The expansion of Google Scholar versus Web of Science: a longitudinal study

de Winter, J.C.F.; Zadpoor, A.A.; Dodou, D.

DOI
10.1007/s11192-013-1089-2

Publication date
2014

Document Version
Final published version

Published in
Scientometrics: an international journal for all quantitative aspects of the science of science, communication in science and science policy

Citation (APA)
de Winter, J. C. F., Zadpoor, A. A., & Dodou, D. (2014). The expansion of Google Scholar versus Web of Science: a longitudinal study. *Scientometrics: an international journal for all quantitative aspects of the science of science, communication in science and science policy*, *98*(2), 1547-1565. https://doi.org/10.1007/s11192-013-1089-2

Scientometrics (2014) 98:1547–1565
DOI 10.1007/s11192-013-1089-2

The expansion of Google Scholar versus Web of Science: a longitudinal study

Joost C. F. de Winter · Amir A. Zadpoor · Dimitra Dodou

Received: 15 May 2013 / Published online: 1 August 2013

Abstract Web of Science (WoS) and Google Scholar (GS) are prominent citation services with distinct indexing mechanisms. Comprehensive knowledge about the growth patterns of these two citation services is lacking. We analyzed the development of citation counts in WoS and GS for two classic articles and 56 articles from diverse research fields, making a distinction between retroactive growth (i.e., the relative difference between citation counts up to mid-2005 measured in mid-2005 and citation counts up to mid-2005 measured in April 2013) and actual growth (i.e., the relative difference between citation counts up to mid-2005 measured in April 2013 and citation counts up to April 2013 measured in April 2013). One of the classic articles was used for a citation-by-citation analysis. Results showed that GS has substantially grown in a retroactive manner (median of 170 % across articles), especially for articles that initially had low citations counts in GS as compared to WoS. Retroactive growth of WoS was small, with a median of 2 % across articles. Actual growth percentages were moderately higher for GS than for WoS (medians of 54 vs. 41 %). The citation-by-citation analysis showed that the percentage of citations being unique in WoS was lower for more recent citations (6.8 % for citations from 1995 and later vs. 41 % for citations from before 1995), whereas the opposite was noted for GS (57 vs. 33 %). It is concluded that, since its inception, GS has shown substantial expansion, and that the majority of recent works indexed in WoS are now also retrievable via GS. A discussion is provided on quantity versus quality of citations, threats for WoS, weaknesses of GS, and implications for literature research and research evaluation.

Keywords Automatic indexing · Citation classic · Citation Index · Historic trend · Most highly cited paper · Strengths and weaknesses

---

Electronic supplementary material The online version of this article (doi:10.1007/s11192-013-1089-2) contains supplementary material, which is available to authorized users.

J. C. F. de Winter (✉) · A. A. Zadpoor · D. Dodou
Delft University of Technology, Delft, The Netherlands
e-mail: j.c.f.dewinter@tudelft.nl

Introduction

Retrieval of publications and their citations is of chief importance in the modern scientific enterprise. In any scientific paper, especially in systematic reviews and meta-analyses, it is important to recover the relevant published sources of information. In addition, citation analyses are broadly used for research assessment of individual scientists, departments, institutions, and countries.

Web of Science (WoS) and Google Scholar (GS) are two popular citation services, and various studies have debated their comparative strengths and weaknesses (e.g., Amara and Landry 2012; Falagas et al. 2008; Franceschet 2010; García-Pérez 2010; Hightower and Caldwell 2010; Kulkarni et al. 2009; Mikki 2010; Mingers and Lipitakis 2010). In this study, we investigate longitudinal trends of citation counts in WoS and GS to shed light on the growth patterns of both citation services.

Philosophies of GS and WoS

WoS and GS have different philosophies. WoS indexes selectively, motivated by the premise that "an essential core of journals forms the literature basis for all disciplines and that most of the important papers are published in relatively few journals" (Thomson Reuters 2013a). In-house editors assess candidate publication outlets using criteria such as timeliness, peer-review process, international diversity of editors and authors, citation impact, and self-citation rate (Thomson Reuters 2013b). WoS currently indexes 12,000 journals, 148,000 conference proceedings, 30,000 books published since 2005, and 46 million records back to 1900 (Thomson Reuters 2013a). WoS requires subscription and is therefore not accessible by the general public.

GS is a free service that uses web crawlers for retrieving scholarly material from journal websites, university repositories, and authors' personal websites. Apart from journals and conferences, GS also retrieves document types that are not indexed in WoS such as working papers, reports, preprints, and theses. Scholarly documents are identified by means of automatic format inspection (title in large font at the front page, author names right below the title, and the presence of a section titled "References" or "Bibliography"). Indexing is done automatically by parsers that identify bibliographic data in the selected documents (Google Scholar 2013). It has been argued that because of its automatic inclusion process, GS is susceptible to errors in metadata (Jacsó 2008) and to indexing of non-scientific works (Cathcart and Roberts 2005; Donlan and Cooke 2005; Jacsó 2005a; Vine 2006; Wleklinski 2005).

Historic development of literature coverage by GS and WoS

Various early studies indicated that GS provided poor coverage of scientific works. Neuhaus et al. (2006), for example, investigated how many of 2,350 articles selected from 47 databases including the American Chemical Society, ERIC, JSTOR, PubMed, and SpringerLink were retrievable from GS and found that coverage of the databases varied between 6 and 100 %. Jacsó (2005a) found that GS retrieved only 16 % of Nature's publications, and concluded that, even when the crawlers have access to a database, they may fail to retrieve documents. Mayr and Walter (2007) reported that GS recovered 79 % of 9,500 titles from journals across different research fields. Meier and Conkling (2008) investigated GS's recovery of engineering records taken from Compendex and reported that 89 % of the records published after 1990 were retrieved by GS, with the corresponding percentages being lower for the 1980s (76 %), 1970s (57 %), 1960s (51 %), and 1950s (33 %), indicating that GS's coverage is poorer for older publications. Further criticism on the poor coverage of scientific works by GS can be found in Jacsó (2005b, c).

More recent longitudinal studies suggest that GS nowadays indexes considerably more documents than it did in the early years after its inception in 2004. Chen (2010) compared GS's coverage of journal articles in eight databases (Emerald, ERIC, JSTOR, Project MUSE, American Chemical Society, Oxford University, SpringerLink, and University of Chicago) for 2010 with the corresponding values reported by Neuhaus et al. (2006) for 2005 and found that GS's coverage had increased from 30–88 % to 98–100 %. Harzing (2013a) analysed the longitudinal development of citation counts of 20 Nobel Prize winners between April 2011 and January 2012 and found that citation counts for chemistry and physics, disciplines traditionally poorly represented in GS, was increasing rapidly. In a follow-up study analysing citation counts between April 2011 and January 2013, Harzing (2013b) again observed a large increase in GS as compared to WoS. For one Nobel Prize laureate in chemistry (E. J. Corey), GS citations amounted to 36 % of his WoS citations in April 2011, 61 % in January 2012, and 74 % in January 2013.

In a longitudinal study comparing the temporal development of the Science Citation Index (SCI; part of WoS) with seven other databases (Chemical Abstracts, Compendex, Cambridge Scientific Abstracts, Inspec, Lecture Notes in Computer Science, MathSciNet, and PubMed Medline) up to 2007, Larsen and Von Ins (2010) reported that the annual growth rate of SCI (2.7 % for all records) was the lowest among the investigated databases, implying that SCI covers a decreasing part of the scientific literature. The coverage of peer-reviewed journal articles by WoS was found to be particularly low for research fields with the highest growth rates, such as computer sciences and engineering.

Relative coverage of the literature by GS and WoS

GS and WoS cover different shares of the scientific literature. Studies have shown that GS provides fewer citation counts than WoS in biology, physics, and chemistry, but more in information technology, human–computer interaction, social sciences, economics, management, engineering, and mathematics (Amara and Landry 2012; Bauer and Bakkalbasi 2005; Bar-Ilan et al. 2007; Bornmann et al. 2009; Bosman et al. 2006; Franceschet 2010; Kousha and Thelwall 2007; Levine-Clark and Gil 2009; Mingers and Lipitakis 2010).

Several studies have investigated the relative coverage of WoS and GS using a citation-by-citation analysis, with relative coverage defined as the percentage of unique records (i.e., records retrieved by only one of the citation services) with respect to the union of the two services. Two years after GS's inception, Bakkalbasi et al. (2006) analysed the citations to articles published in 2003 in two research fields, namely oncology and condensed physics. They found that the percentage of citations unique in GS was smaller than the percentage of unique citations in WoS both in oncology (18 vs. 35 %) and in condensed matter articles (20 vs. 43 %). Others reported a higher number of unique citations in GS than in WoS: 43 versus 24 % in a study of citations to 882 articles in Open Access ISI-indexed journals in biology, chemistry, physics, and computer sciences (Kousha and Thelwall 2007); 37 versus 1 % of the total number of citations to 252 journal articles from the body of work of four psychologists (García-Pérez 2010); 69 versus 4 % in a citation analysis of the work of 29 authors in earth sciences (Mikki 2010); and 33 versus 21 % in an analysis of citations to an informetrics book (Bar-Ilan 2010). In their analysis of 10,000 citations to the works of 25 library and information science faculty members, Meho and Yang (2007) reported 48 % unique citations in GS versus 24 % for the union of WoS and

Scopus. De Groote and Raszewski (2012) investigated citations to 30 nursing articles in WoS, Scopus, and GS, and reported that 1,312 of the 3,497 GS citations were unique, versus 93 of 1,406 citations in WoS.

Aim of this article

Much of the data regarding the coverage of the scientific literature by WoS and GS dates back to several years ago and may therefore need updating. Furthermore, although longitudinal studies on GS's citation counts have been published (e.g., Chen 2010; Harzing 2013a, b), comprehensive longitudinal trends could not be inferred from the existing literature. In particular, no distinction has so far been made between retroactive and actual growth. The former relates to changes in the scope of the citation service independently from the production of new citations, whereas the latter represents the production of new citations as a part of the natural growth of science. Retroactive additions to WoS, for example, are the 2009 extension of its Social Sciences Index backfiles from 1956 to 1900 and the incorporation of ISI Proceedings in 2008. The retroactive growth of GS, on the other hand, is dependent on whether publishers allow crawlers on their websites. It has been reported that Elsevier and the American Chemical Society did not grant access to GS until mid-2007 and 2011, respectively (Burright 2006; Kousha and Thelwall 2008; Neuhaus and Daniel 2008; Vine 2006). Retroactive growth of GS also depends on the effectiveness of GS's crawlers and parsers, and on the number of publications put online (cf. digitalisation of paper archives). In both citation services, retroactive growth of citations to a publication may also occur because of error corrections, such as unification of multiple records corresponding to the same publication.

The aim of this article is to compare the retroactive and actual growth of GS versus WoS. We conducted a longitudinal analysis of citation counts in WoS and GS for articles from diverse research fields for which citation data from mid-2005 were available (i.e., half-to-1 year after the inception of GS on 18 November 2004). Furthermore, a citation-by-citation analysis was carried out to estimate the relative coverage of citations by both services as a function of publication year and type of the citing work.

Method

We chose two types of articles for analysis: (a) highly cited classics, to cover a large number of records and (b) articles spanning a variety of fields, to provide a representative image of the literature. For each article, retroactive and actual growth were estimated and compared between WoS and GS. One of the classic articles was used for a citation-by-citation analysis.

Analysis of citations to a classic article (Garfield 1955) for assessing retroactive and actual growth

A citation analysis was conducted for the classic article by Garfield (1955): "Citation indexes for science: a new dimension in documentation through association of ideas" in Science, for which Jacsó (2005b) in April/May 2005 documented yearly citation data. We extracted the yearly number of citations from both citation services and compared those with the data provided by Jacsó to estimate retroactive and actual growth of WoS and GS citations. Manual editing was done to add/correct the publication year when it was absent or different between WoS and GS. Retroactive growth was calculated as: 100 % * (current number of citations up to 1 May 2005 − number of citations in Jacsó)/(number of citations in Jacsó). Actual growth was calculated as: 100 % * (current number of citations − current number of citations up to 1 May 2005)/(current number of citations up to 1 May 2005). We approximated citations up to 1 May 2005 through linear interpolation between citations up to 2004 and up to 2005). The citations to a reprint appeared in the International Journal of Epidemiology in 2006 (Garfield 2006) were also included in the analysis. Data were extracted on 5 April 2013.

Analysis of citations to a highly-cited classic article (Lowry et al. 1951) for assessing retroactive and actual growth

A citation analysis was conducted for Lowry et al. (1951): "Protein measurement with the Folin phenol reagent" in the Journal of Biological Chemistry, which is claimed to be the most cited article in science (Garfield 1990; Kresge et al. 2005), with citations representing 0.5 % (~300 k/60 M) of all publications in WoS. The yearly numbers of citations were extracted from WoS and GS and compared to the citation counts retrieved for various periods by using Google's general search and Google Blog Search to assess retroactive and actual growth of each citation service. Data were extracted on 8 April 2013.

Analysis of citations to articles from diverse research fields for assessing retroactive and active growth

To assess growth of WoS and GS in diverse fields, we used citation data of articles originally analysed by Pauly and Stergiou (2005). In September 2005, these authors provided citations counts in WoS and GS for 99 articles across 11 research disciplines, plus 15 highly-cited articles earlier analysed by Garfield (1984). From these 114 articles, we selected the articles for which Pauly and Stergiou reported 50 or more citations in at least one of the two citation services and which belonged to the following seven disciplines as defined in Pauly and Stergiou: chemistry, physics, mathematics, molecular biology, psychology, computer sciences, and economy. The articles from the category with the highly-cited works were also selected. The other categories (i.e., ecology, fisheries, oceanography, and geosciences) were not selected because we believed they fit Pauly and Stergiou's personal research interest and may not be sufficiently diverse or representative of science in general.

We excluded two articles (Bradford 1976; Laemmli 1970) for which Pauly and Stergiou reported 65,535 WoS citations (that is the maximum number WoS used to yield; Jacsó 2006), two articles (Einstein 1936; Mancini et al. 1965) that are not indexed (anymore) in WoS, one article (Noyori 1992) for which GS wrongly merges citations with the citations of a book by the same author, one article (Bandura 2001) for which Pauly and Stergiou possibly included citations to a 1994 book chapter (see Jacsó 2006), Lowry et al. (1951) for which Pauly and Stergiou used an outdated citation count provided by Garfield (1984), and Sen (1974) because a comparison of the number of WoS and GS citations in Pauly and Stergiou with the current number of WoS citations up to 2005 showed a large decline which we could not explain. These exclusions left 56 articles for analysis.

For each of the 56 selected articles, the number of citations in WoS and GS was extracted up to April 2013 and up to 13 September 2005 (we approximated citations up to 13 September 2005 through linear interpolation between citations up to 2004 and up to 2005). Retroactive growth was calculated as: 100 % * (current number of citations up to

13 September 2005 − number of citations in Pauly and Stergiou)/(number of citations in Pauly and Stergiou). Actual growth was calculated as: 100 % * (current number of citations − current number of citations up to 13 September 2005)/(current number of citations up to 13 September 2005). Data were extracted between 2 and 14 April 2013.

Citation-by-citation analysis for a classic article (Garfield 1955)

The citations to Garfield (1955) and to its 2006 reprint were manually compared between WoS and GS to identify which of those were unique in one of the citation services and which were common. Each citation was coded according to one of the following document types: journal article, conference paper, book/book chapter, thesis, report, other, or unknown. Common citations between the citation services were defined as: (a) matches between identical publications (e.g., same publication outlet and same volume/issue/page numbers in that outlet), (b) matches between a publication in WoS and the e-print of the publication in GS, or (c) matches between English and non-English titles, as long as those referred to the same publication. Publications with the same authors and titles but different outlets (e.g., a book chapter and a conference article) were considered as separate items. Duplicates were identified in the same manner as common citations, with the difference that whereas common citations were defined between citations services, duplicates were defined within each citation service. When accessible, the reference lists of the unique citations were checked to verify whether Garfield (1955) was indeed cited or whether these citations were false positives, meaning that these records were not citing Garfield (1955). After excluding the duplicates and false positives, the percentages of unique citations in WoS and GS with respect to the union of WoS and GS were calculated for all types of documents combined, and for journals, conferences, and books/book chapters separately. Data were extracted on 5 April 2013.

Results

Analysis of citations to a classic article (Garfield 1955) for assessing retroactive and actual growth

WoS provided a publication year for all citations. GS did not provide a publication year in 108 citations[1] and yielded an incorrect publication year in five citations. We manually retrieved the publication year for these 113 citations. Garfield (1955) yielded 607 citations in WoS and 1,231 citations in GS out of which 228 and 312, respectively, were published in the period 1955–mid-2005 (cf. 215 vs. 95 in Jacsó 2005b).

Figure 1 shows the yearly number of citations as of April 2013 and the number of citations reported by Jacsó (2005c). The corresponding retroactive growth of WoS and GS was 6 and 228 %, respectively, whereas the actual growth was 166 and 295 %. Retroactive growth of GS was particularly large for old citations: in Jacsó (2005b), only 11 citations from 1994 or older were reported, whereas in April 2013, the number of citations up to 1994 was 115, corresponding to a retroactive growth of 945 %.

[1] Note, however, that searching on publication year reveals that GS does have knowledge about the publication year of these articles, possibly because GS keeps track of when it has first retrieved the document.

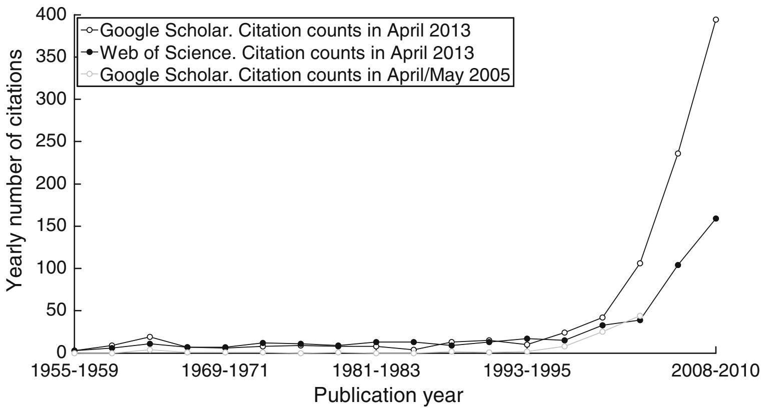

Fig. 1 Yearly number of citations to Garfield (1955) in Web of Science and Google Scholar as reported in Jacsó (2005b) and as extracted by us on 5 April 2013. Data are reported in 3-year bands, except for the first band. April/May 2005 citation counts for Web of Science are not shown, because these numbers were almost identical to the April 2013 counts

Analysis of citations to a highly-cited classic article (Lowry et al. 1951) for assessing retroactive and actual growth

Table 1 shows the cumulative number of citations to Lowry et al. (1951) in WoS for various periods as retrieved from Internet sources, the cumulative numbers attributed to the same periods as of April 2013, and the corresponding percentages of retroactive growth. It can be seen that WoS showed no substantial retroactive growth for any of the investigated periods. For example, the number of citations to Lowry et al. in WoS up to 1988 today (186,456) is virtually the same as it was in 1988 (187,652).

GS showed a large retroactive growth: 987 % for citation counts up to 7 December 2004 (from 18,953 to 205,948, the latter number of citations linearly interpolated between the years 2004 and 2005) and 520 % for citations up to 13 September 2005 (from 33,797 to 209,397, the latter number linearly interpolated between the years 2005 and 2006). GS also showed larger actual growth compared to WoS: 19 versus 6 % since July 2005. Figure 2 illustrates the cumulative number of GS citations to Lowry et al. (1951).

Despite the growth of GS, as of 8 April 2013, WoS still outnumbers GS in the total number of citations (299,989 vs. 247,606). Figure 3 shows that the yearly citation counts to Lowry et al. in both services peaked in the 1980s, that the yearly citations in WoS dropped below the level of GS for the years since 2000, and that since 1993 GS yields a fairly constant yearly number of citations of about 5,000.

Analysis of citations to articles from diverse research fields for assessing retroactive and active growth

Table 2 shows the citation counts for three instances: according to Pauly and Stergiou (2005), according to our analysis up to September 2005, and according to our analysis up to April 2013. The percentages of retroactive and actual growth are also provided.

The mean number of citations across the 56 articles in Pauly and Stergiou (2005) was 2,842 for WoS and 830 for GS. WoS yielded more citations than GS in 39 out of 56 articles. Of the 17 articles for which GS citations exceeded WoS, 9 belonged to economics

Table 1 Number of citations to Lowry et al. (1951) in Web of Science (WoS) as retrieved from Internet sources, corresponding number of citations on 8 April 2013, and percentages of retroactive and actual growth

Period for which citations were counted	Past citation counts	Citation counts on 8 April 2013	% retroactive growth	% actual growth	Source
1961–1972	29,665	30,120	1.5	n/a	Garfield (1974)
In 1988	9,750	9,719	−0.3	n/a	Garfield (1990)
Up to late 1988	187,652	186,456	−0.6	60.9	Garfield (1990)
Up to January 2004	275,669	278,226	0.9	7.8	Kresge et al. (2005)
Up to July 2005	293,328	283,027	−3.5	6.0	Garfield (2005)
Up to January 2006	282,778	284,406	0.6	5.5	Web of Knowledge (2006)

Actual growth was estimated with respect to the total number of citations on 8 April 2013 (299,989)

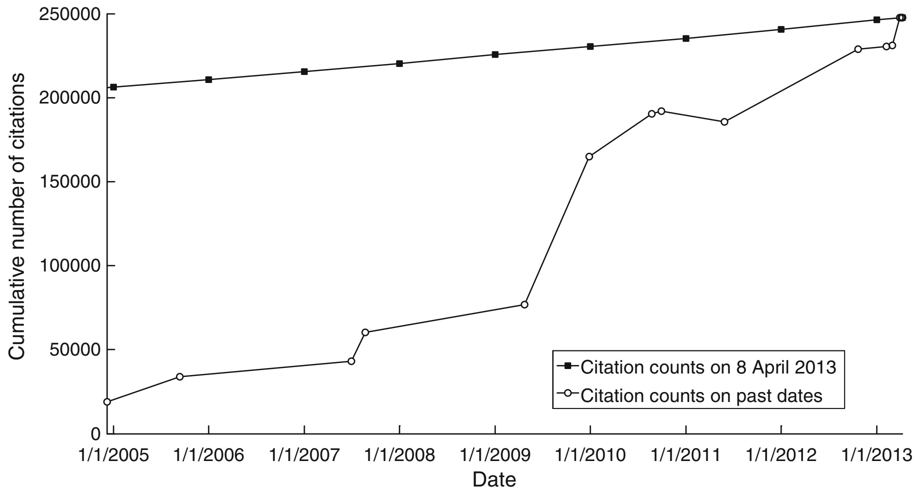

Fig. 2 Cumulative number of Google Scholar citations to Lowry et al. (1951) as retrieved from Internet sources, and cumulative number of citations attritubed to the article on 8 April 2013. Citation counts were retrieved from the following sources: 7-Dec-04, http://schoogle.blogspot.nl/2004/12/quantum-sufficit.html (18,953); 13-Sep-05, Pauly and Stergiou (2005) (33,797); 1-Jul-07, Sharma (2008). Text Book of Bioinformatics. Rastogi Publications. (43,044); 24-Aug-07, https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=lis-medical;18c9b24f.0708 (60,300); 24-Apr-09, http://www.ptt.cc/bbs/PhD/M.1240565836.A.851.html (76,925); 28-Dec-09, http://hi.baidu.com/liujtm/item/896faefdeee35d5dc9f337fb (165,038); 24-Aug-10, http://francisthemulenews.wordpress.com/2010/08/24/el-articulo-cientifico-mas-citado-de-toda-la-historia/ (190,239); 29-Sep-10, http://ipv6.weiming.info/zhuti/Biology/31415349/ (192,100); 29-May-11, http://saypeople.com/2011/05/29/5-very-highly-cited-research-papers-of-all-time/#axzz2Prr4gyh5 (185,501); 21-Oct-12, https://plus.google.com/105232907392515443926/posts/ej1sRjSp4gp (228,752); 7-Feb-13, http://blog.chembark.com/2013/02/07/a-highly-cited-paper/ (230,390); 1-Mar-13, http://www.researchgate.net/post/Value_of_Citations_A_false_indicator_of_creativity_or_a_true_indicator_of_popularity (231,213); 29-Mar-13, own observation in Google Scholar (247,480); 5-Apr-13, own observation in Google Scholar (247,561); 8-Apr-13, own observation in Google Scholar (247,606)

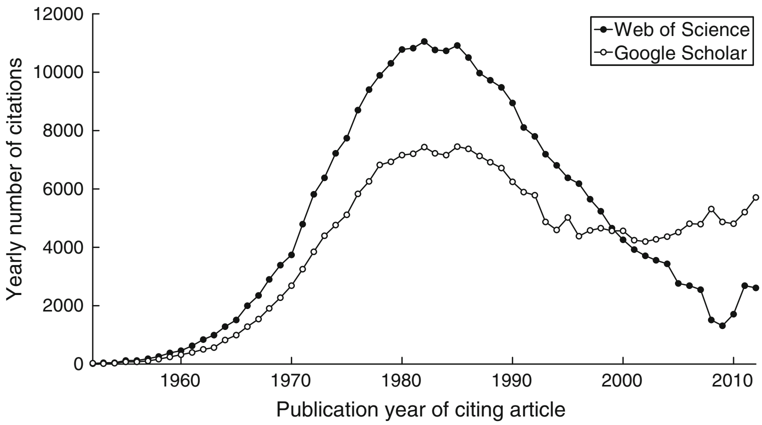

Fig. 3 Yearly number of citations in Web of Science and Google Scholar to Lowry et al. (1951), measured on 8 April 2013

and computer sciences. WoS outnumbered GS in all chemistry and psychology articles and in all but one of the biology articles (Table 3).

The mean numbers of citations in our analysis up to 2005 were 3,270 and 3,135 for WoS and GS, respectively. The articles with WoS citations outnumbering GS decreased from 39 in Pauly and Stergiou to 14. Retroactive growth was considerably larger for GS than for WoS (medians: 169.9 vs. 2.2 %, respectively). GS's largest retroactive growth occurred for articles that were lowly cited by GS in 2005 as compared to WoS, such as chemistry articles (Table 3). The Spearman correlation between GS's retroactive growth percentage and ln(GS/WoS) citations in Pauly and Stergiou was −0.75 ($p < 0.001$, $N = 56$). However, for citation counts up to 2005, WoS still outnumbers GS for 10 of the 13 chemistry articles.

The mean numbers of citations up to 2013 were 3,935 and 4,615 in WoS and GS, respectively. WoS yielded a higher number of citations than GS for a portion of chemistry (6 out of 13) and biology (2 out of 11) articles (Table 3). The articles with more citations in WoS than in GS were older (mean publication year = 1956.1, $SD = 14.0$, $N = 8$) than the articles for which GS outnumbered WoS (mean publication year = 1986.1, $SD = 14.3$, $N = 48$). The medians of the actual growth percentages were similar for both citation services (54.2 and 41.0 % for GS and WoS, respectively).

Citation-by-citation analysis for a classic article (Garfield 1955)

For both citation services, the majority of the citations to Garfield (1955) were from journal articles (90.1 % in WoS vs. 69.6 % in GS), followed by conference proceedings (8.7 vs. 10.6 %), and books or book chapters (1.2 vs. 5.4 %) (Table 4). 14.3 % of GS citations came from types of works not eligible for inclusion in WoS. GS yielded 64 duplicates and 11 false positives. WoS did not yield duplicates and yielded one false positive.

Figure 4 shows the yearly number of unique citations in WoS and GS as well as the intersection and union of both services. The mean yearly number of citations unique in WoS increased from 1.93 for the period 1955–1994 to 4.00 for 1995–2013, the corresponding numbers for GS being 1.53 and 33.79.

Table 2 Web of Science (WoS) and Google Scholar (GS) citations to 56 articles originally analysed by Pauly and Stergiou (2005). (Color figure online)

	Up to September 2005 (original)			Up to September 2005 (our analysis)			Up to April 2013 (our analysis)			% Retroactive growth		% Actual growth	
	Number of citations		Difference	Number of citations		Difference	Number of citations		Difference				
	WoS	GS	ln(GS/WoS)	WoS	GS	ln(GS/WoS)	WoS	GS	ln(GS/WoS)	WoS	GS	WoS	GS
1 Fiske 1925 Chemistry	12878	1529	-2.13	21290	16060	-0.28	22481	18272	-0.21	65.3%	950.4%	5.6%	13.8%
2 Noyori 2002 Chemistry	202	29	-1.94	203	247	0.20	916	1180	0.25	0.4%	753.0%	351.4%	377.0%
3 Duncan 1955 Biostatistics	11124	1465	-2.03	13490	11015	-0.20	14699	15905	0.08	21.3%	651.9%	9.0%	44.4%
4 Singer 1972 Biology	4400	672	-1.88	4440	5014	0.12	5413	7047	0.26	0.9%	646.1%	21.9%	40.5%
5 Scatchard 1949 Chemistry	21973	2307	-2.25	22380	16761	-0.29	23290	18254	-0.24	1.9%	626.5%	4.1%	8.9%
6 Luft 1961 Biology	7926	1340	-1.78	11402	9367	-0.20	11595	9853	-0.16	43.9%	599.1%	1.7%	5.2%
7 Ajie 1990 Chemistry	694	97	-1.97	705	660	-0.07	784	785	0.00	1.6%	580.8%	11.2%	18.9%
8 Bartlett 1959 Chemistry	11807	1513	-2.05	12659	10300	-0.21	13416	12472	-0.07	7.2%	580.8%	6.0%	21.1%
9 Myers 1984 Psychology	1153	217	-1.67	1155	1311	0.13	1234	1547	0.23	0.2%	504.1%	6.8%	18.0%
10 Diederich 1994 Chemistry	94	25	-1.32	93	151	0.49	144	242	0.52	-1.2%	503.5%	55.0%	60.4%
11 Boyd 1984 Psychology	405	73	-1.71	404	432	0.07	436	543	0.22	-0.3%	492.2%	8.0%	25.6%
12 Folch 1957 Chemistry	31378	5468	-1.75	34697	29880	-0.15	42754	42380	-0.01	10.6%	446.5%	23.2%	41.8%
13 Karno 1987 Psychology	138	38	-1.29	139	203	0.38	167	270	0.48	1.1%	435.5%	19.7%	32.7%
14 Diederich 1992 Chemistry	249	46	-1.69	252	239	-0.05	297	310	0.04	1.2%	419.5%	17.9%	29.7%
15 Bandura 1982 Psychology	2274	938	-0.89	2310	4799	0.73	3393	9435	1.02	1.6%	411.6%	46.9%	96.6%
16 Reynolds 1963 Biology	19153	3646	-1.66	22546	18038	-0.22	24137	20521	-0.16	17.7%	394.7%	7.1%	13.8%
17 Bandura 1989 Psychology	490	234	-0.74	493	1141	0.84	991	3180	1.17	0.5%	387.4%	101.2%	178.8%
18 Reed 2000 Chemistry	190	48	-1.38	201	204	0.01	436	496	0.13	6.0%	324.5%	116.5%	143.4%
19 Arrow 1962 Economics	816	1272	0.44	896	4378	1.59	1727	8840	1.63	9.8%	244.2%	92.7%	101.9%
20 Bokhari 1981b Computer Sciences	69	49	-0.34	72	161	0.81	84	215	0.94	3.9%	228.2%	17.2%	33.7%
21 Sen 1976 Economics	272	338	0.22	241	1104	1.52	449	2226	1.60	-11.3%	226.6%	86.2%	101.7%
22 Arrow 1994 Economics	34	79	0.84	37	249	1.91	98	496	1.62	8.2%	214.9%	166.3%	99.4%
23 Bourke 1992 Psychology	82	30	-1.01	80	93	0.15	119	185	0.44	-2.2%	211.6%	48.4%	97.9%
24 Einstein 1949 Physics	71	54	-0.27	149	166	0.10	161	203	0.23	110.1%	206.6%	7.9%	22.6%
25 Jessop 1999 Chemistry	327	102	-1.16	332	307	-0.08	609	745	0.20	1.7%	200.7%	83.2%	142.9%
26 Adleri 1989 Mathematics	127	124	-0.02	128	370	1.06	153	494	1.17	0.9%	198.6%	19.4%	33.4%
27 Cromer 1970 Chemistry	6138	1329	-1.53	6162	3870	-0.47	6297	4088	-0.43	0.4%	191.2%	2.2%	5.6%
28 Arrow 1963 Economics	795	625	-0.24	842	1785	0.75	1377	6266	1.52	5.9%	185.5%	63.6%	251.1%
29 Bagby 1986 Psychology	122	42	-1.07	121	107	-0.13	140	156	0.11	-0.7%	154.3%	15.6%	46.1%
30 Bokhari 1981a Computer Sciences	169	217	0.25	175	529	1.10	215	684	1.16	3.8%	144.0%	22.5%	29.2%

Table 2 continued

	Up to September 2005 (original)			Up to September 2005 (our analysis)			Up to April 2013 (our analysis)						
	Number of citations		Difference	Number of citations		Difference	Number of citations		Difference	% Retroactive growth		% Actual growth	
	WoS	GS	ln(GS/WoS)	WoS	GS	ln(GS/WoS)	WoS	GS	ln(GS/WoS)	WoS	GS	WoS	GS
31 Karmarkar 1984 Mathematics	1021	1054	0.03	1018	2481	0.89	1486	4212	1.04	-0.3%	135.4%	45.9%	69.8%
32 Bagby 1994 Psychology	392	260	-0.41	402	539	0.29	1069	1786	0.51	2.7%	107.3%	165.6%	231.4%
33 Einstein 1935 Physics	2484	2473	0.00	2778	5124	0.61	5325	10922	0.72	11.8%	107.2%	91.7%	113.2%
34 Dynkin 1991 Mathematics	79	75	-0.05	79	147	0.63	99	209	0.75	-0.3%	96.7%	25.6%	41.7%
35 Saitoh 1994 Biology	224	130	-0.54	229	239	0.04	285	336	0.16	2.3%	83.5%	24.4%	40.8%
36 Reed 1998 Chemistry	93	50	-0.62	91	91	0.00	185	203	0.09	-1.8%	82.6%	102.7%	122.4%
37 Hawking 1975 Physics	2282	1843	-0.21	2415	3364	0.33	4334	6722	0.44	5.8%	82.5%	79.4%	99.8%
38 Barron 1993 Computer Sciences	361	553	0.43	465	934	0.70	705	1545	0.78	28.8%	68.9%	51.6%	65.4%
39 Weinberg 1967 Physics	4776	4445	-0.07	4835	7445	0.43	5342	10222	0.65	1.2%	67.5%	10.5%	37.3%
40 Bennett 1993 Physics	1758	1783	0.01	1942	2968	0.42	5205	8517	0.49	10.5%	66.4%	168.0%	187.0%
41 Tsukita 1994 Biology	440	281	-0.45	442	448	0.01	573	663	0.15	0.5%	59.5%	29.6%	48.0%
42 Bennett 2000 Physics	371	372	0.00	412	580	0.34	1084	1629	0.41	11.0%	56.0%	163.1%	180.7%
43 Ferrari 1994 Computer Sciences	19	157	2.11	22	244	2.41	23	275	2.48	15.8%	55.5%	4.5%	12.7%
44 Ferrari 1990 Computer Sciences	86	260	1.11	72	401	1.72	80	480	1.79	-16.6%	54.2%	11.6%	19.7%
45 Venter 2001 Genomics	3230	3338	0.03	3391	5087	0.41	6334	11020	0.55	5.0%	52.4%	86.8%	116.6%
46 Kamath 1992 Mathematics	29	54	0.62	31	82	0.97	35	104	1.09	7.2%	51.5%	12.6%	27.1%
47 Bennett 1996 Physics	886	923	0.04	966	1371	0.35	2232	3367	0.41	9.0%	48.5%	131.1%	145.7%
48 Tsukita 2000 Biology	115	97	-0.17	117	137	0.15	282	362	0.25	2.1%	41.2%	140.3%	164.3%
49 Hendzel 1997 Biology	360	288	-0.22	361	385	0.06	909	1070	0.16	0.1%	33.5%	152.1%	178.3%
50 Tsukita 2001 Biology	303	267	-0.13	307	356	0.15	1050	1391	0.28	1.2%	33.2%	242.4%	291.1%
51 Barron 1986 Computer Sciences	55	65	0.17	63	85	0.29	113	196	0.55	15.4%	30.6%	78.0%	130.9%
52 Chadee 1999 Biology	97	84	-0.14	96	106	0.10	183	218	0.18	-1.2%	26.2%	90.9%	105.7%
53 Barron 1991 Computer Sciences	118	177	0.41	128	217	0.53	191	425	0.80	8.5%	22.7%	49.1%	95.7%
54 Cromer 1965 Chemistry	2640	1449	-0.60	2918	1689	-0.55	2972	1767	-0.52	10.5%	16.6%	1.9%	4.6%
55 Hartle 1983 Physics	1137	1691	0.40	1155	1726	0.40	1572	2553	0.48	1.6%	2.1%	36.1%	47.9%
56 Hawking 1983 Physics	257	355	0.32	280	357	0.24	693	965	0.33	9.0%	0.6%	147.3%	170.1%
Mean	**2842**	**830**	**-0.57**	**3270**	**3135**	**0.38**	**3935**	**4615**	**0.51**	**8.0%**	**242.8%**	**63.9%**	**87.2%**
Median	**366**	**263.5**	**-0.31**	**403**	**534**	**0.27**	**847**	**1286**	**0.41**	**2.2%**	**169.9%**	**41.0%**	**54.2%**

Red bars extending toward the left indicate that WoS yielded a higher number of citations than GS and blue bars extending toward the right indicate that the number of citations in GS was larger than that in WoS. Bars with percentages higher than 500 are of maximum length. Results are sorted in descending order of retroactive growth of Google Scholar.

Table 3 Comparison of Web of Science (WoS) and Google Scholar (GS) citations to 56 articles originally analysed by Pauly and Stergiou (2005), clustered per research field

		Publication year		Number of studies GS>WoS			Mean retroactive growth		Mean actual growth	
	Number of studies	Mean	SD	Up to September 2005 (original)	Up to September 2005 (our analysis)	Up to April 2013 (our analysis)	WoS	GS	WoS	GS
Chemistry	13	1976.9	24.4	0	3	7	8.0%	436.7%	60.1%	76.2%
Psychology	8	1987.3	4.2	0	7	8	0.4%	338.0%	51.5%	90.9%
Biology	11	1985.2	18.4	1	8	9	8.5%	238.3%	73.3%	95.3%
Economics	4	1973.8	14.9	3	4	4	3.2%	217.8%	102.2%	138.5%
Mathematics	4	1989.0	3.6	2	4	4	1.9%	120.5%	25.9%	43.0%
Computer Sciences	7	1988.0	5.4	6	7	7	8.5%	86.3%	33.5%	55.3%
Physics	9	1975.7	21.9	5	9	9	18.9%	70.8%	92.8%	111.6%

Duncan (1955; Biostatistics) and Venter (2001; Genomics) are clustered under Biology

Table 4 Number of citations on 5 April 2013 to Garfield (1955) for Web of Science (WoS) and Google Scholar (GS) as a function of document type

	WoS	GS
All types (incl. duplicates and false positives)	607	1,231
Journals	546 (90.1)	805 (69.6)
Conferences	53 (8.7)	123 (10.6)
Books or book chapters	7 (1.2)	63 (5.4)
Theses	0 (0)	75 (6.5)
Reports	0 (0)	13 (1.1)
Other	0 (0)	43 (3.7)
Unknown	0 (0)	34 (2.9)
Duplicates	0	64
False positives	1	11

The category “Other” includes working papers, book reviews, unpublished manuscripts, proposals, and conference presentations. Percentages in parentheses. Percentages are defined with respect to the total number of citations in the citation service excluding duplicates and false positives

Table 5 shows the number and percentages of unique citations to Garfield (1955) retrieved from WoS and GS. The percentage of unique citations in WoS with respect to the union of WoS and GS is lower for the more recent decades as compared to earlier decades, whereas the opposite is noted for GS. Specifically, across document types, 41.4 and 32.8 % of the citations up to 1994 were unique in WoS and GS, respectively, whereas for the years 1995–2013, GS largely covered WoS, with only 6.8 % of the citations being unique in WoS. For 1995–2013, 57.2 % of the citations were unique in GS.

The same trends hold for journal citations. For 1955–1994, 50.3 % of the journal citations were unique in WoS versus 18.4 % in GS, with the corresponding numbers for 1995–2013 being 6.6 and 45.7 %, respectively. The percentages of unique conference proceedings were relatively stable over the years for both citation services (WoS: 16.7 vs. 14.3 %; GS: 72.2 vs. 61.9 %).

Discussion

We conducted a longitudinal analysis of citation counts to two classic articles and 56 articles published in a variety of research fields, with the aim to estimate retroactive and actual growth percentages of GS versus WoS. Additionally, we carried out a citation-by-

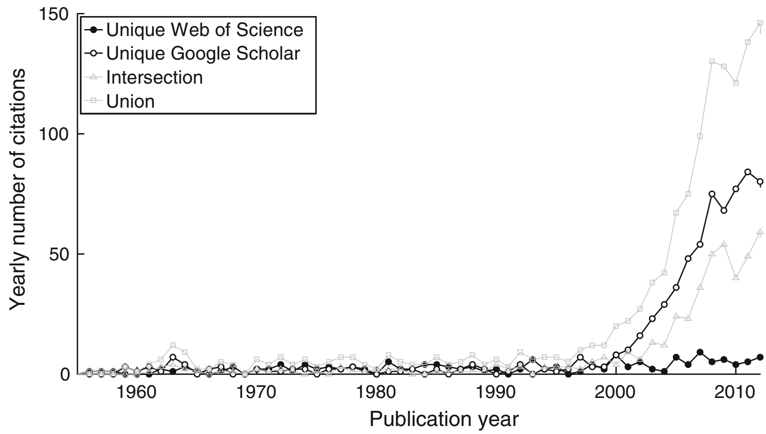

Fig. 4 Yearly unique, common, and union of citations in Web of Science and Google Scholar to Garfield (1955)

Table 5 Number of citations on 5 April 2013 to Garfield's (1955) article for Web of Science (WoS) and Google Scholar (GS), per document type

All types	Number of citations unique in WoS	Number of citations unique in GS	Intersection of WoS and GS	% citations unique in WoS	% citations unique in GS
1955–1994	77	61	48	41.4	32.8
1995–2013	76	642	405	6.8	57.2
Journals					
1955–1994	74	27	46	50.3	18.4
1995–2013	52	358	374	6.6	45.7
Conferences					
1955–1994	3	13	2	16.7	72.2
1995–2013	18	78	30	14.3	61.9
Books or book chapters					
1955–1994	0	14	0	0.0	100.0
1995–2013	6	48	1	10.9	87.3

Percentages of unique citations were calculated as the number of citations unique in one citation service over the union of citations in WoS and GS

citation analysis, to elucidate temporal changes in the relative coverage of citations by the two citation services.

In September 2005, WoS yielded more citations than GS for about two-thirds of the investigated articles. However, GS has demonstrated a striking retroactive growth: for citations up to mid-2005, the number of citations per article increased on average 243 %, and for one article reached 950 %. In contrast, the retroactive growth of WoS citations was smaller than 10 % for three-fourth of the investigated articles. Retroactive growth of GS citations was the largest for articles for which GS used to yield a low number of citations relative to WoS in mid-2005. These were in particular citations to chemistry, psychology, and biology articles. The retroactive growth of GS citations to the two classic articles

(Garfield 1955 and Lowry et al. 1951) since mid-2005 was 228 and 520 %, respectively. GS showed a larger actual growth than WoS since 2005 for the two classic articles, but the mean actual growths were quite comparable between both citation services for the diverse articles from Pauly and Stergiou (2005; means of 64 % in WoS vs. 87 % in GS). The consequence of the differential growth between WoS and GS is that, except for a number of relatively old articles in chemistry and biology, the number of citations up to 2013 in GS is now larger than in WoS for the articles investigated.

The citation-by-citation analysis of Garfield (1955) showed that about two-fifth of the citations up to 1994 are unique in WoS (vs. one-third being unique in GS). For more recent citations (1995–2013), WoS is covered by GS almost entirely, with only 6.8 % of the citations in WoS not being retrievable by GS. Our analyses indicate that WoS and GS recover different portions of the literature: The unique citations of WoS are typically documents before the digital age and conference proceedings not available online, whereas GS retrieves diverse works not eligible for inclusion in WoS.

Note that in our analyses we used WoS instead of Web of Knowledge (WoK) for the sake of comparison with past studies and to make the longitudinal analysis possible. WoK includes more non-English documents than WoS (cf. Chinese Science Citation Database) and has more comprehensive coverage of engineering sciences. However, our analysis showed that 99.4 % of the WoK citations to Lowry et al. (1951) were also retrievable by WoS. Note also that we did not include Scopus because, although an important citation service, we could not recover satisfactory citation data for conducting a longitudinal study.

Quantity versus quality of citations

In our study, we quantified the retroactive and actual growth of WoS and GS based on citations counts to a variety of articles. Of course, quantity does not imply quality. GS includes everything that resembles scholarly work, based on automatic format inspection rather than content inspection, thereby risking inflation of its record. WoS uses a selective inclusion procedure, with the aim to provide a safeguard against low-quality or low-impact material being indexed. It can be argued that WoS's quality control mechanism is especially important today in an era where some scientists seem to suffer from "writing incontinence" (Ioannidis et al. 2010) and where "predatory" (Beall 2010) publishers are on the rise. Using a signal-to-noise metaphor, proponents of WoS may argue that WoS acts as a filter from the high-level of "noise" produced nowadays. Proponents of GS, on the other hand, may argue that GS is the ultimate embodiment of Kilgour's "100 percent availability of information" thesis (Pomerantz 2006). Because GS's inclusion process is automated, it may allow for efficient coverage of publication outlets, free from bias of human decision makers.

Threats for WoS

A potential threat for WoS is that its selective inclusion policy may fail to keep up with rapid developments within the scientific enterprise and knowledge production on the Internet. As mentioned in the introduction, the coverage of peer-reviewed journal articles by WoS is declining as compared to other databases, and is particularly low for research fields with the highest growth rates (Larsen and Von Ins 2010). In an analysis of the publication output of Australian universities, Butler and Visser (2006) reported that publications in chemistry, biology, physics, and health sciences were primarily (between 69 and 85 %) in journals indexed by WoS, whereas the corresponding percentages in social sciences, management, and education ranged between 4 and 19 %. In a comparison of WoS and GS coverage of the work body of renowned scientists in philosophy, computer sciences, and economics, we found that philosopher Michel Foucault counts 1,706 citations in WoS versus 365,188 in GS; computer scientist Herbert Simon counts 21,456 versus 202,862; computer scientist Jeffrey Ullman counts 3,424 versus 81,043; economist Oliver E. Williamson counts 9,873 versus 139,276; and economist Kenneth Arrow counts 10,139 versus 112,505 citations (data extracted by the authors on 27 April 2013). In summary, by sacrificing sensitivity for specificity, WoS risks missing high-impact research.

A second threat for WoS is that in the future, GS may cover all works covered by WoS. We found that for the period 1995–2013, 6.8 % of the citations to Garfield (1955) were unique in WoS, indicating that a very large share of works indexed in WoS is now also retrievable by GS. In line with this observation, based on an analysis of 29 systematic reviews in the medical domain, Gehanno et al. (2013) recently concluded that: "The coverage of GS for the studies included in the systematic reviews is 100 %. If the authors of the 29 systematic reviews had used only GS, no reference would have been missed". GS's coverage of WoS could in principle become *complete* in which case WoS could become a subset of GS that could be selected via a GS option "Select WoS-indexed journals and conferences only".[2] Together with its full-text search and its searching of the grey literature, it is possible that GS becomes the primary literature source for meta-analyses and systematic reviews.

Weaknesses of GS

A weakness of GS is that it has a broad definition of "scholarly", is vulnerable to malign citation manipulation (Beel and Gipp 2010; Labbe 2010; López-Cózar et al. 2012), and yields inaccurate meta-data. In response to a series of critiques earlier raised by Jacsó (2005a, c) about errors in meta-data (e.g., "F Password" as prolific author), Harzing (2008) was "unable to reproduce most of the Google Scholar failures detailed in his paper", suggesting that GS has rectified these failures. Our citation-by-citation analysis, however, showed that GS suffers from a large number of errors including false positive citations (1 % of the GS citations), duplicates (5 % of the GS citations), and lack of publication year (9 % of the GS citations). Errors in GS were particularly prevalent for document types not included in WoS. For example, the publication date was missing in GS for 4 % (29 of 805) of the journal publications, 15 % (11 of 75) of the theses, and 41 % (14 of 34) of the unknown document types. Occasional optical character recognition (OCR) errors occurred as well in GS, giving rise to missing (i.e., false negative) citations. To alleviate metadata errors, GS advises authors and webmasters how to format their online documents (Google Scholar 2013). Some missed citations were identified also in WoS due to errors in the reference lists of the citing articles. For further details about sources of error in GS and WoS, we refer to the supplementary material containing the raw data of the analyses. Another limitation of GS is that it does not provide a comprehensive export function and does not offer refine-search options (e.g., possibility to search among specific types of documents—a function that is currently available only for including or excluding patents) or searching with nested Boolean syntaxes.

[2] It is public information which journals, conferences are indexed in WoS, meaning that such a feature could be easily incorporated in GS.

Implications for evaluation of research

This study shows that GS now covers a large share of the scientific literature, suggesting that GS is an invaluable tool for conducting literature research. The findings of this study may also have certain implications for the use of GS in research evaluation exercises. The fact that GS has expanded notably over last few years and now covers most of the available literature data means that it could be used for meaningful research evaluation particularly in the disciplines that are not comprehensively covered by WoS. Furthermore, since GS services are freely accessible, it could be an attractive and transparent option for funding agencies or budget-deprived institutes (Harzing 2008). However, GS also has limitations compared to WoS for application in research evaluation. First, as pointed out above, it is possible to manipulate the information seen and quantified by GS to inflate one's citation and publication metrics. For example, it is possible to upload a fake or non-peer-reviewed document to a personal website in order to boost one's citation scores (e.g., López-Cózar et al. 2012). Second, the findings of our study show that the data provided by GS contain more duplicates and false positives as compared to WoS. Finally, GS provides no systematic tools for calculating the same type of metrics that can be calculated in WoS. Unless the scientist has created a user profile in GS, it is difficult to calculate his/her h-index, whereas self-citation rates cannot be retrieved either (cf. Couto et al. 2009 for an online tool to calculate self-citation rates in GS). Recently, GS restrained bulk processing of citation data by third party applications, further limiting the options for the use of GS in research evaluation. Illustratively, the Research Excellence Framework (REF), a system for assessing the quality of research in UK higher education institutions, intended to use GS as an additional source of citation information for researchers working in Computer Science and Informatics. However, the REF now reports that "Unfortunately, following discussions with GS, it has not been possible to agree a suitable process for bulk access to their citation information, due to arrangements that GS have in place with publishers" (Research Excellence Framework 2013).

Conclusion

In conclusion, our study showed that GS has exhibited a striking retroactive expansion, considerably increasing its coverage of scientific literature as compared to 1 year after its inception. It is possible that GS fully covers WoS in the foreseeable future. However, improved metadata, more sophisticated search functions, and a stricter control against citation manipulation are challenges for GS yet to be met.

References

Amara, N., & Landry, R. (2012). Counting citations in the field of business and management: Why use Google Scholar rather than the Web of Science. *Scientometrics, 93*(3), 553–581.

Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, *3*(7).

Bandura, A. (2001). Social cognitive theory of mass communication. *Media Psychology, 3*(3), 265–299.

Bar-Ilan, J. (2010). Citations to the "Introduction to informetrics" indexed by WOS, Scopus and Google Scholar. *Scientometrics, 82*(3), 495–506.

Bar-Ilan, J., Levene, M., & Lin, A. (2007). Some measures for comparing citation databases. *Journal of Informetrics, 1*(1), 26–34.

Bauer, K., & Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine, 11*(9).

Beall, J. (2010). "Predatory" open-access scholarly publishers. *The Charleston Advisor, 11*(4), 10–17.

Beel, J., & Gipp, B. (2010). Academic search engine spam and Google Scholar's resilience against it. *Journal of Electronic Publishing, 13*(3), 1–25.

Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A., & 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.

Bosman, J., Van Mourik, I., Rasch, M., Sieverts, E., & 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. Retrieved April 2, 2013, from http://igitur-archive.library.uu.nl/DARLIN/2006-1220-200432/UUindex.html.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry, 72*(1), 248–254.

Burright, M. (2006). Google Scholar: Science & technology. *Issues in Science & Technology Librarianship, 45*.

Butler, L., & Visser, M. S. (2006). Extending citation analysis to non-source items. *Scientometrics, 66*(2), 327–343.

Cathcart, R., & Roberts, A. (2005). Evaluating Google Scholar as a tool for information literacy. *Internet Reference Services Quarterly, 10*(3–4), 167–176.

Chen, X. (2010). Google Scholar's dramatic coverage improvement five years after debut. *Serials Review, 36*(4), 221–226.

Couto, F. M., Grego, T., Pesquita, C., & Verissimo, P. (2009). Handling self-citations using Google Scholar. *International Journal of Scientometrics, Informetrics and Bibliometrics*, *13*(1). Retrieved June 30, 2013, from https://docs.di.fc.ul.pt/jspui/handle/10455/3304.

De Groote, S. L., & Raszewski, R. (2012). Coverage of Google Scholar, Scopus, and Web of Science: A case study of the *h-index* in nursing. *Nursing Outlook, 60*(6), 391–400.

Donlan, R., & Cooke, R. (2005). Running with the devil. Accessing library-licensed full text holdings through Google Scholar. *Internet Reference Services Quarterly, 10*(3–4), 149–157.

Duncan, D. B. (1955). Multiple range and multiple F tests. *Biometrics*, *11*(1), 1–42.

Einstein, A. (1936). Lens-like action of a star by the deviation of light in the gravitational field. *Science, 84*(2188), 506–507.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal, 22*(2), 338–342.

Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics, 83*, 243–258.

García-Pérez, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of *h* indices in psychology. *Journal of the American Society for Information Science and Technology, 61*(10), 2070–2085.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science, 122*(3159), 108–111.

Garfield, E. (1974). Selecting the all-time citations classics. Here are the fifty most cited papers 1961–1972. *Current Contents, 2*, 5–8. Retrieved May 6, 2013, from http://garfield.library.upenn.edu/essays/v2p006y1974-76.pdf.

Garfield, E. (1984). The 100 most-cited papers ever and how we select citation classics. *Current Contents*, *23*, 3–9. Retrieved May 6, 2013, from http://www.garfield.library.upenn.edu/essays/v7p175y1984.pdf.

Garfield, E. (1990). The most-cited papers of all time, SCI 1945–1988. Part 1A. The SCI top 100—will the Lowry method ever be obliterated? *Current Contents*, *7*, 3–14. Retrieved May 6, 2013, from http://www.garfield.library.upenn.edu/essays/v13p045y1990.pdf.

Garfield, E. (2005). The agony and the ecstasy—the history and meaning of the journal impact factor. Chicago: International Congress on Peer Review and Biomedical Publication. Retrieved May 6, 2013, from http://garfield.library.upenn.edu/papers/jifchicago2005.pdf.

Garfield, E. (2006). Citation indexes for science: A new dimension in documentation through association of ideas. *International Journal of Epidemiology, 35*(5), 1123–1127.

Gehanno, J.-F., Rollin, L., & Darmoni, S. (2013). Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC Medical Informatics and Decision Making, 13*(1), 7.

Google Scholar (2013). Inclusion guidelines for webmasters. Retrieved April 8, 2013 from http://scholar.google.com/intl/en/scholar/inclusion.html#overview.

Harzing, A. W. (2008). Google Scholar—a new data source for citation analysis. University of Melbourne. Retrieved June 30, 2013, from http://www.harzing.com/pop_gs.htm.

Harzing, A. W. (2013a). A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics*. doi:10.1007/s11192-013-0975-y.

Harzing, A. W. (2013b). A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel Prize winners. *Scientometrics, 94*(3), 1057–1075.

Hightower, C., & Caldwell, C. (2010). Shifting sands: Science researchers on Google Scholar, Web of Science, and PubMed, with implications for library collections budgets. *Issues in Science and Technology Librarianship, 63*.

Ioannidis, J., Tatsioni, A., & Karassa, F. B. (2010). Who is afraid of reviewers' comments? Or, why anything can be published and anything can be cited. *European Journal of Clinical Investigation, 40*(4), 285–287.

Jacsó, P. (2005a). Google Scholar: The pros and the cons. *Online Information Review, 29*(2), 208–214.

Jacsó, P. (2005b). 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.

Jacsó, P. (2005c). Comparison and analysis of the citedness scores in Web of Science and Google Scholar. In *Digital libraries: Implementing strategies and sharing experiences. Lecture Notes in Computer Science*, vol. 3815 (pp. 360–369). Berlin, Heidelberg: Springer.

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

Jacsó, P. (2008). Google Scholar revisited. *Online Information Review, 32*(1), 102–114.

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

Kresge, N., Simoni, R. D., & Hill, R. L. (2005). The most highly cited paper in publishing history: Protein determination by Oliver H. Lowry. *The Journal of Biological Chemistry, 280*(28), e25–e25.

Kulkarni, A. V., Aziz, B., Shams, I., & Busse, J. W. (2009). Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *Journal of the American Medical Association, 302*(10), 1092–1096.

Labbe, C. (2010). Ike Antkare one of the great stars in the scientific firmament. *ISSI Newsletter, 6*(2), 48–52.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*(5259), 680–685.

Larsen, P. O., & Von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics, 84*(3), 575–603.

Levine-Clark, M., & Gil, E. (2009). A comparative analysis of social sciences citation tools. *Online Information Review, 33*(5), 986–996.

López-Cózar, E. D., Robinson-García, N., & Torres-Salinas, D. (2012). Manipulating Google Scholar citations and Google Scholar metrics: Simple, easy and tempting. arXiv:1212.0638.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *The Journal of Biological Chemistry, 193*(1), 265–275.

Mancini, G., Carbonara, A. O., & Heremans, J. F. (1965). Immunochemical quantitation of antigens by single radial immunodiffusion. *Immunochemistry, 2*(3), 235–254.

Mayr, P., & Walter, A. K. (2007). An exploratory study of Google Scholar. *Online Information Review, 31*(6), 814–830.

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

Meier, J. J., & Conkling, T. W. (2008). Google Scholar's coverage of the engineering literature: An empirical study. *The Journal of Academic Librarianship, 34*(3), 196–201.

Mikki, S. (2010). Comparing Google Scholar and ISI Web of Science for earth sciences. *Scientometrics, 82*(2), 321–331.

Mingers, J., & Lipitakis, E. A. (2010). Counting the citations: A comparison of Web of Science and Google Scholar in the field of business and management. *Scientometrics, 85*(2), 613–625.

Neuhaus, C., & Daniel, H. D. (2008). Data sources for performing citation analysis: An overview. *Journal of Documentation, 64*(2), 193–210.

Neuhaus, C., Neuhaus, E., Asher, A., & Wrede, C. (2006). The depth and breadth of Google Scholar: An empirical study. *Libraries and the Academy, 6*(2), 127–141.

Noyori, R. (1992). Asymmetric catalysis by chiral metal complexes. *ChemTech, 22*(6), 360–367.

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

Pomerantz, J. (2006). Google Scholar and 100 percent availability of information. *Information Technology and Libraries, 25*(2), 52–56.

Research Excellence Framework. (2013). Sub-panel 11: Citation data. Retrieved June 30, 2013, from http://www.ref.ac.uk/subguide/citationdata/googlescholar/.

Sen, A. (1974). On some debates in capital theory. *Economica: Journal of the Econometric Society, 41*(163), 328–335.

Sharma, V. (2008). *Text book of bioinformatics.* Meerut: Rastogi Publications.

Thomson Reuters. (2013a). The Thomson Reuters journal selection process. Retrieved April 4, 2013, from http://thomsonreuters.com/products_services/science/free/essays/journal_selection_process/.

Thomson Reuters. (2013b). Web of Science facts sheet. Retrieved April 29, 2013, from http://thomsonreuters.com/content/science/pdf/Web_of_Science_factsheet.pdf.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., & Beasley, E. (2001). The sequence of the human genome. *Science, 291*(5507), 1304–1351.

Vine, R. (2006). Google Scholar. *Journal of the Medical Library Association, 94*(1), 97–99.

Web of Knowledge. (2006). Retrieved April 29, 2013, from http://wokinfo.com/media/pdf/MostHighlyCitedArticles.pdf.

Wleklinski, J. M. (2005). Studying Google Scholar: Wall to wall coverage? *Online, 29*(3), 22–26.