Dispelling the Myths Behind First-author Citation Counts

Dangzhi Zhao

School of Library and Information Studies, University of Alberta
Edmonton, Alberta, Canada T6G 2J4. Email: dzhao@ualberta.ca

We conducted a full-scale evaluative citation analysis study of scholars in the XML research field to explore just how different from each other author rankings resulting from different citation counting methods actually are, and to demonstrate the capability of emerging data and tools on the Web in supporting more realistic citation counting methods. Our results contest some common arguments for the continued use of first-author citation counts in the evaluation of scholars, such as high correlations between author rankings by first-author citation counts and other citation counting methods, and high costs of using more realistic citation counting methods that are not well-supported by the ISI databases. It is argued that increasingly available digital full text research papers make it possible for citation analysis studies to go beyond what the ISI databases have directly supported and to employ more sophisticated methods.

Introduction

Various methods of allocating credit in the case of multi-authored works have been proposed for counting publications and citations (Egghe, et al, 2000; Oppenheim, 1998; Trueba & Guerrero, 2004; van Hooydonk, 1997). These different methods can result in divergent author rankings (Garfield, 1979; Gauffriau & Larsen, 2005; Lange, 2001; Lindsey, 1980; Long et al., 1980; MacRoberts & MacRoberts, 1989; Smith, 1981; Stokes & Hartley, 1989).

In theory, credit should be allocated among authors of a paper according to their contributions to the paper. In practice, however, there are few studies of scholarly communication, if any, that have used this approach, simply because it is nearly impossible to assess the relative contributions to co-authored papers based solely on the publicly available data such as the sequence of author names listed on the paper (Lindsey, 1982). Some studies give equal credit – full or 1/Nth of full credit with N as the number of co-authors – to all authors regardless of their contributions to the documents. Other studies give full credit only to the author who is considered as having contributed the most to the document. This is often the first author although other means for identifying the main author have also been suggested. Still others propose to assign a fraction of credit proportional to the position of author names in the by-line (Oppenheim, 1998; van Hooydonk, 1997; Trueba & Guerrero, 2004).

First-author counts (a.k.a. straight counts), complete counts and fractional counts are among the most well-known methods for allocating credit among multiple authors. Simply put, when a paper with N authors is cited, with first-author counts, only the number of citations of the first author of this paper increases by 1; with fractional counts, the number of citations of each of the N authors increases by 1/N; and with complete counts, full credit is given to all authors of the paper, i.e. the number of citations of each of the N authors increases by 1.

It is well-known that scientific collaboration has become commonplace, and consequently multi-authored scholarly publications have become the rule rather than the exception. It is also recognized that in some research fields the first author is not always the one who

contributes the most to the article nor is the ordering of author names in the by-line necessarily significant. For example, alphabetic ordering of author names rather than an ordering based on contributions has been found in many studies (Endersby, 1996; Lindsey, 1980; Rudd, 1977; Zhao & Logan, 2002). As a result, fractional counts have theoretically been recognized as the most preferred method for allocating credit in the case of multi-authored works (Lindsey, 1980; van Hooydonk, 1997). In practice, however, first-author counts have been the most popular method in counting citations in the evaluation of scholarly impact, although fractional counts have been frequently used in counting publications in the evaluation of scholarly productivity (Gauffriau & Larsen, 2005; Trueba & Guerrero, 2004).

In the literature, we find the following arguments or reasons for the common practice in citation analysis studies of counting first authors only.

(a) First-author counts are reliable enough in citation analysis and thus can be used as an adequate substitute for complete counts (i.e. counting all authors) because results from first-author counts are highly correlated with those from complete counts. For example, Cole & Cole (1973) found a correlation of .96 in physics, and Lange (2001) found a correlation of at least .87 in psychology in different journals during different time periods. Other examples include Porter (1977) and Long et al. (1980).

(b) The costs of collecting complete counts are prohibitive with the type of support provided by the databases developed by the Institute for Scientific Information (ISI) – the data source for most citation analysis studies to date (Long, et al., 1980). The bias associated with the use of first-author counts is not significant enough for evaluation purposes to justify the use of complete counts at this level of costs (Lange, 2001).

The ISI databases only index first authors of cited documents, which makes counting all authors in citation analysis studies very difficult, unlike counting all authors of publications that are fully indexed in most bibliographical databases. The fact that in some cases information about non-first authors of cited papers can also be found in the ISI databases does not solve this problem. The number of such papers is comparatively small as these are only those cited articles that are indexed as source papers in the ISI databases. With Garfield's Law of Concentration as its operational principle – "the bulk of the information needs in science can be satisfied by a relatively small, multidisciplinary core of journals" (Bensman, 2001, p. 1), the ISI databases will never index all of the papers cited as their source papers (Zhao & Logan, 2002). Remedies to this limitation are possible but are usually costly and have limited applicability. For example, Garfield, the inventor of the ISI databases, suggested that a complete list of publications authored and co-authored by each of the scholars being evaluated be compiled and citations to these publications be obtained from the ISI databases (Garfield, 1977) and then used in the evaluation. This is only feasible for the evaluation of a known group of scholars, and becomes hopeless in answering such questions as "who are the key players in this particular research field as seen from citations?"

(c) Partly due to these high costs, studies have not provided adequate evidence to support arguments for the use of alternative citation counting methods in real-life evaluative citation analysis studies. Most if not all studies that propose and promote these methods employed small or problematic samples of data. Some studies used very small sets of theoretically motivated artificial examples (Egghe, et al, 2000; van Hooydonk, 1997) while others used citations to a set of publications from a few journals within a few years or by a few scholars (Egghe et al., 2000; Lange, 2001; Cronin & Overfelt, 1994). Still others looked at how a set of publications within a period of time in a research field cited themselves, thus disregarding more than 90% of the references made by these

publications (Persson, 2001). In addition, most such studies using real-life data compared first-author counts with complete counts, leaving out the most preferred counting method, namely fractional counts (Cole & Cole, 1973; Lange, 2001; Persson, 2001). Large-scale real-life citation analysis studies using multiple citation counting methods including fractional counts are needed to demonstrate the advantages of the use of more realistic counting methods and the feasibility as well as the procedures and techniques of collecting the different counts.

The present study is such an attempt. Aiming to take a close look at the reasons and arguments reiterated above for the continued reliance on first-author counts in citation analysis studies, it conducts a full-scale evaluative citation analysis study in the XML research field using the three most well-known citation counting methods explained earlier – first-author counts, complete counts and fractional counts. It shows just how different the author rankings resulting from these different citation counting methods actually are, and demonstrates what support emerging data and tools on the Web may provide for more realistic citation counting methods.

Methodology

Data Collection

The research area we analyzed in the present study was XML – eXtensible Markup Language. Although XML has applications in a wide range of areas, the core of XML research belongs to computer science. Thus, we used the NEC Corporation Research Institute's CiteSeer (http://citeseer.ist.psu.edu/), now a joint effort of NEC and the Pennsylvania State University, to collect citing papers on XML and their reference lists. CiteSeer automatically indexes research papers that both fall within the broadly defined computer science field and are publicly available on the Web. Comparing with SCI, one of the ISI databases, CiteSeer is a SCI-like tool freely available on the Web. However, CiteSeer provides more information on cited papers than SCI, including their full titles, all authors, and abstracts as well as full text for those papers available on the Web. Studies have shown that author rankings based on data from CiteSeer are highly correlated with those based on data from SCI when using identical citation counting methods (Zhao & Logan, 2002). This finding implies that using this tool as a data source for citation analysis is just as valid as using SCI. More information about *CiteSeer* can be found in Lawrence et al (1999).

We developed a Java program to search for all documents indexed by CiteSeer in "Header" fields under the term "XML" or "eXtensible Markup Language," and to download all of the records that met the search criteria into a local machine. No citation windows were specified in the present study, indicating that publications from all years were used. The actual search was conducted on December 18, 2001. Since we are here concerned with fundamental issues of citation analysis methodology rather than with the evaluation of a particular research field, using data collected back then poses no problem to our investigation.

Another program was developed in Java to parse these records, and to store the resulting citation information such as titles, authors, publishing sources and years of both citing and cited documents in a data structure that was convenient for later data analysis such as counting citations using multiple methods. Since the existence of duplicates is one of the major differences between traditional databases and CiteSeer (Zhao & Logan, 2002), the citing documents were examined first by another Java program and then manually to remove possible duplicates. Citations made by these duplicates were removed as well.

This way, we collected 312 publications which made 4,578 citations altogether.

Data analysis

Analyzing the perception of the authors of the 312 publications, we ranked the cited authors by the number of citations they received, calculated using first-author counts, fractional counts and complete counts, respectively. The present study however took a simplified approach to fractional and complete counts in that it only took into account the first five authors of each paper rather than all authors. It was hoped that this approach would approximate sufficiently strict fractional and complete counts as publications with more than five authors were not expected to occur too frequently based on the statistics from the present study (Table 1), and even if its approximation was insufficient it would still help us to see beyond the first-author counts that only take into account first authors.

Each of the three author rankings resulting from the three different counting methods was compared with the other two, and for each pair of author rankings of common authors, Pearson's r was calculated to examine the degree of correlation between the two rankings.

Table 1: Distribution of papers by number of authors

Number of authors	Papers retrieved from CiteSeer	
	Number	Percentage
0	4	1
1	83	27
2	77	25
3	78	25
4	36	12
5 or more	34	11

Results

Table 2 presents the highly cited authors ranked by complete counts (list 1), fractional counts (list 2) and first-author counts (list 3) respectively. Only authors whose ranks are 50 or higher are listed in the table for the purpose of brevity although we will analyze the top 100 authors in our discussion below.

There are 45 authors who are common to all three lists of top 100 authors, 76 common to lists 1 and 2, 50 common to lists 1 and 3, and 58 common to lists 2 and 3. In other words, if we want to select 100 highly visible authors, only about half of the selected authors are going to be the same no matter what citation counting method we use, and about three quarters are going to be the same whether we use complete counts or fractional counts.

Table 2: Highly visible authors indicated by number of citations (#c) received

List 1: complete counts			List 2: fractional counts			List 3: first-author counts		
Rank	**Name**	**#c**	**Rank**	**Name**	**#c**	**Rank**	**Name**	**#c**
1	S. Abiteboul	172	1	J. Widom	63.4	1	S. Abiteboul	113
2	J. Widom	163	2	S. Abiteboul	62.3	2	P. Buneman	78
3	D. Suciu	157	3	D. Suciu	47	3	A. Deutsch	67
4	D. Florescu	129	4	J. Clark	43.7	4	T. Bray	64
5	M. F. Fernandez	117	5	P. Buneman	38.2	5	J. Clark	60
6	P. Buneman	102	6	D. Florescu	36.4	6	R. Goldman	46
7	A. Y. Levy	89	7	M. F. Fernandez	31.6	7	D. Florescu	37
8	J. McHugh	86	8	S. DeRose	28.5	8.5	M. F. Fernandez	34
9	A. Deutsch	72	9	J. McHugh	23.6	8.5	S. Ceri	34
10	T. Bray	67	10	T. Bray	23.4	10	J. Shanmugasundaram	30
11.5	R. Goldman	61	11	R. Goldman	22.5	11	J. Robie	29
11.5	W. Fan	61	12	A. Y. Levy	22.4	12	J. McHugh	28
13.5	J. Clark	60	13	M. Murata	21.7	13	Y. Papakonstantinou	26
13.5	S. Cluet	60	14	W. Fan	21.4	14.5	H. Thompson	25
15	S. DeRose	57	15	H. Thompson	20.2	14.5	S. Cluet	25
17	C.SperbergMcQueen	55	16	S. Cluet	19.8	16	S. S. Chawathe	23
17	J. Paoli	55	17	J. Robie	18.6	17	M. Murata	22
17	J. Robie	55	18	C.SperbergMcQueen	18.6	18.5	D. D. Chamberlin	21
19	S. Ceri	52	19	J. Paoli	18.3	18.5	W. Fan	21
20	D. J. DeWitt	51	20	V. Vianu	18.1	20.5	R. G. G. Cattell	17
21	D. Quass	50	21	S. Ceri	17.5	20.5	S. DeRose	17
22	V. Vianu	48	22	A. Deutsch	17.3	23	C. Beeri	16
23	J. Simeon	45	23	C. Brew	16.6	23	T. Milo	16
24	T. Milo	41	24	T. Milo	16.4	23	W. van der Aalst	16
25.5	S. Davidson	40	25	J. Simeon	16.2	27	C. Brew	15
25.5	Y. Papakonstantinou	40	26	Y. Papakonstantinou	14.8	27	H. Hosoya	15
27.5	H. Garcia-Molina	38	27	P. Wadler	14.7	27	O. Lassila	15
27.5	J. L. Wiener	38	28	D. Maier	14.1	27	P. Wadler	15
29	J.Shanmugasundaram	34	29	W. van der Aalst	13.6	27	V. Christophides	15
30	S. Paraboschi	33	30	H. Garcia-Molina	13.3	30	E. Maler	14
31	H. Thompson	32	31	D. Kossmann	12.2	32	A. Bonifati	13
33	C. Brew	29	32	J. D. Ullman	11.9	32	J. Widom	13
33	D. Kossmann	29	33	D. J. DeWitt	11.9	32	T. Berners-Lee	13
33	P. Wadler	29	34	D. Quass	10.9	34.5	D. Brickley	12
35.5	C. Zhang	28	35	D. Megginson	10.5	34.5	M. Hanus	12
35.5	D. Schach	28	36	S. Davidson	10.3	38	D. Fensel	11
39.5	D. Maier	27	37	G. Ghelli	10.3	38	D. Lee	11
39.5	J. Lapp	27	38	M. Hanus	10.2	38	D. Megginson	11
39.5	J. D. Ullman	27	39	S. S. Chawathe	9.72	38	M. J. Carey	11
39.5	M. Murata	27	40	D. D. Chamberlin	9.43	38	V. Apparao	11
39.5	P. Fraternali	27	41	G. Moerkotte	9.37	42	D. Maier	10
39.5	S. S. Chawathe	27	42	O. Lassila	9.08	42	N. Klarlund	10
43.5	E. Damiani	25	43	S. Paraboschi	8.82	42	S. Nestorov	10
43.5	S. Weinstein	25	44	D. Schach	8.78	47	B. Ludascher	9
45	K. Tufte	24	45	J. Lapp	8.58	47	D. Calvanese	9
46.5	D. D. Chamberlin	23	46.5	C. Beeri	8.5	47	E. Baralis	9
46.5	S. Decker	23	46.5	R. Hull	8.5	47	F. Neven	9
49	C. Delobel	22	48	B. C. Pierce	8.33	47	J. Bosak	9
49	D. Fensel	22	49	P. Fraternali	8.32	47	J. E. Hopcroft	9
49	G. He	22	50	S. Weinstein	8.25	47	N. Walsh	9

To determine the correlations between these rankings, Pearson's r's were calculated for the 45 authors common to all three top-100 lists. The r value is 0.917 between author rankings by fractional and complete counts, 0.654 between those by fractional and first-author counts, and 0.648 between those by first-author counts and complete counts. The r's are slightly higher when comparing the rankings for the 2045 authors who are common to all three complete lists of all cited authors using the same three different citation counting methods: 0.954 between fractional and complete counts, 0.77 between first-author counts and fractional counts, and 0.74 between first-author counts and complete counts. Clearly, author rankings by fractional counts and complete counts are highly correlated, but are significantly different from author rankings by first-author counts.

Table 3 highlights some of the authors whose ranks by fractional counts and by first-author counts were significantly different. Authors listed on the left side are significantly under-evaluated by first-author counts whereas those on the right are significantly overestimated by first-author counts comparing with fractional counts. For example, Widom who is ranked at the very top by fractional counts sat only at the 32nd position by first-author counts; similarly, Suciu who was ranked as the 3rd by fractional counts did not even make the list of top 100 authors by first-author counts. Looking from the other direction, Deutsch who was ranked as the 3rd by first-author counts sat at the 22nd position by fractional counts.

Table 3: Examples of authors who are under- or over-estimated by first-author citation counts

SC – First-author counts (i.e. straight counts); **FC** - Fractional counts

Under-evaluated			**Overestimated**		
Author	**Rank**		**Author**	**Rank**	
	SC	**FC**		**SC**	**FC**
Widom	32	1	Deutsch	3	22
Suciu	>100	3	Ceri	8.5	21
DeRose	20.5	8	Shanmugasundaram	10	60.5
Levy	65.5	12	Papakonstantinou	13	26
Sperberg-McQueen	>100	18	Chawathe	16	39
Paoli	100	19	Chamberlin	18.5	40
Vianu	>100	20	Cattell	20.5	>100
Simeon	>100	25	Beeri	23	46.5
Maier	43	28	Lassila	27	42
Garcia-Molina	>100	30	Hosova	27	67
Kossmann	>100	31	Christophides	27	99.5
Ullman	>100	32	Maler	30	52
DeWitt	>100	33			
Quass	>100	34			
Davidson	>100	36			

Discussion

We now discuss the issues summarized in the Introduction section in light of results from the present study.

Correlation between results from first-author citation counts and complete citation counts

The present study did not find as high correlations between results from complete counts and first-author counts as other studies have (Cole & Cole, 1973; Long et al., 1980; Lange, 2001). In fact, the correlation between rankings by first-author counts and by other counts that we measured was too low to justify their use in a scientific study that aims at any degree of accuracy.

Our lower correlations than other studies between author rankings by first-author counts and complete counts can be due to a number of reasons as discussed below.

Different time points. The level of collaboration in scientific research has been increasing in the past few decades. Complete counts take into account authors' contributions through collaboration. Since the collaboration level was not as high during the periods covered in other studies as it is now, the difference they found between results from complete counts and first-author counts was not as significant.

Different disciplines examined. The level of collaboration also varies between disciplines. Thus, it is not unexpected that correlations between results from these two approaches are different depending on the disciplines that are studied: the higher a discipline's collaboration level, the lower the expected correlations. If this is the case, it is reasonable to say that the extent to which first-author counts can be used to substitute for complete counts varies with the discipline being studied. The higher the level of collaboration in a discipline, the more serious the distortion resulting from this substitution is.

Different approaches to calculating correlations. One piece of evidence of this possibility is the difference between the r's for the 45 common top ranked authors and those for all the 2045 common authors. For lower-ranked authors, especially those that received only one citation, it is much more likely that they are ranked relatively the same both by complete counts and by first-author counts. Since these authors are in the majority, their inclusion in the analysis tends to bury the considerable difference between the rankings of highly cited authors, and to thus increase the perceived correlation between these two counting methods. In particular, one would find an even higher correlation if scholars who have never been cited were included in the calculation, burying the significant difference between the rankings of highly-cited scholars even deeper.

If this third possibility is the case, it is clear that substituting complete counts by first-author counts in the evaluation of scholars would introduce significant bias. The reason is that the scholars being evaluated usually are, or at least include, those who have been cited frequently (e.g. scholars who apply for grants or tenure). As we have seen, correlation between results from different citation counting methods will be low for these scholars.

Complete counts vs. fractional counts

As far as the differences between results from complete and fractional counts are concerned, it is quite obvious theoretically that complete counts over-weigh authors of multiple-authored papers. As a result, it can be expected that authors involved in large-group collaborations are more likely to be ranked high by complete counts than other authors, regardless of the quality or impact of their work, and other, less collaborative, authors thus tend to be pushed out of the lists of top-ranked authors.

This is supported by results from the present study.

We identified several large groups of authors among the top-ranked scholars by examining their CVs and publications, as partly shown in Table 2: (a) the research project Lore: A database management system for XML at Stanford University which included Widom, Abiteboul, Quass, Goldman, McHugh, and Gupta; (b) the database group at University of Pennsylvania which was led by Professor S. Davidson and which included Buneman, Deutsch, Fan, Suciu, and Tan; (c) the Verso Project group at INRIA-Rocquencourt, France which was led by Abiteboul and Cluet and which included Bonifati, Florescu, Milo and Vianu; (d) an Italian group which was led by Ceri and included Paraboschi, Fraternali, Damiani, Bonifati, Bertino, Tanca, and Comai; and (e) the Niagara Internet Query group that included Carey, He, Tufte, Zhang, DeWitt and Naughton.

Examples of authors whose ranks by complete counts were pulled up by their groups include He, Carey, and Naughton in the Niagara group, Bertino, and Tanca in the Italian group, and Tan in the database group at the University of Pennsylvania. They did not appear in the list of top 100 authors by fractional counts. Authors such as Bosak and Megginson who were ranked high by fractional counts but were not involved in any of these groups were pushed out of the list of top ranked authors by complete counts.

Thus, fractional counts are preferred to complete counts in the evaluation of scholars. This also makes sense conceptually. Although it is true that the more an author's name appears in the by-lines no matter which position, the more visible the author is, it is not appropriate in the evaluation of scholars to give the same credit to authors of single-authored papers and those of multiple-authored papers as complete counts do.

Possible bias of first-author counts

Possible bias resulting from first-author counts has been discussed theoretically in some of the studies mentioned earlier. Results from the present study can provide some empirical data and add to the discussion.

Obviously, authors who often publish as non-first authors would be under-evaluated by first-author counts and authors who would be overestimated are frequently listed as the first authors of papers. We therefore need to ask the question whether non-first authors tend to make important contributions to the research their papers report. There are two reasons supported by results from the present study that are in favor of a positive answer to this question.

Firstly, key researchers sometimes publish as non-first authors, especially in reporting results from large-group research projects. Collaborative scientific research is often organized by projects. A small number of key scholars come up with the fundamental ideas, develop research proposals, and obtain funding for their research. They then organize a group of scholars to work out the ideas step by step in the direction they previously outlined. In these cases, these key scholars would often publish as non-first

authors of papers reporting each aspect of the research project. Clearly, such scholars have a strong impact on research, not only by producing research ideas and outlining research directions but also by organizing, if not conducting, the research. They play the role of intellectual and / or organizational leaders of the research who are frequently respected scholars in their own right (Griffith & Mullins, 1972).

Widom, Garcia-Molina, Ullman and Davidson appear to be such examples in our study. Widom was an associate professor at the Department of Computer Science and Electrical Engineering, Stanford University. She had frequently (i.e. more than five times) coauthored with many of the top-ranked authors in this study such as Abiteboul, Garcia-Molina, Ullman, Ceri, Quass, Sagiv, Goldman, McHugh, Chawathe, Rajaraman, Wiener, Gupta, and Papakonstantinou. Among these, Quass, Goldman, McHugh, and Gupta were her students at some point, professors Ullman and Garcia-Molina were her colleagues, and Abiteboul used to be a visiting scholar at her department. Many of them, including Abiteboul and her students, were involved in the research project Lore: A database management system for XML, from which many of the highly cited papers in this study were produced. She also had joint research projects with her colleagues Garcia-Molina and Ullman. As another example, Professor S. Davidson has been leading the database group at University of Pennsylvania, of which many of the highly cited authors in the present study were members, including Buneman, Deutsch, Fan, Suciu, and Tan.

Secondly, authors who are listed late in the by-lines are not necessarily those who have made fewer contributions to the papers. As mentioned earlier, alphabetic ordering of author names has been found to occur in many fields (Endersby, 1996; Lindsey, 1980; Rudd, 1977), and appears to be significant in the XML research field as well. Zhao & Logan (2002) identified a striking alphabetical shifting pattern from author rankings by complete counts to those by first-author counts, indicating that authors in the XML research field often were listed alphabetically rather than based on contributions to the papers. The same pattern is noticeable here. Table 2 shows that many authors who are top-ranked by first-author counts and who have names starting earlier in the alphabet shifted down in the ranking by fractional counts (e.g. Deutsch and Ceri) while some authors whose names start late in the alphabet shifted up (e.g. Widom and Suciu).

Cost of collecting citation data for counting non-first authors for the purpose of evaluation

Collaboration has become the mainstream in science, certainly in the XML research field as seen from the present study. Evaluation of scholars using first-author counts would leave out authors' contributions through collaboration and hence would be quite biased and inaccurate. Although there has been a debate on whether first-author counts can adequately replace complete counts when ranking a group of scholars, it has long been recognized and is also supported by results from the present study that different counting methods do produce quite different results in terms of the common top-ranked authors they share. The present full-scale citation analysis study also found that the correlation between author rankings by first-author and complete counts was quite low, thus adding weight to results from previous small-scale studies. Clearly, the continued use of first-author counts in the evaluation of scholars is not so much due to the lack of awareness of the bias resulting from first-author counts, but largely due to the convenience and low cost of calculating first-author counts with the kind of support offered by the ISI databases.

Indeed, we also found from our own experience in working with both data sources that the ISI databases were easier to use than CiteSeer for collecting data for citation analysis studies.

The most challenging part of working with CiteSeer that we experienced was how to download the entire set of search results into a local computer and how to parse citation data in HTML format. As discussed in the Methodology section, we downloaded all the papers on XML indexed by each of the two data sources into a local computer, and then ranked authors of papers that were cited by these papers based on the number of citations they received from these papers. The ISI databases provided us with the option of downloading the entire set of search results into a text file in a data format that was easy to work with computationally. CiteSeer, on the other hand, did not provide any option for viewing the entire set of search results or for downloading search results. Nor, at that time, did we find an Application Programming Interface for CiteSeer. It was quite some work for us who are not trained as programmers to write a Java program that searches the CiteSeer database directly and downloads the search results to a local computer, by-passing the user interface provided by CiteSeer. The downloaded files were all in HTML format, which did not provide rich cues for parsing the citations into elements (e.g. authors, title, publishing year).

As indicated on the CiteSeer website (http://citeseer.ist.psu.edu/oai.html), CiteSeer is now compliant with the Open Archive Initiative (OAI) Protocal for Metadata Harvesting, and provides the option of browsing or downloading its records programmatically from its OAI collection. It is even possible to download all of its OAI records, including citation relationships. However, we found that the citation relationships included in the downloaded files are not complete – only references to documents that are indexed in CiteSeer as source documents are included. Further investigation needs to be done as to how these files can be augmented to include all references in order to be used in citation analysis studies.

Another example of CiteSeer being more difficult to use is the process of removing duplicates of citing papers. Unlike the ISI databases that employ intensive manual processing of mostly journal articles, CiteSeer is fully automated in collecting, parsing, and indexing scholarly papers from the open Web. As a result, duplicates of citing papers occurred quite frequently that were essentially identical publications with minor differences in the spelling of author names, titles, etc. We needed to remove the duplicates first by writing and running a Java program and then manually, which was quite time-consuming. The cost of removing duplicates will obviously increase with the size of the data sample.

Clearly, this problem may not go away easily due to the less-regulated nature of publications on the Web and the automatic nature of CiteSeer. However, the cost of dealing with this problem is significantly lower than the proposed remedies to the limitations of the ISI databases in counting non-first authors (e.g. compiling a complete list of publications by each of the scholars being evaluated).

With the information on all authors of cited works readily available in databases like CiteSeer, the cost of calculating complete counts or even fractional counts is not much higher than computing first-author counts. Although it may still require more technical skill to work with these new databases and it may still be necessary to deal with problems that do not exist in the ISI databases such as the second problem mentioned above, the benefits of counting more than first authors for more balanced and less

biased evaluation of scholarly contributions certainly outweighs the small extra cost and effort.

Conclusion

Counting citations received by scholars has long been used as a way to evaluate the impact of scholars. It has been noted theoretically and supported by small-scale studies that different citation counting methods can produce very different author rankings and that fractional counts are most preferred among the three commonly discussed methods: first-author counts, complete counts, and fractional counts. However, in practice of citation analysis studies, first-author counting is still the most frequently used method. We examined the reasons and arguments for this common practice in the present paper through a full-scale comparative citation analysis study using all of the three citation counting methods just mentioned. This was enabled by taking advantage of data and tools for citation analysis studies increasingly available on the Web, in this case, CiteSeer.

We provide strong empirical evidence for the bias associated with first-author counts, and for the low degree of significance in correlations between author rankings by first-author counts and other counting methods. We also demonstrate the capability of emerging tools such as CiteSeer in supporting various counting methods without adding much cost. These contest some major justifications for the continued use of first-author counts in citation analysis studies, such as the prohibitive costs involved in counting non-first cited authors using the ISI databases. Although it may require more technical skill and effort to work with them, these new tools reduce the costs incurred by all-author counts significantly. More importantly, these tools allow us to go beyond what the ISI databases have supported to explore more sophisticated citation counting methods for more balanced and less biased evaluations of scholarly impact. This may in turn contribute to the emergence of a more effective scholarly communication system. In this system, scholarly work is evaluated based on how many users it has reached in all rather than on how many users it has reached who have published in journals indexed by certain databases such as the ISI databases.

Reference

Bensman, S. J. (2001). Urquhart's and Garfield's Laws: the British controversy over their validity. *Journal of the American Society for Information Science, 52*, 714-724.

Cole, J.R., & Cole, S. (1973). *Social Stratification in Science*. Chicago: University of Chicago Press.

Cronin, B., & Overfelt, K. (1994). Citation-based auditing of academic performance. *Journal of the American Society for Information Science, 45*, 61-72.

Egghe, L. & Rousseau, R. (1990). *Introduction to Informetrics*. New York: Elsevier Science.

Egghe, L., Rousseau, R., & van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: consequences for evaluation studies. *Journal of the American Society for Information Science, 51*, 145-157.

Endersby, J. W. (1996). Collaborative research in the social sciences: Multiple authorship and publication credit. *Social Science Quarterly*, 77, 375-392.

Garfield, E. (1979). *Citation indexing — its theory and application in science, technology, and humanities*. New York: John Wiley & Sons.

Gauffriau, M., & Larsen, P. O. (2005). Counting methods are decisive for rankings based on publication and citation studies. *Scientometrics*, *64*(1): 85-93.

Griffith, B.C. & Mullins, N.C. (1972). Coherent social groups in scientific change. *Science, 177*, 959-964.

Lange, L. L. (2001). Citation counts of multi-authored papers – First-named authors and further authors. *Scientometrics*, *52*, 457-470.

Lawrence, S., Giles, C. L. & Bollacker, K. (1999). Digital libraries and autonomous citation indexing. *IEEE Computer, 32*(6): 67-71.

Lindsey, D. (1980). Production and citation measures in the sociology of science: The problem of multiple authorship. *Social Studies of Science, 10*, 145-162.

Lindsey, D. (1982). Further evidence for adjusting for multiple authorship, *Scientometrics, 4,* 389-395.

Long, J. S., McGinnis, R. & Allison, P. D. (1980). The problem of junior-authored papers in constructing citation counts. *Social Studies of Science, 10*, 127-143.

MacRoberts, M. H. & MacRoberts, B. R. (1989). Problems of citation analysis: a critical review. *Journal of the American Society for Information Science, 40*, 342-349.

Oppenheim, C. (1998). Fractional counting of multiauthored publications. *Journal of the American Society for Information Science, 49*, 482-484.

Persson, O. (2001). All author citations versus first author citations. *Scientometrics, 50* (2): 339-344

Porter, A. L. (1977). Citation analysis: queries and caveats. *Social Studies of Science, 7*: 257-267.

Rudd, E. (1977). The effect of alphabetical order of author listing on the careers of scientists. *Social Studies of Science, 7,* 268-269.

Smith, L. C. (1981). Citation analysis. *Library Trends, 30*, 83-106.

Stokes, T. D. & Hartley, J. A. (1989). Coauthorship, social structure and influence within specialties. *Social Studies of Science, 19*, 101-125.

Trueba, F.J. & Guerrero, H. (2004). A robust formula to credit authors for their publications. *Scientometrics*, *60*, 181-204

Van Hooydonk, G. Fractional counting of multiauthored publications: Consequences for the impact of authors. *Journal of the American Society for Information Science, 48*, 944-945.

Zhao, D. & Logan, E. (2002). Citation analysis of scientific publications on the Web: A case study in XML research area. *Scientometrics, 54*, 449-472.