Web of Science with the Conference Proceedings Citation Indexes – The Case of Computer Science

Judit Bar-Ilan

barilaj@mail.biu.ac.il
Department of Information Science, Bar-Ilan University, Rmat Gan, 52900 (Israel)

Abstract
In September 2008 Thomson Reuters added to the ISI Web of Science (WOS) the Conference Proceedings Citation Indexes in Science and in the Social Sciences and Humanities. This paper examines how this change affects the publication and citation counts of highly cited computer scientists. Computer science is a field where proceedings are a major publication venue. The results show that most of the highly cited publications of the sampled researchers are journal publications, but these highly cited items receive more than 40% of their citations from proceedings papers. The paper also discusses issues related to double-counting, i.e., when a given work is published both in a proceedings and later on as a journal paper.

Introduction
The citation indexes established by Eugene Garfield are selective databases. In his book, Garfield (1979) explains that there are several reasons for selecting which journals to index: it is impractical or even impossible to list all scientific journals and it is not economically feasible to index all journals. Thus there has to be a selection process, which is based on the core journals in each discipline (Bradford, 1934). However, when considering a multidisciplinary database, it turns out that much more than the core is covered, because according to Garfield's law of concentration (Garfield, 1977), the "tail of the literature of one discipline consists, in a large part, of the cores of the literature of other disciplines" (Garfield, 1979, p.23). Due to this finding, the core literature of all scientific disciplines is estimated to be covered by no more than 1,000 journals. In 1979, the Science Citation Index already covered more than 3,000 journals, which is far more than the core. ISI included additional journals based on quality judgements. Today the Science Citation Index Expanded accessible through the Web of Science (WOS), indexes more than 6,650 journals (Thomson Reuters, 2008), and the list is constantly expanding. As a comparison, Elsevier's Scopus lists almost 12,000 active journal titles in the areas of health, life, multidisciplinary and physical sciences (Elsevier, 2008).

Until recently, the Web of Science included journal papers only with a few exceptions. One of the more notable exceptions was the Lecture Notes in Computer Science (LNCS) series – most volumes of these series are proceedings of computer science conferences. Just as an example, the series published more than 550 volumes in 2008 alone, and the huge majority of these volumes are proceedings of computer science conferences. WOS indexes, as of January 2008, 139,302 items from LNCS and an additional 25,999 items from its sub-series Lecture Notes in Artificial Intelligence (LNAI). The earliest indexed records from the LNCS series that are indexed by the Science Citation Index Expanded are from 1981, and between 1999 and 2005 the LNCS was even included in the Journal Citation Reports. It is not quite clear why LNCS were indexed by ISI, because the huge majority of the items published in LNCS are conference papers, and LNCS does not cover the most prestigious computer science proceedings. The proceedings of the two major societies (ACM and the IEEE Computer Society) are published by the societies and are not part of the LNCS series.
However, the above discussion is of little value, since in September 2008, ISI merged into WOS all the records from its two proceedings citation databases (Thomson Reuters, 2008), Conference Proceedings Citation Index- Science (CPCI-S) and Conference Proceedings

Proceedings of ISSI 2009, pages 399–409.
Edited by B. Larsen and J. Leta.

Citation Index- Social Science & Humanities (CPCI-SSH) covering proceedings from 1990 to present. When searching WOS, one can exclude these two databases from the search, but the citation counts are from the whole database and they include citations received also from proceedings publications indexed by ISI.

In this paper we examine the effects of this recent change on computer science publications. It is well-known that in computer science proceedings publications are considered to be at least as important as journal publications (see for example Kling & McKim, 1999; Goodrum, McCain, Lawrence & Giles, 2001 or Visser & Moed, 2005). Thus it is expected that the inclusion of citations to and from proceedings will have a huge influence of the number of publications and on the number of citations to publications. Here we considered the most highly cited publications of a sample of highly cited computer scientists (as defined by ISI Highly Cited Researchers - http://hcr3.isiknowledge.com/).

Literature review

Drott (1995) examined the republication rate of papers published in the proceedings of the ASIS conferences and found that the republication rate was much lower than expected in other disciplines. He created typography of conference paper functions:

- Self improvement – as a venue to report initial results, which are later extended and rewritten in the form of a journal paper
- Group contribution – as a means of sharing information.
- Final product – no intention to republish. Goodrum et al. (2001) suggest an additional category: a substitute for journal publication. This category is based upon discussions with computer scientists who view conference proceedings as sufficient and do not feel the need to republish the results in journals. In our opinion this fourth category is well covered in "final product".

Glänzel, Schlemmer, Schubert and Thijs (2005) found, based on data extracted from the 1994-2002 volumes of the ISI Proceedings database's Science and Technology edition, that about half of the papers indexed there belonged to the field of engineering. In their categorization, computer science is viewed as part of engineering. Their results support the prevailing view that proceedings publications have great importance in computer science.

Moed and Visser (2007) produced an extensive report on the need and feasibility of extending WOS with proceedings publications in the field of computer science. They explored this possibility for Dutch computer scientists. The WOS source and citation data was expanded with proceedings published in LNCS, ACM and IEEE computer science conferences. The expanded database increased the coverage of the publications of Dutch computer scientists from 25% to 35%. The results were shown to Dutch computer scientists, who claimed that even with this extended coverage some of the important conferences were missed. The internal coverage of the external database (i.e., the percentage of citations in the items indexed by the database that referred to items in the extended database) was 51%, which is only a moderate coverage compared to about 80% internal coverage in WOS for physics and chemistry. The citation impact of proceedings series was found to be more variable than annual volumes of journals, but the citation links between recurring conferences was found to be statistically similar to journal self-citation rates.

Bar-Ilan (2006) studied the citations to works of Michael Rabin, a prominent computer scientist and found that among the top 12 most highly cited items based on sources indexed by WOS, there are two technical reports and one conference publications, emphasizing that other means of publications (sometimes called grey literature) are of high importance in computer science not only in terms of quantity, but also in terms of visibility (citation count). She also developed the notion of several manifestations of a work based on the FRBR specifications (IFLA, 1998). In computer science rather often several almost identical versions of the same work are published, first as a technical report or a preprint, then as a proceedings paper and later as a journal paper and/or a chapter in an edited book. A similar notion of a concept is also mentioned by Moed and Visser (2007). This line of thought will be further developed in the current paper.

Goodrum et al. (2001) compared highly cited items in the ISI's Science Citation Index with Citeseer (now at http://citeseer.ist.psu.edu/ - not updated anymore or its follow-up beta Citeseer^x at http://citeseerx.ist.psu.edu/), a computer science citation database where items are indexed and citations are extracted automatically. With one exception, all of the overlapping items between the 25 top cited items in both databases were to books or book chapters, and none of the items in either list were proceedings papers.

Meho and Rogers (2008) compared the citation counts retrieved from WOS and Scopus for 22 top researchers in the area of human-computer interaction (HCI), which can be viewed as a subfield of computer science. The data from WOS was retrieved before the proceedings databases were merged into WOS. The results show that Scopus provided considerable better coverage mainly due to the indexing of ACM and IEEE proceedings. However the wider coverage did not significantly alter the rankings of the scientists based on citation counts.

Lisée, Larivière and Archembault (2008) studied conference proceedings in general. They showed that conference papers age faster than journal paper, found that conference publications constitute about 8% of the references in engineering papers, and about 20% of the references in computer science papers. The data for the study was derived from the ISI Citation Indexes (without the Proceedings Citation Indexes). Frolich and Resler (2001) used the Science Citation Index to determine the citation counts for all publications of the University of Texas Institute for Geophysics, and found that papers in what they called "mainstream journals" receive on average considerably more citations than papers published in proceedings.

The rest of the paper is organized as follows: first we describe the research objectives and the data collection and analysis. In the results section the number of publications with and without proceedings papers and the percentage of citations from proceedings papers in the h-core are analyzed. Finally we discuss the problem of "republications", i.e., proceedings papers that are published later on as journal papers.

Methods

All of the above-mentioned studies examined ISI citation patterns before the merger with the proceedings citation databases. This merging created an entirely new situation where citations are extracted from a much larger database, and thus it is of great importance to study and understand the effects of this change. It should be noted that the study conducted by Moed and Visser (2007) also created an expanded database for computer science, with the papers from the most important conference series added to the ISI database. The effects of this change were studied for Dutch computer scientist irrespective of their scientific standing, whereas here we consider highly cited computer scientists, and study the changes in the number of publication, number of citations and the sources of citations.

In the current study we asked the following questions:

- What is the effect of the inclusion of proceedings papers on publication counts?
- What percentage of the top-cited items are conference publications?
- What is the percentage of the citations of the top-cited items that come from conference proceedings?
- What proportion of the conference publications are later republished as journal papers?

As a case study we chose computer science, where the importance of conference proceedings is known to be high. More specifically, the current study concentrated on highly-cited computer scientists.

Data collection and analysis

The list of highly cited computer scientists was retrieved from the ISI Highly Cited database (http://hcr3.isiknowledge.com/home.cgi). As of mid-December 2008 the list of highly-cited computer scientists was comprised of 339 researchers. Out of this list a random sample of size 50 was created. In a few cases the names were highly ambiguous and even after refining the publications by subject area to computer science related fields it was difficult to tell whether the remaining list contained only the publications of the specific researchers and whether most of his/her publications were included. In these cases the next researcher in the alphabetic list was selected instead of the researcher chosen for the random sample.

For each selected researcher we downloaded his/her list of publications. Researchers with middle initials sometimes appear without their middle initials, thus we searched both variations, i.e., when searching for the indexed publications of Barbara H. Liskov, the query was Liskov BH OR Liskov B in the author field. For each researcher two queries were submitted, once to all databases (SCI Expanded, SSCI, A&HCI, CPCI-S and CPCI-SSH) and once without the two proceedings databases (CPCI-S and CPCI-SSH). In each case the results were filtered to include relevant subject areas only (computer science and its subcategories, electronic and electrical engineering, applied mathematics and telecommunications). The number of journal articles and proceedings papers as categorized by WOS (under document type) was recorded. It must be noted that the document type categorization is not perfect. In our context we noticed special problems with the Lecture Notes in Computer Science, which is sometimes categorized as a journal and sometimes as proceedings. In addition many of the 1990-1992 LNCS records are indexed twice (once as a journal and once as a proceedings paper), inflating both the publication and the citation counts.

When searching for author, WOS does not only search for author but for editor as well, and retrieves all items published in volumes/proceedings edited by the researcher. For example, 508 records were retrieved for Oscar Ibarra, but he authored only 206 items according to WOS. We could not find any systematic way to exclude edited, but not authored items through the WOS interface, but this was easily done on the files downloaded from WOS' Marked List feature in text-delimited form. Scopus does not include items edited but not authored when searching for an author, and has separate codes for authors and for editors. All items edited but not authored by the selected researcher were excluded.

Next we determined the researcher's h-index (Hirsch, 2005) using the Citation Report option and removing items that were edited but not authored by the specific researcher. For each of the h_i items (h_i – is the h-index of researcher i), we recorded the number of journal articles and the number of proceedings papers that cited the item, based on the information provided by the “Analyze Results” option of WOS. We also noted the number of proceedings papers among the h_i items.

Data for the 50 selected researchers were collected during the second half of December 2008. In the next section we provide descriptive statistics derived from the collected data.

Results

Publication counts

Table 1 displays the publication counts of the selected scientists with and without the Conference Proceedings Citation Indexes (CPCIs). We see that on the average the publication counts increase by more than 70% with the addition of the new databases. We also see extremely huge variations between the researchers. These initial numbers are slightly misleading, since most of the highly cited scientists served also as editors of proceedings.

The fifth column in Table 1 displays the publication counts from the expanded database after the edited items are removed. Note that in some cases the number in the last column is lower than what appears in the second column. This happens because even before the expansion of the database, some proceedings were indexed (notably Lecture Notes in Computer Science), thus even before the expansion edited items that appeared in such proceedings were attributed to the researcher when searching for his/her name in the author field.

The sixth and seventh columns display the number of journal and proceedings papers after the removal of edited and not authored items. Note that the sum of the journal and proceedings papers is usually somewhat lower than the corresponding number in the fifth column of Table 1. This is because besides journal and proceedings papers there are additional document types (e.g. editorials or review articles). On average 38% of the publications are proceedings papers (standard deviation: 18%), and 52% are journal articles (standard deviation 17%). When trying to interpret these numbers, one has to take into account that the proceedings data are only from 1990 and onwards, whereas the journal data are from 1965 and onwards and some (or even most) of these highly cited researchers were active before 1990, thus probably the actual percentage of proceedings papers is higher than what we see from the WOS data.

Table 1. Publication counts of the selected scientists with and without the Conference Proceedings databases

Name	No. publ. without data from the CPCIs	No. publ. with data from the CPCIs	Increase in publ. count	No. publ. with data from the CPCIs, excluding edited, but not authored items	No. journal papers	No. conf. papers
Aazhang, Behnaam	64	148	131.3%	148	40	107
Abiteboul, Serge	91	151	65.9%	92	36	47
Alur, Rajeev	220	256	16.4%	137	33	94
Apt, Krzysztof R.	76	84	10.5%	69	46	19
Beeri, Catriel	68	73	7.4%	44	30	11
Biglieri, Ezio	99	180	81.8%	180	63	90
Blum, Manuel	36	41	13.9%	41	28	13
Cheriton, David R.	17	32	88.2%	32	16	16
Coppersmith, Don	129	141	9.3%	102	67	25
Courcelle, Bruno	91	98	7.7%	98	62	33
Culik, Karel	122	131	7.4%	131	115	14
Dill, David L.	55	101	83.6%	101	22	71
Dolev, Daniel	65	126	93.8%	99	56	39
Dwork, Cynthia	78	91	16.7%	58	31	25
Eyuboglu, M. Vedat	15	17	13.3%	17	14	2
Forney, G. David	60	73	21.7%	73	39	19
Galil, Zvi	151	213	41.1%	143	106	24
Garcia-Molina, Hector	103	217	110.7%	180	66	110
Girard, Jean Yves	56	64	14.3%	64	23	37
Goldreich, Oded	102	136	33.3%	136	77	53
Goldwasser, Shafi	32	172	437.5%	56	19	29
Hagenauer, Joachim	46	107	132.6%	96	32	53
Ibarra, Oscar H.	334	508	52.1%	206	119	67
Karmarkar, Narendra K.	12	20	66.7%	20	11	8
Kesselman, Carl	35	65	85.7%	65	42	20
Leveson, Nancy G.	49	70	42.9%	70	31	23
Liskov, Barbara H.	38	59	55.3%	59	25	31
Lovasz, László	135	327	142.2%	152	117	25
Maier, David	62	82	32.3%	82	47	30
Manber, Udi	40	120	200.0%	51	32	12
Marsan, Marco Ajmone	231	360	55.8%	168	70	88
Megiddo, Nimrod	138	152	10.1%	102	74	18
Micali, Silvio	51	72	41.2%	72	34	38
Overmars, Mark H.	118	221	87.3%	156	93	52
Parrow, Joachim	176	208	18.2%	32	17	13
Peleg, David	150	188	25.3%	188	97	86
Pozar, David M.	113	121	7.1%	121	89	15
Preparata, Franco P.	133	206	54.9%	143	106	22
Raghavan, Prabhakar	87	167	92.0%	149	65	79
Ramamritham, Krithi	25	215	760.0%	45	15	27
Reps, Thomas	78	124	59.0%	111	35	71
Salton, Gerard	119	123	3.4%	123	62	6
Seshadri, Nambi	34	65	91.2%	65	25	37
Sloane, Neil J.A.	181	192	6.1%	192	151	13
Soloway, Elliot	56	66	17.9%	66	29	13
Tarjan, Robert E.	151	170	12.6%	170	128	32
Vardi, Moshe Y.	300	347	15.7%	205	69	124
Volakis, John L.	219	308	40.6%	308	149	93
Wing, Jeanette M.	147	159	8.2%	47	22	16
Zadeh, Lotfi A.	632	1452	129.7%	137	42	59
Average	**112.4**	**176.38**	**73.0%**	**108.04**	**56.34**	**40.98**
Standard deviation	**102.9**	**207.7**	**121.7%**	**59.4**	**37.1**	**31.5**

Table 2. Citation data of the selected scientists – journal citations and proceedings citations

Name	Times Cited (total)	h-index	No. proc papers in h-core	No. citations to the h-core	No. citations to the h-core from journal articles		No. citations to the h-core from proceedings papers	
Aazhang, Behnaam	4067	25	0	3768	1363	36.2%	2359	62.6%
Abiteboul, Serge	962	17	7	687	402	58.5%	204	29.7%
Alur, Rajeev	3696	24	11	3123	835	26.7%	2258	72.3%
Apt, Krzysztof R.	1066	15	3	850	599	70.5%	224	26.4%
Beeri, Catriel	706	12	1	600	400	66.7%	182	30.3%
Biglieri, Ezio	2587	21	3	2188	954	43.6%	1192	54.5%
Blum, Manuel	1946	19	1	1838	1018	55.4%	762	41.5%
Cheriton, David R.	781	11	0	743	511	68.8%	206	27.7%
Coppersmith, Don	1665	21	5	1428	725	50.8%	438	30.7%
Courcelle, Bruno	1921	22	3	1532	948	61.9%	552	36.0%
Culik, Karel	1312	20	1	741	510	68.8%	191	25.8%
Dill, David L.	2177	14	5	1944	584	30.0%	1338	68.8%
Dolev, Daniel	2044	21	0	1793	837	46.7%	904	50.4%
Dwork, Cynthia	1333	15	2	1047	490	46.8%	535	51.1%
Eyuboglu, M. Vedat	950	11	0	933	462	49.5%	426	45.7%
Forney, G. David	6188	32	0	5866	3144	53.6%	2273	38.7%
Galil, Zvi	2026	25	0	1220	822	67.4%	329	27.0%
Garcia-Molina, Hector	2408	25	7	1639	827	50.5%	767	46.8%
Girard, Jean Yves	1169	10	3	1138	595	52.3%	521	45.8%
Goldreich, Oded	2747	25	4	2225	950	42.7%	1233	55.4%
Goldwasser, Shafi	2675	13	4	2529	895	35.4%	1590	62.9%
Hagenauer, Joachim	2304	14	1	2188	882	40.3%	1245	56.9%
Ibarra, Oscar H.	1895	19	0	985	655	66.5%	276	28.0%
Karmarkar, Narendra K.	1480	9	0	1473	1164	79.0%	255	17.3%
Kesselman, Carl	2849	15	4	2730	667	24.4%	2027	74.2%
Leveson, Nancy G.	999	15	0	843	473	56.1%	318	37.7%
Liskov, Barbara H.	1221	15	0	1136	740	65.1%	372	32.7%
Lovasz, László	3668	29	2	2834	1939	68.4%	787	27.8%
Maier, David	1064	16	1	929	567	61.0%	325	35.0%
Manber, Udi	1154	15	4	1002	453	45.2%	520	51.9%
Marsan, Marco Ajmone	1140	16	3	830	482	58.1%	294	35.4%
Megiddo, Nimrod	2100	22	1	1627	1092	67.1%	465	28.6%
Micali, Silvio	3383	19	3	3155	1091	34.6%	1992	63.1%
Overmars, Mark H.	1856	18	1	1268	483	38.1%	727	57.3%
Parrow, Joachim	1221	9	2	1179	444	37.7%	718	60.9%
Peleg, David	1608	21	3	938	511	54.5%	401	42.8%
Pozar, David M.	3576	32	0	2847	2220	78.0%	489	17.2%
Preparata, Franco P.	3143	25	0	2453	1838	74.9%	405	16.5%
Raghavan, Prabhakar	1688	20	3	1286	651	50.6%	579	45.0%
Ramamritham, Krithi	70	6	1	55	23	41.8%	28	50.9%
Reps, Thomas	1538	16	4	1189	582	48.9%	568	47.8%
Salton, Gerard	3843	27	1	3429	1816	53.0%	1464	42.7%
Seshadri, Nambi	4972	18	3	4866	1676	34.4%	3123	64.2%
Sloane, Neil J.A.	5050	37	0	3397	2295	67.6%	670	19.7%
Soloway, Elliot	839	15	0	719	481	66.9%	198	27.5%
Tarjan, Robert E.	10014	51	0	8448	5435	64.3%	2775	32.8%
Vardi, Moshe Y.	2288	23	7	1393	661	47.5%	764	54.8%
Volakis, John L.	1817	22	0	833	655	78.6%	136	16.3%
Wing, Jeanette M.	1004	12	2	935	357	38.2%	545	58.3%
Zadeh, Lotfi A.	21807	28	1	21516	13403	62.3%	6909	32.1%
Average	**2680.3**	**19.6**	**2.1**	**2286.3**	**1232.1**	**53.7%**	**957.2**	**42.1%**
STdev	**3225.9**	**7.9**	**2.4**	**3145.9**	**1960.0**	**14.2%**	**1134.8**	**15.7%**

Citation counts

In Table 2 we display the data related to the citations, total citations, the h-index, the number of proceedings papers in the h-core, the number and percentage of journal citations to the h-core and number and percentage of proceedings citation to the h-core. The h-core, as defined by Burrell (2007) is the set of papers that are included in the computation of the h-index, where the h-index (Hirsch, 2005) of an author is the unique number such that h of his/her papers have h or more citations, and the rest of the papers have at most h citations. The sum of the percentages of the journal and proceedings papers in the h-core is less than 100%, because of the additional document types that are not recorded in the table.

The average h-index of these highly cited researchers is 19.6. Out of the papers in the h-core only 10.9% on the average are proceedings papers, i.e., most of the highly cited papers (at least in this sample) are journal articles. There are 16 scientists in the list for whom there are no proceedings papers at all in the h-core. The highest number of proceedings papers in the h-core is 11, which is 45.8% of the size of the h-core for Rajeev Alur. There are only three cases where the publication with the highest number of citations is a proceedings paper. On the other, hand, citations from conference papers constitute on average 42% of the citations to the papers in the h-core. For 18 researchers (36%) more than 50% of the citations to their h-core are from proceedings publications. Thus we see that even though conference papers are considered very important in computer science, the journal papers receive more citations on average than proceedings papers, at least for our sample. However, the results indicate that proceedings are a major source for citations, and the extension of WOS with the Proceedings Citation Indexes has a major impact on citation counts.

Republication of proceedings papers

Proceedings papers can be seen as a first publication of a result (sometimes a technical report or a preprint precedes it). Results published in a conference can be later republished (perhaps in an extended form) as a journal paper. To study the extent of this behaviour, we tested the publication lists of half of the sampled researchers, and looked for identical or almost identical titles of proceedings papers and journal articles. The results are displayed in Table 3. Note that we looked for near identical titles only; if the titles of the journal articles were considerably different we were not able to identify them as republications of the original. The findings are based on the titles only, and it is quite possible that the republished items are considerable extensions of the originals. In addition the journal coverage is from 1965 and onwards (our institutional subscription only provides access from 1965 and onwards), while the proceedings coverage is only from 1990 and onward. However, even if this is the case the republication of works has considerably effect of publication and citation counts – this point is further discussed in the next section of the paper.

We see from Table 3, that on average almost a quarter of the proceedings papers are republished as journal articles. In a few cases the item published first (the proceedings paper) received more citations than the journal paper that was published at a later time. In computer science there are considerable publication delays of journal articles which can explain why the proceedings papers are cited. However, if we consider the average citations per publication with and without the first published proceedings paper (only for items that were republished later as journal papers), we see that the average citation counts increase by approximately 15% on average if the first published proceedings papers are removed from the calculations.

Table 3. The extent of republication

Name	Publ. count	Proc. papers	Republ. items	% out of total publ.	% out of proc. papers	Ave. cit. per publ.	Ave. cit per publ. without first published proceedings papers
Aazhang, Behnaam	148	107	6	4.1%	5.6%	27.48	29.04
Alur, Rajeev	137	94	15	10.9%	16.0%	26.73	29.91
Beeri, Catriel	44	11	1	2.3%	9.1%	15.98	17.28
Blum, Manuel	41	13	2	4.9%	15.4%	47.46	58.03
Coppersmith, Don	105	28	10	9.5%	35.7%	17.20	18.55
Courcelle, Bruno	98	33	7	7.1%	21.2%	19.60	21.12
Dolev, Daniel	99	39	8	8.1%	20.5%	21.18	21.46
Dwork, Cynthia	58	25	4	6.9%	16.0%	20.05	24.19
Galil, Zvi	143	24	17	11.9%	70.8%	14.17	16.62
Garcia-Molina, Hector	180	110	4	2.2%	3.6%	13.38	14.08
Goldreich, Oded	136	53	19	14.0%	35.8%	20.20	24.08
Goldwasser, Shafi	56	29	4	7.1%	13.8%	47.77	65.60
Karmarkar, Narendra K.	20	8	0	0.0%	0.0%	74.00	82.22
Leveson, Nancy G.	70	23	2	2.9%	8.7%	14.27	14.69
Lovasz, László	153	26	6	3.9%	23.1%	23.97	24.84
Manber, Udi	51	12	3	5.9%	25.0%	22.63	25.39
Megiddo, Nimrod	102	18	5	4.9%	27.8%	20.59	21.85
Micali, Silvio	72	38	9	12.5%	23.7%	47.22	70.32
Peleg, David	188	86	34	18.1%	39.5%	8.55	10.68
Raghavan, Prabhakar	149	79	11	7.4%	13.9%	11.32	11.85
Salton, Gerard	123	6	2	1.6%	33.3%	31.24	32.28
Sloane, Neil J.A.	192	13	7	3.6%	53.8%	26.30	27.28
Tarjan, Robert E.	170	32	11	6.5%	34.4%	58.91	63.23
Vardi, Moshe Y.	205	124	19	9.3%	15.3%	11.22	12.13
Wing, Jeanette M.	47	16	3	6.4%	18.8%	21.36	22.82
Average	**111.48**	**41.88**	**8.36**	**6.9%**	**23.2%**	**26.51**	**30.38**
Standard deviation	**54.56**	**35.79**	**7.65**	**4.3%**	**16.0%**	**16.26**	**20.30**

Discussion

The results show that proceedings publications have a major effect on the publication and citation counts of highly cited computer scientists. In this section we discuss the implications of republication of works, which as we saw above is rather prevalent in computer science. Its extent is probably even greater than what can be seen in Table 3.

The FRBR (IFLA, 1998) defines four entities for describing products of intellectual or artistic endeavour:
Work – a distinct intellectual or artistic creation, an abstract entity.
Expression – realization of a *work.* FRBR views "variant texts incorporating revisions or updates to an earlier text are viewed simply as *expressions* of the same *work"* (FRBR, 1998, p. 16). But on the other hand, rewritings, abstracts and summaries are considered to represent new *works*.

Manifestation - the physical embodiment of an *expression* of a *work*. When production involves changes in the physical form, it results in a new *manifestation*.
Item – a single exemplar of a *manifestation*.
Moed and Visser (2007) used the term *concept* to describe something similar to the FRBR entity *work*. Bar-Ilan (2005) used the term *manifestation* to describe different versions of the concept, i.e., republication of a result in different publication venues. However a closer examination of the FRBR taxonomy shows that *manifestation* is not the appropriate terminology for what we are trying to describe here. Taylor (2007) would almost definitely views each republished item as a different *work*, because usually the journal publication extends and revises the proceedings paper. However, we prefer to talk about different *expressions* of the same *work*.

Publishing several *expressions* of the same *work* has far-reaching effects on both publication and citation counts. When the ISI indexed journal articles only, the extent of indexing several *expressions* of the same *work* was very small. Now the situation changed considerably. The extension of WOS results in a direct increase in publication counts and also in citation counts because of the increase in the number of source items. Such increase takes place also when ISI increases its journal coverage, but extending journal coverage does not necessarily increase the number of *expressions* of a *work* indexed by the database. However, the addition of proceedings papers as source items to the database does not simply increase the quantity of the source items, but results in an increase in multiple *expressions* of the same *work* in the database. Although, usually journal articles based on previously published proceedings papers are not identical to the proceedings paper, still there is a huge overlap in the reference lists of the two publications. Thus, multiple expressions of a *work* not only have a positive effect on the publication counts of the authors, but they also have a positive effect on the citation counts of items referenced in the publications. Thus it seems that researchers living in the "publish or perish" and "get cited" world should all benefit from multiple *expressions* of a *work*.

However, there is another point to be considered. It is not always enough to get cited, sometimes authors want (or need for promotion) to have highly-cited papers that count when computing the h-index. In this case multiple *expressions* can have an adverse effect, because instead of citing a single *expression* of a *work*, the referring author can chose from several *expressions*, resulting in a dispersal of citations among the different *expressions*. It can happen that none of the *expression*s becomes part of the h-core, whereas combined citation counts from all the *expressions* of the same *work* would have increased the h-index.

If we opt for counting citations to a single *work*, instead of a specific *expression* we come across interesting, new problems: how do we differentiate between *works* and *expressions*? Does a thoroughly extended and revised paper, where errors in the proceedings paper are corrected constitute a new *work*? Or (as it sometimes happens) when the set of authors involved in the proceedings paper is not identical to the set of authors of the journal article, are we still speaking of multiple *expressions*? These questions were of minor relevance when the citation database for a very high percentage of the items contained only a single *expression* of each *work*. Now that WOS is extended with the Proceedings Citation Indexes and Scopus also covers a considerable number of proceedings we will have to deal with these questions.

Conclusion

This paper studied the effects of the extension of WOS with the Conference Proceedings Citation Indexes. As a case study we examined the publications and the citations of a set of highly cited computer scientists. The results based on WOS show a considerable increase in both publication and citation counts.

We recommend further studies in this area examining the effects of conference publications in Scopus and comparing the researchers' publication lists with the items indexed by the Web of Science.

We also raised some theoretical questions regarding works, expressions and manifestations of products of intellectual endeavours.

References

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, 1553-1566.

Burrell, Q. L. (2007). On the h-index, the size of the Hirsch core and Jin's A-index. *Journal of Informetrics*, 1, 170-177.

Elsevier (2008). *Scopus Info – Scopus in Detail – What does it Cover?* Retrieved December 29, 2008 from: http://info.scopus.com/detail/what/titles.asp.

Garfield, E. (1977). The mystery of the transposed journal lists – wherein Bradford's Law of Scattering in generalized according to Garfield's Law of Concentration. In *Essays of An Information Scientist,* vol 1. (pp. 222-223). Philadelphia: ISI Press.

Garfield, E. (1979). *Citation Indexing – Its Theory and Application in Science, Technology and Humanities*. New York: John Wiley & Sons.

Glänzel, W., Schlemmer, B., Schubert, A. & Thijs, B. (2005). Proceedings literature as additional data source for bibliometric analysis. In P. Ingwersen and B. Larsen (Eds.), *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics* (ISSI 2005) (pp. 159-167).Stockholm: Karolinska University Press.

Goodrum, A. A., McCain, K. W., Lawrence, S. & Giles, C. L. (2001). Scholarly publishing in the Internet age: A citation analysis of computer science literature. *Information Processing and Management*, 37 (661-675).

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

IFLA (1998). Functional requirements for bibliographic records. Final report. *UBCIM Publications – New Series*, Vol. 19. Retrieved January 10, 2008, from http://www.ifla.org/VII/s13/frbr/frbr.pdf

Kling, R. & McKim, G. (1999). Scholarly communication and the continuum of electronic publishing. *Journal of the American Society for Information Science*, 50(10), 890-906.

Lisée, C., Larivière, V. & Archembault, É. (2008). Conference proceedings as a source of scientific information. *Journal of the American Society for Information Science and Technology,* 59(11), 1776-1784.

Meho, L. I. & Rogers, Y. (2008). Citation counting, citation ranking and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology,* 59(11), 1711-1726.

Moed, H. F. & Visser, M. S. (2007). *Developing Bibliometric Indicators of Research performance in Computer Science: An Exploratory Study*. Research Report to the Council of Physical Sciences of the Netherlands Organisation for Scientific Research (NWO). CWTS Report 2007-1. Retrieved January 10, 2008, from http://www.cwts.nl/pdf/NWO_Inf_Final_Report_V_210207.pdf

Taylor, A. G. (2007). Understanding FRBR: What It Is And How It Will Affect Our Retrieval Tools? Westport, CN: Libraries Unlimited.

Thomson Reuters (2008). *Web of Science Help*. Retrieved December 29, 2008 from: http://images.isiknowledge.com/help/WOS/h_database.html.

Visser, M. S. & Moed, H. F. (2005). Developing bibliometric indicators of research performance in computer science. In P. Ingwersen and B. Larsen (Eds.), *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics* (ISSI 2005) (pp. 275-279).Stockholm: Karolinska University Press.