Available online at www.sciencedirect.com
ScienceDirect
Journal of Informetrics 1 (2007) 26–34
Journal of INFORMETRICS An International Journal
www.elsevier.com/locate/joi

Some measures for comparing citation databases

Judit Bar-Ilan[a,*], Mark Levene[b], Ayelet Lin[a]

[a] *Department of Information Science, Bar-Ilan University, Ramat Gan 52900, Israel*
[b] *School of Computer Science and Information Systems, Birkbeck University of London, Malet Street, London WC1E 7HX, UK*

Received 29 June 2006; received in revised form 3 August 2006; accepted 3 August 2006

Abstract

Citation analysis was traditionally based on data from the ISI Citation indexes. Now with the appearance of Scopus, and with the free citation tool Google Scholar methods and measures are need for comparing these tools. In this paper we propose a set of measures for computing the similarity between rankings induced by ordering the retrieved publications in decreasing order of the number of citations as reported by the specific tools. The applicability of these measures is demonstrated and the results show high similarities between the rankings of the ISI Web of Science and Scopus and lower similarities between Google Scholar and the other tools.

Keywords: Similarity measures; Rankings; Citation databases

1. Introduction

Citation analysis is a major subfield of informetrics. Until recently the only comprehensive tool for carrying out empirical research in this area was the ISI Citation Indexes (see for example White's (2001) discussion on CAMEOs). This situation has changed, at first in individual disciplines (like CiteSeer in computer science), and now with the introduction of Elsevier's Scopus and Google Scholar.

Citation data is heavily influenced by the coverage of the specific database, since it can take into account only citations from items indexed by it. The three major tools: Web of Science (the Web version of the ISI Citation Indexes), Scopus and Google Scholar were compared and reviewed in several publications from different aspects (for example: Bauer & Bakkalbasi, 2005; Deis & Goodman, 2005; Jacso, 2005a, 2005b; Noruzi, 2005; Bar-Ilan, 2006). CiteSeer and SCISearch (a different interface of the ISI Science Citation Index) were compared by Goodrum, McCain, Lawrence, and Giles (2001). The above-mentioned studies provided numbers and descriptive statistics as a means for comparing between the different tools.

With the existence of multiple citation databases it becomes necessary to compare them systematically both from the scientometric and the informetric points of view. Descriptive statistics and specific examples are not sufficient for systematic comparison of the different citation databases. In this paper we introduce a set of measures for comparing the different citation databases. The measures compute the similarities between the rankings induced by the number of citations a publication receives in the specific database (i.e. the most cited item is ranked number 1, the second most

* Corresponding author. Tel.: +972 523667326; fax: +972 3 5353937.
E-mail addresses: barilaj@mail.biu.ac.il (J. Bar-Ilan), M.Levene@dcs.bbk.ac.uk (M. Levene), lineyal@netvision.net.il (A. Lin).

1751-1577/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2006.08.001

cited is ranked number 2, etc.). The use of these measures and statistical analysis of the results is demonstrated on a subset of the highly cited Israeli researchers, as defined in ISI's Highly Cited database (ISI HighlyCited.com, 2002) supplemented by the three recent Israeli Nobel prize winners.

The measures are defined in Section 2, the data collection and empirical settings appear in Section 3. In Section 4 the results are displayed and analyzed, and Section 5 concludes the paper.

2. The measures

The rankings were compared using four basic measures that complement each other. In this section the measures are defined. Each of the measures is defined for a pair of databases (A and B), where A and B can WoS (Web of Science), Scopus or Google Scholar. The measures introduced here were applied to comparing rankings of search engine rankings (Bar-Ilan, Mat-Hassan, & Levene, 2006; Bar-Ilan, Levene, & Mat-Hassan, 2006; Bar-Ilan, Keenoy, Yaari, & Levene, submitted for publication).

2.1. Overlap and footrule

Overlap (O) is defined as follows:

$$O = \frac{|\mathrm{PUBL_A} \cap \mathrm{PUBL_B}|}{|\mathrm{PUBL_A} \cup \mathrm{PUBL_B}|}$$

where PUBL_X is the set of publications retrieved from database X. The measure O does not take into account the rankings, it only measures the proportion of the publications retrieved from both databases out of the total number of publications retrieved by either of them.

Footrule, F, is the normalized Spearman footrule. Spearman's footrule (Diaconis & Graham, 1977; Dwork, Kumar, Naor, & Sivakumar, 2001) can be computed for two permutations, and thus it can be applied only for the publications that are ranked in both databases. Each such publication is given its relative rank in the set of publications retrieved from both databases. Suppose for the moment that there are no ties in the rankings (i.e. no two publications receive exactly the same number of items). This is an unrealistic assumption and we will deal with it in Section 3. The result of the re-rankings is two permutations σ_1 and σ_2 on $1 \ldots Z$ where $|Z|$ is the number of overlapping publications. After these transformations Spearman's footrule is computed as

$$Fr^{|Z|}(\sigma_1, \sigma_2) = \sum_{i=1}^{|Z|} |(\sigma_1(i) - \sigma_2(i))|$$

When the two rankings are identical on the set Z, $Fr^{|Z|}$ is zero, and its maximum value is $|Z|^2$ when $|Z|$ is even, and $(|Z|+1)(|Z|-1)$ when $|Z|$ is odd. When the result is divided by its maximum value, $Fr^{|Z|}$ will be between 0 and 1, independent of the size of the overlap. This measure is undefined for $|Z| = 0,1$. Thus we compute the *normalized Spearman's footrule*, NFr, for $|Z| > 1$

$$NFr = \frac{Fr^{(|Z|)}}{\max Fr^{(|Z)}}$$

NFr ranges between 0 and 1; it attains the value 0 when the relative ranking of the publications in the set Z is identical. Since we are interested in similarity measures, we define F as

$$F = 1 - NFr$$

The weakness of this measure is that it totally ignores the non-overlapping elements and only takes into account the relative rankings, thus for example if $|Z| = 2$, and these two publications are ranked at ranks 1 and 2 in database A, while in database B they are ranked at 9 and 10 (and the first eight publications are not ranked in database A), the value of F will be 1, just like the case where both A and B rank these two publications at ranks 1 and 2, respectively.

2.2. Fagin measure

Spearman's footrule is a very useful measure to compare the ordering in two permutations (Diaconis & Graham, 1977; Dwork et al., 2001). However when comparing two sets of ranked results the underlying sets are often not identical. Fagin, Kumar, and Sivakumar (2003) extended Spearman's footrule in such a way that the measure does not require that both rankers rank exactly the same set of items. They developed the measure for comparing search engine rankings, but here we modify the description to fit the current setting. Suppose that exactly k publications were retrieved from both databases (not necessarily the same publications). The number of citations each publication receives induces a natural ranking on these items. Suppose for the moment that there are no ties in the rankings (i.e. no two publications receive exactly the same number of items). This is an unrealistic assumption and we will deal with it in Section 3. Each publication that was retrieved from A, but not from B is artificially assigned rank $k+1$ (similarly for the publications retrieved from B and not from A). The rationale for this artificial rank is that if A indexes the specific item its rank would be $k+1$ or more (note that if it is not indexed by A it would not be ranked at all).

Let Z be the set of publications retrieved by both databases, S the set of publications retrieved only from A and T the set of publications retrieved only from B, σ_1 the ranking of the publications retrieved from A and σ_2 the ranking of the publications retrieved from B then

$$F^{(k+1)}(\sigma_1, \sigma_2) = \sum_{i \in Z} |\sigma_1(i) - \sigma_2(i)| + \sum_{i \in S} ((k+1) - \sigma_1(i)) + \sum_{i \in T} ((k+1) - \sigma_2(i))$$

This measure has to be normalized so that when the two rankings are on identical sets in identical order the measure equals 1, and when there is no overlap between the sets, the measure is 0. Thus

$$G^{(k+1)} = 1 - \frac{F^{(k+1)}}{\max F^{(k+1)}}$$

where $\max F^{(k+1)} = k(k+1)$.

In case the number of retrieved elements from both databases is not identical, we introduce the following modification of the measure: suppose k_1 publications were retrieved from database A and k_2 from database B, then

$$F^{(k_1 k_2)}(\sigma_1, \sigma_2) = \sum_{i \in Z} |\sigma_1(i) - \sigma_2(i)| + \sum_{i \in S} ((k_2+1) - \sigma_1(i)) + \sum_{i \in T} ((k_1+1) - \sigma_2(i))$$

and

$$G^{(k_1,k_2)} = 1 - \frac{F^{(k_1,k_2)}}{\max F^{(k_1,k_2)}}$$

where

$$\max F^{(k_1,k_2)} = \frac{k_1(k_1+1)}{2} + \frac{k_2(k_2+1)}{2}$$

This measure unlike the footrule, takes into account the non-overlapping publications as well, but in our opinion it gives too much weight to these elements. Suppose that, for two lists of ten ranked results, $|Z| = 5$, and A ranks z_1 at position 1, z_2 at position 2 ... and z_5 at position 5; B ranks z_1 at position 5, z_2 at position 4 ... and z_5 at position 1. In this case $G^{(10,10)}$ equals 0.618. Now if B ranks the items in Z exactly like A, $G^{(10,10)}$ increases only slightly to 0.727. The amount of change in G for a given overlap is rather small, since G is mainly determined by the size of the overlap.

2.3. Inverse rank measure

Our last measure attempts to correct this problem, by giving more weight to identical or near identical rankings among the top ranking publications. This measure tries to capture the intuition that identical or near identical rankings among the top publications indicate greater similarity between the rankings induced by the databases. First, let

$$N^{(k_1,k_2)}(\sigma_1 \sigma_2) = \sum_{i \in Z} \left| \frac{1}{\sigma_1(i)} - \frac{1}{\sigma_2(i)} \right| + \sum_{i \in S} \left| \frac{1}{\sigma_1(i)} - \frac{1}{(k_2+1)} \right| + \sum_{i \in T} \left| \frac{1}{\sigma_2(i)} - \frac{1}{(k_1+1)} \right|$$

where k_1, k_2, S, T, Z and σ_i are as before. This measure has to be normalized as well, thus

$$M^{(k_1,k_2)} = 1 - \frac{N^{(k_1,k_2)}}{\max N^{(k_1,k_2)}}$$

where

$$\max N^{(k_1,k_2)} = \sum_{i=1}^{k_1} \left(\frac{1}{i} - \frac{1}{k_2+1} \right) + \sum_{i=1}^{k_2} \left(\frac{1}{i} - \frac{1}{k_1+1} \right)$$

Considering the same two cases as before (five overlapping elements, opposite versus identical rankings), the M values will be 0.386 and 0.905, respectively, emphasizing the importance of similarity in rankings in the top positions. Now suppose that $|Z| = 5$ as before, but the overlapping elements are ranked 6, 7, 8, 9 and 10 by both A and B. In this case $G^{(10,10)}$ is 0.182 (compared with 0.727 when the overlapping elements were identically ranked in the top positions) while $M^{(10,10)}$ is 0.149 (compared with 0.905 when the overlapping elements were identically ranked in the top positions)—showing the larger weight given by the M measure to overlapping elements in the top positions.

3. Data collection

To demonstrate the feasibility of the measures, we applied them to the publications of the highly cited Israeli scientists (ISI HighlyCited.com, 2002), as defined by the ISI (ISI HighlyCited.com). This list is based on citations to items indexed by ISI and were published between 1981 and 1999. The list is comprised of 44 names. Rather interestingly it does not include the three Israeli Nobel prize winners in the last two years (Robert Aumann, Aaron Ciechanover and Avram Hershko). These three names were added to the list. We had disambiguation problems with a few of the names, and as a result we excluded eight names.

Scopus only provides full citation data of items from 1996 and onwards. In order to have a "fair" comparison, only publications from 1996 and onwards were considered. Note that this is a different period from the period for which the researcher was included in the list of highly cited authors, thus it may well be the case that during the period under consideration the researcher will have few or no publications.

We only considered highly cited papers of the highly cited researchers, and thus only items with 20 or more citations in the specific database were retrieved. There were two reasons for this decision: (1) the data had to be carefully cleansed (especially from Google Scholar) and by considering only the most highly cited items, we were able to carry out the cleansing in reasonable time and (2) as noted in the previous section we had to find a solution for "ties", i.e. two publications that received the same number of citations in the specific database. Among the more highly cited items there were fewer occurrences of ties. Note that as a result of the decision to retrieve publications with twenty or more citations only, we do not have the information whether an item retrieved from database A, and not retrieved from database B is indexed by B (but received less than 20 citations from sources indexed by B) or is not indexed at all by B. This is the usual assumption when applying the measures discussed in Section 2.

In the final list for analysis we only included scientists whose publications appeared in all three databases and had at least three items published from 1996 onwards with 20 or more citations. Thus we had to exclude 16 additional names: two researchers had no highly cited publications in any of the three databases, one had 60 highly cited papers in WOS, two in Google Scholar and none in Scopus (a physicist), 12 (computer scientists and/or mathematicians, and one pharmacologist) had a considerable number of highly cited publications indexed by Google Scholar, but less than three by either WOS or Scopus (usually both). These differences are due to the fact that Google indexes books, proceedings and technical reports as well, but WOS excludes these types of publications almost entirely and Scopus indexes them only in a limited fashion. There was only a single case where the scientist (a hydrologist) was excluded because of the lack of enough highly cited items indexed by Google Scholar. The final list was comprised of 22 scientists.

Table 1 lists these scientists together with the number of items with more than 20 citations from each database and the total number of citations received by these items. The searches were carried out during the second half of January 2006. There is no clear "winner", but it seems that Google Scholar retrieves more items and citations in computer science and less in chemistry. The differences between the Web of Science and Scopus are not as significant. An interesting case is Ehud Duchovni—he is a high energy physicist, a member of the Opal and Atlas groups conducting experiments at

Table 1
The list of scientists examined, the number of highly cited publications and the total number of citations these publications received in each citation database

Scientist	Affiliation	Discipline	WOS		Scopus		Google Scholar	
			Items	Total citations	Items	Total citations	Items	Total citations
Alon, Noga	Tel Aviv U.	Mathematics, computer science	8	220	9	274	30	1438
Aurbach, Doron	Bar Ilan U.	Materials science	40	2073	40	2067	18	562
Chet, Ilan	Weizmann Inst.	Plant & animal science	17	552	17	593	16	567
Ciechanover, Aaron	Technion	Molecular biology & genetics	38	6194	41	6309	35	5202
Cohen, Irun R.	Weizmann Inst.	Immunology	37	2210	40	2357	34	1910
Dekel, Avishai	Hebrew U.	Space sciences	27	1618	18	1162	14	1220
Duchovni, Ehud	Weizmann Inst.	Physics	63	2352	29	1038	47	1891
Geiger, Benjamin	Weizmann Inst.	Molecular biology & genetics	44	3836	45	3853	42	3282
Goldreich, Oded	Weizmann Inst.	Computer science	8	326	5	246	39	2957
Harel, David	Weizmann Inst.	Computer science	3	112	4	264	23	3090
Hershko, Avram	Technion	Molecular biology & genetics	21	3743	21	3705	20	2888
Jortner, Joshua	Tel Aviv U.	Chemistry	28	1760	25	1436	15	621
Kanner, Joseph	Agricultural Research Organization	Agricultural sciences	4	195	4	201	3	88
Kerem, Batsheva	Hebrew U.	Molecular biology & genetics	18	969	17	943	14	673
Mechoulam, Raphael	Hebrew U.	Pharmacology	30	2110	33	2393	30	1543
Oren, Moshe	Weizmann Inst.	Molecular biology & genetics	66	7021	65	7227	61	6156
Piran, Tsvi	Hebrew U.	Space sciences	38	3016	28	1807	30	2943
Procaccia, Itamar	Weizmann Inst.	Physics	12	439	13	510	13	476
Shamai, Shlomo	Technion	Computer science	12	658	14	1083	22	1961
Sharir, Micha	Tel Aviv U.	Engineering, computer science	4	114	7	175	20	782
Sklan, David	Hebrew U.	Agricultural sciences	11	311	13	346	4	102
Turkel, Eli	Tel Aviv U.	Mathematics	3	90	4	124	4	138

CERN in Geneva. There are more than one hundred members in these groups, and all of them coauthor each publication. The Web of Science indexes all the authors, while Scopus does not. Another Israeli member of these groups is Giora Mikenberg (also in the list of highly cited researchers). The list of highly cited items he authored is highly similar to the list of Ehud Duchovni on Web of Science, but there were no items retrieved from Scopus—probably due to the fact that his name is further down on the list (M vs. D!) and therefore the Opal and Atlas publications were not attributed to him on Scopus.

The measures described in Section 2 can only be applied to ranked lists without ties. When the ranking is induced by the number of citations received there are often ties. Ties were resolved in the following way: suppose items x and y were retrieved by database A and have exactly the same citation count.

Case 1. *x* and *y* were also retrieved by B

a) x received more citations than y in database B. In this case $\text{rank}_A(x)<\text{rank}_A(y)$ (recall that the lower rank numbers correspond to higher citation counts).
b) y received more citations than x in database B. In this case $\text{rank}_A(y)<\text{rank}_A(x)$ (recall that the lower ranks correspond to higher citation counts).
c) x and y were tied also in B. In this case the decision is arbitrary, but consistent in both lists, i.e. either $\text{rank}_A(x)<\text{rank}_A(y)$ and $\text{rank}_B(x)<\text{rank}_B(y)$ or $\text{rank}_A(y)<\text{rank}_A(x)$ and $\text{rank}_B(y)<\text{rank}_B(x)$.

Case 2. Only one of the items, say x, is retrieved by B, then $\text{rank}_A(x) < \text{rank}_A(y)$.

Case 3. Neither x nor y are retrieved by B, then the decision is arbitrary.

This tie-resolution algorithm can be easily extended to the case where more than two items are tied. Note that the tie-resolution is dependent on the database B, thus different rankings may result when comparing the results of A to B or to C.

4. Results

The four measures introduced in Section 2 were computed for each researcher and for each pair of databases. The results are displayed in Table 2. Note that values above 0.7 indicate high similarity; this threshold was set arbitrarily and is similar to the accepted threshold for "high correlation".

We see that there is complete agreement between Scopus and the Web of Science on the ranked lists of Avram Hershko and Joseph Kanner. In Kenner's case the list is only four items long, but in Hershko's case we are comparing lists of length 21. When looking at the number of citations on a per item basis, we see that the Web of Science recorded slightly more citations than Scopus, for example for the top ranked item, the Web of Science reported 1919 citations, whereas Scopus reported 1909 citations. There was almost total agreement on Eli Turkel's list as well, the only difference was that WoS listed only three publications with more than 20 citations, and Scopus listed four. Checking WoS we observed that the fourth item on Scopus is indeed number four on the WoS list as well, but was not retrieved because it received only 18 citations.

There was much less agreement between Google Scholar and the other two databases. One of the most striking results is Oded Goldreich. Scopus listed 5 items and Google Scholar 39 items, with only two overlapping items (publications that were retrieved by both database) and these two items appeared in opposite order in the two lists

Table 2
Similarity values for the ranked lists from Web of Science, Scopus and Google Scholar

Researcher	Web of Science – Scopus				Web of Science – Google Scholar				Scopus – Google Scholar			
	O	*F*	*G*	*M*	*O*	*F*	*G*	*M*	*O*	*F*	*G*	*M*
Alon	0.545	0.778	0.725	0.584	0.276	0.750	0.457	0.468	0.267	0.750	0.466	0.703
Aurbach	0.951	0.942	0.972	0.982	0.415	0.639	0.652	0.552	0.415	0.694	0.663	0.618
Chet	0.889	0.922	0.961	0.884	0.833	0.839	0.872	0.842	0.737	0.796	0.889	0.816
Ciechanover	0.927	0.981	0.969	0.989	0.775	0.921	0.928	0.916	0.762	0.926	0.941	0.922
Cohen	0.925	0.942	0.964	0.963	0.821	0.840	0.873	0.922	0.762	0.855	0.874	0.920
Dekel	0.667	0.938	0.767	0.828	0.323	0.760	0.468	0.574	0.231	1	0.412	0.647
Duchovni	0.394	0.728	0.535	0.377	0.528	0.825	0.736	0.589	0.357	0.700	0.480	0.304
Geiger	0.978	0.973	0.987	0.980	0.870	0.918	0.926	0.919	0.891	0.907	0.920	0.912
Goldreich	0.444	1.000	0.792	0.911	0.119	0.833	0.222	0.157	0.073	0	0.146	0.112
Harel	0.750	0.500	0.600	0.353	0.130	1	0.237	0.134	0.174	0.750	0.322	0.280
Hershko	1	1	1	1	0.952	0.880	0.941	0.949	0.952	0.880	0.932	0.946
Jortner	0.893	0.878	0.891	0.875	0.448	0.643	0.686	0.701	0.444	0.528	0.639	0.604
Kanner	1	1	1	1	0.750	1	0.867	0.928	0.750	1	0.867	0.928
Kerem	0.944	0.972	0.895	0.791	0.778	0.918	0.787	0.774	0.824	0.939	0.855	0.897
Mechoulam	0.853	0.962	0.961	0.966	0.781	0.821	0.892	0.904	0.765	0.799	0.898	0.889
Oren	0.985	0.961	0.961	0.913	0.866	0.904	0.911	0.885	0.881	0.909	0.925	0.953
Piran	0.585	0.910	0.710	0.716	0.457	0.827	0.639	0.691	0.400	0.797	0.616	0.792
Procaccia	0.923	0.944	0.964	0.978	0.786	0.800	0.887	0.853	0.733	0.800	0.881	0.853
Shamai	0.857	0.750	0.884	0.863	0.545	0.639	0.754	0.790	0.636	0.878	0.851	0.921
Sharir	0.571	0.500	0.800	0.519	0.091	1	0.216	0.447	0.125	0.500	0.271	0.304
Sklan	0.500	0.875	0.787	0.871	0.250	0.500	0.507	0.663	0.308	0.750	0.565	0.700
Turkel	0.750	1	1	1	0.400	1	0.533	0.331	0.600	1	0.600	0.377
Average	0.788	0.884	0.869	0.834	0.554	0.830	0.681	0.681	0.549	0.780	0.682	0.700
S.D.	0.198	0.147	0.136	0.200	0.278	0.134	0.242	0.249	0.276	0.220	0.247	0.262

Table 3
Results of the statistical tests for the measures O, G and M[a]

		O	G	M
WoS-Scopus	Mean	0.788	0.869	0.834
	S.D.	0.198	0.136	0.200
WoS-GS	Mean	0.554	0.681	0.681
	S.D.	0.278	0.242	0.249
Scopus-GS	Mean	0.549	0.682	0.700
	S.D.	0.296	0.247	0.262
F		38.266***	21.674***	8.433**
WoS-Scopus vs. WoS-GS		***	***	**
WoS-Scopus vs. Scopus-GS		***	***	*
WoS-GS vs. Scopus-GS		–	–	–

(*), (**) and (***) indicate the strength of the significance, (–) indicates that no differences were found. Levels of significance: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.
[a] WoS-Scopus vs. WoS-GS checks whether the significant F-value was caused by differences in the WoS-Scopus vs. WoS-GS measures.

(that is why $F = 0$). The G and M values are very low because of the extremely small overlap and the disagreement in the ordering of the overlapping elements. Note that the items listed by Google Scholar were checked against Oded Goldreich's list of publication and/or the publisher's site, non-existing publications were removed and publications listed more than once were collated. We observe a similar pattern when comparing David Harel's list in WoS versus Google Scholar. In this case there were three overlapping elements (all the elements listed by WoS), but in his case there was total agreement on the relative ranking, i.e. the items that were ranked 1, 2 and 3 respectively on WoS, were ranked 3, 5 and 6 on Google Scholar. Thus in this case $F = 1$, but the G and M values are low, because Google Scholar ranked 23 publications versus 3 by WoS and there was no agreement on the top ranked items of Google Scholar (these were not listed by WoS). Note that the most cited item according to Google Scholar is a book and WoS does not index books.

From looking at the averages in Table 2, it seems that the WoS and Scopus rankings are rather similar, whereas the Google Scholar ranking is considerably different. However, the standard deviations are considerable. Thus we decided to run some statistical tests. We ran the repeated measure ANOVA test for each query, for each database and for each measure, to compare the rankings of the databases for the 22 researchers. This test measures the variability of the database rankings (i.e. whether the similarity between the rankings of database A and B is significantly different from the similarity between the rankings of database A and C); see for example (Grimm & Yarnold, 2005) If the results of this F-test (ANOVA) is significant then it means that the differences between the three pairs WoS-Scopus, WoS-GS (GS stands for Google Scholar) and Scopus-GS cannot be explained by random errors, but there are consistent differences. If the F-value is significant, one can run additional tests; in our case the appropriate tests were Bonferroni-adjusted post-hoc tests to determine the differences between which two pairs are responsible for the significance of the F-test. There are cases where the differences between more than two pairs are significant. The tests were run on all four measures, however the footrule (F) did not fulfill the test assumptions (sphericity), and thus here we report only the significance of the statistical tests for three measures, O, G and M (see Table 3).

The results indicate that the significant differences for the set of researchers tested for this study were caused by the considerable differences in the rankings of Google Scholar as compared to either the Web of Science or Scopus.

Google Scholar (and to some extent Scopus as well) indexes proceedings and books as well, while WoS indexes mainly journal papers. For three researchers: Alon, Goldreich and Harel the number of items indexed by Google Scholar was considerably higher than the number of items indexed by the other two databases. For these researchers we compared the rankings induced by the three databases, when considering journal papers only. The results appear in Tables 4 and 5.

As can be seen in Table 5, the differences remain considerable even when only journal publications are taken into account.

Table 4
Number of journal papers indexed by each database

Scientist	Affiliation	Discipline	WOS		Scopus		Google Scholar	
			Items	Total citations	Items	Total citations	Items	Total citations
Alon, Noga	Tel Aviv U.	Mathematics, computer science	8	220	8	243	20	210
Goldreich, Oded	Bar Ilan U.	Computer science	8	302	4	220	7	776
Harel, David	Weizmann Inst.	Computer science	3	112	4	264	9	1613
Sharir, Micha	Technion	Engineering, computer science	4	114	7	175	11	457

Table 5
Similarity values for the ranked lists from Web of Science, Scopus and Google Scholar when considering journal papers only

Researcher	Web of Science – Scopus				Web of Science – Google Scholar				Scopus – Google Scholar			
	O	*F*	*G*	*M*	*O*	*F*	*G*	*M*	*O*	*F*	*G*	*M*
Alon	0.600	0.778	0.778	0.615	0.400	0.750	0.615	0.545	0.400	0.750	0.641	0.803
Goldreich	0.571	1.000	0.905	0.965	0.400	0.750	0.556	0.370	0.222	0.000	0.400	0.287
Harel	0.750	0.500	0.600	0.353	0.333	1.000	0.533	0.358	0.444	0.750	0.760	0.869
Sharir	0.571	0.500	0.800	0.519	0.154	1.000	0.377	0.601	0.200	0.500	0.422	0.362

5. Conclusions

In this paper we introduced a set of measures for comparing rankings of different citation databases induced by the number of citations the tested publications receive in each database.

The results indicate that Scopus and the Web of Science are comparable in terms of the rankings induced. Note that the measures were computed only for a small set of cases, in order to be able to generalize the results, larger-scale, discipline-specific tests should be carried out.

Some of the differences are caused by the differing indexing strategies of the databases. Google Scholar does not have a clear policy, but unlike WoS it indexes books and proceedings as well. These types of publications are often cited more than journal papers, especially in computer science. Had we compared the rankings only on journal papers, the similarity measures.

References

Bar-Ilan, J. (2006). H-index for Price medalists revisited. *ISSI Newsletter*, *2*(1), 3–5.

Bar-Ilan, J., Levene, M., & Mat-Hassan, M. (2006). Methods for evaluating dynamic changes in search engine rankings – A case study. *Journal of Documentation*, *62*(6).

Bar-Ilan, J., Keenoy, K., Yaari, E., & Levene, M. (submitted for publication). User rankings of search engine results.

Bar-Ilan, J., Mat-Hassan, M., & Levene, M. (2006). Methods for comparing search engine results. *Computer Networks*, *50*(10), 1448–1463.

Bauer, K., & Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*, *11*(9). Retrieved June 16, 2006, from http://www.dlib.org/dlib/september05/bauer/09bauer.html.

Deis, L. F., & Goodman, D. (2005). Web of Science (2004 version) and Scopus. *The Charleston Advisor*, *6*(3), 5–21. Retrieved June 16, 2006, from http://www.charlestonco.com/comp.cfm?id=43.

Diaconis, P., & Graham, R. L. (1977). Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society, Series B (Methodological)*, *39*, 262–268.

Dwork, C., Kumar, R., Naor, M., & Sivakumar, D. (2001). Rank aggregation methods for the Web. In *Proceedings of the 10th World Wide Web Conference* (pp. 613–622).

Fagin, R., Kumar, R., & Sivakumar, D. (2003). Comparing top k lists. *SIAM Journal on Discrete Mathematics*, *17*(1), 134–160.

Goodrum, A. A., McCain, K. W., Lawrence, S., & Giles, L. C. (2001). Scholarly publishing in the Internet age: A citation analysis of computer science literature. *Information Processing and Management*, *37*(6), 661–675.

Grimm, L. G., & Yarnold, P. R. (2005). *Reading and understanding multivariate statistics*. Washington, DC: American Psychological Association.

ISI HighlyCited.com (2002). ISI Highly Cited Researchers – Country: Israel. Retrieved June 16, 2006, from http://hcr3.isiknowledge.com/browse_author.pl?page=0&link1=Browse&valueCategory=0&valueCountry=81&submitCountry.x=18&submitCountry.y=6.

ISI HighlyCited.com. About ISI HighlyCited.com. Retrieved June 16, 2006, from http://hcr3.isiknowledge.com/popup.cgi?name=hccom.

Jacso, P. (2005). 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). Comparison and analysis of the citedness scores in Web of Science and Google Scholar. In Proceeding of Digital Libraries: Implementing Strategies and Sharing Experiences, *Lecture Notes in Computer Science*, 3815, 360–369.

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

White, H. D. (2001). Author-centered bibliometrics through CAMEOs: Characterizations automatically made and edited online. *Scientometrics*, *51*(3), 607–637.