

Functions of Uni- and Multi-citations: Implications for Weighted Citation Analysis

Dangzhi Zhao[†], Alicia Cappello & Lucinda Johnston

School of Library and Information Studies, University of Alberta, Edmonton,
Alberta T6G 2J4, Canada

Abstract

Purpose: (1) To test basic assumptions underlying frequency-weighted citation analysis: (a) Uni-citations correspond to citations that are nonessential to the citing papers; (b) The influence of a cited paper on the citing paper increases with the frequency with which it is cited in the citing paper. (2) To explore the degree to which citation location may be used to help identify nonessential citations.

Design/methodology/approach: Each of the in-text citations in all research articles published in Issue 1 of the *Journal of the Association for Information Science and Technology (JASIST)* 2016 was manually classified into one of these five categories: Applied, Contrastive, Supportive, Reviewed, and Perfunctory. The distributions of citations at different in-text frequencies and in different locations in the text by these functions were analyzed.

Findings: Filtering out nonessential citations before assigning weight is important for frequency-weighted citation analysis. For this purpose, removing citations by location is more effective than re-citation analysis that simply removes uni-citations. Removing all citation occurrences in the Background and Literature Review sections and uni-citations in the Introduction section appears to provide a good balance between filtration and error rates.

Research limitations: This case study suffers from the limitation of scalability and generalizability. We took careful measures to reduce the impact of other limitations of the data collection approach used. Relying on the researcher's judgment to attribute citation functions, this approach is unobtrusive but speculative, and can suffer from a low degree of confidence, thus creating reliability concerns.

Practical implications: Weighted citation analysis promises to improve citation analysis for research evaluation, knowledge network analysis, knowledge representation, and information retrieval. The present study showed the importance of filtering out nonessential citations before assigning weight in a weighted citation analysis, which may be a significant step forward to realizing these promises.

Originality/value: Weighted citation analysis has long been proposed as a theoretical solution to the problem of citation analysis that treats all citations equally, and has attracted increasing

Citation: Dangzhi Zhao,
Alicia Cappello &
Lucinda Johnston (2017).
Functions of Uni- and
Multi-citations:
Implications for Weighted
Citation Analysis.
Vol. 2 No. 1, 2017
pp 51–69
DOI: 10.1515/jdis-2017-0003
Received: Nov. 16, 2016
Accepted: Nov. 20, 2016



[†] Corresponding author: Dangzhi Zhao (Email: dzhao@ualberta.ca).

Research Paper

research interest in recent years. The present study showed, for the first time, the importance of filtering out nonessential citations in weighted citation analysis, pointing research in this area in a new direction.

Keywords Weighted citation analysis; Citation function; Research evaluation; In-text citation analysis; Re-citation analysis

1 Introduction

Citation analysis is used in research evaluation exercises around the globe, directly affecting the work and lives of millions of researchers and the expenditure of billions of dollars. It is therefore crucial to address the problems and limitations that plague it. Central amongst critiques of the current practices of citation analysis has long been that it treats all citations equally, regardless of whether they are crucial to the citing paper or perfunctory. This problem is especially troublesome when tracing or assessing research impact.

Weighting citations by how they are used in the citing paper has long been proposed as a theoretical solution to this problem (Herlach, 1978; Narin, 1976; Voos & Dagaev, 1976). By weighting citations, it is hoped that essential citations could be assigned greater weight than perfunctory ones so that citation analysis can focus on more profound influences and organic relationships. In practice, however, it has not been studied closely at a large-scale until recently. Increasingly available digital full-text documents and advances in text processing technologies are now making it feasible to conduct large-scale studies on weighted citation analysis. As a result, interest in these types of studies is growing. Studies have experimented with weighting citations by the frequency with which they occur in the text (e.g. Ding et al., 2013; Hou, Li, & Niu, 2011; Tang & Safer, 2008; Zhu et al., 2015), by the citation impact of citing papers (Ding & Cronin, 2011), and by the location and context in which they are cited (Boyack, Small, & Klavans, 2013; Jeong, Song, & Ding, 2014). It has been found that frequency-weighted citation rankings can outperform traditional citation rankings of top authors, and that in-text citation frequency was the best of many full-text features to help spot citations that were considered crucial to the citing papers by their authors (Zhu et al., 2015).

Frequency-weighted citation analysis assigns a weight to citations based on the frequency with which they appear in the text of the citing paper. Clearly, this practice assumes that the more frequently a reference is mentioned in the text, the more influential it is to the citing paper. The present study is a preliminary test of this basic assumption, which underlies frequency-weighted citation analysis. Its result is expected to provoke further discussions and studies in this area. Further studies



will be important for assessing and improving the practice of weighted citation analysis which have been attracting more interest as a solution to one of the fundamental concerns regarding citation analysis.

1.1 Research Questions

If the signal to be detected in citation analysis is the direct and substantial flow of knowledge from the cited to the citing papers, perfunctory citations can be considered a source of noise. This noise is quite serious as a high incidence of perfunctory citations (40% or more) has been repeatedly observed in previous studies (Small, 1982). For example, Teufel, Siddharthan, and Tidhar (2006) found that only a fifth of references are essential for the citing papers, and Moravcsik and Murugesan (1975) noted that 40% references were perfunctory, frequently simply copied from other papers without ever having been read (Dubin, 2004).

There are two approaches to dealing with noise: filter out the noise, or amplify the signal. The ultimately best approach is likely some combination of the two.

The signal amplification approach has been used by almost all frequency-weighted citation counting schemes found in the literature. This approach assigns a weight of N (or a function of N such as N^2) to a citation that appears N times in a citing paper. The assumption is clearly that the more frequently a reference is mentioned in the text, the more significant it is to the citing paper.

Compared to the signal amplification approach, the noise filtration approach, which was introduced by Zhao and Strotmann (2015a; 2016), attempts to make the fundamental qualitative distinction between references that represent real use by, core impact on, or organic connection with, the citing paper (which it aims to retain for analysis) and those that are merely mentioned in passing as related work or background information (which it aims to remove). By only counting core connections in knowledge networks, this approach can help research evaluation become more sensitive to the essential impact of research. It can also better capture “aboutness” of documents, the essence of subject indexing in knowledge representation and retrieval. Knowledge representation and retrieval systems that make use of citation links can therefore benefit from improved precision in computer-aided subject indexing and in their “more like this” features (Zhao & Strotmann, 2015a). In addition, the signal amplification required to counter the strong noise created by perfunctory citations (40% or more) tends to be so strong (N^2 is the minimal power of N required) that it can cause serious distortions (Zhao & Strotmann, 2016). Filtering out this noise before applying necessary signal amplification can avoid this technical problem.

The key, and difficult, question is how to identify and filter out perfunctory citations. Zhao and Strotmann (2015a, 2016) proposed a simple method for this:



Research Paper

re-citation analysis, which focuses on re-citations (i.e. references that appear more than once in the text of a citing paper) by filtering out uni-citations (i.e. references that appear only once in the text of a citing paper). The basic assumption of re-citation analysis is that papers are likely to be cited multiple times in a publication that relies heavily on them, while perfunctory citations should appear only once in a citing paper.

In order to test the assumptions of frequency-weighted citation analysis, the research questions addressed in the present study are as follows:

- 1) *Do uni-citations correspond to citations that are nonessential to the citing articles?*
- 2) *Does the influence of a cited document on the citing paper increase with the frequency with which it is cited in the citing paper?*

Results from investigating these questions will be important for evaluating the validity of the signal amplification approach to frequency-weighted citation analysis. Results will also be important for assessing whether the potential of the noise filtration approach for improving citation analysis for research evaluation, knowledge network analysis, knowledge representation, and information retrieval (Zhao & Strotmann, 2015a) can be realized by re-citation analysis. To shed light on other directions that may realize this potential, the following question will also be addressed.

- 3) *To what degree can citation location be used to help identify nonessential citations?*

To address these questions, we identify a set of typical functions of citations in the citing paper, and examine how citations of different in-text frequencies or from different locations in the text are distributed by these functions. Details will be provided in the Methodology section below.

1.2 Related Studies

Citation analysis examines citation patterns and networks in scholarly literature through statistical analysis and network visualization. It is applied widely in the social sciences to trace knowledge flow, to evaluate research impact, to study the characteristics of scholarly communities and knowledge networks, and to create citation link-based knowledge representation and retrieval systems (Borgman & Furner, 2002; Hall, Jaffe, & Trajtenberg, 2005; Zhao & Strotmann, 2015b).

The basic assumption underlying citation analysis is that a citation represents the citing author's use of the cited work, and that it therefore indicates that the citing and cited works are related in subject matter or methodological approach (Garfield,



1979; White, 1990). The total number of citations that a document, or any aggregate of documents (e.g. author oeuvre, journal), receives (or a score derived from it, e.g. *h*-index) is therefore used to assess its impact on research in research evaluation. Citation links are used to signify knowledge flow from the cited to the citing group and, along with scores derived from these links, to measure the relatedness between documents, or their aggregates, in the study of knowledge networks and the representation and retrieval of related documents (Borgman & Furner, 2002; Zhao & Strotmann, 2015b).

The assumptions of citation analysis are believed to be in line with Merton's normative view of science (Garfield, 1979; Merton, 1942; White, 1990). Like other activities of science, citation behavior is assumed to be governed by a set of norms which require authors to cite documents that have influenced them in developing their current works in order to give credit where credit is due (Edge, 1979; Griffith, 1990; Peritz, 1992; Tranöy, 1980). Although citations for reasons other than giving due credit exist (Cronin, 1984; Edge, 1979), citation analysis has generally been found to produce valid results because it is based on a statistical analysis of the collective perceptions of large numbers of citing authors, most of whom do adhere to the norms, most of the time (Small, 1977; White, 1990). This is especially true with citation network analysis and citation link-based knowledge representation and retrieval, as even non-normative citations will not refer to unrelated works.

Researchers do cite for various reasons and citations do serve many different functions in citing papers, however. Beginning in the 1970s, a great deal of research has been done on citer motives, citing behaviors, and citation functions. It was at this time that the use of citation analysis in research evaluation caused concerns that citations may not represent the actual use of the cited documents, and that citation counts that do not take into account citers' motives, citing behavior, and citation functions may not reflect the impact or merit of the cited documents (Brooks 1985, 1986; Case & Higgins, 2000; Chubin & Moitra, 1975; Garfield, 1962; Liu, 1993; Moravcsik & Murugesan, 1975; Shadish et al., 1995; Vinkler, 1987; White & Wang, 1997). These studies have also been reviewed in various contexts and for different purposes (e.g. Borgman & Furner, 2002; Bornmann & Daniel, 2008; Tabatabaei, 2013). Tabatabaei (2013) did a thorough review of studies on citer motives, citing behaviors, and citation functions in order to develop a coding scheme for assessing the contribution of information science to other disciplines, as reflected by the functions of highly-cited *Journal of the Association for Information Science and Technology (JASIST)* papers in the citing articles. Bornmann and Daniel (2008) summarized a number of citation behavior studies, and provided a unified typology of citation motivations: citations of the affirmational, assumptive, conceptual, contrastive, methodological, negational, perfunctory, or persuasive type. Small (1982)



Research Paper

identified five typical distinctions in citation classification schemes: (1) negative or refuted, (2) perfunctory or noted only, (3) compared or reviewed, (4) used or applied, and (5) substantiated or supported by the citing work.

In order to assign different weights to citations of different functions, which would improve citation analysis and information retrieval results, studies have explored how textual properties, including citation frequency and citation location in the citing papers, may be used to automatically differentiate citations of different functions or importance to citing papers.

Chubin and Moitra (1975) considered cited references being cited multiple times in a citing paper (i.e. multi-citations) as the most affirmative. Voos and Dagaev (1976) stated that “we do not believe that there can be much argument with the premise that an author who is cited more than once in an article might have more relevance, and/or importance than an author who is cited only once in an article” (pp. 20–21). Herlach (1978) found that multi-citations are about 30% more topically relevant to the citing paper than uni-citations. Bonzi (1982) confirmed results from Herlach (1978) and Voos and Dagaev (1976) that multi-citations can be used as a good predictor of importance or relevance to the citing paper. Tang and Safer (2008) found that giving high importance to multi-citations may help improve citation-based rankings. Zhu et al. (2015) also found that in-text citation frequency was the best feature to help spot citations that were considered crucial to the authors of a citing paper, and that frequency-weighted citation ranking can outperform traditional citation ranking of top authors, at least in the research field they studied.

The structure of scientific articles reporting original research results has been, to a large degree, standardized over the years to include “Introduction,” “Methods,” “Results,” “Discussion,” and “Conclusion” sections (Doumont, 2010). This structure reflects the progression of most research projects (Doumont, 2010), facilitates more effective and efficient use of research articles, and has been recommended by many style manuals and required by most scientific journals (McCain & Turner, 1989). Bertram (1972) suggested that citation level or significance is predictable through the identification of the section of the article in which a citation appears. Although some later studies (e.g. Hanney et al., 2005) found no significant difference in terms of citation location for citation importance, many studies found that citations located in the Methodology, Results, Discussion, or Conclusion sections may play a more significant or meaningful role than those located in the introductory sections (Bonzi, 1982; Cano, 1989; Tang & Safer, 2008; Voos & Dagaev, 1976).

McCain and Turner (1989) considered both citation location and citation frequency in the calculation of a Utility Index, which can be more effective in citation analysis (Ding et al., 2013; Herlach, 1978). Herlach (1978) noted that a paper that has been cited in the Introduction or Literature Review, and subsequently mentioned in the



Methodology or Discussion sections, will likely have made a more significant contribution to the citing article than one which has been mentioned only once in the entire article. Tang and Safer (2008) also emphasized other factors that may affect the impact of citation frequency on citation significance such as the “pond effect” (p. 262).

2 Methodology

We collected all research articles in a single issue of the *Journal of the Association for Information Science and Technology (JASIST)* 2016, Volume 67 Issue 1, and coded all in-text citations as to their function based on the context in which they appear. There were 14 articles and 1,473 in-text citations in total.

Previous studies on citer motives and citation functions either asked citing authors, through questionnaires and/or interviews, or analyzed the citation context of each citation occurrence. Both approaches have pros and cons (Case & Higgins, 2000; Harwood, 2008; Prabha, 1983; Shadish et al., 1995), and the present study chose to use the latter, which relies on the researchers’ judgment or interpretation instead of the citing authors’ motivational claims. This approach is unobtrusive but speculative, and can suffer from a low degree of confidence and accuracy, thus creating reliability concerns. However, these concerns are more relevant when the researchers do not have a clearly defined coding scheme. Another limitation of this approach is its scalability and generalizability. The scale of this type of studies has always been small due to the time consuming nature of manually coding large numbers of citation occurrences in research articles.

As a case study of the Library and Information Science (LIS) field, as represented by a single issue of *JASIST*, the present study suffers from the limitation of scalability and generalizability. However, we took careful measures to reduce the impact of other limitations. We examined research articles in the field of LIS, a field that we understand well and can therefore assess more confidently and accurately. We used a clearly defined coding scheme developed in a dissertation for a purpose similar to ours, i.e. assessing the contribution of information science to other disciplines as reflected by the functions of highly cited *JASIST* papers in the citing articles (Tabatabaei, 2013). The detailed explanation and coding examples for each of the categories of citation functions in the scheme also helped with the accuracy and consistency of coding in the present study.

This coding scheme has five categories: Applied, Contrastive, Supportive, Reviewed, and Perfunctory. Tabatabaei (2013) defined these categories as follows (p. 153) and provided further detailed interpretation of the scheme and coding examples (pp. 154–176). We can reasonably consider citations in the first three



Research Paper

categories (i.e. Applied, Contrastive, or Supportive) as having substantial influence on the citing paper and those in the last two categories (i.e. Reviewed or Perfunctory) as nonessential citations.

- 1) **Applied:** When a citing paper borrowed or adopted a significant element from a cited paper and used it in developing its own theme or study; or when the whole cited paper inspired the citing paper to develop a significant element; or when a citing paper built upon a cited paper, expanded, or furthered a cited paper's study or even modified a cited paper's method or approach; the contribution of the cited paper to the citing paper was coded under the main category of Applied.
- 2) **Contrastive:** When a citing paper contrasted its data, method, model, theory, findings, etc., with what was used, documented, reported, or found in a cited paper, the contribution of the cited paper to the citing paper was coded under the main category of Contrastive.
- 3) **Supportive:** Citing authors made references to cited papers to establish the legitimacy of their topics, to substantiate an assumption or a claim, to justify their central arguments, data, or methods, to confirm their findings, or to support an assertion, an opinion, a method, or a result.
- 4) **Reviewed:** Describing or reviewing relevant and similar prior studies always comprises a significant number of references in a paper. Citing authors usually provide their readers with some background reading to set the stage for the research area or problem. Sometimes citing authors introduce readers to the origin of an idea or concept discussed in their paper. This type of citation illustrates the history or state of the art of the research problem that is investigated in the citing paper, or reviews the current state of knowledge or research area in a subject field related to the citing paper. Usually citing authors acknowledge the pioneering achievements of other researchers and discuss a range of previous researchers' views on the topic. In sum, reviewed citations provide the readers with contextual information necessary to understand the broad context of the study or the significance of the research questions or problems of the citing paper.
- 5) **Perfunctory:** A citation has little importance, significance, or contribution to the theme, analysis, or results of the citing paper. Citing authors make these perfunctory references to the cited papers without additional comments. Usually more than one citation is mentioned in the same context, the cited paper was apparently not very relevant to the citing paper's immediate concern or theme, and the citing author made no attempt to compare or analyze the cited paper's contribution to the citing paper (Bornmann & Daniel, 2008).



The 14 source articles were processed in a random order. Each of the in-text citations were classified into one of the above five function categories by two coders, independently. Both coders are second-year students of the Master of Library and Information Studies (MLIS) program. The two coders agreed completely on 48% of the in-text citations. For another 37%, they agreed that the citations were either perfunctory or reviewed but did not agree on which of these two functions was more appropriate. In other words, if we consider both perfunctory and reviewed citations as a single category labeled “nonessential,” the inter-coder reliability was 85%.

For each in-text citation, the following raw data were recorded into a spreadsheet: article number; author(s); year of publication; location of the in-text citation within the source article (using the exact terminology of the source article); whether the author was self-citing or not; category of citation function (according to Tabatabaei (2013)’s coding system); and the page, column, paragraph or line in which the in-text citation was found.

To confirm the accuracy of the raw data, we first sorted the data by article number and then by author name(s). Where entries were similar or identical, the spelling and sequence of names, as well as the publication date, were checked against the corresponding article’s in-text citation and reference list. In this way, misspellings and typos were identified and corrected.

At this point, a new data column, labeled “Section,” was added in which the headings recorded in the Location column (as labeled by the individual citing authors) were each assigned to one of the following sections that represent the typical overall structure of scientific papers (Doumont, 2010; Suppe, 1998): Introduction, (Theoretical) Background, Related Work/Literature Review, Methodology, Findings/Results, and Discussion/Conclusion. In cases where author-labeled headings, such as “algorithm description,” “data analysis,” “experiments,” and “limitations,” did not translate directly into the sections we identified, the content under the author-labeled heading was carefully examined to determine its purpose within the overall structure of the article, and therefore which of our five sections was most appropriate. For example, article 5 did not include a Methodology heading, however, it had two author-labeled headings: “Algorithm Description” and “System Performance Analysis.” These sections contained detailed information about the algorithm the authors used in their experiment and how they analyzed the resulting data, i.e. their “methods” for obtaining and analyzing their data. As these two headings also came before the “Experimental Results” heading, we assigned these two headings to our Methodology section. It should be noted that not all content under a particular heading had in-text citations. For example, there were no in-text citations under the Results heading in article 10.



Research Paper

A separate data table was created, which counted the number of times a specific reference appeared throughout the article in each of the five function categories: Applied, Contrastive, Supportive, Reviewed, or Perfunctory. The overall count of each reference was considered the “citation frequency.” Each reference was then put into one of the five citation frequency categories: uni-citation, 2 citations, 3 citations, 4 citations, and 5+ citations. For example, the cited reference Shenton and Dixon (2003) in article 3 appears four times as Contrastive, once as Supportive, and once as Reviewed, for a total citation frequency of six. It was thus assigned to the frequency category “5+ citations.” At this stage, each reference was assigned to the function category of the highest impact. For the example just mentioned, the cited reference was assigned to the Contrastive function, its highest impact function category.

The data were then ready to be analyzed according to the section, function, and frequency, which we did using the pivot tables function in Microsoft Excel.

3 Results and Discussion

3.1 Overall

Most references were cited only once in the text as clearly shown by Table 1, which presents the distribution of in-text citations by the frequency with which they appear. Among the 1,473 in-text citation occurrences, 531 (36%) were uni-citations. The other 942 citation occurrences represented only 278 unique citations. Among the 809 unique citations, 66% were uni-citations.

Table 1. Distribution of in-text citations by the frequency with which they appear.

| Citation frequency | 1 | 2 | 3 | 4 | 5+ | All |
|------------------------------|-----|-----|-----|-----|-----|--------------|
| # of unique cited references | 531 | 130 | 64 | 37 | 47 | 809 |
| % of unique cited references | 66% | 16% | 8% | 5% | 6% | 100% |
| Total in-text occurrences | 531 | 260 | 192 | 148 | 342 | 1,473 |
| % of the in-text occurrences | 36% | 18% | 13% | 10% | 23% | 100% |

As shown in Table 2, most citation occurrences (67%) that we examined functioned as either perfunctory or reviewed, and only a small percentage of references were essential to the citing paper (e.g. only 16% were categorized as Contrastive or Applied citations). This result is in line with findings from previous studies (e.g. Teufel et al., 2006).

Table 2. Distribution of all in-text citation occurrences by function.

| | Perfunctory | Reviewed | Supportive | Contrastive | Applied | Total |
|-----------------------------------|-------------|----------|------------|-------------|---------|--------------|
| # of in-text citation occurrences | 230 | 759 | 249 | 79 | 156 | 1,473 |
| % of in-text citation occurrences | 15.6% | 51.5% | 17.0% | 5.0% | 11.0% | 100% |

3.2 Uni-citations Corresponding to Nonessential Citations

Table 3 and Figure 1 show the number and percentage of unique references in the five frequency categories divided by the five function categories. As explained earlier, the function of each unique reference that was cited multiple times in the citing paper is represented by the one that has the highest impact.

Table 3. Distribution of unique cited references by in-text frequency and function.

| Frequency | Perfunctory | Reviewed | Supportive | Contrastive | Applied | Total |
|--------------|------------------|------------------|------------------|----------------|------------------|-------------------|
| 1 citation | 108 (21%) | 255 (48%) | 102 (19%) | 12 (2%) | 54 (10%) | 531 (100%) |
| 2 citations | 13 (10%) | 59 (45%) | 30 (23%) | 11 (9%) | 17 (13%) | 130 (100%) |
| 3 citations | 2 (3%) | 30 (47%) | 12 (19%) | 7 (11%) | 13 (20%) | 64 (100%) |
| 4 citations | 0 (0%) | 9 (24%) | 7 (19%) | 10 (27%) | 11 (30%) | 37 (100%) |
| 5+ citations | 0 (0%) | 9 (19%) | 10 (21%) | 10 (21%) | 18 (39%) | 47 (100%) |
| Total | 123 (15%) | 362 (45%) | 161 (20%) | 50 (6%) | 113 (14%) | 809 (100%) |

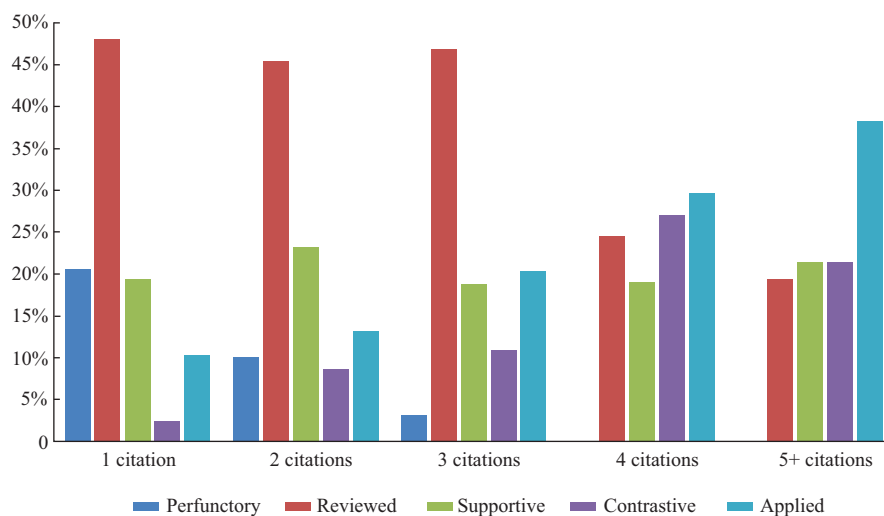


Figure 1. Percentage of unique cited references of different in-text frequency numbers by function.

Clearly, most uni-citations (69%) were nonessential (i.e. perfunctory or reviewed) to the citing papers. However, re-citation analysis that removes all uni-citations from the analysis, would have excluded 31% of uni-citations unfairly because they—as Supportive, Contrastive or Applied citations—had a substantial influence on the citing paper. Considering that 36% of all in-text citation occurrences were uni-citations (Table 1) and 67% were nonessential citation occurrences (Table 2), re-citation analysis would filter out 37% of all nonessential in-text citation occurrences, but it would also remove 34% of all in-text citation occurrences that



Research Paper

represent a substantial influence on the citing papers. It appears that removing uni-citations from a citation analysis is not an effective approach to filtering out nonessential citations.

3.3 In-text Citation Frequency Corresponding to Likelihood of Citations Having Substantial Influence on Citing Papers

Table 3 and Figure 1 also show that the likelihood of citations serving a function that indicates a more significant influence on the citing paper does seem to increase with in-text citation frequency. For example, 39% and 19% of the references that appeared five or more times in a citing paper functioned at least once as Applied or nonessential (Perfunctory or Reviewed), respectively, as compared to 10% and 69% for uni-citations.

This trend can be seen even more clearly in Table 4 and Figure 2, which show the percentage of unique citations in the five frequency categories appearing in each of the following function categories (from left to right in each frequency group in Figure 2): Perfunctory only (all in-text occurrences of a cited reference were perfunctory), Nonessential (all occurrences of a cited reference were either Perfunctory or Reviewed), Applied (at least one occurrence of a cited reference was Applied), Applied or Contrastive (at least one occurrence of a cited reference was either Applied or Contrastive), and Applied or Contrastive or Supportive (at least one occurrence of a cited reference was Applied or Contrastive or Supportive). For example, among the 37 unique references that were each cited four times in a citing paper, eight (22%) were cited in a nonessential function every single time (of the four times), and 21 (57%) were each cited at least once as Applied or Contrastive. Since these categories are not mutually exclusive, the percentages do not add up to 100% and the sum of the numbers is different from the total.

Table 4. Distribution of unique cited references by in-text frequency and levels of influence.

| Levels of influence | Citation frequency | | | | |
|--------------------------------------|--------------------|------------|-----------|-----------|-----------------|
| | 1 | 2 | 3 | 4 | 5+ |
| Perfunctory only | 108 (20%) | 13 (10%) | 2 (3%) | 0 (0%) | 0 (0%) |
| Perfunctory or reviewed only | 363 (68%) | 53 (41%) | 20 (31%) | 8 (22%) | 5 (11%) |
| Applied | 54 (10%) | 17 (13%) | 13 (20%) | 11 (30%) | 18 (38%) |
| Applied or contrastive | 66 (12%) | 28 (22%) | 20 (31%) | 21 (57%) | 28 (60%) |
| Applied or contrastive or supportive | 168 (32%) | 58 (45%) | 32 (50%) | 28 (76%) | 38 (81%) |
| Total unique | 531 | 130 | 64 | 37 | 47 |

Clearly, the percentage of cited references that only had a nonessential function (first two bars in each frequency category in Figure 2) decreases, and the percentage of cited references that had a substantial influence (next three bars) increases



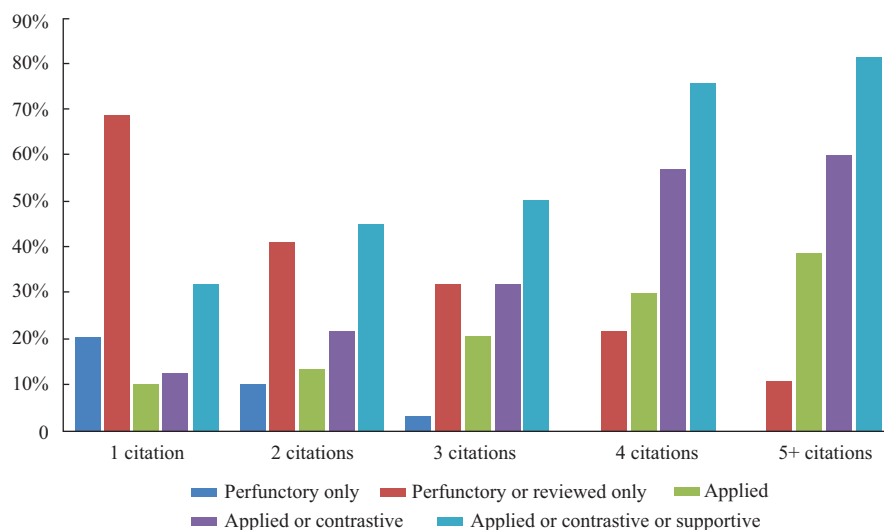


Figure 2. Percentage of unique cited references of different in-text frequency by level of influence.

with in-text frequency. This result supports the assumption underlying the signal amplification approach to frequency-weighted citation analysis. However, 41%, 31%, 22%, and 11% of cited references that each appeared twice, 3 times, 4 times, and 5+ times, respectively, were cited purely in a nonessential function. That means that a large percentage (31%) of cited references that appeared more than once in the citing text would be weighted higher than their true value to the citing paper when using frequency-weighted citation counting. And this problem of overweighting would be even more serious for cited references in the high frequency groups or when N^2 instead of N is used as the weight. This result shows that identifying and filtering out nonessential citations is also important for the commonly used signal amplification approach to frequency-weighted citation analysis, due to a high incidence of nonessential citations that appear more than once in the citing paper.

3.4 Other Factors that Might Help Identify Nonessential Citations

Since filtering out nonessential citations is so important to weighted citation analysis, and as removing uni-citations does not seem to be effective for this purpose, we were curious about whether there are other factors that might help identify nonessential citations. We explored one factor here: citation location, the section (Introduction, Methodology, etc.) in which an in-text citation appears.

Table 5 and Figure 3 show how in-text citations functioned within each section overall. For example, 25% of all in-text citation occurrences contained in the Methodology section functioned as Applied.



Research Paper

Table 5. In-text citations by function and location.

| Section | Perfunctory | Reviewed | Supportive | Contrastive | Applied | Total |
|---------------------------------------|------------------|------------------|------------------|----------------|------------------|---------------------|
| Introduction | 125 (32%) | 205 (52%) | 30 (8%) | 18 (5%) | 14 (4%) | 392 (100%) |
| Background | 5 (3%) | 175 (94%) | 5 (3%) | 1 (1%) | 0 (0%) | 186 (100%) |
| Related Studies/ Literature Review | 61 (21%) | 212 (72%) | 14 (5%) | 4 (1%) | 3 (1%) | 294 (100%) |
| Methodology | 9 (2%) | 120 (32%) | 126 (34%) | 24 (6%) | 92 (25%) | 371 (100%) |
| Results/Findings | 9 (8%) | 23 (19%) | 42 (36%) | 9 (8%) | 35 (30%) | 118 (100%) |
| Discussion/Conclusion | 21 (19%) | 24 (21%) | 32 (29%) | 23 (21%) | 12 (11%) | 112 (100%) |
| Total | 230 (16%) | 759 (52%) | 249 (17%) | 79 (5%) | 156 (11%) | 1,473 (100%) |

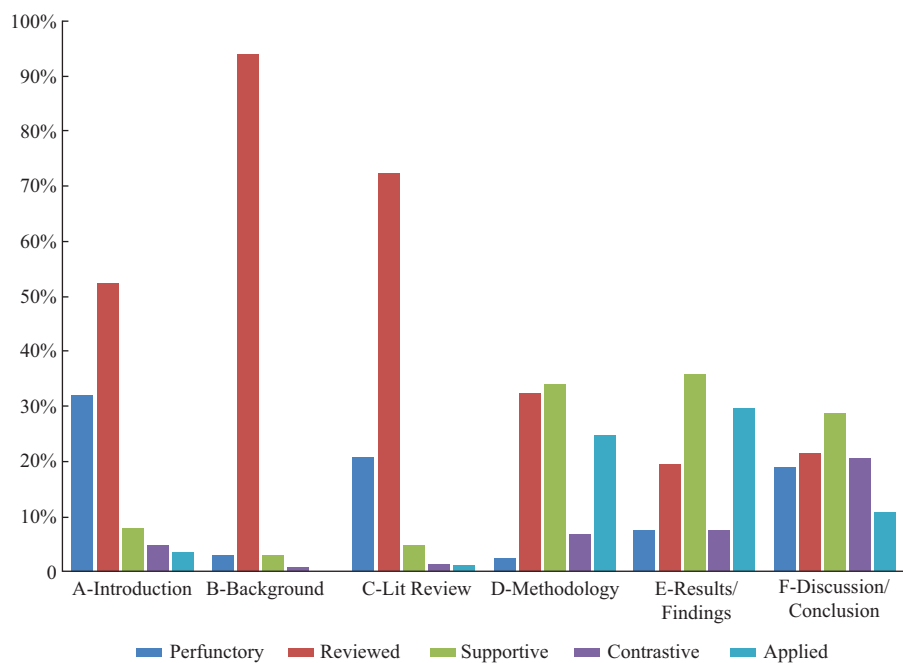


Figure 3. Percentage of in-text citations by location.



As seen from these data, 97% of in-text citation occurrences in the Background section, 93% in the Related Studies/Literature Review section, and 84% found in the Introduction section, functioned as nonessential (i.e. either Perfunctory or Reviewed). In comparison, only 34%, 27%, and 40% in each of the Methodology, Results/Findings, and Discussion/Conclusion sections, respectively, functioned as nonessential. Our data support findings from previous studies that citations in the Methodology, Results, Discussion, or Conclusion sections may play a more significant or influential role than those located in Introduction section.

If we remove (or ignore) all in-text citation occurrences in the Background and Related Studies/Literature Review sections, 46% of citation occurrences that functioned as nonessential would be filtered out, and only 5.6% of all citation occurrences that had a substantial influence (i.e. Supportive, Contrastive and Applied combined) would be lost. This 46% filtration rate with a 5.6% error rate is much better than the 37% filtration rate with a 34% error rate provided by removing uni-citations, as mentioned earlier.

The Introduction section is less straightforward due to the lower percentage of nonessential citations in this section than in the Background and Literature Review sections. Removing all citations in this section would improve the filtration rate to 79%, but the error rate would be more than tripled from 5.6% to 18%. If we only remove uni-citations there, the filtration rate would be improved to 60%, with only a slight increase in the error rate to 7.9% since 93% of uni-citations in this section are nonessential.

Clearly, removing citations by location is more effective for filtering out nonessential citations than removing uni-citations as in re-citation analysis. Removing all citation occurrences in the Background and Literature Review sections, and all uni-citations in the Introduction section, appears to provide a good balance between filtration and error rates.

4 Conclusion

Central amongst critiques of the current practices of citation analysis has long been that it treats all citations equally, be they crucial to the citing paper or perfunctory. Weighting citations by how they are used in the citing paper has therefore long been proposed as a theoretical solution to this problem and has attracted increasing research interest in recent years. The present study tested the basic assumptions underlying frequency-weighted citation analysis through a case study in the LIS field. All in-text citations in 14 research articles published in *JASIST* 2016 issue 1 were manually coded as to their function in the citing papers using a predefined coding scheme. As a case study of the LIS field, the present study suffers from the limitation of scalability and generalizability. However, we took careful measures to reduce the impact of other limitations of the data collection approach we applied.

Results from the present study support the assumption underlying the signal amplification approach to weighted citation analysis that the likelihood of citations having substantial influence on citing papers increases with their in-text citation frequency. However, a large percentage of multi-citations was found to play purely a nonessential role in the citing paper, and would be over-weighted by frequency-weighted citation counting. This finding underscores the importance of filtering out



Research Paper

nonessential citations before assigning weight in order to improve the accuracy and effectiveness of frequency-weighted citation analysis.

Removing citations by location was found to be more effective for filtering out nonessential citations than re-citation analysis that simply removes uni-citations. It was found that uni-citations correspond to nonessential citations only to a degree, and that re-citation analysis would suffer from too large an error rate to be effective. In comparison, removing all citation occurrences in the Background and Literature Review sections and all uni-citations in the Introduction section, appears to provide a good balance between filtration and error rates.

Weighted citation analysis promises to improve citation analysis for research evaluation, knowledge network analysis, knowledge representation, and information retrieval, as mentioned earlier in this paper and explained in detail in Zhao and Strotmann (2015a). Results from the present study showed, for the first time, the importance of filtering out nonessential citations before assigning weight in a weighted citation analysis, pointing research in this area in a new direction. Future studies are invited to explore effective ways to filter out nonessential citations, and to evaluate the differences that filtering out nonessential citations before assigning weight can make in the areas that weighted citation analysis promises to improve.

Author Contributions

D. Zhao (dzhao@ualberta.ca) proposed the research problems, designed the research framework and methodology, analyzed the data, and wrote the manuscript. A. Cappello (cappello@ualberta.ca) coded functions and locations of in-text citations, produced the tables and charts included in the paper, wrote part of the methodology section, edited the manuscript, and formatted it as per the journal's requirements. L. Johnston (lucinda.johnston@ualberta.ca) recorded the in-text citations, coded their functions, locations and self-citations, cleaned the data, produced tables and charts for preliminary analyses, and wrote part of the methodology section.

References

- Bertram, S. (1972). Citations counts. In A. Pitemick (Ed.), Fourth Annual Meeting, American Society for Information Science, Western Canada Chapter (pp. 61–67). Vancouver: University of British Columbia.
- Bonzi, S. (1982). Characteristics of a literature as predictors of relatedness between cited and citing works. *Journal of the American Society for Information Science*, 33(4), 208–216.
- Borgman, C.L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36(1), 3–72.
- Bornmann, L., & Daniel, H.-D. (2008). What co-citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80.
- Boyack, K.W., Small, H., & Klavans, R. (2013). Improving the accuracy of co-citation clustering using full text. *Journal of the American Society for Information Science and Technology*, 64(9), 1759–1767.



- Brooks, T.A. (1985). Private acts and public objects: An investigation of citer motivations. *Journal of the American Society for Information Science*, 36(4), 223–229.
- Brooks, T.A. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science*, 37(1), 34–36.
- Cano, V. (1989). Citation behavior – Classification, utility, and location. *Journal of the American Society for Information Science*, 40(4), 284–290.
- Case, D.O., & Higgins, G.M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, 51(7), 635–645.
- Cronin, B. (1984). The citation process. The role and significance of citations in scientific communication. London: Taylor Graham.
- Chubin, D.E., & Moitra, S.D. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social Studies of Science*, 5(4), 423–441.
- Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. *Information Processing and Management*, 47(1), 80–96.
- Ding, Y., Liu, X., Guo, C., & Cronin, B. (2013). The distribution of references across texts: Some implications for citation analysis. *Journal of Informetrics*, 7(3), 583–592.
- Doumont, J. (Ed.). (2010). *English communication for scientists*. Cambridge: NPG Education. Retrieved on September 22, 2016, from <http://www.nature.com/scitable/ebooks/english-communication-for-scientists-14053993>.
- Dubin, D. (2004). The most influential paper Gerard Salton never wrote. *Library Trends*, 52(4), 748–764.
- Edge, D. (1979). Quantitative measures of communication in science: A critical review. *History of Science Cambridge*, 17(36), 102–134.
- Garfield, E. (1962). Can citation indexing be automated? *Essays of an Information Scientists: 1962–1973*, 84–90. Retrieved on September 22, 2016, from <http://www.garfield.library.upenn.edu/allvols.html>.
- Garfield, E. (1979). *Citation indexing – Its theory and application in science, technology, and humanities*. New York: John Wiley & Sons.
- Griffith, B.C. (1990). Understanding science: Studies of communication and information. In C.L. Borgman (Ed.), *Scholarly Communication and Bibliometrics* (pp. 33–45). Newbury Park: Sage Publications, Inc.
- Hall, B.H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, 36 (1), 16–38.
- Hanney, S., Frame, I., Grant, J., Buxton, M., Young, T., & Lewison, G. (2005). Using categorizations of citations when assessing the outcomes of health research. *Scientometrics*, 65(3), 357–379.
- Harwood, N. (2008). Citers' use of citees' names: Findings from a qualitative interview-based study. *Journal of the American Society for Information Science and Technology*, 59(6), 1007–1011.
- Herlach, G. (1978). Can retrieval of information from citation indexes be simplified? Multiple mention of a reference as a characteristic of the link between cited and citing article. *Journal of the American Society for Information Science*, 29(6), 308–310.
- Hou, W., Li, M., & Niu, D. (2011). Counting citations in texts rather than reference lists to improve the accuracy of assessing scientific contribution. *BioEssays*, 33(10), 724–727.



Research Paper

- Jeong, Y.K., Song, M., & Ding, Y. (2014). Content-based author co-citation analysis. *Journal of Informetrics*, 8(1), 197–211.
- Liu, M. (1993). The complexities of citation practice: A review of citation studies. *Journal of Documentation*, 49(4), 370–408.
- McCain, K.W., & Turner, K. (1989). Citation context analysis and aging patterns of journal articles in molecular genetics. *Scientometrics*, 17(1), 127–163.
- Merton, R.K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, 1(1), 115–126.
- Moravcsik, M.J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, 5(1), 86–92.
- Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Washington, DC: Computer Horizons.
- Peritz, B.C. (1992). On the objectives of citation analysis: Problems of theory and method. *Journal of the American Society for Information Science*, 43(6), 448–451.
- Prabha, C.G. (1983). Some aspects of citation behavior – A pilot-study in business administration. *Journal of the American Society for Information Science*, 34(3), 202–206.
- Shadish, W.R., Tolliver, D., Gray, M., & Gupta, S.K.S. (1995). Author judgements about works they cite: Three studies from psychology journals. *Social Studies of Science*, 25(3), 477–498.
- Small, H. (1977). A co-citation model of a scientific specialty: A longitudinal study of collagen research. *Social Studies of Science*, 7(2), 139–166.
- Small, H. (1982). Citation context analysis. In B.J. Dervin & M.J. Voigt (Eds.), *Progress in Communication Sciences*, 3, (pp. 287–310). Norwood: Ablex.
- Suppe, F. (1998). The structure of a scientific paper. *Philosophy of Science*, 65(3), 381–405. Retrieved on September 22, 2016, from <http://www.jstor.org/stable/188275>.
- Tabatabaei, N. (2013). *Contribution of information science to other disciplines as reflected in citation contexts of highly cited JASIST papers*. Montreal. (McGill University Ph.D. dissertation)
- Tang, R., & Safer, M.A. (2008). Author-rated importance of cited references in biology and psychology publications. *Journal of Documentation*, 64(2), 246–272.
- Teufel, S., Siddharthan, A., & Tidhar, D. (2006). Automatic classification of citation function. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (pp. 103–110). Stroudsburg, PA: Association for Computational Linguistics.
- Tranöy, K.E. (1980). Norms of inquiry: Rationality, consistency requirements and normative conflict. In *Rationality in Science* (pp. 191–202). Springer Netherlands.
- Vinkler, P. (1987). A quasi-quantitative citation model. *Scientometrics*, 12(1), 47–72.
- Voos, H., & Dagaev, K.S. (1976). Are all citations equal? Or Did we op. cit. your idem? *Journal of Academic Librarianship*, 1(6), 19–21.
- White, H.D. (1990). Author co-citation analysis: Overview and defense. In C.L. Borgman (Ed.), *Scholarly Communication and Bibliometrics*, 84, 106. Newbury Park: Sage.
- White, M.D. & Wang, P.L. (1997). A qualitative study of citing behavior: Contributions, criteria, and metalevel documentation concerns. *Library Quarterly*, 67(2), 122–154.
- Zhao, D., & Strotmann, A. (2015a). Re-citation analysis: Promising for research evaluation, knowledge network analysis, knowledge representation and information retrieval? *Proceedings of the 15th International Society for Scientometrics and Informetrics Conference*, June 30–July 3, 2015, Istanbul, Turkey.



- Zhao, D., & Strotmann, A. (2015b). Analysis and visualization of citation networks. Williston, VT: Morgan & Claypool Publishers.
- Zhao, D., & Strotmann, A. (2016). Dimensions and uncertainties of author citation rankings: Lessons learned from frequency-weighted in-text citation counting. *Journal of the Association for Information Science and Technology*, 67(3), 671–682.
- Zhu, X., Turney, P., Lemire, D., & Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology*, 66(2), 408–427.



This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

