Is Self-citation Biased? An Investigation via the Lens of Citation Polarity, Density, and Location

Lina Zhou[1]
Department of Business Information Systems and Operations Management
The University of North Carolina at Charlotte
9201 University City Blvd., Charlotte, NC 28223
Phone: 1(704)6871976
E-mail: lzhou8@uncc.edu

Uchechukwuka Amadi
Department of Technology Innovation Management
Northcentral University
2488 Historic Decatur Rd., Suite 100, San Diego, CA 92106
E-mail: amadi1@umbc.edu

Dongsong Zhang
Department of Business Information Systems and Operations Management
The University of North Carolina at Charlotte
9201 University City Blvd., Charlotte, NC 28223
E-mail: dzhang15@uncc.edu

Final Version Accepted by *Information Systems Frontier*

[1] Corresponding author.

Is Self-citation Biased? An Investigation via the Lens of Citation Polarity, Density, and Location

ABSTRACT

Traditional citation analysis methods have been criticized because their theoretical base of statistical counts do not reflect the motive or judgment of citing authors. In particular, self-citations may give undue credit to a cited article or mislead scientific development. This research aims to answer the question of whether self-citation is biased by probing into the motives and context of citations. It takes an integrated and fine-grained view of self-citations by examining them via multiple lenses — polarity, citation density, and location. In addition, it explores potential moderating effects of citation level and associations among location contexts of citations to the same work for the first time. We analyzed academic publications across different topics and disciplines using both qualitative and quantitative methods. The results provide evidence that self-citations are free of bias in terms of citation density and polarity uncertainty, but they can be biased with respect to positivity and negativity of citations. Further, they revealed impacts of self-citing behavior on some citation patterns involving citation density, location concentration, and associations. The examination of self-citing behavior from those fresh perspectives shed new lights on the nature and function of self-citing behavior.

Keywords: citation analysis, self-citation, bias, polarity, density, location

INTRODUCTION

Citation analysis has become an integral part of the decision making process in academic and scientific life as a basis for ranking authors, journals, and institutions, and even for making promotion decisions. In the past decades, a number of IS studies have employed citation analysis to provide the IS community a clear, up-to-date overview of research evolution, citation trends, scholarly discourse, and to determine journal rankings and how the field should be further advanced, and so on (Dwivedi, Lal, Mustafee, & Williams, 2009; Lowry, Moody, Gaskin, Galletta, Humpherys, Barlow, & Wilson, 2013; Wang, Liang, Jia, Ge, Xue, & Wang, 2016; Kapoor, Tamilmani, Rana, Patil, Dwivedi, & Nerur, 2018). Citations are extensively used to measure the quality and diffusion of information (Catalinia, Laceterab, & Oettlc, 2006). Explicit reference to prior literature is a way of acknowledging intellectual and cognitive influences of peer scholars on collective construction of new knowledge in a field (Maričić, Spaventi, Pavičić, & Pifat-Mrzljak, 1998). The counts of paper citations and/or downloading have been traditionally used as indicators of the visibility and impact of a paper (O'Leary, 2008). However, self-citation has incurred criticisms to the count based citation analysis due to its subjectivity to self-manipulation.

Self-citation can be manifested in many different forms (Ioannidis, 2015). Simply, self-citation is a type of citation in which the citing article and the referenced work have at least one author in common (Glänzel, Debackere, Thijs, & Schubert, 2006). In other words, self-citation generally refers to a situation whereby authors reference their own work. Statistics repetitively shows that average self-citation rates are significant, which range between 10 and 20 percent of all citations, depending on the discipline and its stage of development (see also (Hyland, 2003)). Inclusion of self-citations can lead to an increase of the h-index by 24 percent on average in several sub-areas

of informatics (Couto, Pesquita, Grego, & Veríssimo, 2009). On one hand, self-citations are perceived as self-serving and may not objectively reflect the assessment of value of the cited work (McCain & Turner, 1989). This raises concerns about the value of self-citations compared with citations to the works of others (Swales, 1986). On the other hand, counter-arguments are made that self-citation demonstrates authors' worth and previous effort and their continuance of research/publications in a particular subject area (Sammarco, 2008). Additionally, few differences were detected between the motivations for self-citation and those for citation to others' work (Bonzi & Snyder, 1991), and there were no meaningful changes in authors' impact based on the h-index after removing self-citations (Hirsch, 2005; Rad, Shahgholi, & Kallmes, 2012). Central to such a debate is the question about whether self-citations are biased.

Against the backdrop, this research investigates possible bias associated with self-citation via three lenses. One is *uncertainty* in citation polarity (e.g., positive vs. negative). The polarity of citation indicates the citing authors' position to the work cited. It has been argued that statistical counts do not reflect the motive or judgment of a citing author (Shadish, Tolliver, Gray, & Sen Gupta, 1995), and the count alone may give an ambiguous or misleading position of scientific developments (Moravcsik & Murugesan, 1975). Citation polarity, which manifests the citing motive and opinion, has the potential to address the above limitation. A second lens is the *location* that shows where the referenced work is cited in a citing article based on meta-structure of a paper (e.g., introduction). Citation location can provide context for understanding a citing relationship such as the role of a citation (Lutz & Hans-Dieter, 2008; Maričić et al., 1998). For instance, a citation located at the introduction section usually sets the stage for new contributions, while a citation in the method section implies utility of a research method (Cano, 1989). A third angle is citation *density* that indicates the extent to which a pair of citing and cited work are related.

This study makes multifold contributions to the bibliometrics research. First, it offers three new lenses for examining the potential bias of self-citation, namely polarity, density, and location of citations. These lenses, used both individually and as a whole, can enable a more complete and in-depth understanding of self-citing behavior, compared with the commonly used citation counts. Second, it introduces polarity uncertainty as a new citation measure to address a common situation where the same work is referenced more than once in a citing article. The measure affords a complete picture of the citing authors' position to the cited work. Additionally, compared with related previous studies that focused on citation count at an article level, this study takes a finer-grained view by measuring the density of citing-cited relationships. Further, the analysis of location correlations and association patterns of more than one citation to the same work in a citing article provides insights into the overall value of the cited work for the first time. Third, it uncovers the moderating effect of self-citing behavior on the location distribution of cited work. Fourth, this is the first study that explores the moderating effect of citation level (the level of citation count) on the influence of self-citation, which addresses the question of whether self-citation behavior is context-dependent. Fifth, the design of the data collection method presented in this study is particularly a good fit for understanding self-citing behavior. Previous citation studies collected either all references included in each citing paper (e.g., Cano, 1989; Catalinia et al., 2006) or one citing paper per source paper (e.g., Maričić et al., 1998). Our pool of article collection covers all citations to each cited paper, which allowed us to examine self-citing motives and effects thoroughly.

BACKGROUND AND RELATED WORK

Citation analysis

Citation analysis is a branch of bibliometrics research that systematically analyzes citations to and from documents (Diodato, 1994). Citation analysis has formed the basis for the assessment of quality of research, journals, and/or authors in the world of scholarship (Moravcsik & Murugesan, 1975) across science and technology fields (Eugene Garfield, 1972).

The citations of a paper are usually quantified in terms of citation count across different major citation indexing resources such as Web of Knowledge, Elsevier Scopus, and Google Scholar (Zhang, Ding, & Milojević, 2013). A citation count indicates the frequency of citations to a paper by other articles (Fooladi et al., 2013). The citation counts of individual articles are further used to derive the impact of a journal and a research scholar. For instance, the impact factor of a journal is measured as the average number of articles published in the journal from two preceding years that are cited in the year under review (Fooladi et al., 2013; Hodge & Lacasse, 2011; Whitehouse, 2001). H-index measures the number of a scholar's publications each having at least *h* citations and his/her remaining publications each having fewer than h+1 citations (Hirsch, 2010; Waltman & Eck, 2012; C.-T. Zhang, 2013). The application of h-index has been extended to journals and groups of researchers (Hodge & Lacasse, 2011; Waltman & Eck, 2012). However, there has been decade-long discussion about effectiveness of the citation analysis practice that largely hinges on citation counts without looking into the content or context of citations (Cano, 1989; Harzing & van der Wal, 2009). One of the areas that has attracted much attention is self-citation (Cole & Cole, 1971; Hodge & Lacasse, 2011) that can be manipulated to inflate the citation counts or research impact (Bartneck & Kokkelmans, 2010).

Self-citation

Citation in itself is not just an acknowledgement of existence of previous work, but also an establishment of connection (relationship) to the work cited (Osareh, 1996). The relationship

between cited work and a citing article has been a subject of discourse in citation analysis (Osareh, 1996).

Self-citation is relatively under studied despite its consequences in research assessment (Rodríguez-Ruiz, 2009). Models show that "each additional self-citation increases the number of citations from others by about one after one year, and by about three after five years" (Fowler & Aksnes, 2007). Handling self-citation has been factored into the validity of citation measures (see also (Rodríguez-Ruiz, 2009)). There have been calls for corrective measures that exclude self-citations in determining the real impact of an author in order to have an unbiased comparison among authors (Straub & Anderson, 2009). For instance, measures have been developed to exclude or discount self-citations from the computation of h-index (Brown, 2009; Ferrara & Romero, 2013). On the other hand, evidence from certain discipline has shown that the effect of self-citation in the h-index is minimal (Rad et al., 2012). Therefore, the question of whether self-citation is manipulative and gives undue credit to research has yet to be fully understood. Previous developments do not address a fundamental research question regarding self-citation: *Is self-citation biased?*

Citation Content Analysis

Citing constitutes an author's interpretation of the cited work, which is a process of meaning creation and symbol making (Small, 1978). An in-depth analysis of citations, which involves content and context analysis, is absolutely essential for assessing the intrinsic value of the original research (Rodríguez-Ruiz, 2009). Citation content analysis aims to "describe the contextual relationship between citing and cited works, investigate the social and intellectual interaction between authors, investigate the relative contribution of individual and collective norms to citing behavior, and understand the nature and function of such behavior" (Zhang et al., 2013).

Citation content analysis is empirical (Moravcsik & Murugesan, 1975), which generally can be accomplished using two basic types of approaches − inductive and deductive methods. The inductive approach involves the development of categories, level of abstraction, and coding (Elo & Kyngas, 2008; Mayring, 2000). The identification of such categories requires reading through the material and developing headings as appropriate to describe various aspects of the work. The deductive approach involves establishing connections with the text through the analysis of concepts, hypotheses, methods, and so on. This is done by reviewing the content of text to determine correspondence to the developed categories (Mayring, 2000). The inductive methods are often used to categorize authors' citation motivation, and the deductive methods to prove or analyze such categorization (Zhang et al., 2013). We chose a deductive approach to examining self-citations by building upon existing categorizations of citation motivations.

Citation Classification and Polarity

The deductive approach to citation content analysis requires the establishment of categories, and more fundamentally the basis of categorization. The potential bias of self-citation, if any, can be rooted in the author's or citer's motives or intention for citing a work. In the remainder of this section, we first introduce citation motives or behavior, then review related classification schemes, and finally introduce a polarity based classification scheme.

Citation Motivations. An author makes his/her citing decision for various reasons and under various circumstances (Shadish et al., 1995). Hence it is very important to understand particular circumstances surrounding citations or why and in what context the paper is cited (Brooks, 1986; Nicolaisen, 2007; Rodríguez-Ruiz, 2009). For instance, Garfield (1972) identified a list of 15 motives or reasons for citing, including paying homage to pioneers, giving credit to related work, providing background reading, identifying methodology and equipment, and so on, criticizing

previous work, correcting one's own work, correcting work of others, substantiating claims, alerting to forthcoming work, providing leads to poorly disseminated, poorly indexed, or uncited work, authenticating data and classes of facts, identifying original publication in which an idea or concept was discussed, identifying the original publication describing an eponymic concept or term, disclaiming work or ideas of others, and disputing priority claims of others (see also (Case & Higgins, 2000)).

Citation Classification. There exist several classification schemes of citations, which are summarized in Table 1. Muravcsik and Murugesan (1975) revealed that there were more conceptual than operational citations, and more evolutionary (60%) than juxtapositional citations. Additionally, perfunctionary accounted for two-fifths, and negational one-seventh (14%) of the citations. Among the six sub-categories of Chubin and Moitra (1975), the first four belonged to the affirmative category, and the last two belonged to the negational category. Based on their results of a case study, about 5% of citations were negational (all were partial) and 95% were affirmative that spread across all of its sub-categories. Brooks (1985) showed that persuasiveness played a dominant role in motivating citations, and conversely, negative credit and social consensus (citing because many other works cited the same work) were largely overlooked. Swale (1986) introduced the dimension of "extensive vs. short" based on the length of citation and germane post-citation discussion. The scheme also incorporated zero as the third category to extend two of the dimensions from (Moravcsik & Murugesan, 1975), namely evolutionary vs. juxtapositional and confirmative vs. negational.

Scheme	Basis	#categories	Dimensions/categories
Muravcsik and Murugesan (M&M) (1975)	Context of citation	4	conceptual vs. operational organic vs. perfunctionary evolutionary vs. juxtapositional

			confirmative vs. negational
Chubin and Moitra (1975)	Refinement of *confirmative or negational* in M&M	6	basic citation subsidiary citation additional information perfunctory citation partial citation total citation
Brooks (1985)	Citer motivations	7	currency scale negative credit operational information persuasiveness positive credit reader alert social consensus
Swale (1986)	Citation length, germane post-citation discussion, and M&M	3	extensive vs. short evolutionary vs. juxtapositional vs. zero confirmative vs. negational vs. zero

Table 1. A summary of classification schemes of citations

The current classification schemes covers a wide range of motivational contexts of citations. Nonetheless, they also highlight the lack of standards in assessing authors' motives in support of evaluative citation analysis. To overcome the issue, we propose to use polarity, a common core to the above-mentioned schemes, as the basis for categorizing citations.

Polarity. Polarity culminates citing authors' judgment on a cited work. Our survey of various citation classification schemes reveals a convergence of citing motives into distinct polarity values. For instance, positive assessment captures a number of specific categories (see Table 1) such as giving credit to relative work (Eugene Garfield, 1972), confirmative (Moravcsik & Murugesan, 1975; Swales, 1986), affirmative (additional information) (Chubin & Moitra, 1975), and positive credit (Brooks, 1985); negative assessment represents categories such as criticizing previous work,

disclaiming work or ideas of others, disputing priority claims of others (Eugene Garfield, 1972), and negational (partial and total) (Chubin & Moitra, 1975; Moravcsik & Murugesan, 1975; Swales, 1986); and neutral statement describes categories such as zero (Swales, 1986). Accordingly, our conceptualization of the polarity of citations consists of three categories:

- Positive citation − uphold the view of what is referenced in the cited work such as concepts and method in a favorable light.
- Negative citation − dispute, fault or attempt to invalidate facts, concepts, information or method of what is referenced in the cited work.
- Neutral citation − the citing author noted the view or use of other authors' position about a cited work without stating his/her own view.

Although many studies of citation classification schemes referred to the citing author's polarity to the work cited, they did not employ it as the basis for their work. In other words, the polarity of citation has been one area of unconscious application of evaluative citation analysis. This study explicitly employed polarity as a dimension to investigate self-citing behavior.

SELF-CITATIONS AND RESEARCH QUESTIONS

Self-citations can boost the citation count of a source paper. A study showed that 13% of astronomy and 10-15% of linguistics citations were self-citations (Swales, 1986). Author self-citations account for approximately 6.7 percent of articles in high-impact general medicine journals (Kulkarni, Aziz, Shams, & Busse, 2011). As a result, self-citation has been recognized as one of the major limitations of traditional statistical methods for citation analysis (Whitehouse, 2001). The latter are also subject to various other manipulative citing behaviors such as citations

not related to the paper (Rodríguez-Ruiz, 2009) and mere perfunctory citations (Moravcsik & Murugesan, 1975).

The above-mentioned limitations are rooted in the inability of count based citation analysis (Butler & Visser, 2006) to take into account the citing author's intention or motives for citing (Shadish et al., 1995; Zhang et al., 2013). In other words, such a practice neglects the opinions of those who cite the papers. Consequently, traditional methods may inflate the impact of a cited paper by failing to account for disparate opinions that citing authors may express on the cited work. Take the paper (Bowers, 2009) as an example. Among others, two of its citing articles (e.g., Plaut & McClelland, 2010; Lehky, Kiani, Esteky, & Tanaka, 2011) held divergent positions on the work; and a third one (e.g., Quiroga & Kreiman, 2010) held an opposite opinion on the cited work to that of (Plaut & McClelland, 2010), which in turn triggered counter-arguments by the original author in a fourth citing article (i.e., Bowers, 2010). In the traditional count based method, the four citing articles, including one self-citation, would contribute 4 towards the total citation count of the cited work. However, this number can be misleading due to the mixed viewpoints held by different citing authors. A controversial paper tends to attract much criticism or support by other scholars, which is to its advantages.

We address the limitations of traditional methods for citation analysis by investigating self-citations from the perspectives of polarity, citing location, and citing density. Incorporating the polarity of authors' viewpoints into the analysis of self-citations can help assess the impact of cited work more precisely by measuring the degree of support. Polarity uncertainty refers to citing authors' difficulty in evaluating the quality of cited work and the implications of the cited work for future research and practice. This is resulted from authors' inability to describe the cited work, particularly in terms of contributions and limitations, in a coherent manner. Authors are expected

to have a lower uncertainty about their own work than non-authors. In addition, self-citations are not only perceived as a kind of self-serving behavior (McCain & Turner, 1989), but also rhetorically reflect the promotional strategies of individuals (Hyland, 2003). Thus, self-citations could be viewed as reporting with positive bias. Our first set of research questions is as the following:

RQ1-1. Do self-citations reduce polarity uncertainty of the cited work?

RQ1-2. Do self-citations lead to positive bias to the cited work?

The density of citing can be measured precisely as the number of citations to a referenced work in a citing article. The density of citing is important for understanding the strengths and influence of the cited on the citing paper. Presumably, the higher the citation density, the stronger is the influence of the cited work on the citing article. However, this aspect has been under explored in traditional count-based citation analysis. Following similar rationales for the influence of self-citation on citation count, the behavior of citing own work may lead to an elevated citing density, which motivates our next research question:

RQ2. Do self-citations increase citation density?

Cano (1989) introduced citation location as a bibliometric element. Nevertheless, there is no consensual definition of reference location in the literature. Cano (1989) roughly categorized the areas of citation into beginning (up to the 15th percentile based on the length of the citing paper), middle (from the 20th to the 75th percentile), and end sections (from the 80th percentile). The study also found that citation concentration in a citing article steadily declined after the 15th percentile location mark. A pragmatic approach was developed by mapping locations to particular sections of a citing paper's structure (Maričić et al., 1998; McCain & Turner, 1989). Nevertheless,

differences exist in paper meta-structure. For instance, McCain & Turner (1989) grouped results and discussion as one location, whereas Maričić et al. (1998) kept them separate.

By analyzing the location distribution of self-citations, we introduce a structural element into the pure citation-counting approach to self-citation. Citations serve differing roles at various locations of a paper. For instance, citations in the introduction section are typically used to set the stage for new research to be investigated; those in the method section are used to inform the method design; and citations appearing in the discussion are to relate own findings to those of the cited papers and provide alternative explanations for any possible differences. We extended the taxonomies of location (Maričić et al., 1998; McCain & Turner, 1989) by incorporating two new elements: related work and conclusion. These location elements, driven by evolving publishing practice, serve distinctive functions. Citations in the related work are used to present the background literature in relation to which their new contribution is to be examined and to establish the need for the present work based on critical assessments of bodies of other relevant work; and citations in conclusion are used to highlight main findings, convey the larger significance of one's research, and recommend directions for future research. Therefore, location is a useful lens for examining self-citing behavior. The question of whether self-citations follow different patterns from non-self-citations with respect to citing location becomes interesting:

RQ3. Do self-citations follow different location patterns from non-self-citations?

A work may be cited with more than one motive. In other words, the same work may serve multiple functions in a citing article. One of such manifestations is that citations to the same work occur at different locations of a citing article. Accordingly, it would be interesting to understand the patterns of location associations of two or more citations to the same work. It is reasonable to assume that self-citers both have access to and develop a thorough understanding of the full body

of the cited work, ranging from research motivations, through method design, to findings and conclusions. This may not be the case for other citers. Thus, we are interested in exploring the following question:

RQ4. Do self-citations change the association patterns between different location elements?

Citation level is derived from citation count by discretizing the latter into three levels: low, medium, and high. The effect of citation count or citation level on self-citation is much less understood than that in the opposite direction. By viewing citation level as a contextual factor, we aim to gain an understanding of whether the context may influence the effects of self-citation and location. Specifically, we are interested in the interplays between citation level and location and between citation level and density, respectively, which remain unexplored in the literature. So our last set of research questions is as follows:

RQ5-1. Does citation level moderate the effect of self-citation on citation density?

RQ5-2. Does citation level moderate the effect of location on citation density?

METHOD

To answer the research questions, we employed mixed methods for citation content analysis (Zhang et al., 2013) that combined qualitative interpretation of symbolic representations embedded in citations and quantitative analysis of bibliographical references. These two complementary analyses can potentially help resolve some ambiguities and inconsistencies inherent in citation counting.

Paper Collection

The process of paper collection started with selecting cited papers, and then moved onto collecting all the citing articles to each of the selected cited work. To improve the generality of our research findings, our paper collection strategies aimed to diversify paper selection while meeting our research needs. In particular, the primary strategy was in accordance with common article search strategies of citing authors: topic-oriented and discipline-oriented strategies. Within each strategy, we adopted additional strategies to increase the coverage and diversity of paper collection.

The topic-oriented search strategy collects articles (i.e., cited work) based on topics. To this end, we selected two topics — deception detection (DD) and mobile user authentication (MA), for two main reasons: 1) the topics were relatively specific, which was conducive to effective citation content analysis, and 2) the two topics represented different dominant research methods, with deception being behavioral science research and authentication being design science research. In addition, the former topic has a much longer history of research than the latter.

The discipline-oriented strategy started with selecting disciplines, and more specifically, core journals of the disciplines. We selected three distinct disciplines: information systems (IS), applied linguistics (AL), and electrical engineering (EE) to "represent a broad cross-section of academic practice" (Hyland, 2003). Among them, AL is a representative of a soft field, EE represents a hard field, and IS falls in between. We identified some leading journals in each of the fields by consulting with experienced researchers in respective disciplines and by cross-referencing journal impact factors.

To allow for sufficient time to accumulate citation, we selected papers that were published within a selected timeframe. Given that the maximal citation echo is usually observed within six years after publication (Ferligoj, Maricˇic´, Pifat, & Spaventi, 1988), we implemented the

discipline-oriented strategy by collecting papers published 6 years prior to the year of data collection. For implementation of the topic-oriented strategy, we extended the timeframe to be 10 years. This is because top journal publications are expected to accumulate citations faster than publications on a specific topic (in any journal). We used Web of Science as the source of citation data primarily due to its ease of use and few duplicate entries (Sanderson, 2008).

To further diversify the collected papers, we selected papers at varying citation levels at the time of data collection. The strategy also allowed us to answer the research questions related to citation level. Based on the distribution of citation counts, the collected papers were divided into three citation levels: high (> 40), medium (15~40), and low levels (<15). We performed random stratified sampling by choosing papers within one standard deviation from the median at each citation level. The full texts of the citing articles were downloaded from ISI's Science Citation Index Extended, Social Science Citation Index, and Art & Humanities Citation Index.

Strategy	Topic/ Discipline	Citation level	Citation Count (total)	Cited Work	Citation density (total)	Citation density (mean [std])
Topic-oriented	DD	H	107	2	251	2.35 [2.185]
		M	64	3	105	1.64 [1.074]
		L	30	3	60	2 [2.407]
	MA	H	45	1	72	1.6 [1.074]
		M	25	1	57	2.28 [2.747]
		L	18	2	27	1.5 [0.618]
Discipline-oriented	AL	H	43	1	84	1.95 [1.479]
		M	21	1	37	1.76 [1.261]
		L	7	1	11	1.57 [0.976]
	EE	H	66	1	112	1.7 [1.189]
		M	28	1	40	1.43 [1.372]
		L	10	1	19	1.9 [1.524]
	IS	H	52	1	80	1.73 [1.122]
		M	26	1	37	1.42 [1.172]
		L	9	1	21	2.33 [2.958]
Total			551	21	1023	1.86 [1.652]

Table 2. Descriptive statistics of the selected papers and their citations

The descriptive statistics of the selected papers is reported in Table 2. As shown from the table, the dataset contained a total of 1,023 citations, resulting from 551 citing-cited pairs, and the average citation density was greater than 1. The maximum citation density for the high, medium, and low citation levels in the dataset were 13, 13, and 14, respectively. The statistics reported in the table were based on cleaned datasets that we prepared for citation analysis.

Data Cleaning and Analysis

We observed a number of different types of noise from the collected dataset, including citation error, redundancy, non-English (English title and abstract only), commentary, response to an article authored by other scholars, and editorial of a special issue where the cited work was published. Two common types of citation errors include mixed use of authors' first names and last names and articles listed in the references but not actually cited in the paper. Redundant articles occurred when two articles were nearly identical but both contributed to the citation count. Although a commentary expresses opinions and offers explanations, it tends to be short and devoid of the structural elements of a full article. The last two types of noise are prone to biases. Like a book review, a response article is devoted to discussing another article, which leads to an inflation of citation density. More importantly, the response that represents a different school of thoughts tends to be highly criticizing. On the contrary, the editorial of a special issue has a tendency of promoting the cited work published in the same issue. All those noise data were removed prior to citation content analyses.

In addition to noise, we also resolved ambiguities associated with author names and publications. For instance, a citing article referenced more than one article from the same group of authors, and those articles were published in the same year.

Citation content analysis was used to analyze all in-text citations of each cited paper across different locations of each citing paper to account for citation density and polarity uncertainty. Five coders, including two faculty researchers, two graduate students who were working on their thesis research, and one undergraduate student who had assisted with multiple research projects, performed manual analysis of the citation pairs independently. Each of the citing articles was first analyzed by two of the coders, then the inconsistent results were adjudicated by one of the researchers who had published across different disciplines.

Variables

This research involved five variables, including self-citation, citation location, density, and level, and polarity uncertainty.

- Self-citation was the situation whereby authors cited their own work or their co-authored work. It was coded as a binary variable.
- Citation location was operationalized as structural elements of a citing work where a source paper was cited. It consisted of five possible values: introduction, related work, method, discussion, and conclusion. In view of cross-disciplinary differences in terms of paper meta-structure, the scope of related work also covered theoretical foundation and hypotheses development, and the scope of discussion also covered results.
- Citation density was operationalized as the frequency of c_{ij}, reference *i* being cited in article *j*, as shown in equation (1).

$$d(c_{ij}) = \sum_{l\in\{i,r,m,d,c\}} f(c_{ij}, l), \qquad (1)$$

where $\{i, r, m, d, c\}$ denotes the set of locations; $f(c_{ij}, l)$ denotes the frequency count of c_{ij} in location *l*.

- Citation polarity was measured by three variables: polarity uncertainty, positivity, and negativity. Polarity uncertainty was measured by adapting Shannon's index of entropy (Shannon, 1948). Uncertainty is closely associated with the notion of diversity, but it is extremely difficult to explicitly measure diversity (Tang, Suganthan, & Yao, 2006; Zhou, Zhang, & Sung, 2013). Most of other measures such as Gini-Simpson indices (Blau, 2000) and disagreement statistics assume that different categories (e.g., polarity values) are exclusive. As a result, they have ignored the situation where more than one citation to the same work in a citing article could carry different polarity values. Entropy (Shannon, 1948) was selected in this study because it allowed for modeling polarity as a multi-valued attribute of a citation. The polarity uncertainty of c_{ij}, $u(c_{ij})$, and the overall polarity uncertainty of cited work i, $u(c_i)$, are defined in equations (2) and (3), respectively.

$$u(c_{ij}) = -\sum_{k\in\{+,\mp,-\}} p(c_{ij},k)\ln p(c_{ij},k) \quad (2)$$

$$u(c_i) = \frac{1}{N_i}\sum_{j\epsilon\, N_i} \mathrm{u}(c_{ij}), \quad (3)$$

where $\{+,\mp,-\}$ denote the polarity values of positive, neutral, and negative, respectively, and $p(c_{ij},k)$ is the probability of c_{ij} taking polarity k. According to the equations, $u(c_{ij})$ would be ln(|k|), if $p(c_{ij},k)$ was uniformly distributed across k; the value would be 0 if only a single value of k was present; and $u(c_i)$ was measured as the geometric mean of $u(c_{ij})$ across all citing articles $j \in N_i$ of i, where N_i is the citation count of i.

- Positivity $ps(c_{ij})$ was operationalized as the normalized total frequency counts of positive citations and half of neutral citations, and negativity $ng(c_{ij})$ was measured as

the normalized frequency count of negative citations. These definitions are shown in equations (4) and (5).

$$ps(c_{ij}) = \frac{1}{d(c_{ij})}\left(f(c_{ij},'+') + \frac{1}{2}f(c_{ij},'\mp')\right) \quad (4)$$

$$ng(c_{ij}) = \frac{1}{d(c_{ij})}f(c_{ij},'-') \quad (5)$$

where $f(c_{ij},\cdot)$ denotes the frequency count of c_{ij} with polarity value, and $d(c_{ij})$ is the normalization factor.

RESULTS

The overall self-citation ratio in the collected data was 9.98%, and the ratio in topic-oriented papers (13.5%) was higher than that in discipline-oriented papers (6.1%). We first tested our assumption about the influence of self-citation on citation count by performing a paired-sample t-test of citation count between before (*M*=26.24, *std*=19.06) and after removing self-citations (*M*=23.57, *std*=17.8). The results reveal a significant reduction in citation count after removing self-citations (p<.001), which confirmed our assumption.

The descriptive statistics of polarity uncertainty, positivity, negativity, and citation density with (i.e., Y) and without inclusion of self-citations (i.e., N) are reported in Table 3. The descriptive statistics of citation density with respect to the location of citations for self- and non-self-citations are reported in Table 4.

Topic/ Discipline	Including self-citation?	Polarity uncertainty	Positivity	Negativity	Citation density
DD	Y	.139 [.280]	.823 [237]	.035 [.147]	2.07 [1.958]
	N	137 [.283]	.808 [.245]	.04 [.157]	1.96 [1.861]
AU	Y	.161 [.302]	.73 [.231]	.016 [.078]	1.77 [1.687]
	N	.153 [.299]	.702 [.231]	.019 [.089]	1.79 [1.803]
IS	Y	.212 [.333]	.527 [.22]	.099 [.279]	1.7 [1.423]

	N	.16 [.292]	.521 [.217]	.102 [.282]	1.71 [1.438]
AP	Y	.146[.289]	.613 [.233]	.047 [.187]	1.86 [1.366]
	N	.12 [.272]	.608 [.236]	.049 [.191]	1.84 [1.389]
EE	Y	.211[.334]	.5 [.229]	.148 [.292]	1.64 [1.269]
	N	.20 [.322]	.497 [.241]	.16 [.306]	1.55 [1.052]

Table 3. Descriptive statistics (mean [std]) of polarity measures and citation density

To answer RQ1, we performed tests on three different dependent variables, namely polarity uncertainty, positivity, and negativity, by comparing their values with and without inclusion of self-citations. The results of a paired-sample T-test on polarity uncertainty of citations before (M=.159, std=.297) and after removing self-citations (M=.153, std=.293) did not yield any significant difference (p>.05). In addition, we also compared the polarity uncertainty of citations between self-citations (M=.215, std=.326) and non-self-citations, which yielded a similar result (p>.05). However, the tests on positivity and negativity showed different results in that self-citing behavior increased the positivity (p<.05) and lowered the negativity (p<.01) on the cited work. The results suggest a negative response to RQ1-1 and a positive response to RQ1-2.

To answer RQ2, we performed a paired-sample t-test of citation density between before (M=1.86, std=1.652) and after removal of self-citations (M=1.80, std=1.596). The analysis did not yield any significant difference (p > .05). Similarly, an independent sample t-test between the density of self-citations (M=2.38, std=2.032) and non-self-citations did not yield any significant difference (p>.05).

To answer RQ3, we first compared the location distribution between self-citations and non-self-citations, as shown in Table 4 and Figure 1. Compared with non-self-citations citations, self-citations appeared to have a higher concentration in method and discussion, but lower concentration in introduction and related work. A paired sample T-test confirmed the difference between concentrations of self-citations and other citations in method (p<.1) and discussion (p<.1).

Topic/ Discipline	Self-citation	I*	R	M	D	C
DD	Y	1 [.98]	.23 [.43]	.92 [1.41]	.62 [.852]	.04 [.196]
	N	.93 [.821]	.23 [.784]	.25 [.674]	51 [1.03]	.04 [.224]
AU	Y	.31 [.63]	.92 [.641]	.15 [.376]	.31 [.48]	0 [0]
	N	.48 [.704]	.87 [.977]	.12 [.366]	.32 [.791]	0 [0]
IS	Y	0 [0]	.5 [.707]	.5 [.707]	0 [0]	.5 [.707]
	N	.56 [.747]	.72 [.934]	.15 [.567]	.26 [.62]	.01 [.108]
AP	Y	1.33 [1.528]	.33 [.577]	.67 [.651]	0 [0]	0 [0]
	N	.5 [.938]	.79 [1.0]	.15 [.629]	.29 [.575]	.1 [.352]
EE	Y	.67 [.651]	.17 [.577]	.67 [.651]	.83 [2.329]	0 [0]
	N	1.03 [.805]	.27 [.537]	.13 [.398]	.12 [.571]	0 [0]

**: All the location elements are represented by their first letter initials.*

Table 4. Descriptive statistics (mean [std]) of citation density by location

To answer RQ4, we performed pairwise correlation analyses of citation density between different location elements both with and without inclusion of self-citations. The results, as reported in Table 5, reveal a negative correlation in citation density between introduction and related work (p<.001), and a positive correlation between introduction and discussion (p<.001), between method and discussion (p<.001), and between method and conclusion (p<.001), irrespective of whether self-citations were included or not. However, related work and conclusion were only positively correlated (p<.05) when self-citations were excluded, and discussion and conclusion (p<.05) and related work and method (p<.05) were only correlated when self-citations were included.

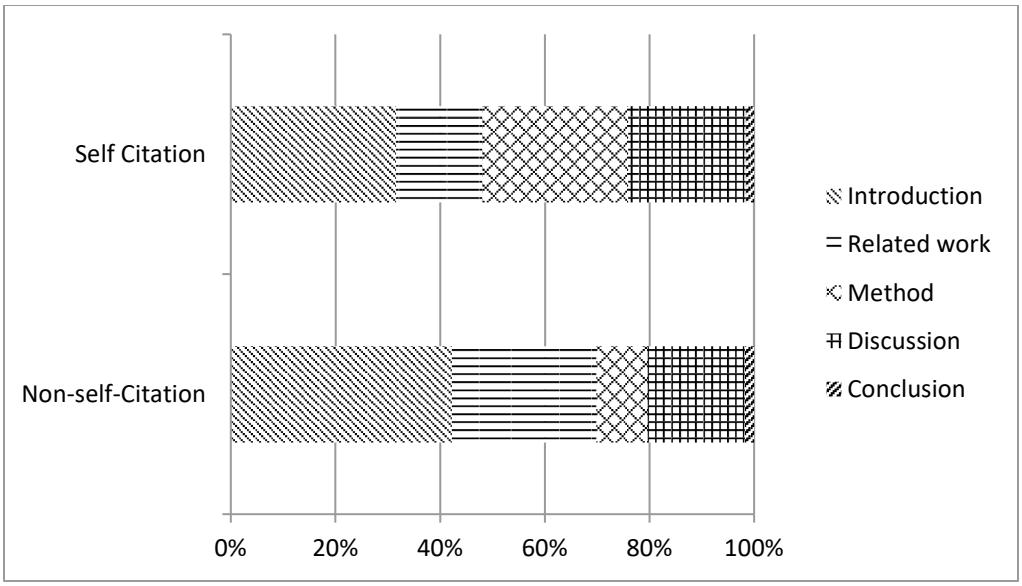

Figure 1. Location distribution of citations

To gain a further understanding of the association among two or more location elements, we also conducted association rule mining. The derived association rules (i.e., $location_l \rightarrow location_r$) reveal the co-occurrence patterns of a cited work across different locations of the same citing articles. The quality of association rules was assessed by three measures: 1) support: a measure of how many articles cited the same source paper at one or more locations, which was expressed as a percentage of the whole dataset; 2) confidence: defined as the ratio of co-occurrence frequency of specific location pairs to the frequency of the $location_l$ in the dataset; and 3) lift: defined as the ratio of the confidence to the support of $location_r$ of a rule, measuring how much better a rule is at predicting citation in $location_2$ than just assuming citations in $location_2$ in the first place. We set the minimum lift to be 1 and the minimum confidence to be 0.3. In view of data sparsity (61% cited work was cited only once), we set the minimum support to be 0.03 when self-citations were included, and to be 0.02 when they were excluded, respectively.

without with self-citation	I	R	M	D	C

I		-.204**	-.025	.155**	-.067
R	-.222**		-.063	0	.09*
M	.06	-.093*		.187**	.172**
D	.143**	-.017	.189**		.129**
C	-.077	.075	.117**	.098*	

***: significant at .01 level; *: significant at .05 level*

Table 5. Correlations of citation density between different location elements

Association rules	With self-citation			Without self-citation		
	lift	confidence	Support	Lift	confidence	support
M, D→ I	1.23	0.7	3.4%	1.15	0.67	2.4%
I, M→ D	2.36	0.53	3.4%	2.44	0.52	2.4%
M→ D	1.48	0.33	4.9%	1.42	0.31	3.6%
I, D→ M	2.06	0.31	3.4%	—	—	—
R, D→ I	—	—	—	1.1	0.64	2.8%

—: The rule did not meet the cut-off thresholds for association rules.

Table 6. Location association rules with and without inclusion of self-citations

Table 6 reports a list of extracted association rules. The top 3 rules were similar irrespective of whether self-citations were included or not. The first rule can be interpreted as: if a paper was cited in both method and discussion sections of the same citing article, there was about a 70% chance (or 67% when self-citations were removed) that the same paper was also cited in introduction. The last two rules differed between the two settings, with the rule (I, D→ M) only applicable when self-citations were included, while the rule (R, D→ I) was only applicable when self-citations were excluded.

The descriptive statistics of citation density for each combination of location and citation level is presented in Table 7. To answer RQ5-1, we analyzed the effect of self-citation on citation density at each citation level separately. The results showed that citation density was higher when self-citations were included than not only when the citation level was high ($p < .01$). However, no

effect of self-citation on citation density was detected when the citation level was medium or low ($p > .05$). Thus, there was a moderating effect of the citation level.

Citation level	I	R	M	D	C
H	0.76 [.846]	.54 [.933]	.26 [.711]	.37 [.83]	.02 [.168]
M	.8 [.843]	.35 [.723]	.18 [.546]	.32 [.898]	.03 [.172]
L	.66 [.816]	.54 [.744]	.2 [.619]	.39 [.934]	.07 [.302]

Table 7. Descriptive statistics of citation density (mean [std]) by location and citation level

To answer RQ5-2, we analyzed the effect of location on citation density at each citation level separately. We first tested the main effect of location on citation density due to scattered evidence in the literature. The analysis yielded a significant main effect of location ($p<.001$). A post-hoc contrast analysis showed that citation density in introduction ($p<.001$), related work ($p<.001$), and discussion ($p<.01$) was higher than that in method; citation density in introduction was higher than that in related work and discussion ($p<.001$); citation density in related work was higher than that in discussion ($p<.05$); and citation density in conclusion was lower than that in all of the other location elements ($p<.05$ or less). We then repeated the analysis for each citation level individually. The results consistently yielded a significant effect of location ($p < .001$). Nevertheless, we also observed some differences in the results of contrast analyses among citation levels. For instance, for highly cited papers, there was higher concentration in introduction than in any other location ($p < .001$); for papers that were either moderately or less frequently cited, their concentrations did not show any difference between introduction and related work ($p>.05$). Similarly, higher citation density in related work than in discussion was only observed for highly cited paper ($p<.01$) but not for cited work at other citation levels ($p>.05$). Interestingly, the difference in citation density between conclusion and method increased with the citation level, from $p<.05$ for the low level,

through $p<.01$ for the medium level, and to $p<.001$ for the high level. These results provided evidence for the moderating effect of citation level on the impact of location on citation density.

DISCUSSION

This findings of this empirical study suggest that self-citations do not have an influence on polarity uncertainty or citation density of the cited work. Nevertheless, in addition to citation count, self-citations show a positive effect on the positivity and a negative effect on negativity of the cited work. Self-citation influence citation concentration in some locations such as method and discussion but not others, while the effect of self-citing behavior is positively moderated by the citation level. This study also uncovered the correlation/association patterns of citation locations and their subjectivity to the influence of self-citing behavior. These findings as a whole suggest that self-citations are not biased if citation analysis is solely based on count or related aggregated measures, but they could induce bias if polarity, citation concentration, location, and/or association are taken into consideration. By examining self-citations from multiple new perspectives, we gain broad and new insights into the nature and function of self-citing behavior.

This study discovers some interesting behavioral differences between citing one's own work and work of others. For instance, self-citations are more likely to occur in method than citations of others' work. If a reference is cited in both introduction and discussion, it it more like to be cited simultaneously in method if it is a self-citation than a citation of others' work. Based on the differing utility of citation location (Cano, 1989), the current findings point to the high utility of one's own prior work in discovery of new knowledge. For instance, authors likely extend their previously tested methods and experimental material in a new study, and/or use their prior related work to lay the ground work for a framework in a new study. The rapid expansion of scientific

literature limits scholars' ability to understand the content of each reference thoroughly, which might lead to an increasing use of perfunctory or periphery citations of little use to the production of new information. Self-citation is less subject to the above issues than non-self-citations because researchers presumably are familiar with their own work and give due credit of new contributions to their own previous research.

The findings of this research have significant implications for citation analysis research and practice. First, despite a growing interest in citation motivations, few studies have empirically investigated motivations of self-citation. The current findings offer an explanation to divergent viewpoints about the impact of self-citation in literature. Additionally, the differences between self-citing and other-citing behaviors discovered in this study challenge the findings of a previous study (Snyder & Bonzi, 1998). Further, the proposed scheme of citation polarity and measures of polarity address the lack of consensus on citation classification and of quantification of motivations for citation in literature.

Second, the citation location, as a relatively new bibliometric parameter, has received little attention in the study of self-citation. This research fills the gap by revealing both differences and similarities between patterns of self-citations and non-self-citations from a location perspective. One interesting difference is that the second most concentrated location for self-citation is method, but is discussion for non-self-citation. The location of a citation depends on some factors that imply a degree of impact of a cited paper on the citing one (Cano, 1989; Maričić et al., 1998). Thus, the finding suggest differing motives between self-citations and non-self-citations. In addition, this is the first study that examines the correlation/association of location context of two or more citations to the same work in a citing article. The association rules can not only reveal interesting citing

behavioral patterns, but also illustrate the high utility of a citation across different functions such as offering both conceptual and operational value to developing new knowledge in a citing paper.

Third, the examination of citation level in this study goes beyond treating it as a dependent variable as in traditional citation analysis. The results show that the effect of self-citation and that of location are generally more pronounced when the citation level is higher. The findings suggest that the patterns of citation location become evident once citation count reaches a critical mass.

Fourth, this research provides additional insights into the interaction between polarity and citation location. A comparison of polarity uncertainty among different locations revealed that polarity uncertainty was the lowest in introduction and the highest in discussion section. This finding is in line with functions of citations in the two specific locations. Earlier studies of citation location (e.g., Cano, 1989; Maričić et al., 1998) failed to take into account negative opinions of citing authors on the cited work. Out of 1,023 citing articles in our sample, 66 (6.45%) contained at least one negative citation, which is not negligible. In a positive light, negative citation is one of the ways in which researchers debate and advance their field. On the other hand, it was found in an earlier study (Catalinia et al., 2006) that receiving a negative citation was associated with a slightly faster drop in citations to the paper in the long run. Interestingly, our search of negative self-citations in our dataset resulted in only 3 cases, implying that self-citations contribute to the longevity of citations to the papers.

Fifth, the citation analysis in this study employed a mixture of qualitative and quantitative methods. The mixed method triangulation can better support citation analysis for two main reasons. First, citation content analysis can help address some challenging issues in identification of citing motives such as context-dependency. For example, “Unfortunately, Rosenfeld, Soskins, Bosh, and Ryan, (2004) and Mertens and Allen (2008) showed that the original form of the P300-

based CIT was vulnerable to countermeasure…" (Labkovsky & Rosenfeld, 2012). Although 'unfortunately' in the sentence had a negative connotation, the polarities of the above two citations were both positive because of their recognitions of a limitation of an earlier protocol. Second, statistical evidence enhanced the reliability of our research finding. This study strengthens signs of shift in research approaches from citation count to a more in-depth interpretation of the meaning of citing past literature.

CONCLUSION AND FUTURE RESEARCH

This study empirically examined self-citations from perspectives of polarity uncertainty, and the location, density, and level of citations. The findings suggest that self-citations are free of bias in terms of polarity uncertainty and citation density, but they are subject to bias in terms of positivity and negativity. In addition, self-citing behavior leads to some changes in citation patterns with respect to citation density, location concentration, and association, while keeping the overall citation patterns largely intact. Furthermore, depending on the citation level, the effect of self-citing behavior and location concentration of citations could vary. Therefore, the findings of this research provide an integrated view of potential biases induced by self-citing behavior by probing into the motives and/or context of self-citations from multiple perspectives.

This research is worth continuing in several directions. First, manual citation content analysis is both time-consuming and labor-intensive. Polarity mining techniques have made it possible to determine the polarity of citations automatically (e.g., Athar, 2011; Catalinia et al., 2006), which bode well for expediting the data analysis process and scaling up the data collection. Nevertheless, addressing context-dependency and neutral polarity of citations remains technically challenging. Second, the potential bias of other forms of self-citing behavior can be explored at an aggregated

level of researcher and journal beyond individual articles. For instance, coercive or forced self-citations where journal editors force authors to cite articles that appear in their journals (Straub & Anderson, 2009; Rodriguez-Ruiz, 2009), and where peer reviewers request authors to cite their work (Thombs, Levis, Razykov, Syamchandra, Leentjens, Levenson, & Lumley, 2015), often cause criticism for their potential bias. Additionally, given interdisciplinary similarities and differences in how self-citations (Hyland, 2003; Snyder & Bonzi, 1998) and citations in general are used (White, 2004), it is worth replicating the study with datasets collected from a wider range of scientific disciplines and topics to validate the generality of self-citing patterns more fully. Third, there might be inter-disciplinary and inter-topic differences in terms of self-citing behavior, which can be followed up in a future investigation. Furthermore, we did not consider possible changes of self-citation patterns with paper aging, which can shed additional light on the potential bias of self-citing behavior.

ACKNOWLEDGEMENTS

The authors would like to thank the researchers and students who provided assistance with encoding the content and context of paper citations. This research is supported in part by the National Science Foundation. Any opinions, findings or recommendations expressed here are those of the authors and are not necessarily those of the sponsors of this research.

REFERENCES

Athar, A. (2011). *Sentiment analysis of citations using sentence structure-based features*. Paper presented at the Proceedings of the ACL 2011 Student Session, Portland, Oregon.

Bartneck, C. and Kokkelmans, S. (2010). Detecting h-index manipulation through self-citation analysis. *Scientometrics* 87(1): 85-98.

Blau, J. R. (2000). Group Enmity and Accord: The Commercial Press in Three American Cities. *Social Science History, 24*(2), 395-413.

Bonzi, S., & Snyder, H. W. (1991). Motivations for citation: A comparison of self citation and citation to others. *Scientometrics, 21*(2), 245-254.

Bowers, J. S. (2009). On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. *Psychol Rev, 116*(1), 220-251.

Bowers, J. S. (2010). More on grandmother cells and the biological implausibility of PDP models of cognition: a reply to Plaut and McClelland (2010) and Quian Quiroga and Kreiman (2010). *Psychol Rev, 117*(1), 300-306; discussion 289-390, 297-309, 306-308.

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.

Brown, R. J. C. (2009). A simple method for excluding self-citation from the h-index: the b-index. *Online Information Review, 33*(6), 1129-1136.

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

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.

Catalinia, C., Laceterab, N., & Oettlc, A. (2006). The incidence and role of negative citations in science. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 112*(45), 13823-13826.

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.

Cole, J., & Cole, S. (1971). Measuring the quality of sociological research: Problems in the use of the Science Citation Index. *The American Sociologist, 6*, 23-29.

Couto, F. M., Pesquita, C., Grego, T., & Veríssimo, P. (2009). Handling self-citations using Google Scholar. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics, 13*(1).

Diodato, V. (1994). *Dictionary of bibliometrics*. New York: Haworth Press.

Dwivedi, Y., Lal, B., Mustafee, B., & Williams, M. (2009). Profiling a decade of information systems frontiers' research. *Information Systems Frontiers*. *11*(1), 87-102.

Elo, S., & Kyngas, H. (2008). The qualitative content analysis process. *J Adv Nurs, 62*(1), 107-115.

Ferrara, E. and Romero, A. E. (2013). Scientific impact evaluation and the effect of self-citations: Mitigating the bias by discounting the h-index. *Journal of the American Society for Information Science and Technology* 64(11): 2332-2339.

Ferligoj, A., Maricˇic´, S., Pifat, G., & Spaventi, J. (1988). Cluster analysis of citation histories from an institutional setting. In N. Tudor-Sˇilovic´ & I. Mihel (Eds.), *Information research: Research methods in library and information science*. London: Taylor Graham.

Fooladi, M., Salehi, H., Yunus, M. M., Farhadi, M., Chadegani, A. A., Farhadi, H., & Ebrahim, N. A. (2013). Does Criticisms Overcome the Praises of Journal Impact Factor? *Asian Social Science, 9*(5).

Fowler, J. H. & Aksnes, D. W. (2007). "Does self-citation pay?" *Scientometrics*, *72*(3): 427-437.

Glänzel, W., Debackere, K., Thijs, B., & Schubert, A. (2006). "A concise review on the role of author self-citations in information science, bibliometrics and science policy." *Scientometrics 67(2)*: 263-277.

Garfield, E. (1972). Citation Analysis as a Tool in Journal Evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science, 178*(4060), 471-479.

Harzing, A.-W., & van der Wal, R. (2009). A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology, 60*(1), 41-46.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A, 102*(46), 16569-16572.

Hirsch, J. E. (2010). An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics, 85*(3), 741-754.

Hodge, D. R., & Lacasse, J. R. (2011). Evaluating Journal Quality: Is the H-Index a Better Measure Than Impact Factors? *Research on Social Work Practice, 21*(2), 222-230.

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

Ioannidis, J. P. A. (2015). A generalized view of self-citation: Direct, co-author, collaborative, and coercive induced self-citation. *Journal of Psychosomatic Research,* 78(1), 7-11.

Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in Social Media Research: Past, Present and Future. *Information Systems Frontiers, 20*(3), 531-558.

Kulkarni, Aziz, Shams, & Busse (2011). Author Self-Citation in the General Medicine Literature. *PLOS ONE***,** 6(6), e20885.

Labkovsky, E., & Rosenfeld, J. P. (2012). The P300-Based, Complex Trial Protocol for Concealed Information Detection Resists Any Number of Sequential Countermeasures Against Up to Five Irrelevant Stimuli. *Appl Psychophysiol Biofeedback, 37*(1), 1-10.

Lehky, S. R., Kiani, R., Esteky, H., & Tanaka, K. (2011). Statistics of visual responses in primate inferotemporal cortex to object stimuli. *J. Neurophysiol., 106*(3), 1097-1117.

Lowry, P. B., Moody, G. D., Gaskin, J., Galletta, D. F., Humpherys, S. L., Barlow, J. B., & Wilson, D. W. (2013). Evaluating journal quality and the association for information systems senior scholars' journal basket via bibliometric measures: do expert journal assessments add value? *MIS Quarterly., 37*(4), 993-1012.

Lutz, B., & Hans-Dieter, D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation, 64*(1), 45-80.

Maričić, S., Spaventi, J., Pavičić, L., & Pifat-Mrzljak, G. (1998). Citation context versus the frequency counts of citation histories. *Journal of the American Society for Information Science, 49*(6), 530-540.

Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 1*(2).

McCain, K., & Turner, K. (1989). Citation context analysis and aging patterns of journal articles in molecular genetics. *Scientometrics, 17*(1-2), 127-163.

Moravcsik, M. J., & Murugesan, P. (1975). Some Results on the Function and Quality of Citations. *Social Studies of Science, 5*(1), 86-92.

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology, 41*(1), 609-641.

O'Leary, D. E. (2008). The relationship between citations and number of downloads in Decision Support Systems. *Decision Support Systems, 45*(4), 972-980.

Osareh, F. (1996). Bibliometrics, Citation Analysis and Co-Citation Analysis: A Review of Literature I *Libri* (Vol. 46, pp. 149).

Peritz, B. (1983). A classification of citation roles for the social sciences and related fields. *Scientometrics, 5*(5), 303-312.

Plaut, D. C., & McClelland, J. L. (2010). Locating object knowledge in the brain: comment on Bowers's (2009) attempt to revive the grandmother cell hypothesis. *Psychol Rev, 117*(1), 284-288.
Quiroga, R., & Kreiman, G. (2010). Measuring sparseness in the brain: comment on Bowers (2009). *Psychol Rev, 117*, 291–297.
Rad, A. E., Shahgholi, L., & Kallmes, D. (2012). Impact of Self-citation on the H Index in the Field of Academic Radiology. *Academic Radiology, 19*(4), 455-457.
Rodríguez-Ruiz, Ó. (2009). The citation indexes and the quantification of knowledge. *Journal of Educational Administration, 47*(2), 250-266.
Sanderson, M. (2008). Revisiting h measured on UK LIS and IR academics. *Journal of the American Society for Information Science and Technology, 59*(7), 1184-1190.
Shadish, W. R., Tolliver, D., Gray, M., & Sen Gupta, S. K. (1995). Author Judgements about Works They Cite: Three Studies from Psychology Journals. *Social Studies of Science, 25*(3), 477-498.
Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal, 27*(3), 379-423.
Small, H. G. (1978). Cited Documents as Concept Symbols. *Social Studies of Science, 8*(3), 327-340.
Snyder, H., & Bonzi, S. (1998). Patterns of self-citation across disciplines (1980-1989). *Journal of Information Science, 24*(6), 431-435.
Straub, D. W., & Anderson, C. (2009). Journal Self - Citation V I: Forced Journal Self - Citation – Common, Appropriate, Ethical? *Communications of the Association for Information Systems, 25*, 57–66.
Swales, J. (1986). Citation Analysis and Discourse Analysis. *Applied Linguistics, 7*(1), 39-56.
Tang, E. K., Suganthan, P. N., & Yao, X. (2006). An analysis of diversity measures. *Machine Learning, 65*(1), 247-271.
Thombs, B. D., Levis, A. W., Razykov, I., Syamchandra, A., Leentjens, A. F. G., Levenson, J. L., & Lumley, M. A. (2015). Potentially coercive self-citation by peer reviewers: A cross-sectional study. *Journal of Psychosomatic Research, 78*(1), 1-6.
Waltman, L., & Eck, N. J. v. (2012). The inconsistency of the h-index. *J. Am. Soc. Inf. Sci. Technol., 63*(2), 406-415. doi:10.1002/asi.21678
Wang, N., Liang, H., Jia, Y., Ge, S., Xue, Y., & Wang, Z. (2016). Cloud computing research in the IS discipline: A citation/co-citation analysis. *Decision Support Systems, 86*, 35-47.
White, H. D. (2004). Citation analysis and discourse analysis revisited. *Applied Linguistics, 1*, 89–116.
Whitehouse, G. H. (2001). Citation rates and impact factors: should they matter? *Br J Radiol, 74*(877), 1-3.
Zhang, C.-T. (2013). The h'-Index, Effectively Improving the h-Index Based on the Citation Distribution. *PLoS ONE, 8*(4), e59912.
Zhang, G., Ding, Y., & Milojević, S. (2013). Citation content analysis (CCA): A framework for syntactic and semantic analysis of citation content. *Journal of the American Society for Information Science and Technology, 64*(7), 1490-1503.
Zhou, L., Zhang, D., & Sung, Y.-w. (2013). The effects of group factors on deception detection performance. *Small Group Research, 44*(3), 272-297.