus.

3. The GS h-index is between 0.7 and 1.3 times the average of the h indices based on WOS and Scopus.

We call the first group *low*, the second *high* and the third the same. The *high* group is comprised entirely of mathematicians and computer scientists. All, except one of the mathematicians and all except two of the computer scientists belong to this group. The exceptions Leviatan, Shamai, and Turkel belong to the *same* group. Thus there seems to be a discipline specific bias here. For computer science this can be explained by the prevalence of peer-reviewed conference proceedings publications that are highly cited and highly valued. Often the computer science researchers do not submit full versions of their conference papers to journals. Thus proceedings are the major citing and cited venue for computer science. WOS, with the exception of the *Lecture Notes in Computer Science* and the *Lecture Notes in Artificial Intelligence* series, does not index computer science proceedings. Scopus indexes a much larger set of computer science conference proceedings (the list can be downloaded from http://www.info.scopus.com/detail/what/titlelistinfo.htm) but still seemingly it covers fewer publications in computer science than GS. GS also indexes Technical Reports, which are sometimes highly cited items as well. For example, the most highly cited publication of Amir Pnueli, is "The temporal logic of programs", a Technical Report, which was cited 1094 times according to Google Scholar (as of November 11, 2006).

All three high-energy physicists are in the *low* group, which is rather surprising, especially since physicists usually submit their preprints to arxiv.org and Google Scholar indexes arxiv.org extensively.

H-index only provides partial information, if a researcher has an h-index h, we can be sure that his publications received at least h^2 citations, but the actual number could be much higher. In Table 2 we display the number of citations to the top-h publications for each researcher based on WOS, Scopus and GS respectively.

Even for researchers with comparable h-indices for WOS, Scopus and Google Scholar, there can be differences in the citation counts. Note that the citation counts were calculated only for the top *h* documents (for the h of the specific database and the author). For example for Avram Hersho, the citation count for Google Scholar is about 20% lower than for the other two databases. On the other hand, in some cases the citation counts for GS are much higher than for the other databases. Consider, for example David Harel: only 188 citations in WOS versus 3374 citations in GS. His h-index in WOS is 9 versus 33 in GS, thus some of the increase in the number of citations can be expected. A researcher with h-index 9 is expected to receive at least 81 citations to the top cited publications. Harel actually received 188 citations based on WOS, i.e. 2.32 times more than the minimum for an h-index of 9. Harel's h-index, according to GS is 22, thus he should have received a minimum of 484 citations to these 22 items. The actual number of citations is 6.97 times higher than the minimum.

Table 2. H-index and the number of citations to the h most highly cited publications according to WoS,
Scopus and Google Scholar for highly cited Israeli researchers for 1996–2006

Researcher	h-WOS	# cits. WOS	h-Scopus	# cits. Scopus	h-GS	# cits. GS
Alexander, Gideon	32	1,624	30	1,383	20	938
Alon, Noga	14	363	17	540	27	1,694
Aumann, Robert J.	8	101	6	81	11	407
Aurbach, Doron	29	2,062	29	2,127	19	1,041
Beeri, Catriel	3	26	3	52	8	347
Chet, Ilan	21	722	21	772	20	804
Ciechanover, Aaron	33	6,084	34	7,013	30	5,239
Cohen, Irun R.	29	2,247	32	2,738	26	1,651
Dagan, Gedeon	13	403	14	469	12	469
Dekel, Avishai	25	1,818	25	1,801	24	1,933
Dolev, Daniel	5	184	7	221	18	1,372
Duchovni, Ehud	32	1,624	29	1,352	15	550
Geiger, Benjamin	34	4,241	33	3,855	31	3,439
Gohberg, Israel	8	111	8	90	11	243
Goldreich, Oded	12	444	14	527	32	3,569
Harel, David	9	188	10	429	22	3,374
Hershko, Avram	21	4,228	21	4,256	21	3,373
Hochberg, Yosef	4	143	4	153	7	200
Jortner, Joshua	26	1,935	26	1,734	21	1,074
Kanner, Joseph	9	281	9	293	6	153
Kerem, Batsheva	20	1,108	19	1,103	18	992
Kotler, Burt P.	10	290	10	224	9	188
Leviatan, Yehuda	6	77	5	62	5	53
Lubotzky, Alex	5	47	5	43	10	247
Mechoulam, Raphael	28	2,470	29	2,771	26	1,943
Mikenberg, Giora	31	1,545	10	270	4	186
Moran, Shlomo	7	86	6	88	11	423
Netzer, Hagai	28	1,955	28	1,968	18	1,157
Oren, Moshe	47	6,926	49	7,399	38	5,460
Peleg, David	8	113	11	181	21	953
Piran, Tsvi	32	3,200	32	3,223	33	3,976
Pnueli, Amir	11	216	8	236	23	2,243
Procaccia, Itamar	18	582	19	625	18	688
Shainberg, Isaac	8	128	10	152	9	154
Shamai, Shlomo	16	876	17	1,398	20	2,232
Sharir, Micha	13	281	15	361	17	646
Shelah, Saharon	10	139	10	139	15	476
Sklan, David	18	486	20	538	15	334
Turkel, Eli	9	184	9	221	11	330
Wigderson, Avi	8	107	13	329	23	1,126

Thus we see that the average number of citations the top h items received in GS (153.4 citations) is much higher than the average number of citations per item when the calculations are based on WOS (20.9 citations). One has to take into account that the sources and the validity of the citations in GS were not examined in this study.

Because of inconsistencies noticed in the publication data, it can be expected that the citation counts are not perfect either. Examining the citing items for GS was beyond the scope of the current study.

Conclusions

The findings show that it matters which citation tool is used to compute the h-index of scientists. Also there seems to be disciplinary differences in the coverage of the databases. The differences in citation counts create a dilemma for science policy makers and promotion committees.

We recommend further studies to explore the relative strengths and weaknesses of the currently available citation tools. We also recommend to further explore the capabilities and limitations of Google Scholar, especially in terms of the citing items.

References

BAKKALBASI, N., BAUER, K., GLOVER, J., WANG, L. (2006), Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3 **:** 7. Retrieved November 4, 2006, from http://www.bio-diglib.com/content/3/1/7

BANKS, M. G. (2006), An extension of the Hirsch index: Indexing scientific topics and compounds. *Scientometrics*, 69 (1) : 161–168.

BAR-ILAN, J. (2005), Expectations versus reality – Search engine features needed for Web research at mid 2005. *Cybermetrics*, 9, paper 2. Retrieved November 4, 2006, from http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.html

BAR-ILAN, J. (2006), An ego-centric citation analysis of the works of Michael O. Rabin based on multiple citation indexes. *Information Processing and Management*, 42 (6) : 1553–1566.

BAR-ILAN, J. (2006B), H-index for Price medalists revisited. *ISSI Newsletter*, 2 (1) : 3–5.

BAR-ILAN, J., LEVENE, M., LIN, A. (2007), Some measures for comparing citation databases. *Journal of Informetrics*, 1 (1) : 26–34.

BATISTA, P. D., CAMPITELI, M. G., KINOUCHI, O., MARTINEZ, A. S. (2006), Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68 (1) : 179–189.

BAUER, K., BAKKALBASI, N. (2005), An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*, 11(9). Retrieved February 23, 2006, from http://www.dlib.org/dlib/september05/bauer/09bauer.html

BELEW, R. K. (2005), Scientific impact quality and quantity: Analysis of two sources of bibliographic data. Retrieved November 4, 2006, from http://arxiv.org/abs/cs.IR/0504036

BORNMANN, L., DANIEL, H. (2005), Does the h-index for ranking of scientists really work? *Scientometrics*, 65 : 391–392.

BRAUN, T., GLANZEL, W., SCHUBERT, A. (2005), A Hirsch-type index for journals. *The Scientist*, 19 (22) : 8.

BRAUN, T., GLANZEL, W., SCHUBERT, A. (2006), A Hirsch-type index for journals. *Scientometrics*, 69 (1) : 169–173.

BURRELL, Q. L. (2006), Hirsch's h-index: a preliminary stochastic model. In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 26–28.

BURNHAM J. F. (2006), Scopus database: A review. *Biomedical Digital Libraries*, 3 (1). Retrieved November 4, 2006, from http://www.bio-diglib.com/content/3/1/1

BUTLER, L., MCALLISTER, I. (2006), The Hirsch index: Is it applicable to the Social Sciences? In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 31–32.

COSTAS, R., BORDONS, M. (2006), H-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 189–191.

CRONIN, B., MEHO, L. (2006), Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57 (9) : 1275–1278.

DEIS, L. F., GOODMAN, D. (2005), Web of Science (2004 Version) and Scopus. *The Charleston Advisor* 6 (3). Retrieved November 4, 2006, from http://www.charlestonco.com/comp.cfm?id=43

EGGHE, L. (2006A), *Dynamic h-index: The Hirsch Index in Function of Time.* Preprint. Retrieved November 4, 2006, from http://doclib.luc.ac.be/dspace/bitstream/1942/980/1/dynamic+h.pdf

EGGHE, L. (2006B), Theory and practise of the g-index. *Scientometrics*, 69 (1) : 131–152.

EGGHE, L., ROUSSEAU, R. (2006), An informetric model for the Hirsch-index. *Scientometrics*, 69 (1) : 1221–129.

GARDNER, S., ENG, S. (2005), Gaga over Google? Scholar in the Social Sciences. *Library Hi-Tech News*, 22 (8) : 42–45.

GARFIELD, E. (1963), Science Citation Index. *Science Citation Index 1961*, volume 1, v-xvi. Retrieved October 31, 2006, from http://garfield.library.upenn.edu/papers/80.pdf

GILES, J. (2005), Start your engines. *Nature News*, 554–555.

GLÄNZEL, W. (2006), On the h-index – A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67 (2) : 315–321.

GLÄNZEL W., PERSSON, O. (2005), H-index for Price medalists. *ISSI Newsletter*, 1 (4) : 15–18.

GOOGLE SCHOLAR (2005), *Support for Scholarly Publishers*. Retrieved October 31, 2006, from http://scholar.google.com/intl/en/scholar/publishers.html

HIRSCH, J. E. (2005A), An index to quantify an individual's scientific research output. Retrieved November 4, 2006, from http://arxiv.org/abs/physics/0508025

HIRSCH, J. E. (2005B), An index to quantify an individual's scientific research output. *PNAS*, 102 (46) : 16569–16572.

IGLESIAS, J. E., PECHARROMAN, C. (2006), *Scaling the h-index for Different Scientific ISI Fields*. Preprint. Retrieved November 4, 2006, from http://arxiv.org/ftp/physics/papers/0607/0607224.pdf

ISI HIGHLYCITED.COM (NO DATE_A), *About HighlyCited.com*. Retrieved November 4, 2006, from http://hcr3.isiknowledge.com/popup.cgi?name=hccom

ISI HIGHLYCITED.COM (NO DATE_B), *How Do We Identify Highly Cited Researchers?* Retrieved November 4, 2006, from http://hcr3.isiknowledge.com/isi_copy/howweidentify.htm

JACSO, P. (2004), Scopus. *Peter's Digital Reference Shelf*, September 2004. Retrieved November 4, 2006, from http://www.galegroup.com/servlet/HTMLFileServlet?imprint=9999&region=7&fileName=reference/archive/200409/scopus.html

JACSO, P. (2005A), As we may search – Comparison of major features of Web of Science, Scopus and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89 (9) : 1537–1547.

JACSO, P. (2005B), Visualizing overlap and rank differences among web-wide search engines. *Online Information Review*, 29 (5) : 554–560.

JACSO, P. (2006), Deflated, inflated and phantom citation counts. *Online Information Review*, 30 (3) : 297–309.

KELLY, C. D., JENNIONS, M. D. (2006), The h-index and career assessment by numbers. *TRENDS in Ecology and Evolution*, 21 (4) : 167–170.

KOUSHA, K., THELWALL, M. (2006A), Google Scholar Citations and Google Web/URL citations: A multi-discipline exploratory analysis. In: *Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting*, Nancy (France). Retrieved November 4, 2006, from http://eprints.rclis.org/archive/00006416/

KOUSHA, K., THELWALL, M. (2006B), Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 72–73.

LAGUARDIA, C. (2005), E-Views and Reviews: Scopus vs. Web of Science, *Library Journal*, January 15, 2005.

LIANG, L. (2006), H-index sequence and h-index matrix: Constructions and applications. *Scientometrics*, 69 (1) : 163–169.

MILLER, C. W. (2006), *Superiority of the h-index over the Impact Factor for Physics*. Preprint. Retrieved November 4, 2006, from http://arxiv.org/PS_cache/physics/pdf/0608/0608183.pdf

MOED, H. F. (2005), *Hirsch Index is a Creative and Appealing Construct but Be Cautious when Using It to Evaluate Individual Scholars*. Retrieved November 4, 2006, from
http://www.cwts.nl/hm/Comments_on_Hirsch_Index_2005_12_16.pdf

MULLEN, L. B., HARTMAN, K. A. (2006), Google Scholar and the Library Web Site: The early response by ARL Libraries. *College & Research Libraries*, 67 (2) : 106–122.

NEUHAUS, C., NEUHAUS, E., ASHER, A., WREDE, C. (2006), The depth and breadth of Google Scholar: An empirical study. *portal: Libraries and the Academy*, 6 (2) : 127–141.

NORUZI, A. (2005), Google Scholar: The new generation of citation indexes. *LIBRI*, 55 (4) : 170–180.

PAULY, D., STERGIOU, K. I. (2005), Equivalence of results from two citation analyses: Thomson's ISI Citation Index and Google's Scholar service. *Ethics in Science and Environmental Politics*, 2005, 33–35.

PAYNE, D. (2004), Google Scholar welcomed. *News from The Scientist*, 5(1):20041123-01. Retrieved October 31, 2006, from http://www.the-scientist.com/news/20041123/01/

PURVIS, A. (2006), The h-index: Playing the numbers game. *TRENDS in Ecology and Evolution*, 21 (8) : 422.

RAHM, E., THOR, A. (2005), Citation analysis of database publications. *SIGMOD Record*, 34 (4) : 48–53.

REED ELSEVIER (2004), *Scopus Comes of Age*. Retrieved October 31, 2006, from http://www.reed-elsevier.com/index.cfm?Articleid=1075

ROUSSEAU, R. (2006A), *A Case Study: Evolution of JASIS' Hirsch Index*. Preprint, Retrieved November 4, 2006, from http://eprints.rclis.org/archive/00005430/01/Evolution_of_h_JASIS_rev.pdf

ROUSSEAU, R. (2006B), *Simple Models and the Corresponding h- and g-index*. Preprint. Retrieved November 4, 2006, from http://eprints.rclis.org/archive/00006153/01/Rousseau_Dalian.pdf

SAAD, G. (2006), Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. *Scientometrics*, 69 (1) : 117–120.

SCHUBERT, A., GLÄNZEL, W. (2006), A systematic analysis of Hirsch-type indices for journals, preliminary results. In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 251–253.

SCOPUS (NO DATE), *About Scopus*. Retrieved October 31, 2006, from http://www.info.scopus.com/about/

SIDIROPOULOS, A., KATSAROS, D., MANOLOPOULOS, Y. (2006), *Generalized h-index for Disclosing Latent Facts in Citation Networks*. In Proceedings of LinkKDD06. Retrieved November 4, 2006, from http://arxiv.org/PS_cache/cs/pdf/0607/0607066.pdf

SMITH, A. (2006), Google Scholar as a cybermetric tool: A comparison with the New Zealand PRBF research assessment. In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 139–141.

THOMSON SCIENTIFIC (NO DATE), *Web of Science*. Retrieved October 31, 2006, from
http://scientific.thomson.com/products/wos/

VAN RAAN, A. F. J. (2005), Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 69 (1) : 117–120.

VAUGHAN, L., SHAW, D. (2006), Comparison of citations from ISI, Google and Google Scholar: Seeking Web indicators of input. In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 155–156.

VISSER, M. S., MOED, H. F. (2005), Developing bibliometric indicators of research performance in computer science. In: *Proceedings of ISSI 2005*, vol. 1, 275–279.

VISSER, M. S., MOED, H. F. (2006), How important are conference papers for measuring research performance in Computer Science? In: *Book of Abstracts, 9th International Science & Technology Indicators Conference*, Leuven, Belgium, 161–162.