Comparative Investigation of Self-Citation Patterns in Information Science: A Pilot Study

Joel DesArmo, University of Wisconsin-Milwaukee

Abstract
This study is an exploratory investigation into the citation patterns in the field of Information Retrieval. Preliminary findings are based on a comparison of self-citation rates in JASIST, a leading journal in Information Retrieval, and a journal in an unrelated field. Although the sample size is small, the preliminary results suggest that there may be irregularities in the self-citation rates among Information Retrieval authors. The exploratory study found that self-citation rate in JASIST is nearly three times that of a journal in an unrelated field. Further study with a more robust sample is warranted to confirm the findings herein and to shed light on the motivations for self-citations.
Keywords: self-citation; citation patterns; scholarly communication

1   Introduction

Considering the long history of written scholarly communication, citation analysis is a relatively new science. Likely the first published citation analysis research appeared in a 1927 issue of the journal *Science* in which the authors used citation analysis to inform libraries which chemistry journals they should purchase (Gross et al., 1927). The Science Citation Index was launched in 1961 and began the avalanche of citation analysis. A good amount of the mountain of citation analysis research has been conducted in the area of self-citation. The present study seeks to investigate the patterns of self-citation in the field of Information Retrieval, thereby laying the groundwork for future studies investigating the motivations behind the patterns, which can ultimately lead to departments creating policies to guide the most appropriate inclusion of self-citations within their faculty's scholarly works.

2   Literature Review

There is a significant body of literature related to citation analysis of scholarly work. Vinkler (1998, p.107) stated, "As referencing can be assumed as a peer evaluation process resulted in referencing some papers and neglecting others, citations obtained can be applied for assessing international impact of scientific publication activity". It can be argued therefore that self-citation dilutes the peer-review process, as repeatedly acknowledging one's own work "stuffs the ballot box" with votes for one's self.

Anderson investigated the influence scholars have as perceived by peers in their field. The study surveyed 876 Danish researchers, primarily from social science fields, but also including computer science, natural science and medicine. They asked the researchers to list the three most influential researchers in their field. They found the correlation between frequency of author citation and the perceived level of impact by peers to be low (Andersen, 2000). One potential contributing factor to this discrepancy could be the role of self-citations in boosting citation counts for papers which is not reflected in the perceived scholarly impact felt by authors' peers.

Findings have both taken issue with self-citations as well as defended self-citations. Bonzi and Snyder analyzed 51 authors who self-cited across several natural science disciplines and found no difference between the self-citations and the citations of other authors (Bonzi & Snyder, 1991).

First described as "the Matthew Effect" in the seminal piece by Robert Merton in 1968, the concept was laid out that success will lead to further success (Merton, 1968). Frequent citation breeds more citation bringing academic prestige and the benefits that go along with such prestige, such as career advancement and even supplemental opportunities such as paid speaking engagements as a domain expert or expert consulting opportunities. Because of the rewards, the motivations for citation, and particularly self-citation, must be questioned. Merton, himself, self-cites 7 times in his landmark 1968

paper in which he originally coined the term "Matthew Effect." *The Matthew Effect in Science* has a total of 51 references and notes, of which 8 are merely endnotes leaving 43 references; a self-citation rate of 16% (Merton, 1968). Scholars clearly have a potential motive for engaging in self-citation, for if their work has little or no citations according to the Matthew Effect the likelihood that the uncited scholar will continue to be the uncited scholar is great.

3   Method

The present study will seek to address the following research question:
RQ1. How does the self-citation rate in the field of Information Retrieval compare to the field of Business?

The hypothesis for this study is as follows:
H$_1$. There is no significant difference in the self-citation rates between the fields of Information Retrieval and Business.

The present study is a small pilot study to pave the way for more robust comparative analysis of the self-citation patterns and motivations of the field of Information Retrieval. As a pilot study, a small data set was collected and analyzed. The study consulted with Web of Science Journal (WoS) Citation Reports (JCR) Social Sciences 2011 edition to select two academic journals from which to collect self-citation data. Using the author's personal domain knowledge, the *Journal of the American Society for Information Science and Technology* (JASIST) was selected as a representative high impact journal in the field of Information Retrieval.

The additional journal was originally meant to in a professional field and Education was selected out of convenience. Once again, the Web of Science JCR was consulted and journals listed in the Education & Educational Research field were selected. The list was sorted by impact factor and the journal with the largest impact factor, the *Academy of Management Learning & Education* (AMLE), was selected for inclusion in this study. However after data collection began, it was determined that the actual nature of AMLE was Business and not Education. This was taken as a serendipitous change of plans, but was viewed as still in alignment with the original design of comparing a professional field. JASIST has an impact factor of 2.081, while the AMLE has an impact factor of 4.800.

4   Results

Self-citation rates were calculated for each individual article as well as for each journal and can be seen in Table 1. The shaded colors in the columns of Table 1 represent the different volumes inspected as described in the Method section of this paper. The mean self-citation rate for JASIST was 11.04%, while the mean rate for AMLE was 4.31%. This number is skewed however, because of articles with zero citations. For example, if an article has 10 citations (C) and zero self-citations (S), it has a self-citation rate of $\dfrac{S}{C}$, which is equal to zero divided by 10 = 0%. Similarly, if an article has 100 citations and zero self-citation, it has the exact same self-citation rate of 0%. Clearly this skews the rate. Therefore, the journal's author self-citation rate was also calculated as the sum of all the self-citations per article for JASIST divided by the sum of all the total citations per article for JASIST. When considering the rate for the journal as a whole, the result is slightly lower at 9.66%. The calculation is the same for AMLE and the journal's author self-citation is 3.32%.

Article #	JASIST			Academy of Management Learning & Education		
	Self-citation	Total citation	Selfie-Rate	Self-citation	Total citation	Selfie-Rate
1	6	34	17.65%	4	83	4.82%
2	8	70	11.43%	0	80	0.00%
3	1	29	3.45%	0	73	0.00%
4	1	22	4.55%	0	61	0.00%
5	15	67	22.39%	4	98	4.08%
6	2	12	16.67%	0	84	0.00%
7	2	33	6.06%	1	45	2.22%
8	0	44	0.00%	15	42	35.71%
9	2	136	1.47%	5	60	8.33%
10	2	31	6.45%	2	62	3.23%
11	3	60	5.00%	0	57	0.00%
12	1	26	3.85%	3	80	3.75%
13	2	20	10.00%	5	76	6.58%
14	2	35	5.71%	1	69	1.45%
15	8	85	9.41%	4	87	4.60%
16	2	14	14.29%	0	152	0.00%
17	13	52	25.00%	1	124	0.81%
18	6	17	35.29%	1	52	1.92%
MEAN	4.22	43.72	11.04%	2.56	76.94	4.31%
MODE	2	#N/A		0	80	
MEDIAN	2	33.5		1	74.5	
TOTAL	76	787	9.66%	46	1385	3.32%
TOTAL AVG RATE factors in the zeros						
Percentage with zero self-citations	5.56%			33.33%		

Table 1. Descriptive Statistics

Central tendency statistics were calculated as well for the simple number of self-citations per article. The mean number of self-citations per article for JASIST is 4.22 and for AMLE it is 2.56, nearly one half that of JASIST. The mode self-citation per article for JASIST is 2 and for AMLE the mode is 0. The median self-citations per article for JASIST is also 2, while for AMLE the median is 1, which is between AMLE mean of 2.56 and AMLE mode of 0.

Furthermore, it is quite noteworthy that the percentage of articles appearing in JASIST without any self-citations is 5.56% of all articles. In contrast, the number of articles appearing in AMLE without any self-citations is 33.33%. This represents a factor of 6 times as many articles in AMLE without any self-citations compared to JASIST.

A two-tailed independent measures t-test statistical analysis indicates that there is a significant difference in the self-citation rates between the fields of Information Retrieval and Business, $t(34) = 2.249$, $p = 0.031 < .05$. Results of the t-test proved to be statistically significant. The null hypothesis is rejected and therefore this study finds that there is a statistically significant difference in the self-citation rates between the fields of Information Retrieval and Business. The SPSS output can be viewed in Table 2.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SelfCitationRate	Equal variances assumed	1.898	.177	2.249	34	.031	.06611	.02940	.00636	.12596
	Equal variances not assumed			2.249	33.565	.031	.06611	.02940	.00633	.12589

Table 2. T-test results

5   Conclusions

The findings of this pilot study demonstrate that there may be self-citation irregularities in the field of Information Retrieval. There are number of factors that could be contributing to the self-citation rate discrepancy between the fields of Information Retrieval and Business found in this pilot study. It could be that the field of Information Retrieval simply has fewer authors from which to draw citations and therefor the result is an increase in self-citation as was observed in this study. It could also be that faculty in Library and Information Science are under more pressure from administrators or colleagues to acquire high citation counts, perhaps in order to enhance the reputation of their department or even the field as a whole since an increase in the academic reputation of any one researcher may boost the reputation of the institution at which that researcher works or even could bring additional prestige to an under-exposed field as a whole. Further study is needed to discern the motivations behind self-citations in Information

Science. Furthermore, the preliminary findings of this pilot study suggest that an in-depth investigation into the self-citation patterns within Information Science is warranted.

References

Andersen, H. (2000). Influence and reputation in the social sciences–how much do researchers agree? *Journal of Documentation*, *56*(6), 674–692.

Bonzi, S., & Snyder, H. W. (1991). Motivations for citation: A comparison of self citation and citation to others. *Scientometrics*, *21*(2), 245–254. doi:10.1007/BF02017571

Gross, P. L., Gross, E., & others. (1927). *College libraries and chemical education*. Science.

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*(3810), 56–63.

Vinkler, P. (1998). Comparative investigation of frequency and strength of motives toward referencing. The reference threshold model. *Scientometrics*, *43*(1), 107–127.

Table of Tables