# A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar

Massimo Franceschet

Department of Mathematics and Computer Science – University of Udine
Via delle Scienze, 206 – 33100 Udine (Italy)
massimo.franceschet@dimi.uniud.it

**Abstract.** Given the current availability of different bibliometric indicators and of production and citation data sources, the following two questions immediately arise: do the indicators' scores differ when computed on different data sources? More importantly, do the indicator-based rankings significantly change when computed on different data sources? We provide a case study for computer science scholars and journals evaluated on Web of Science and Google Scholar databases. The study concludes that Google Scholar computes significantly higher indicators' scores than Web of Science. Nevertheless, citation-based rankings of both scholars and journals do not significantly change when compiled on the two data sources, while rankings based on the h index show a moderate degree of variation[1].

## 1 Introduction

Bibliometrics has become a standard tool of science policy and research management in the last decades. In particular, academic institutions increasingly rely on citation analysis for making hiring, promotion, tenure, and funding decisions (Weingart, 2005).

Citation analysts retrieve production and citation data from bibliographic and citation sources and compute performance indicators to measure the quality of research of the bibliometric unit under evaluation. The databases of the Institute for Scientific Information (ISI) have been the most generally accepted data sources for bibliometric analysis. The ISI was founded by Eugene Garfield in 1960 and acquired by Thomson (today Thomson-Reuters) in 1992, one of the world's largest information companies. ISI maintains Web of Knowledge, an online academic database which provides access to many resources, in particular to Web of Science (WoS), that includes journal publications and citations covering Sciences, Social Sciences and Arts and Humanities. A major alternative to Web of Science that is growing in popularity is Google Scholar, a freely accessible service provided by Google Inc. While Web of Science database contains mainly

[1] Draft version of the original paper published in Scientometrics 83(1), 243-258, 2010. DOI: http://dx.doi.org/10.1007/s11192-009-0021-2

journal publications, Google Scholar finds different types of sources, including journal papers, conference papers, books, theses and reports (Meho and Yang, 2007).

Common evaluation criteria that characterize quality of research at scholar level are productivity and impact. Traditional bibliometric indicators, like number of published papers and number of received citations, aim to separately capture these criteria; recently proposed indexes, e.g., the h index (Hirsch, 2005), try to measure, in a single note, more aspects of quality. The h index of a scholar is the highest number h of papers published by the scholar that have each received at least h citations. The index is meant to capture both productivity and impact in such a way that it is hard to increase it, as well as to rig it, over a certain threshold. Bornmann and Daniel describe opportunities and limitations of the h index (Bornmann and Daniel, 2007). The following are acknowledged limitations of the index and corresponding proposals to correct the mentioned flaws:

- it disadvantages small but highly-cited paper sets too strongly. Egghe (2006) proposes the g-index to account for this problem. Given a set of articles ranked in decreasing order of the number of citations that they received, the g-index is the largest number such that the top g articles received together at least $g^2$ citations;
- it puts newcomers at a disadvantage since both publication output and citation rates will be relatively low. To solve this problem Hirsch (2005) proposes the m quotient which is computed by dividing the h index by the scientific age of the author;
- it does not account for the number of authors in a paper. To address this problem Batista et al. (2006) suggest to adjust the original h index by dividing it by the mean number of researchers in the h publications that determine the h index. The new index is named individual h index;
- it allows scientists to rest on their laurels since the index never decreases and it might increase even if no new papers are published. In order to address this issue Katsaros et al. (2006) propose the contemporary h index. The contemporary h index adds an age-related weighting to each cited article, giving less weight to older articles.

There are also different proposals to assess journal performance. The traditional one is the impact factor (roughly, the number of recent citations per paper) (Garfield, 1979). Additional ones are the journal h index (Braun et al., 2006) and prestige-oriented metrics (Bollen et al., 2006).

Computer science is an original discipline combining science and engineering. Research in computer science includes two main flavors: Theory, developing conceptual frameworks for understanding computations, algorithms, data structures and other aspects of computing; Systems, building software artifacts and assessing their properties. The hybrid nature of computer science is part of its attraction but also complicates the evaluation (Choppy et al., 2008; Computing Research Association, 1999). A distinctive feature of computer science publication is the importance of selective conferences. Journals have their role, but do not necessarily carry more prestige. Moreover, publications are not the only scientific contributions. Artifacts such as software can be as important as publications. A notable example is Google search engine, that originates from the academic world (Brin and Page, 1998).

The present contribution addresses the following two questions:

1. do the values of bibliometric indicators for computer science scholars and journals differ when computed on Web of Science and Google Scholar?
2. do the indicator-based rankings of computer science scholars and journals significantly change when computed on Web of Science and Google Scholar?

The first question might appear rhetorical to the reader. Since Google Scholar finds more types of publications than Web of Science both at the cited (or target) and at the citing (or source) level, one expects that the values of the indicators on Google Scholar are higher than those on Web of Science. However, the proportion of the increment for the different indicators is to be investigated. The second question is significant in the cases, like academic competitions for scholar hiring and promotion or journal selection, where the ranking and not the actual score is important. If indicators' scores are different but indicator-based scholar rankings are similar, then either data source might be used for bibliometric analysis.

Our study considers 13 bibliometric indicators to measure scholar performance, namely: papers, cited papers, papers per year, papers per single author, citations, citations per year, citations per single author, citations per paper, as well as h index and four variants: g index, m quotient, contemporary h index, and individual h index. We compute the mentioned indicators both on Web of Science and Google Scholar data sources for the publications of a group of well respected Italian computer science researchers working in different sub-fields. We compare both the indicators' scores and the indicator-based rankings computed on the two data sources under investigation. Moreover, we contrast the Web of Science and Google Scholar citation-based and h index-based rankings for two computer science journal samples: (i) the top-20 computer science journals according to current Thomson Scientific impact factor, and (ii) the top-20 computer science journals in subject category *theory and methods* according to the total number of received citations as recorded in Web of Science database.

The rest of the paper is as follows. In Section 2 we revise related literature. In Section 3 we describe the method and the tools we have used to collect, store, and analyse the data of our study. Section 4 describes the outcomes of our study. Finally, Section 5 compares our findings with previous ones and discusses the implications of our results.

## 2 Related literature

There are many studies that compare citation data retrieved on different data sources. Table 1 displays those studies we have found in the literature. The first column shows the publication reference of the study, the second column contains the set of data sources that are compared, while the research field of

| **reference** | **compared sources** | **field** |
|---|---|---|
| (Goodrum et al., 2001) | CiteSeer and Web of Science | computer science |
| (Zhao and Logan, 2002) | CiteSeer and Web of Science | computer science |
| (Whitley, 2002) | Chemical Abstracts and Web of Science | chemistry |
| (Bauer and Bakkalbasi, 2005) | Web of Science, Scopus, and Google Scholar | JASIST papers |
| (Jacsò, 2005) | Web of Science, Scopus, and Google Scholar | Current Science papers and Eugene Garfield papers |
| (Noruzi, 2005) | Web of Science and Google Scholar | webometrics |
| (Pauly and Stergiou, 2005) | Web of Science and Google Scholar | different disciplines |
| (Bakkalbasi et al., 2006) | Web of Science, Scopus, and Google Scholar | oncology and condensed matter physics |
| (Kousha and Thelwall, 2007) | Web of Science and Google Scholar | different disciplines |
| (Saad, 2006) | Web of Science and Google Scholar | business sciences |
| (Norris and Oppenheim, 2007) | CSA Illumina, Web of Science, Scopus, and Google Scholar | social sciences |
| (Bar-Ilan et al., 2007) | Web of Science, Scopus, and Google Scholar | different disciplines |
| (Meho and Yang, 2007) | Web of Science, Scopus, and Google Scholar | library and information science |
| (Kousha and Thelwall, 2008) | Web of Science and Google Scholar | different disciplines |
| (Shaw and Vaughan, 2008) | Web of Science and Google Scholar | information science |
| (Meho and Rogers, 2008) | Web of Science, Scopus and Google Scholar | human-computer interaction |
| (Bar-Ilan, 2008) | Web of Science, Scopus, and Google Scholar | different disciplines |
| (Sanderson, 2008) | Web of Science, Scopus and Google Scholar | library and information science and information retrieval |
| (Bornmann et al., 2009) | Web of Science, Scopus, Google Scholar, and Chemical Abstracts | chemistry |

**Table 1.** Literature comparing citation data over different data sources

the publications considered in the study is given in the third column. The papers are sorted in chronological order.

However, only a few of these studies compare the ranking of the bibliometric units under investigation according to bibliometric indicators different from citation count. In particular, all these exceptions focus on the h index:

- Saad compares the h-index at the author level using bibliometric data of productive consumer scholars retrieved from Web of Science and Google Scholar (Saad, 2006). The author finds that the h index computed on Web of Science and that computed on Google Scholar correlate at 0.82. The median h index is 11 on Google Scholar and it is 9 on Web of Science; the h index on Google Scholar is higher than that on Web of Science in 39 cases over 55;
- Bar-Ilan compares the h index for highly cited Israeli researchers and three recent Israeli Nobel price winners computed on Web of Science, Scopus and Google Scholar (Bar-Ilan, 2008). Except for few cases the differences in the h values between Scopus and Web of Science are not significant. The differences between Google Scholar and the other databases are much more considerable. In particular, the author notices that Google Scholar computes higher h values for almost all mathematicians and computer scientists in the group;
- Meho and Rogers analyse the h index of researchers in the field of human-computer interaction computed on Scopus, Web of Science, and Google Scholar (Meho and Rogers, 2008). The authors find that Google Scholar computes higher h values (20.6 on average) than Scopus (12.3 on average) and Web of Science (8.0 on average). Despite this, there is a significant correlation (Spearman 0.96) between the h index ranking on Google Scholar and that of the union of the other two data sources;
- Sanderson computes the h index for UK library and information science and information retrieval academics on Scopus, Web of Science and Google Scholar (Sanderson, 2008). The author finds that scholars who publish in more computer science related forums have a significantly higher Google Scholar h index than their Scopus and Web of Science h scores. For academics with computer science research focus, the average Google Scholar h index is 15.2, while the average h score on Scopus is 7.9 and on Web of Science (cited reference search) is 6.8. These differences have an impact on the rankings of the scholars. Indeed, the (Kendall) correlation of the three variables is moderate: Google Scholar h and Web of Science h correlate at 0.51, Google Scholar h and Scopus h correlate at 0.69, and Scopus h and Web of Science h correlate at 0.64.

## 3 Method and tools

In this section we describe the method and the tools we have used to collect, store, and analyse the data of our case study.

### 3.1 Data sources, data samples and bibliometric indicators

We collected bibliographic and citation data from two well-known data sources: Google Scholar[2] and Thomson Scientific Web of Science[3]. While Web of Science database contains mainly journal publications, Google Scholar finds different types of sources, including journal papers, conference papers, books, theses and reports (Meho and Yang, 2007).

The scholar sample consists of the publications of a group of Italian well respected computer science scholars working at the Department of Mathematics and Computer Science of the University of Udine. We opted for this sample mainly to address the problem of homonymy for scholars. The sampled computer science researchers work in the department of the writing author. This gave him the possibility to carefully check the association of bibliographic items with authors either by using his domain knowledge or by directly consulting the local scholars. The sampled scholars produced 324 journal papers that received in total 1378 citations from other journal papers (data from Web of Science). Moreover, the authors in the sample produced 1776 publications (including journal papers, conference papers, books, reports and theses) that received an overall amount of 10690 citations from other publications (data from Google Scholar). Scholars in the sample published in different topics of computer science, including human-computer interaction, programming languages, image processing and computer vision systems, information and database systems, formal methods and applications, artificial intelligence and expert systems, algorithms, graphics modelling and virtual reality, web and multimedia systems.

We computed the following bibliometric indicators to measure scholar performance:

1. *papers* (abbreviated as pap). The number of papers published by the author.
2. *cited papers* (cp). The number of papers with at least one citation.
3. *papers per year* (ppy). The number of papers divided by the academic age. The academic age of a scholar is the age of the eldest paper published by the scholar.
4. *papers per author* (ppa). The number of papers per single author. This is computed by dividing each paper unit by the number of authors of that paper and summing the results over all papers.
5. *citations* (cit). The number of citations received by papers of the author.
6. *citations per year* (cpy). The number of citations divided by the academic age.
7. *citations per author* (cpa). The number of citations per single author. This is computed by dividing each citation count of a paper by the number of authors of that paper and summing the results over all papers.
8. *citations per paper* (cpp). The number of citations divided by the number of papers.

[2] `http://scholar.google.com`

[3] `http://scientific.thomson.com/products/wos/`

9. *h index* (h). The highest number h of papers that have each received at least h citations (Hirsch, 2005).
10. *g index* (g). The highest number g of papers that received together at least $g^2$ citations (Egghe, 2006).
11. *m quotient* (m). The h index divided by the academic age (Hirsch, 2005).
12. *contemporary h index* (hc). An age-weighted h index obtained by giving more weight to recent papers (Katsaros et al., 2006). In particular, citations to papers published k years ago are weighted 4/(k+1). The h index is then computed as usual on the weighted citation counts.
13. *individual h index* (hi). The h index divided by the mean number of authors in the set of papers contributing to the h index (Batista et al., 2006).

These bibliometric indicators can be clustered in three main categories:

1. *paper-based metrics*, including papers, cited papers, papers per year and papers per author. These metrics focus on productivity of a scholar;
2. *citation-based metrics*, including citations, citations per year and citations per author. These metrics concentrate on impact of a scholar;
3. *hybrid metrics*, including citations per paper and h-type indicators. These indicators aim to capture both productivity and impact in a single figure.

Moreover, papers per year, citations per year, m and contemporary h index are *age-centric metrics*: they favour scholars with a high research performance compared to their academic age, for instance brilliant young researchers. On the other hand, papers per author, citations per author and individual h index are *author-centric* indicators: they prefer researchers that publish alone or with few co-authors.

In addition, we used the following two journal samples:

– the top-20 computer science journals according to the current Thomson Scientific impact factor. For these journals, we analyzed the published articles since ever and the citations they have received until December 2008. We refer to this sample as the journal sample A;
– the top-20 computer science journals in subject category *theory and methods* according to the total number of received citations recorded in Web of Science. For these journals, we analyzed the published articles during years 2005 and 2006 and the citations they have received until December 2008. We refer to this sample as the journal sample B.

For each journal in the samples, we computed the total number of citations received by its articles and the h index (Braun et al., 2006).

### 3.2 Data collection

Data collection was complicated by the well known *name problem*: scholars are usually recorded using initials and surname, e.g., "M. Franceschet", but sometimes the full name is used, e.g., "Massimo Franceschet". Moreover, journals are stored either using the full journal name, like "ACM Transactions on Graphics" or using some abbreviation, like "ACM T Graphic". Using the abbreviated name for a target increases the probability of homonymy, but using the full name may cut off those bibliographic items that contain only the abbreviated form of the name. To address the name problem, we decided to sample scholars from the department of the writing author, whose research publications are known to the writing author. Moreover, in order to retrieve journal papers from Google Scholar, we wrote queries containing both the full journal name and its ISO abbreviation. For Web of Science we wrote the full journal name as recorded in Thomson Scientific Journal Citation Report.

Next, we computed the bibliometric indicators on the extracted bibliographic data. For Google Scholar, we took advantage of Publish or Perish[4], a free software that queries Google Scholar and computes several citation statistics for scholars and journals. The statistics computed by Thomson Scientific for Web of Science database do not cover all indicators we are using in this paper. Moreover, to the best of our knowledge, there is no automatic tool similar to Publish or Perish working on data contained in Web of Science. For these reasons, we implemented a software tool that computes all statistics we need on the basis of the data given by Web of Science. The outcome of this step is a *bibliometric matrix* for each data source under consideration. Such a matrix contains, for each bibliometric unit, a row with the values of the different bibliometric indicators for that unit. We stored the bibliometric matrices in XML format. This allowed to query the data using XQuery and transform them into an HTML web page with XSLT (Harold and Means, 2004).

### 3.3 Analysis of indicators' scores and correlations

We performed two types of investigation. The first is an analysis of the values of the bibliometric indicators computed on the two data sources. We computed, for each indicator $I$ included in the bibliometric matrix for data source $X$, the average $\mu(I, X)$ of the scores of $I$ over all sampled bibliometric units (scholars or journals). Moreover, we computed the ratio $\rho(I) = \mu(I, GS)/\mu(I, WOS)$. We expect that the values for the index $\rho(I)$ are greater than one for all indicators. This because Google Scholar uses a broader universe of cited and citing publications with respect to Web of Science.

A second investigation aims to analyse how the indicator-based rankings of bibliometric units computed on Google Scholar and on Web of Science differ. Indeed, a difference in the values of an indicator over two data sources does not necessarily imply a difference in the rankings of the units according to that indicator over these data sources. We investigated the statistical correlation between the same indicator computed on the two considered data sources using two standard non-parametric correlation methods: Spearman rank correlation coefficient and Kendall rank correlation coefficient (Moore, 2006). For each correlation

[4] Available at `http://www.harzing.com/pop.htm`

| | pap | cp | ppy | ppa | cit | cpy | cpa | cpp | h | g | m | hc | hi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **GS** | 136.62 | 94.15 | 4.97 | 56.79 | 822.31 | 39.92 | 347.86 | 5.84 | 14.31 | 23.31 | 0.76 | 10.31 | 5.29 |
| **WoS** | 24.92 | 16.00 | 1.66 | 10.47 | 106.00 | 6.77 | 47.31 | 4.48 | 5.00 | 8.38 | 0.34 | 3.54 | 1.81 |
| **ratio** | 5.48 | 5.88 | 2.99 | 5.42 | 7.76 | 5.89 | 7.35 | 1.31 | 2.86 | 2.78 | 2.27 | 2.91 | 2.93 |

**Table 2.** Mean scores for bibliometric indicators and ratio between them for computer science scholars

method, we computed a *correlation vector* whose elements are the method correlation coefficients of the indicators. The correlation vector mean is an overall measure of the degree of association of the two data sources according to the used correlation method. Finally, we analysed how much the correlation methods of Spearman and Kendall are associated by computing the Pearson correlation of the two method correlation vectors. For all statistical computations we took advantage of the free software R (R Development Core Team, 2009).

## 4 A case study for computer science scholars

In this section we apply the method proposed in Section 3 to computer science literature. Section 4.1 compares bibliometric indicators for computer science scholars, while Section 4.2 analyses citations of computer science journals.

### 4.1 Computer science scholars

Table 2 contains, for each indicator, the mean score computed on Google Scholar, the mean score computed on Web of Science, and the ratio between the former and the latter. Figure 1 depicts the boxplots for the three main indicators: papers, citations and h index. Notice the presence of one outlier in the number of papers of Web of Science and in the number of citations of Google Scholar (it refers to the same author).

As expected, the values of the indicators computed on Google Scholar are higher compared to those computed on Web of Science. The reason is that Google Scholars finds a broader set of bibliographic items of different types (including conference papers, which are particulary important in the field of computer science), while Web of Science is limited to papers of journals with relatively high impact factor. This holds both for cited (or target) papers and for citing (or source) papers. In particular, we observe that:

– Google Scholar finds more than 5 times the papers of Web of Science. This is shown by all paper-based metrics except papers per year, whose ratio is lower (3.99). The latter lower ratio is justified as follows. Recall that paper per year is defined as the ratio between papers and academic age of the scholar. We noticed that Google Scholar finds older papers with respect to Web of Science: the average academic age for scholars found by Google Scholar is 19.62 years, while it is 15 years in case of Web of Science;

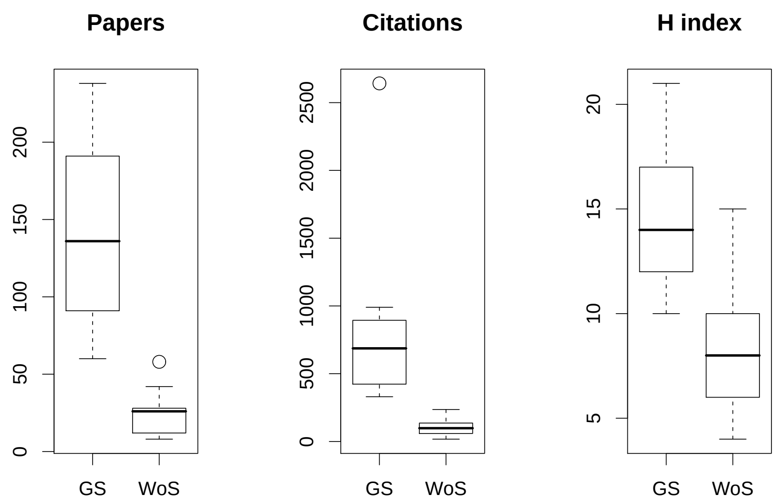

**Fig. 1.** Box plots for papers, citations, and h index

- Google Scholar finds almost 8 times the citations of Web of Science. This is shown by all citation-based metrics except citation per year, whose ratio is lower (5.89) because the academic age of scholars computed on Google Scholar is higher than that computed on Web of Science;
- the ratio for citations (7.76) is higher than the ratio for papers (5.48). This is confirmed by the ratio of citations per paper, which is bigger than one (1.31). This because both the citing and the cited universes are involved in the count of the total number of citations accrued to an author, whereas only the cited universe is relevant for the count of the total number of papers published by an author.
- h-type indexes have similar ratios that are less but close to 3. Notice that m quotient has a slightly lower ratio because of the higher academic age found by Google Scholar.

In the following our aim is to investigate if the above observed differences in the values of the computed indicators have an impact on the relative rankings of scholars. Table 3 summarizes the results for Spearman and Kendall correlation tests we have performed; the indicators are sorted in decreasing order with respect to Spearman correlation coefficient. The alternative hypothesis of the tests states that there is a significant positive correlation between the two data sources. The table gives the correlation coefficient and the corresponding p-value. The linear models (the scatterplots with linear interpolation) for papers, citations and h index are given in Figure 2.

The Spearman correlation tests for citations, citations per year/author, papers, cited papers, h g, hi are significant at the 0.01 level. For these indicators, we may reject the null hypothesis that the two samples are uncorrelated. The Kendall test is more restrictive: the indicators with a significant correlation at

|  | Spearman |  | Kendall |  |
|---|---|---|---|---|
| **index** | **cor** | **p-value** | **cor** | **p-value** |
| citations per year | 0.95 | 0 | 0.82 | 6e-06 |
| citations | 0.92 | 0 | 0.82 | 6e-06 |
| g index | 0.89 | 2e-05 | 0.81 | 0.0001 |
| citations per author | 0.87 | 4e-05 | 0.69 | 0.0003 |
| individual h index | 0.84 | 0.0002 | 0.70 | 0.0005 |
| papers | 0.69 | 0.005 | 0.49 | 0.01 |
| cited papers | 0.67 | 0.006 | 0.48 | 0.01 |
| h index | 0.65 | 0.009 | 0.52 | 0.01 |
| papers per year | 0.63 | 0.01 | 0.38 | 0.04 |
| citations per paper | 0.62 | 0.01 | 0.49 | 0.01 |
| papers per author | 0.53 | 0.03 | 0.38 | 0.04 |
| m quotient | 0.50 | 0.04 | 0.31 | 0.07 |
| contemporary h index | 0.38 | 0.1 | 0.31 | 0.1 |
| **mean** | **0.70** |  | **0.56** |  |

**Table 3.** Spearman and Kendall correlation tests for computer science scholars

the 0.01 level are citations, citations per year/author, hi and g only. We partitioned the indicators in the following three groups on the basis of the Spearman correlation coefficient:

1. citations, citations per year, citations per author, g, and individual h have a fairly good correlation ($\geq 0.84$);
2. papers, cited papers, papers per year, citations per paper, and h have a moderate correlation (between 0.62 and 0.69);
3. m, contemporary h, papers per author have a weak correlation ($\leq 0.53$).

Summing up, citation-based metrics are generally well correlated, paper-based metrics have a moderate correlation, and h-type indexes do not show a common behaviour distributing over the three correlation groups we have identified. In particular, the h index shows a modest correlation. The average Spearman correlation is 0.70 and the average Kendall correlation is 0.56. Moreover, the correlation between Spearman and Kendall outcomes is 0.97, showing that the two correlation methods mostly agree.

The reader might wonder why the citation correlation between Google Scholar and Web of Science is so relevant while the raw citation counts computed on the same data sources differ so strongly. It is interesting to investigate the reasons of this apparent discrepancy. To this end, for each author, we analysed the distribution of citations to papers retrieved with Google Scholar and classified the top-cited items retrieved by Google Scholar in journal papers, conference papers and books. The results are summarized in Table 4.

The 2nd and 3rd columns in the table are the h index (h) and the number of papers (p). Moreover, we computed the ratio $r_1 = h/p$; this is the fraction of author's papers that belong to the author's Hirsch core (the set of papers

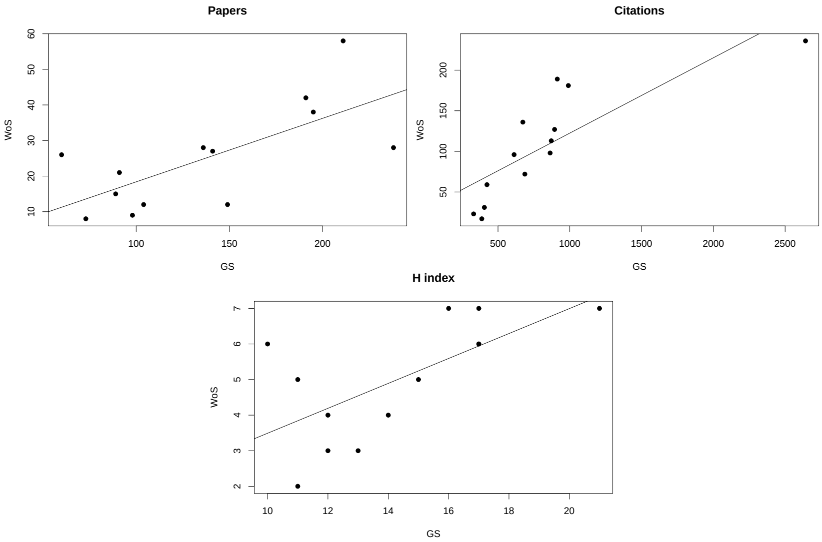


**Fig. 2.** Linear models papers, citations, and h index

determining the author's h index). Notice that, despite the h index and the number of papers vary between authors, the ratio $r_1$ is quite stable (with one outlier), with a mean of 0.10. This means that, on average, the size of the Hirsch core is one-tenth of the size of the set of all papers. The 5th and 6th table columns are the number of citations to papers in the Hirsch core (hc) and the number of citations to all papers (c). The ratio $r_2 = hc/c$ is the fraction of citations to papers in the Hirsch core. A comparison of ratios $r_1 = h/p$ and $r_2 = hc/c$ gives an indication of the skewness (asymmetry) of the distribution of citations. If citations were equally distributed among the $n$ papers, each paper having $k$ citations, then $r_1 = k/n$ and $r_2 = k^2/(n \cdot k) = k/n = r_1$. If the distribution is right skewed (positively asymmetric), then there are few highly cited papers and many poorly cited ones, and $r_1 < r_2$. As the reader can check from Table 4, the citation distribution for each author is right skewed: on average, the papers in the Hirsch core, that represent 10% of the papers, collect 63% of citations. Moreover, the number of citations to papers in the Hirsch core is significantly correlated with the number of citations to all papers (Pearson 0.97, p = 3e-08). Therefore, we have that citations to papers in the Hirsch core are a representative sample of all citations. The remaining table columns are the percentages of journal papers (%J), conference papers (%C), and books (%B) in the Hirsch core. Notice that the fraction of journal papers is significant for each author and, on average, one item over two in the Hirsch core is a journal paper. All in all, we have an explanation for the above mentioned discrepancy: the top-cited papers found by Google Scholar well represent the citations to the set of all papers and among them journal papers, which are indexed by Web

| author | h | p | h/p | hc | c | hc/c | %J | %C | %B |
|---|---|---|---|---|---|---|---|---|---|
| LC | 20 | 228 | 0.09 | 670 | 1318 | 0.51 | 60% | 40% | 0% |
| AD | 14 | 170 | 0.08 | 461 | 786 | 0.59 | 50% | 50% | 0% |
| GF | 16 | 241 | 0.07 | 577 | 1140 | 0.51 | 62% | 25% | 13% |
| MF | 13 | 71 | 0.18 | 267 | 379 | 0.70 | 31% | 62% | 7% |
| FH | 23 | 290 | 0.08 | 2227 | 2970 | 0.75 | 61% | 30% | 9% |
| ML | 12 | 117 | 0.10 | 233 | 417 | 0.56 | 42% | 58% | 0% |
| MM | 12 | 116 | 0.10 | 322 | 482 | 0.67 | 50% | 50% | 0% |
| SM | 10 | 106 | 0.09 | 651 | 817 | 0.80 | 70% | 30% | 0% |
| AM | 17 | 133 | 0.13 | 539 | 962 | 0.56 | 47% | 29% | 24% |
| CP | 15 | 173 | 0.09 | 383 | 670 | 0.57 | 33% | 67% | 0% |
| AP | 18 | 221 | 0.08 | 572 | 1103 | 0.52 | 56% | 33% | 11% |
| VR | 11 | 99 | 0.11 | 574 | 702 | 0.82 | 54% | 46% | 0% |
| CT | 17 | 162 | 0.10 | 680 | 1022 | 0.67 | 35% | 65% | 0% |
| **mean** | **15** | **164** | **0.10** | **627** | **982** | **0.63** | **50%** | **45%** | **5%** |

**Table 4.** For each author in the scholar sample, we show: **h** = Hirsch index; **p** = number of papers; **h/p**: the fraction of papers in the Hirsch core; **hc** = citations to papers in the Hirsch core; **c** = citations to all papers; **hc/c**: the fraction of citations to papers in the Hirsch core; **%J** = percentage of journal papers in the Hirsch core; **%C** = percentage of conference papers in the Hirsch core; **%B** = percentage of books in the Hirsch core. Source: Google Scholar

of Science, represent a substantial fraction. Hence, despite the difference in the actual citation counts, the highly cited scholars on Google Scholar are also highly cited on Web of Science, and viceversa.

Finally, recall that the h index is also the size of the Durfee square contained in the Ferrers diagram of the citation distribution (Anderson et al., 2008) (see Figure 3). The citations contained in the Ferrers diagram that are outside the Durfee square belong either to the upper triangle (the citations to papers in the Hirsch core that are outside the Durfee square, whose number is denoted by $uc$) or to the lower triangle (the citations to papers outside the Hirsch core, whose number is denoted by $lc$) of the Ferrers diagram. Of course, we have that $c = h^2 + uc + lc = hc + lc$. Hence $uc = hc - h^2 \geq 0$ and $lc = c - hc \geq 0$. The quotient $l = (uc + lc)/c = (1 - h^2/c)$ is the fraction of citations that fall outside the Durfee square. It is a relative value between 0 and 1 that quantifies the information about the citation distribution that is lost when using the h index instead of taking into account the whole citation distribution. The loss of information is low when $l$ is close to 0, and it is high when $l$ is close to 1. An absolute measure for the same purpose is defined by Hirsch (2005) as the quotient $a = c/h^2 \geq 1$. Going back to Table 4, we have that, on average on all scholars, $h^2 = 225$, $uc = 402$, $lc = 355$, $c = h^2 + uc + lc = 982$, $l = 0.77$, and $a = 4.36$. Notice that 77% of the citations are outside the Durfee square and thus are not taken into account when computing the selective h index. Moreover, the information loss in the upper part of the Ferrers diagram (highly cited papers)

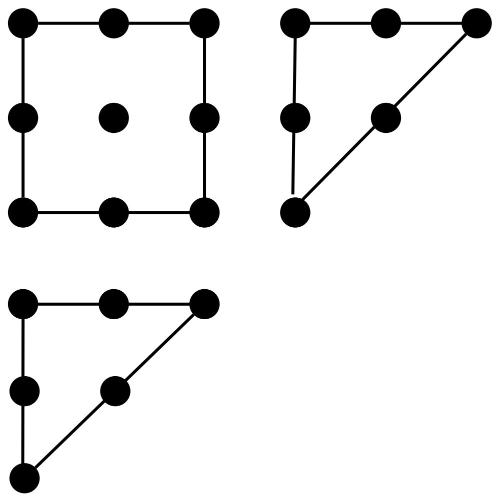

**Fig. 3.** The Ferrers diagram is obtained by ranking the citation counts of publications in decreasing order and plotting each citation count $c$ as a row with $c$ solid circles. In the example shown in the figure we have 6 publications with 6, 5, 4, 3, 2, and 1 citations, respectively. The Durfee square is the largest-sized square contained within the Ferrers diagram. It contains 9 citations. The size of the Durfee square is the h index (3). The figure also shows the upper and lower triangles of the diagram, each containing 6 citations

is bigger than the information loss in the lower part (lowly cited papers) and the area of the Durfee square is smaller than both the area of the upper and lower Ferrers triangles. Finally, the value $a = 4.36$ is coherent with the empirical findings of Hirsch, who found a value for this quotient between 3 and 5 (Hirsch, 2005).

### 4.2 Computer science journals

In this section we make a citation analysis for two relevant samples of computer science journals. Table 5 contains the mean scores for citations and h index computed on journals of both samples. We observe the following:

- Google Scholar scores are higher than Web of Science ones, as noticed for scholars;
- the ratio between citations found by Google Scholar and citations found by Web of Science is, for both journal samples, lower than the same ratio for scholars, that was 7.76. The same is noticed for the h index scores. This happens because, in the case of scholars, Google Scholar finds more citing (source) publications as well as more cited (target) papers. However, in the case of journals, the set of target papers is fixed (the papers published by

| Sample A | cites | h index | Sample B | cites | h index |
|---|---|---|---|---|---|
| **GS** | 56283.25 | 104.35 | **GS** | 1970.30 | 19.35 |
| **WoS** | 17087.50 | 48.15 | **WoS** | 399.35 | 7.60 |
| **ratio** | 3.29 | 2.18 | **ratio** | 4.93 | 2.55 |

**Table 5.** Mean indicator scores and ratio between them for computer science journals

| Sample A | Spearman | | Kendall | | Sample B | Spearman | | Kendall | |
|---|---|---|---|---|---|---|---|---|---|
| **index** | **cor** | **p-value** | **cor** | **p-value** | **index** | **cor** | **p-value** | **cor** | **p-value** |
| cites | 0.84 | 0 | 0.64 | 1e-05 | cites | 0.70 | 0.0003 | 0.51 | 0.0007 |
| h index | 0.78 | 2e-05 | 0.61 | 9e-05 | h index | 0.61 | 0.002 | 0.49 | 0.002 |

**Table 6.** Spearman and Kendall correlation tests for computer science journals

the journals under consideration), and the possible differences are only in the set of source publications;

- the ratios for both citations and h for sample B are higher than the same ratios for sample A. This can be explained with the fact the sample A contains some journals with significant overlap with other disciplines, like Bioinformatics, Cognitive Brain Research, and Journal of the American Medial Informatics Association. For these journals, we noticed that the number citations found by Web of Science is relatively higher. For instance, Web of Science finds 19769 citations to papers of Cognitive Brain Research, which is close the number of citations that Google Scholar retrieves, which is 23109.

Table 6 summarizes the results for Spearman and Kendall correlation tests for citations and h index over both journal samples. The corresponding linear models are shown in Figures 4 and 5. Notice that:

- in all cases, the correlation is significant at level 0.01;
- the correlation for citations in both journal samples (Spearman 0.84 and 0.70) is lower than the observed correlation for citations in the scholar sample (Spearman 0.92);
- the correlation for h index in sample A (Spearman 0.78) is higher than the same correlation for scholars (Spearman 0.65), which in turns is close to the correlation for h index in sample B (Spearman 0.61);
- Web of Science and Google Scholar are better correlated on sample A than on sample B. This can be explained again with the fact that sample B contains pure computer science journals, while sample A admits computer science journals with strong relationships to other disciplines, like Biology, Neuroscience, and Medicine. As observed above, for these inter-disciplinary journals, the gap between the citations found by Web of Science and Google Scholar is less important and hence the statistical correlation is higher.

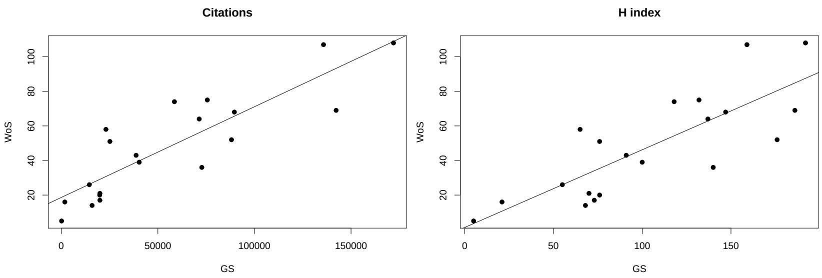


**Fig. 4.** Linear models for journal sample A

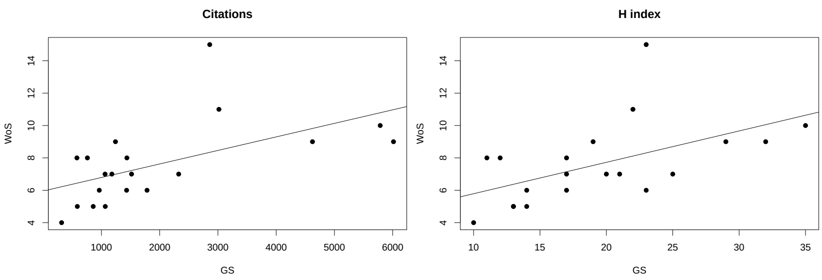


**Fig. 5.** Linear models for journal sample B

## 5 Conclusion

We made a comparison of bibliometric indicators' scores and indicator-based computer science scholar and journal rankings computed on Web of Science and Google Scholar. Our main conclusions are:

– Google scholar computes significantly higher indicators' scores than Web of Science. The increment for scholars varies with the indicator and it is roughly 5 times for paper-based indicators, 8 times for citation-based indicators and 3 times for h-type indicators. The increment for journal indicators is lower but still significant. The noticed increase for the h index confirms the findings of Meho and Rogers on a sample of human-computer interaction researchers (Meho and Rogers, 2008), those of Sanderson on a sample of computer science academics (Sanderson, 2008), and those of Bar-Ilan on mathematicians and computer scientists (Bar-Ilan, 2008);
– there is a significant correlation of citation-based rankings between the two sources. However, paper-based and h-based rankings show a weaker degree of association. Our result for the h index – Spearman correlation at 0.65 – contrasts with the outcome of Meho and Rogers who found a significant Spearman correlation at 0.96 (Meho and Rogers, 2008). On the other hand, our result for the h index – Kendall correlation at 0.52 – almost perfectly

matches the findings of Sanderson who found a Kendall correlation at 0.51 on a sample of computer science academics (Sanderson, 2008).

Our general advice is to use Google Scholar when the user is interested in finding papers and corresponding citations of *computer science* scholars and journals. The main advantages of using Google Scholar are freedom and a broader universe of cited and citing items. In particular, Google Scholar finds conference papers, which are a fast vehicle to transfer ideas in the quickly evolving field of computer science (Choppy et al., 2008). The main drawback of Google Scholar is that the consistency and accuracy of data is admittedly lower compared to that of Web of Science and other commercial citation-enhanced databases (Jacsò, 2005). Therefore, the time needed to obtain meaningful data might be significantly higher than the time spent to get the data with fee-based data sources. For instance, in their large-scale study about citations of library and information science scholars, Meho and Yang experienced that collecting meaningful data from Google Scholar took 30 as much time as collecting usable data from Web of Science (Meho and Yang, 2007).

A somewhat surprising finding of the present study is that the citation-based rankings of scholars and journals in the field of computer science are similar when compiled on Google Scholar and Web of Science. We observed that, for the sampled scholars, the statistical association between the two data sources holds because of the substantial percentage of highly cited scholar publications that are contained in journals (on average 50%). Hence, when the user is interested in comparing different scholars or journals on the basis of the citations they received, and when scholars to be compared publish in journals with significant frequencies, either Web of Science or Google Scholar can be used (the former might be preferred because it allows a quicker and more flexible analysis). On the other hand, the correlation between the two data sources based on the popular h index is weaker than the correlation observed for citations. In this case, great care must be taken when selecting the data source for the analysis. Our advice here is to perform a (time-consuming) join of the publications and citations contained in the two databases and use the combined universe to compute the h index for scholars and journals.

# Bibliography

[Anderson et al., 2008]Anderson, T. R., Hankin, R. K. S., and Killworth, P. D. (2008). Beyond the Durfee square: enhancing the h-index to score total publication output. *Scientometrics*, 76(3):577–588.

[Bakkalbasi et al., 2006]Bakkalbasi, N., Bauer, K., Glover, J., and Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 7. Retrieved December 20, 2008 from `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1533854`.

[Bar-Ilan, 2008]Bar-Ilan, J. (2008). Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2):257–271.

[Bar-Ilan et al., 2007]Bar-Ilan, J., Levene, M., and Lin, A. (2007). Some measures for comparing citation databases. *Journal of Informetrics*, 1(1):26–34.

[Batista et al., 2006]Batista, P. D., Campiteli, M. G., and Konouchi, O. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1):179–189.

[Bauer and Bakkalbasi, 2005]Bauer, K. and Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*, 11(9). Retrieved December 20, 2008, from `http://www.dlib.org/dlib/september05/bauer/09bauer.html`.

[Bollen et al., 2006]Bollen, J., Rodriguez, M. A., and de Sompel, H. V. (2006). Journal status. *Scientometrics*, 69(3):669–687.

[Bornmann and Daniel, 2007]Bornmann, L. and Daniel, H.-D. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology*, 58(9):1381–1385.

[Bornmann et al., 2009]Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A., and Daniel, H.-D. (2009). Convergent validity of bibliometric Google Scholar data in the field of chemistry citation counts for papers that were accepted by Angewandte Chemie International Edition or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. *Journal of Informetrics*, 3(1):27–35.

[Braun et al., 2006]Braun, T., Glänzel, W., and Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, 69(1):169–173.

[Brin and Page, 1998]Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117.

[Choppy et al., 2008]Choppy, C., van Leeuwen, J., Meyer, B., and Staunstrup, J. (2008). Research evaluation for computer science. In *European Computer Science Summit*. Retrieved December 20, 2008, from http://www.informatics-europe.org/ECSS08/papers/Research_evaluation_CACM.pdf.

[Computing Research Association, 1999]Computing Research Association (1999). Best practices memo – Evaluating computer scientists and engineers for promotion and tenure. *Computing Research News*. Retrieved December 20, 2008, from http://www.cra.org/reports/tenure_review.html.

[Egghe, 2006]Egghe, L. (2006). Theory and practice of the g-index. *Scientometrics*, 69(1):131–152.

[Franceschet, 2009]Franceschet, M. (2009). A comparison of bibliometric indicators for computer science scholars and journals on web of science and google scholar. *Scientometrics*. In press. Available at http://dx.doi.org/10.1007/s11192-009-0021-2.

[Garfield, 1979]Garfield, E. (1979). *Citation indexing: its history and applications in science, technology and humanities*. Wiley, New York.

[Goodrum et al., 2001]Goodrum, A. A., McCain, K. W., Lawrence, S., and Giles, C. L. (2001). Scholarly publishing in the internet age: A citation analysis of computer science literature. *Information Processing & Management*, 37(5):661–675.

[Harold and Means, 2004]Harold, E. R. and Means, W. S. (2004). *XML in a Nutshell*. O'Reilly, 3rd edition.

[Hirsch, 2005]Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Science of the United States of America*, 102(46):16569–16572.

[Jacsò, 2005]Jacsò, P. (2005). As we may search. Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9):1537–1547. Retrieved December 20, 2008, from http://www.ias.ac.in/currsci/nov102005/1537.pdf.

[Katsaros et al., 2006]Katsaros, C., Manolopoulos, Y., and Sidiropoulos, A. (2006). Generalized h-index for disclosing latent facts in citation networks. Retrieved December 20, 2008, from http://arxiv.org/abs/cs.DL/0607066.

[Kousha and Thelwall, 2007]Kousha, K. and Thelwall, M. (2007). Google scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58(7):1055–1065.

[Kousha and Thelwall, 2008]Kousha, K. and Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: a comparison between four science disciplines. *Scientometrics*, 74(2):273–294.

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

[Meho and Yang, 2007]Meho, L. I. and Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science vs. Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13):2105–2125.

[Moore, 2006]Moore, D. (2006). *Basic Practice of Statistics*. WH Freeman Company, 4th edition.

[Norris and Oppenheim, 2007]Norris, M. and Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences literature. *Journal of Informetrics*, 1(2):161–169.

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

[Pauly and Stergiou, 2005]Pauly, D. and Stergiou, K. I. (2005). Equivalence of results from two citation analyses: Thomson ISI citation index and Google scholar service. *Ethics in Science and Environmental Politics*, 5:33–35.

[R Development Core Team, 2009]R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[Saad, 2006]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.

[Sanderson, 2008]Sanderson, M. (2008). Revisiting h measured on UK LIS academics. *Journal of the American Society for Information Science and Technology*, 59(7):1184–1190.

[Shaw and Vaughan, 2008]Shaw, D. and Vaughan, L. (2008). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics*, 74(2):317–330.

[Weingart, 2005]Weingart, P. (2005). Impact of bibliometrics upon the science system: inadvertent consequences? *Scientometrics*, 62(1):117–131.

[Whitley, 2002]Whitley, K. M. (2002). Analysis of SciFinder Scholar and Web of Science citation searches. *Journal of the American Society for Information Science and Technology*, 53(14):1210–1215.

[Zhao and Logan, 2002]Zhao, D. Z. and Logan, E. (2002). Citation analysis using scientific publications on the web as data source: A case study in the XML research area. *Scientometrics*, 54(3):449–472.