ⓜ CYBER*metrics*

VOLUME 13 (2009): ISSUE 1. PAPER 2

Handling self-citations using Google Scholar 📄

Francisco M. Couto, Catia Pesquita, Tiago Grego, and Paulo Veríssimo

Informatics Department, Faculty of Sciences, University of Lisbon, Portugal
E-mail:**fcouto@di.fc.ul.pt**

Abstract

The increasing use of citation impact indexes for evaluation and comparison not only of individual researchers but also of institutions, universities and even countries has prompted the development of new citation metrics. Currently, the number of publications and citations is widely accepted as an easy and balanced way to compare scientists. Calculation of such statistics depends on the availability of a comprehensive database of publications and their citations. Google Scholar aims at providing such a service and is currently the most widely used freely available search engine for scientific and academic literature. However, the citations generally used to calculate citation statistics include self-citations, which deviates from the intention of using citations as a reflection of research impact.

To the best of our knowledge, there are no available tools for calculating citation statistics that account for self-citations. We present a web-based service CIDS (Citation Impact Discerning Self-citations), that takes into account self-citations. An assessment of CIDS in a research team has been shown that both the number of citations and the h-index is sensitive to self-citations at the individual level, the h-index increasing 24% on average when considering them. However, self-citation is highly variable among individuals and its contribution highly variable. We conclude that at the individual and research unit level, self-citations are not dismissible when calculating citation statistics. Even the h-index is influenced by self-citation and comparing individuals without taking them in account can produce misleading results.

CIDS is available at: **http://xldb.fc.ul.pt/tools/cids/**.

Keywords

Self-citation, Bibliometrics, Citation Analysis, h-index

1. Introduction

Quantification of research impact and relevance is becoming increasingly important for individuals, academic institutions and even countries. It is often used as a means of evaluation for university faculty recruitment and promotion, attribution of funding and awards (Ball, 2005; Holden, 2006). Citation analysis can be used to estimate research impact, assuming that citations are references to work that has influenced the author (MacRoberts & MacRoberts, 1989). Several citation indicators have been used, such as the number of highly cited papers, total number of published papers, the mean number of citations per paper and the total number of citations. Recently a new metric was proposed by Hirsch, the h-index (Hirsch, 2005), defined as follows:

A scientist has index **h** if **h** of his or her **Np** papers have at least **h** citations each and the other **(Np − h)** papers have **≤h** citations each.

An h-index of 20 means that the author has published 20 papers each having at least 20 citations. The value of h is likely to increase linearly with time, i.e. a scientist's h-index will never decrease, it can only increase as more papers are published and cited.

Several aspects can influence the h-index, such as the disciplinary field. For instance, in Computer Science the highest h-index is 70 [1], while in the life sciences the highest h-index is 191 (Hirsch, 2005). The h-index has received a lot of attention from the scientific community and caused a lot of discussion (Ball, 2005). However, its appeal is clear: it contributes to the ranking of scientists using a single value that is straightforward to calculate and fairly robust. Its advantages and pitfalls have been reviewed in (Bar-Ilan, 2008; Costas & Bordons, 2007; Glänzel, 2006; Rousseau, 2008).

The h-index definition does not include any reference on how to discern self-citations. A self-citation is usually defined as a citation where the citing and the cited paper share at least one author, but it can be split into three kinds:

1. the papers share the first author;

2. one paper's first author is co-author on the other;

3. the papers share only co-authors (Schreiber, 2007).

Self-citations make up a significant number of the total citations, between 10 to 30% depending on the research area (Aksnes, 2003; Hyland, 2003). Self-citation allows authors to connect present work to related and meaningful past works, providing a means to build upon and discuss previous findings without too much repetition. It can also be used to present the author as an expert in the area, by referring the authors' previous works. In addition, authors also often resort to self-citation because it is easier to cite a paper they know well, than search the literature for another relevant paper on the subject. For the same reason there is a tendency to cite more work done by colleagues from the same group, university, institution, country or network of collaborators. However, there are less legitimate uses of self-citation, since it is a means to gratuitously increase one's citation counts and purposefully increase the visibility of older work.

The h-index is sensitive to such manipulations, since the papers near h citations are obvious targets for

self-citing. There is even an online tool that highlights the papers where an increase in citations will increase h-index [2]. Although Hirsch states that self-citations have little effect on h-values (Hirsch, 2005), other studies show that removing self-citations can account for an average decrease of the h-index by as much as 12% (Kelly & Jennions, 2006) or as little as 1% (Cronin & Meho, 2006). The influence of self-citations in the h-index is more relevant for the younger scientists with few publications, where coincidentally the h-index may be more frequently used as a means of evaluation or comparison with fellow competitors (Schreiber, 2007). Self-citation is influenced by various factors, including total number of citations (poorly cited papers have a higher self-citation rate), number of authors (multiple author publications receive more self-citations), the area of research (where citation norms, the accumulativeness at the level of individual research and the number of researchers in the area causes variation) and the year of publication (since self-citations make up the majority of the citations a paper receives in its first year but decreases with time) (Aksnes, 2003). Therefore removing self-citations when calculating the h-index would result in a more reasonable quantification of research impact.

The importance of research impact indicators is unquestionable and several tools have been made available for their computation. We developed a web-based service CIDS (Citation Impact Discerning Self-citations) to calculate citation statistics with and without self-citations. It uses Harzing's Publish or Perish [3] output and filters out self-citations according to Google Scholar to produce a final corrected output.

Source	Link	Access
CiteSeerX	**http://citeseerx.ist.psu.edu/**	free
Google Scholar	**http://scholar.google.com**	free
getCITED	**http://www.getcited.org/**	free
CiteULike	**http://www.citeulike.org/**	free
The DBLP Computer Science Bibliography	**http://www.informatik.uni-trier.de /~ley/db/**	free
Thomson/Reuters Web of Science	**http:/scientific.thomson.com/isi/**	commercial
Scopus from Elsevier	**http://www.scopus.com/**	commercial

Table 1. The most commonly used sources of bibliographic data

The rest of the paper is organized as follows: Section 2 presents an overview of the related work, Section 3 introduces our proposed approach, Section 4 describes the results and finally Section 5 expresses our main conclusions.

```
Cites,Authors,Title,Year,Source,Publisher,ArticleURL,CitesURL
27,"D Rebholz-Schuhmann, H Kirsch, F Couto","Facts from text-is text
mining ready to deliver",
2005,"PLoS Biol","medicine.plosjournals.org",
"http://medicine.plosjournals.org/perlserv?request=get-document&
amp;doi=10.1371/journal.pbio.0030065",
"http://scholar.google.com/scholar?num=100&hl=en&
lr=&cites=1955700969208253187
```

Figure 1. Data returned by HPP for one of the publications analyzed

Input: pubs, list of Google Scholar identification numbers for each publication returned by HPP (e.g. in Figure 1 the number is 1955700969208253187);
Output: *citation_ inc_self*, number of citations for each publication; *citations_ex_ self*, number of citations excluding self-citations for each publication;
1: **for all** *pub* **in** *pubs* **do**
2:      *citations_inc_self*[*pub*] = get citations(*pub*)
3:      *authors* = *get_authors*(*pub*)
4:      *citations_exc_self*[*pub*] = *get_citations_exc_authors*(*pub*, authors*)
5: **end for**

Figure 2. Number of Citations Calculation

```
<paper>
 <authors>D Rebholz-Schuhmann, H Kirsch, F Couto</authors>
 <title>Facts from text-is text mining ready to deliver</title>
 <year>2005</year>
 <source>PLoS Biol</source>
 <publisher>medicine.plosjournals.org</publisher>
<citations_excluding_selfcitation>23</citations_excluding_selfcitation>
<citations_including_selfcitation>27</citations_including_selfcitation>
</paper>
```

Figure 3. XML returned by CIDS containing the citation statistics of one of the publications analyzed

```
<statistics>
  <papers>229</papers>
  <including_selfcitation>
    <cites_paper>10.4192139737991</cites_paper>
```

```
      <h_index>25</h_index>
      <citations>2386</citations>
    </including_selfcitation>
    <excluding_selfcitation>
      <cites_paper>7.4235807860262</cites_paper>
      <h_index>21</h_index>
      <citations>1700</citations>
    </excluding_selfcitation>
  </statistics>
```

Figure 4. XML returned by CIDS containing the citation statistics of one of the researchers in our study

2. Related Work

In order to calculate citation indicators one obviously needs to have access to the publications and their citations. Table 1 presents sources of bibliographic data that include web-search engines providing listings of scientific publications and their citations. Some of these sources are freely available and others subscription-based, and some of them with restricted scope, such as DBLP which is limited to computer science.

2.1 Web of Science

The Web of Science provides access to bibliographic data from approximately 8,700 multidisciplinary research journals. Web of Science also provides cited reference searching. However, Web of Science access is subscription-based, which means that its access his restricted to institutions that subscribe it (Harzing & Wal, 2007).

The main disadvantage of Web of Science to calculate the citation impact is that it only includes journal articles published in Thomson/Reuters (previously called Thomson ISI) indexed journals (Roediger, 2006). It misses all the citations to books, book chapters, dissertations, thesis, technical reports, conference papers, workshop papers, and journal articles not indexed by Thomson/Reuters. Thomson/Reuters analyzes more than 2,000 journals per year and only 10 to 12% are selected to be included in the Web of Science (Testa, 2004). For example, Thomas E. Nisonger showed that more than 70% of his citations were not captured by Web of Science (Nisonger, 2007). To overcome this problem Web of Science provides the cited reference function that includes citations from journals that are listed by Thomson/Reuters, however, the number of listed journals is also limited (Meho & Yang, 2007).

Another problem of Web of Science is the duplication of entries for the same article derived from small variations in its citations. A study showed that Web of Science had only 60% of the articles using a unique entry, while Google Scholar had 85% (Belew, 2005).

2.2 CiteSeer

CiteSeer is a public digital library and search engine for scientific and academic literature, primarily in computer and information science. First developed in 1997 at the NEC Research Institute in Princeton, it was the first to provide automated citation indexing and citation linking using the method of autonomous citation indexing. CiteSeer aims to improve the dissemination, feedback and overall accessibility of scientific literature.

CiteSeer can no longer be used as an accurate source of current research, since it has not been comprehensively updated since 2000 due to limitations of the system's capabilities. Nonetheless, a new generation of CiteSeer is recently available as CiteSeerX (**http://citeseerx.ist.psu.edu/**), which was developed to continue the project using a more efficient architecture.

2.3 Google Scholar

An alternative to CiteSeer was launched in 2004, Google Scholar. It is a freely available search engine that indexes academic literature and consists of full-text journal articles, technical reports, preprints, theses, books, and other documents. Google has agreements with publishers to allow Google Scholar to access their private repositories, but the exact listing of publishers is undisclosed.

Although it's been generally acclaimed, Google Scholar is still in beta. Its frequency of update is unknown, making it difficult to pinpoint how up-to-date and comprehensive Google Scholar is. Its search results are certainly not totally sound and complete, since it includes some non-scholarly citations, and it does not include some journals (Meho & Yang, 2007). Nonetheless, Google Scholar has a larger coverage of citations than Web of Science for citations published after 1990. This is not true for older publications since they have not yet been published on the web (Daniel & Stergiou, 2005). Moreover the easy access to publications provided by Google Scholar overcomes some of the inconveniences of the commercial bibliographic sources (Bosman et al., 2006).

Google Scholar has a feature, named "cited by" which is of particular relevance, since it gives the number of citations the paper has, and links to them. Previously, citation indexing was only provided by expensive commercial services like Scopus by Elsevier and Web of Science. However, Google Scholar offers no option to filter self-citations. In addition Google Scholar does not have an API nor does it present its results in XML, making further analysis a difficult task.

2.4 Harzing

Once access to publications and their citations is available, the next step is to use that information to calculate the citation impact. Harzing's Publish or Perish [3] (HPP) is a software tool that calculates citation statistics based on the raw citations it retrieves from Google Scholar. The results are presented on-screen but can also be copied to a blank document or saved in various output formats, for further reference or analysis. HPP calculates the several statistics including: total number of papers, total number of citations, average number of citations per paper, average number of citations per author, average number of papers per author, average number of citations per year, Hirsch's h-index and related parameters, Egghe's g-index and others.

However it does not discern self-citations in its calculations. And although outputting results into several formats (BibTex, CSV, EndNote and RIS) it does not output to XML, making the retrieval of its information more cumbersome. Nevertheless HPP is a very user-friendly tool for citation statistics, providing a valuable resource to

authors.

Authors:	D Rebholz-Schuhmann, H Kirsch, F Couto
Title:	Facts from text-is text mining ready to deliver
Year:	2005
Source:	PLoS Biol
Publisher:	medicine.plosjournals.org
Citations excluding self-citation:	23
Citations including self-citation:	27

Figure 5. Output of CIDS presenting the citation statistics of one of the publications analyzed

	Including Self-citations	Excluding Self-citations
Number of papers:	229	
Citations per papers:	10.4192139737991	7.4235807860262
H-index:	25	21
Citations:	2386	1700

Figure 6. Output of CIDS presenting the citation statistics of one of the researchers in our study

3. CIDS

We have developed CIDS (Citation Impact Discerning Self-citations) to address the issue of self-citation when calculating citation impact statistics. CIDS aims at being a straightforward service that helps authors interested in ascertaining their citation impact with and without self-citations. It is available as an online service where users upload their HPP results and receive an e-mail detailing several citations statistics both with and without self-citations.

Users are requested to install HPP and use it to retrieve their citation statistics and publications listing. Following the recommendation of Bornmann & Daniel (2007) it is important that there is human intervention at this point since automatic searches cannot with a 100% certainty rule out papers by a different scientist of the same name. Thus, the user has to read the title of each paper and select the ones that were written by him or her and not by another person with the same name. The user then uses the "Copy" feature of HPP to create a text document with the results (see Figure 1). This document is the input of CIDS web-service. It includes citation statistics as well as a list of all publications the user decided to include.

Figure 2 presents the algorithm to calculate the number of citations, which uses the following functions:

> *get_citations*(*pub*) returns the number of citations shown in
> **http://scholar.google.com/scholar?cites=** followed by *pub*. For example,
> **http://scholar.google.com/scholar?cites=1955700969208253187** shows 27 citations.

> *get_authors*(*pub*) returns the list of authors of *pub*. For example, for
> *pub*=1955700969208253187 it returns
> **'D%20Rebholz-Schuhmann,%20H%20Kirsch,%20F%20Couto'**.

> *get_citations_exc_authors*(*pub*,*authors*) is similar to *get_citations*(*pub*), it only adds at the end of
> the link the string **&q=-author:** and the list of authors. This way, Google Scholar excludes from
> the citations the ones that have at least one of its authors in the list of authors given as input. For
> example,
> **http://scholar.google.com/scholar?cites=1955700969208253187&
> q=-author:"D%20Rebholz-Schuhmann,%20H%20Kirsch,%20F%20Couto"** shows 23
> citations.

After identifying the self-citations we can calculate citation statistics. CIDS calculates the following statistics:

- total number of papers;
- total number of citations (including and excluding self-citations);
- average number of citations per paper (including and excluding self-citations);
- Hirsch's h-index (including and excluding self-citations).

The output is an XML file including a listing of all publications, detailing authors, title, year of publication, source, publisher and citations including and excluding self-citations, as well as the statistics (see Figures 3 and 4). The XML file can be manually verified in a web browser, using a developed XSL Web style sheet (see Figures 5 and 6).

4. Assessment

CIDS was developed to be used by our research group, which develops work in several sub-areas of Informatics (Computer Science and Engineering) including Distributed Systems, Human Computer Interaction and Multimedia, Information Management, Security and Dependability. The research lab had at the time of the study approximately 72 collaborators, 23 of which hold a doctoral degree, 12 are PhD students and 18 are MSc students or junior researchers. Our lab is funded by both national and international projects, and as such, accurate calculation of citation impact is a matter of importance, for both the individual researchers and the research unit as a whole. In September 2007, 21 out of 23 PhD holders and/or senior staff members participated in the assessment of CIDS by providing their HPP output files.

	Maximum	Minumum	Average	Median
Total number of papers	229	4	47.2	20.0
Total citations	2386	4	335.5	74.0
Self-citations	686	0	106.0	49.0
Citations per paper including self-citations	14.2	1.0	5.4	3.9
Citations per paper excluding self-citations	10.0	0.1	3.4	2.3
Self-citation rate	97.4%	0.0%	40.5%	35.1%
h-index including self-citations	25	1	7.6	5.0
h-index excluding self-citations	21	1	6.0	4.0

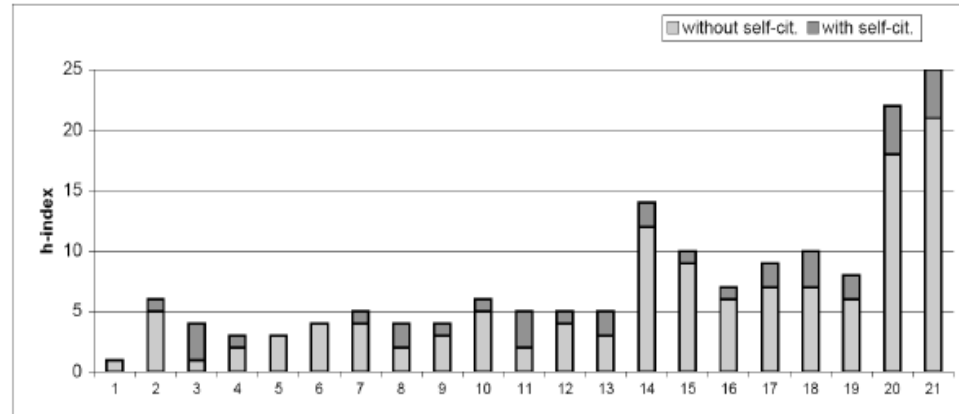**Table 2. Average and median citation statistics**

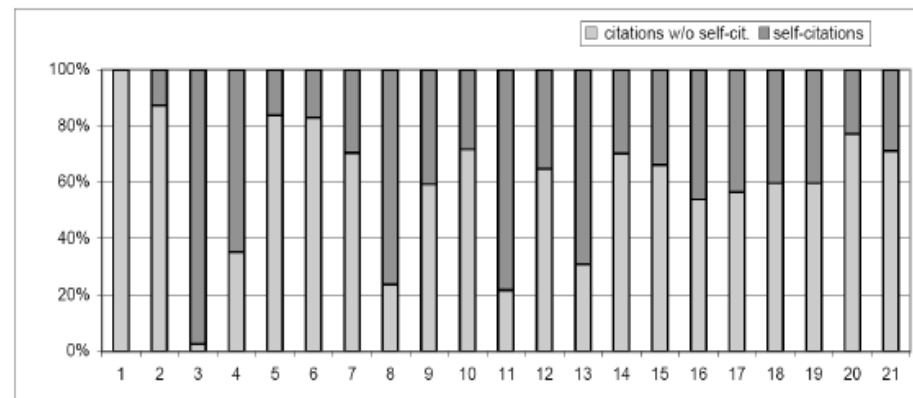**Figure 7. Contribution of self-citations for each researcher's h-index**

Figure 8. Composition of each researcher's citations

	Google Scholar	Thomson/Reuters
Total number of papers	329	18
Total citations	3070	44
Citations per paper	9.33	2.44
h-index	29	4
Self-citation rate	34.2%	24.2%

Table 3. Google Scholar and Thomson/Reuters citation statistics for senior scientist

	Google Scholar	Thomson/Reuters
Total number of papers	16	10
Total citations	140	11
Citations per paper	8.75	3.67
h-index	7	2
Self-citation rate	27.9%	9.1%

Table 4. Google Scholar and Thomson/Reuters citation statistics for junior scientist

4.1. Results

In this assessment each researcher was attributed a number corresponding to his or her ranking by ascending number of publications, i.e., researcher 1 has the fewer publications and researcher 21 the most. An overview of the results obtained is given in Table 2. Average and median values are given because of the high variability among the researchers, for instance one researcher has 4 papers while another has 229. Self-citations make up 31% of the total citations of the research unit as a whole and the average self-citation rate is 40%. Self-citations contribute to an average increase of 24% to the h-index of each researcher as shown in Figure 7. Self-citation is highly variable at individual level as seen in Figure 8. Note that the total number of papers is underestimated since HPP does not account for offline papers present in Google Scholar results and Google Scholar cannot provide full coverage of all authors publications. Offline papers are papers which content was not found online but were cited by other online papers. On average, researchers from our lab have published 11 more papers than

what HPP finds.

From the research team involved in this assessment we selected a senior and a junior researcher to compare their citation statistics obtained by using Google Scholar with the ones obtained by using Thomson/Reuters. Tables 3 and 4 show these statistics for the senior and junior researcher, respectively. The citation report provided by Thomson/Reuters includes the self-citations, so we could not obtain the citation statistics without using self-citations from Thomson/Reuters. It was only possible to obtain the list of citations with or without self-citations. The Thomson/Reuters results only include the journals that are listed by Thomson/Reuters, since when using the cited reference function we could not obtain the citation report.

For both cases all the papers found in Thomson/Reuters were also found by Google Scholar.

4.2. Discussion

The results presented are influenced directly by the ability of Google Scholar of producing a complete listing of author's publications and their citations. Google Scholar can fail to produce this complete listing due to lack of coverage, and in this case study the average loss is of 35%. Currently, Google Scholar limits the number of returned publications per search to 1000 (i.e. 10 pages if 100 publications are returned per page) which can pose a problem for highly cited and highly published authors. More importantly, Google Scholar limits the number of authors to 4, not showing additional author's names, making the identification of self-citation in multiple authorship papers not fully accurate. Thus, a suitable Google Scholar API is much required to get a more complete and exact application.

By analyzing the Thomson/Reuters and Google Scholar results we can verify that Thomson/Reuters is more incomplete. The h-index difference in Google Scholar between the senior and the junior is 22, and in Thomson/Reuters is just 2. The self-citation rate is also larger in Google Scholar, which can be explained by the large number of self-citations to preliminary work normally presented in conference papers. The results are also influenced by HPP performance. For instance, HPP does not retrieve Google Scholar citations of offline articles. These citations refer to articles Google Scholar was not able to find online but to which there are citations in online papers. As an example, a Google Scholar search for papers by one of the researchers results in 298 hits, whereas HPP only retrieved 229.

Computer Science has a rate of self-citations above average (Aksnes, 2003) but maximum h-indexes below average [1] when compared to other fields. In the studied group, self-citation was somewhat above the numbers reported for the same general field in a different country (24%) (Aksnes, 2003) and average citations per paper were above the value for Computer Sciences as given by Thomson/Reuters for the period of 1993-2003 (2,3%) [4]. However there is considerable variability amongst researchers, not only on self-citation rates and average citations per paper but also on total number of papers published. The top 4 publishing authors have published on average 146 papers, have an average h-index of 16 (13 without self-citations) and 33% self-citations. The bottom 4 have published on average 8 papers, have an average h-index of 3 (2 without self-citations) and 43,7 % self-citations. An extreme showcase of the influence of self-citations on the h-index is presented by researchers 1 and 3. Researcher 1 has 4 papers, 4 citations and an h-index of 1, whereas researcher 3 has 9 papers, 38 citations and an h-index of 4. However, researcher 1 has no self-citations and researcher 3 has 37 self-citations. When removing self-citations from the h-index calculations, they both have an h-index of 1. For the average scientist, self-citations are expected to have a significant influence on citation impact calculations, since their effect is only diluted for highly cited scientists.

At the level of the research unit or department, self-citations are also important, particularly if one calculates the impact of that group by the average h-index. In our case, it dropped from 7,6 to 6 by excluding self-citations. An alternative to the average h-index for comparing groups of researchers is the Prathap's h_2 (Prathap, 2006): *A department has at least h if at least h of its researchers have h-index at least h.* This however is only helpful to compare departments working in the same research area, due to different citations patterns. Comparison of several departments according to their h_2 is already available [5]. However, h_2 is also affected by the number of people in the research unit since it can never have a h_2 higher that the number of people in it. Our group has a meta-index of 7 with self-citations and 6 without them, in this case the average h-index and h_2 capture equally well the groups' joint impact.

As a final remark, h-index does not take into account the *scientific age* of the authors, and thus penalizes younger researchers who have not yet had time to publish more. Comparison between research units should also take this into account, since a team with older researchers with high h-index is not necessarily stronger than a team of younger and promising researchers with lower h-indexes. Therefore, current experience of hiring and promotion committees suggests that the h-index, same as with number of publications and citations, should be just one piece of the evaluation puzzle.

5. Conclusions

Self-citation is a useful component of a scientific publication, as we discussed earlier. However, assessing research impact is affected by self-citations, since they almost never reflect the impact of that paper in the community. Considering the increasing number of evaluations that resort to citation impact statistics, authors feel tempted to deliberately cite themselves to increase their citation indicators. If at large scale, self-citations do not represent a big influence on citation statistics, at the scale of the individual researcher or research unit, self-citations can play an important role. In this case study, we have shown that removing the self-citations from h-index calculations decreases the latter on average by 24%. Moreover, comparing individuals based on their h-index including self-citations would be misleading and ultimately unfair, given the high variability in self-citations at the individual level.

Removing self-citations when making such comparisons is more accurate. However, dismissing self-citations may not be as straightforward as it seems. An exception can be considered in the case of papers with a large number of authors, where a self-citation of the third kind can be reflecting impact. This problem is becoming more critical since the average number of authors per paper seems to be increasing in several areas (Aksnes, 2003). An obvious solution would be to consider only first author citations, however one issue still remains: should papers with many authors have their impact divided by the number of authors, considering that the work was very distributed?

Future work on CIDS will include a direct use of Google Scholar, without using HPP. This feature will soon be available in a second version of CIDS. Also it would be interesting to further distinguish self-citations by type and account for the multiple authorship effect. We will also improve the coverage of papers analyzed by submitting websites where the contents of offline papers can be found for inclusion in Google Scholar.

6. Acknowledgements

Notes

[1] **http://www.cs.ucla.edu/~palsberg/h-number.html**

[2] **http://www-ihm.lri.fr/~roussel/moulinette/h/h.cgi**

[3] **http://www.harzing.com/pop.htm**

[4] **http://www.in-cites.com/analysis/03-sixth-com.html#Baselines**

[5] **http://www.cs.utah.edu/~shirley/hindex/**

References

Aksnes, W.(2003). A macro study of self-citation. **Scientometrics**, 56(2):235–246

Ball, P (2005). Index aims for fair ranking of scientists. **Nature**, 436:900, 2005.

Bar-Ilan (2008). Which h-index? a comparison of wos, scopus and google scholar. *Scientometrics*, 74(2):257–271, 2008.

Belew, R. (2005). Scientific impact quantity and quality: Analysis of two sources of bibliographic data. **arXiv**:**cs/0504036v1**.

Bornmann, Lutz and Daniel, Hans-Dieter (2007). What do we know about the h index? **Journal of the American Society for Information Science and Technology**, 58:1381–1385

Bosman, J, Mourik, I, Rasch, M van, Sieverts, E and Verhoeff, H. (2006). Scopus reviewed and compared. the coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar. *Utrecht University Library*, 2006. **http://igitur-archive.library.uu.nl/DARLIN /2006-1220-200432/Scopus%20doorgelicht%20%26%20vergeleken%20-%20translated.pdf**

Costas, R. and Bordons, M (2007). The h-index: advantages, limitations and its relation with other bibliometric indicators at the micro-level. **Journal of Informetrics**, 1(3):193–203

Cronin, Blaise and Meho, Lokman (2006). Using the h-index to rank influential information scientistss: Brief communication. **Journal of the American Society for Information Science and Technology**, 57(9):1275–1278

Daniel, P. and Stergiou, K. (2005). Equivalence of results from two citation analyses: Thompson isi's citation index and google's scholar service. **Ethics in Science and Environmental Politics**, 2005, 33–35

Glänzel, W. (2006). On the opportunities and limitations of the h-index. **Science Focus**, 1(1):10–11, 2006.

Harzing, A. and Wal, R. (2008). Google Scholar as a new source for citation analysis. **Ethics in Science and Environmental Politics**, e008:5, 2007.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. **Proceedings of the National Academy of Sciences**, 102:16569.

Holden, C. (2006). Random samples: Data point - impact factor. **Science Magazine,** 309(5738):1181

Hyland, Ken (2003). Self-citation and self-reference: Credibility and promotion in academic publication. **Journal of the American Society for Information Science and Technology**, 54(3):251–259

Kelly, C. and Jennions, M (2006). The h index and career assessment by numbers. **TRENDS in Ecology and Evolution**, 21(4):251–259

MacRoberts, M. and MacRoberts, B. (1989). Problems of citation analysis: A critical review. **Journal of the American Society for Information Science and Technology**, 40(5):342–349

Meho, L. and Yang, K (2007). A new era in citation and bibliometric analyses: Web of science, scopus, and google scholar. **Journal of the American Society for Information Science and Technology**, 58:1–21

Nisonger, T. (2007). Citation autobiography: an investigation of ISI database coverage in determining author citedness. **College & Research Libraries**, 65(2):152–163

Prathap, G. (2006). Hirsch-type indices for ranking institutions' scientific research output. **Current Science**, 91(11):1439

Roediger, H. (2006). The h index in science: A new measure of scholarly contribution. **APS Observer: The Academic Observer**, 19

Rousseau, R (2008). Reflections on recent developments of the h-index and h-type indices. *COLLNET Journal of Scientometrics and Information Management*, 2(1):1–8.

Schreiber, M. (2007). Self-citation corrections for the Hirsch index. **EPL**, 78:30002

Testa, J. (2004). The Thomson Scientific journal selection process, Thomson Reuters Essay, 2004. **http://thomsonreuters.com/business_units/scientific/free/essays/journalselection/**