The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0024-2535.htm

Author self-citation pattern in science

Mohammad Reza Davarpanah
Library and Information Science Department, Faculty of Education, Ferdowsi University of Mashhad, Mashhad, Iran, and
Farzaneh Amel
Faculty of Education and Psychology, Ferdowsi University of Mashhad, Mashhad, Iran

Received 23 June 2008
Revised 20 August 2008
Accepted 12 September 2008

Abstract

Purpose – The purpose of this paper is to study the author self-citation behavior in four disciplines: electronic engineering, general and internal medicine, organic chemistry and plant sciences.

Design/methodology/approach – By using SCI and random stratified method 1,000 articles were analysed as a sample in the four disciplines during 2004-2006.

Findings – It was found that about 60 per cent of the articles in the four disciplines' literature contained at least one self-citation. Four disciplines all showed skewed distributions of articles citation rates, either self-citation or other citations. Organic chemistry articles had the highest self-citations than the other disciplines. Share of self-citation decreases with growing time window. The expected self-citation rate increased with growing number of citation, co-authorship and author productivity.

Originality/value – The outcomes of this study suggest that self-citation indicators should be used as supplementary indicators in evaluative bibliometrics.

Keywords Electronic engineering, Sciences, Authorship, Medical sciences, Literature

Paper type Research paper

Introduction

Citations indicators have been increasingly applied in the context of science policy and research evaluation. The basic assumption underlying such applications is that citations can be regarded as a measure of scientific quality or impact (Aksnes, 2006). The scientific community uses bibliometric data, including citation counts of articles and impact factors of journals in which the articles were published, to judge the importance of articles, and assess the productivity of faculty members and the scientific merit of their work (Gami *et al.*, 2004). Whenever citations are used as indicators to evaluate scientific research, self-citations are often considered problematic. Author self-citation refers to citing one's previous publications in a new publication. Author self-citation exists when the citing and cited papers have at least one author in common. Although authors may have good reasons to cite their own works, these citations do not necessarily reflect the importance of their work or its impact on the rest of scientific community (Fowler and Aksnes, 2006). Author self-citations may misrepresent the importance of individual articles, skew the calculation of journal impact factors, and bias perceptions of the importance of a publication.

Several study have analysed author self-citation. First Garfield and Sher used a quantitative approach to author self-citations; they found that the share of authors self-citations in basic research amounts to 20 per cent on average (Garfield and Sher, 1964). According to Trimble (1986) about 15 per cent of all citations in astronomical papers published during January 1983 were self-citations, in the sense that the cited and citing paper had at least one author in common. Dimitroff and Arlitsch (1995) found that 50 per cent of the articles in the library and information science literature contained at least one self-citation. Bonzi and Snyder (1998), in a case study including the social

0024-2535
DOI 10.1108/00242530910952846

sciences and the humanities, found an average self-citation rate of 11 per cent , varying from 16 per cent in the physical sciences (chemistry and geology) to 3 per cent in the social sciences. These studies have shown that self-citation rates tend to vary between disciplines. In a macro study Aksnes (2003) investigated the role of self-citation in the scientific production of Norway (1981-1996). He found the highest share of self-citation among the least cited papers. There was a strong positive correlation between the number of self-citations and the number of authors of the publications. Also, the share of self-citation showed significant variations among different scientific disciplines.

Gami *et al.* (2004) reported that nearly one-fifth of all citations to articles about diabetes mellitus in clinical journals in the year 2000 were author self-citations. The frequency of self-citation was not associated with the quality of publications. Articles published in highly cited journals had a smaller proportion of author self-citations than articles published in less-cited journals. In a study, Glanzel (2004) showed that the ageing of self-citations is much faster than that of foreign citations. In the third after the year of publication, the expected number of self-citation for a given number of foreign citations becomes practically stationary. The results of the study showed that a citation window not smaller than three years and not larger than four years is sufficient for reliable bibliometric analyses since the share of self-citations is for such citation windows within acceptable limits. Thijs, Glanzel and Noyons (2006) suggested that both national and subject specific peculiarities influence the share of self-citations at the macro level, at this level of aggregation – there is practically no need for excluding self-citations. The results showed a quite complex situation at the meso-level, therefore they suggested the usage of both indicators, including and excluding self-citations.

Hypotheses

- The use of self-citations, as compared to the use of citations to others, is significantly different among the individual disciplines.
- Share of self-citation decreases with growing time window.
- There is a relationship between the number of self-citations and author publishing productivity.
- There is a relationship between the number of citation and the number of self-citation; and
- There is positive correlation between the number of self-citations and the number of co-authors of the articles.

Methodology

As a basis for our analysis we used science citation index (SCI) database provided by Thomson Scientific (Formerly Institute for Scientific Information, [ISI]). This database contains bibliometric information for science, engineering, agriculture, and medical sciences areas. One discipline were selected randomly from each of this broad areas. Organic chemistry was selected from the science, electronic engineering from the engineering, plant sciences from the agriculture, and general and internal medicine from the medical sciences. We applied the 2007 edition of the SCI (web of knowledge, http://isiknowledge.com) with data covering 2004-2006. In this way the citation window applied three years. Only regular articles were included in the analysis. On June, 2007 the database contained in total 10,177 articles in the four disciplines. From this, 3,455 (34 per cent) articles were uncited (Table I). The uncited articles removed from our study.

The remaining 6,722 articles considered as the statistical population of this study: including 4,623 in organic chemistry, 536 in electronic engineering, 172 in plant sciences, and 1,391 in general and internal medicine. Out of this by using random stratified method 1,000 articles were chosen as a sample in the four disciplines: 354 in organic chemistry, 226 in electronic engineering, 118 in plant sciences, and 302 in general and internal medicine. For each article we calculated that at least one author (first author or co-author) is also an author (first author or co-author) of the citing paper. All citations, whether monographic or journal, were included.

Findings

The share of self and other citation among science disciplines

This hypothesis was formulated to know how the share of self and other citation varies according to scientific fields. For each publication we collected data on field assignments. As basis for the classification, we used the ISI categories. Citation data were collected based on time cited field. Table II shows that the number of self and other citations varies among the disciplines. The lowest self-citation percentage is in general and internal medicine (16.04 per cent), while the highest percentage is in organic chemistry (68.28 per cent). Organic chemistry articles had double the mean proportion of author self-citations compared with general and internal medicine. As is evident from the table, the number of other citations in the general and internal medicine is greater than either the organic chemistry, plant sciences, or electronic engineering. As a proportion of all citations, there are also large differences among these groups. Over all 41.25 per cent of all citations are self-citations; 58.75 per cent are other citations in the four fields. Based on the data in Table III author self-citations per publication ranged from 1 to 32, mean 1.83, and other citation ranged from 1 to 63, mean 4.10. On average about 60 per cent of the articles in the four disciplines have at least one self-citation. The data also revealed that 18 per cent of the articles in plant sciences, 24.4 per cent in electronic engineering, 10.9 per cent in General and internal medicine and 14.4 per cent in organic chemistry have received only self-citations.

Analysis of variance (ANOVA) was used to measure differences in the citation rates among the disciplines (Table IV). The general and internal medicine were significantly

Field	Number of article	Number of uncited article	Number of cited article
Plant sciences	361	189 (52.4)	172 (47.6)
Electronic engineering	1,315	779 (59.3)	536 (40.7)
General and internal medicine	2,369	978 (41.3)	1,391 (58.7)
Organic chemistry	6,132	1,509 (24.6)	4,623 (75.4)
Total	10,177	3,455 (33.95)	6,722 (66.05)

Table I. Number of cited and uncited articles in four fields during 2004-2006

Fields	Number of articles	Time-cited	Self-citation	Other citation
Plant sciences	118	552	167 (30.25%)	385 (69.75%)
Electronic engineering	226	914	263 (28.77%)	651 (71.23%)
General and internal medicine	302	1,771	284 (16.04%)	1487 (83.96%)
Organic chemistry	354	2,300	1,570 (68.26%)	730 (31.74%)
Total	1,000	5,537	2,284 (41.25%)	3,253 (58.75%)

Table II. Self and other citations in disciplines

Disciplines	Self citation No. of self citation	Frequency of article	%	Other citation No. of other citation	Frequency of articles	%
Plant sciences	0	45	38.1	0	22	18.6
	1-2	53	45	1-2	48	40.7
	3-4	12	10.2	3-4	19	16.1
	5-6	5	4.3	5-6	14	11.9
	7-more	3	2.4	7-more	15	12.7
	Total	118	100	Total	118	100
Electronic engineering	0	97	42.9	0	55	24.4
	1-2	98	43.4	1-2	106	46.9
	3-4	20	8.9	3-4	31	13.7
	5-6	7	3.1	5-6	15	6.6
	7-more	4	107	7-more	19	8.4
	Total	226	100	Total	226	100
General and internal medicine	0	154	51	0	33	10.9
	1-2	115	38.1	1-2	113	37.4
	3-4	24	8	3-4	59	19.5
	5-6	7	2.3	5-6	31	10.3
	7-more	2	0.6	7-more	66	21.9
	Total	302	100	Total	302	100
Organic chemistry	0	102	28.8	0	51	14.4
	1-2	151	42.6	1-2	115	32.5
	3-4	55	15.5	3-4	65	18.4
	5-6	29	8.2	5-6	43	12.1
	7- more	17	4.9	7- more	80	22.6
	Total	354	100	Total	354	100

Table III. Distribution of self citation and other citation

lower in self-citation and greater in other citation than the other disciplines. Clearly, there is a much greater tendency to self-cite in the organic chemistry than in the other three disciplines. Over all, we find considerable variations in the self and other citation rates among different scientific fields. Within disciplines, there are significant differences in self-citations or citations to others, as measured by the *t*-test.

In the continuous self-citation impact factor (ScIF) for each discipline were calculated by developing the following formula:

$$ScIF = \frac{Sc1 + Sc2 + Sc3}{Tc1 + Tc2 + Tc3}$$

Pervious studies (Glanzel *et al.*, 2004) have shown that self-citation indicators become quite stable in a period of three-four years after publication. A citation window of three years is thus sufficient for self-citation studies. The factor is defined as the recorded number of self-citations within a certain year and discipline (for example, 2006) to the time cited of the articles during the same and two proceeding years (2006, 2005, 2004), divided by the number of such items. In the formula SC is self-citation in first, second, and third year; TC is time-cited in three years.

Based on the above formula self-citation impact factor for each discipline were measured (Table V). For example ScIF in plant sciences is calculated as follows:

$$ScIF = \frac{30 + 37 + 100}{76 + 149 + 327} 0.302$$

As it is obvious from the table self-citation impact factor is high in organic chemistry and low in general and internal medicine. Table VI shows the result of ANOVA test on ScIF. There is association between ScIF (Table V) and self-citation rate (Table II) in each discipline. Threfore, the ScIF formula is reliable.

Share of self-citation decreases with growing time window

Under this hypothesis we analyze how the number and share of self-citations varied with time after publication. Often in bibliometric analysis, citations are collected during the first three years after publication (Glanzel *et al.*, 2004). In this way the citation window applied three years (2004-2006). For each article we identified the self-citations by citing years, that is, by year after the article was published. The results of the analysis are shown in Figure 1 and Table IV.

As is evident from the figure the distribution is skewed. Data in Table VII shows that there is a significant negative correlation between citation window and self-citations. It means that the average values of self-citations observed in four disciplines decrease over time (from 2004-2006). Across all disciplines, self-citations tend to be for more recent works than citations to other works.

	Sum of squares	df	Mean square	F	Sig.
Self-citation					
Between groups	229.709	3	76.570	19.146	0.000
Within groups	3,983.155	996	3.999		
Total	4,212.864	999			
Other citation					
Between groups	653.822	3	217.941	5.739	0.001
Within groups	37,821.769	996	37.974		
Total	38,475.591	999			

Table IV. ANOVA test

Fields	SCIF
Plant sciences	0.302
Electronic engineering	0.287
General and internal medicine	0.160
Organic chemistry	0.317

Table V. ScIF for each disciplines

	Sum of squares	df	Mean square	F	Sig.
Between groups	4.914	3	1.638	12.578	0.000
Within groups	129.579	995	0.130		
Total	134.493	998			

Table VI. ANOVA test for ScIF

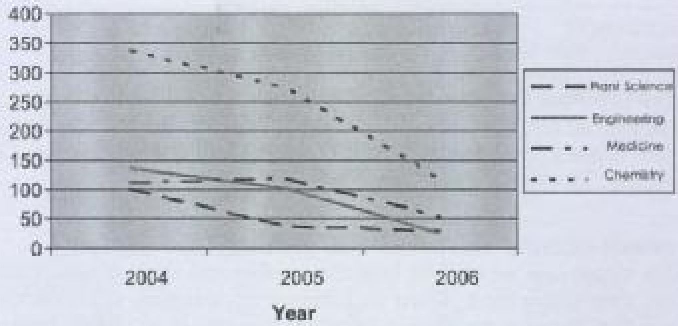

Figure 1.
Self-citations and time window

Table VII.
Correlation between self-citations and time

		Year	Self-citation
Year	Pearson correlation	1	0.089[a]
	Sig. (2-tailed)	–	0.005
	N	1,000	1,000
Self-citation	Pearson correlation	0.089[a]	1
	Sig. (2-tailed)	0.005	–
	N	1,000	1,000

Note: [a]Correlation is significant at the 0.01 level (2-tailed)

Relationship between the number of self-citations and author publishing productivity

Authors who publish a lot have more opportunities to cite their own previous work. Thus, we include variable for publications during the study years, because this may increase the opportunity for self-citations. Therefore, the number of self-citations and the number of publications during the study period were collected. To assess association between the both variables, Pearson's correlation coefficient test was used (Tables VIII). As the data shows there is a significant positive correlation between self-citations and publications. In other words, the more papers one has published the more self-citations one can gives, and the persons with the highest productivity of papers accordingly tended to have the highest number of self-citations.

Relationship between the number of citation and the number of self-citation

For each article in each discipline we calculated the numbers of citations and shares of self-citations. The results of correlation are shown in Table IX. As we can see the number of self-citations increases with total number of citations. For instance, papers cited 10 times on average received 2.25 self-citations.

Correlation between the number of self-citations and the number of co-authors of the articles

It is generally known that the total number of self-citations increase with number of authors; because there are more authors to cite themselves. In order to test this assumption we calculated the number of authors and number of self-citations for the

		Co-authors	Publications
Plant sciences			
Co-authors	Pearson correlation	1	0.451[a]
	Sig. (2-tailed)	–	0.000
	N	118	118
Publications	Pearson correlation	0.451[a]	1
	Sig. (2-tailed)	0.000	–
	N	118	118
Organic chemistry			
Co-authors	Pearson correlation	1	0.172[a]
	Sig. (2-tailed)	–	0.001
	N	354	354
Publications	Pearson correlation	0.172[a]	1
	Sig. (2-tailed)	0.001	–
	N	354	354
Electronic engineering			
Co-authors	Pearson Correlation	1	0.371[a]
	Sig. (2-tailed)	–	0.000
	N	226	226
Publications	Pearson Correlation	0.371[a]	1
	Sig. (2-tailed)	0.000	–
	N	226	226
General and internal medicine			
Co-authors	Pearson correlation	1	0.244[a]
	Sig. (2-tailed)	–	0.000
	N	302	302
Publications	Pearson correlation	0.244[a]	1
	Sig. (2-tailed)	0.000	–
	N	302	302

Note: [a]Correlation is significant at the 0.01 level (2-tailed)

Table VIII. Correlation between co-authors and publishing productivity in each discipline

		Self-citations	Time-cited
Self-citations	Pearson correlation	1	0.524[a]
	Sig. (2-tailed)	–	0.000
	N	1,000	1,000
Time-cited	Pearson correlation	0.524[a]	1
	Sig. (2-tailed)	0.000	–
	N	1,000	1,000

Note: [a]Correlation is significant at the 0.01 level (2-tailed)

Table IX. Correlation between self-citations and time-cited

publications. Pearson's correlation coefficient (Table X) showed that there is a significant positive correlation between number of authors and self-citations in the four disciplines. This means that the number of self-citations increase with number of authors. It is also obvious from Table XI.

Conclusions

The findings showed that the four disciplines had significant differences in citation patterns, either self-citations or other citations. The share of author self-citations in the four disciplines comprise 41.25 per cent of all citation to articles published in 2004-2006. On average 60 per cent of the articles in the four disciplines literature contained at least one self-citation. The greatest frequency of self-citation is in organic chemistry. Although, self-citations are characterized by large variations among different scientific disciplines, but, in comparison of previous studies such as Garfield and Sher (1964), Trimble (1986), Snyder and Bonzi (1998) the rate of self-citation in the four studied disciplines is excessive. Differences that do exist would appear to reflect the norms of scholarly writing in a discipline. The reasons why authors do so are not clear, but, if the citation expresses reward, self-citations distort necessarily the system as such.

The share of self-citations decreases with growing time window. The chance of receiving other citations may increase with growing citation window. There is a strong positive correlation between the number of self-citations and the number of citations and number of co-authors of the publications as already observed earlier by Aksnes (2003). Therefore, the number of self-citations is a mathematical function of other citations and number of authors. It can be conclude that co-authorship influences self-citation patterns. Thus, multi-authorship increases above all the probability of citations.

Four disciplines all showed skewed distributions of cited and uncited rates. The situation becomes problematic if a paper receives only self-citation. The study revealed that on average 56 per cent (40 per cent uncited + 16 per cent only self-cited) of the papers in the four disciplines are uncited or only self-cited. Here, this question will come to the mind that in this case, can bibliometrics immediately conclude on the quality of these papers. It can be said that the less frequently cited or even uncited does not reveals about quality or the standing of its authors in the community. "Uncited

		Self-citations	Co-authors
Self-citations	Pearson correlation	1	0.205[a]
	Sig. (2-tailed)	–	0.000
	N	1,000	1,000
Co-authors	Pearson correlation	0.205[a]	1
	Sig. (2-tailed)	0.000	–
	N	1,000	1,000

Note: [a]Correlation is significant at the 0.01 level (2-tailed)

Table X.
Correlation between self-citations and co-authors

Number of authors	Mean of citations	Mean of other citation	Mean of self-citation
1-2	5.124	4.105	1.018
3-4	4.983	3.612	1.371
5-6	6.231	4.602	1.629
7-8	6.775	4.937	1.837
9-10	5.192	2.807	2.384
11-12	10.666	7.2	3.466
13-14	4	3.333	0.666
15-16	4	4	0
17-18	4.666	3	1.666
19-20	9.5	6.5	3
23-24	1.5	0.5	1
28-29	17	9	8

Table XI.
Number of authors and mean of citations

papers by Nobel prize winners may just serve as an example" (Glanzel, 2004). The only use of citation counts can therefore be as misleading for measure of scientific quality of papers; and the basic assumptions of citation analysis "that the number of citations reflects an articles influence" may be false.

To the extent that assessors of the importance of a publication rely on bibliometric indices based on citation counts, author self-citation may artificially inflate an articles importance to the general scientific community. The outcome of this study confirms that self-citation indicators should be used as supplementary indicators in evaluative bibliometrics. For this purpose the mathematical model "SCIF" introduced in this study can be applicable for recommending a reasonable rate of self-citation.

References

Aksnes, D.W. (2003), "A macro-study of self-citation", *Scientometrics*, Vol. 56, No. 2, pp. 235-46, available at: http://english.nifustep.no/norsk/publikasjoner/a_macro_study_of_self_citation (accessed 11 November 2006).

Aksnes, D.W. (2006), *Citation Rates and Perceptions of Scientific Contribution*, available at: http://english.nifustep.no/norsk/publikasjoner/citation_rates_and_perceptions_of_scientific_contribution_2 (accessed 11 November 2006).

Dimitroff, A. and Arlitsch, K. (1995), "Self-citations in the library and information science literature", *Journal of Documentation*, Vol. 51 No. 1, pp. 44-56.

Fowler, J.H. and Aksnes, D.G (2006), "Does self-citation pay?", available at: http://jhfowler.ucsd.edu/does_self_citation_pay.pdf (accessed 2 December 2006).

Gami, A.S., Montori, V.M., Wilczynski, N.L. and Haynes, R.B. (2004), "Author self-citation in the diabetes literature", *CMAJ*, Vol. 170 No. 13, pp. 1925-7, available at: www.cmaj.ca/cgi/reprint/170/13/1925?ijkey=0ad479e8dfd23291fc089ab8e3c2555b86d31eb5 (accessed 12 November 2006).

Garfield, E. and Sher, I.H. (1964), *The Use of Citation Data in Writing the History of Science*. available at: www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf (accessed 19 December 2006).

Glanzel, W., Thijs, B. and Schlemmer, B. (2004), "A bibliometric approach to the role of author self-citations in scientific communication", *Scientometric*, Vol. 59, No. 1, pp. 63-77, available at: www.steunpuntoos.be/WG_Papers/Scientometrics_59_1_63.pdf (accessed 24 November 2006).

Snyder, H. and Bonzi, S. (1988), "Patterns of self-citation across discipline", *Journal of Information Science*, Vol. 24, No. 6, pp. 431-5.

Thijs, B., Glanzel, W. and Noyons, C.M. (Eds) (2006), *The Influence of Author Self-Citations on Bibliometric Meso-Indicators. The Case of European Universities*, available at: http://cat.inist.fr/?aModele=afficheN&cpsidt=17549543 (accessed 24 November 2006).

Trimble, V. (1986), "A note on self-citation rates in astronomical papers", available at: http://articles.adsabs.harvard.edu/cgi-bin/nph-iarticle_query?1986PASP...98.1347T&data_type=PDF_HIGH&whole_paper=YES&type=PRINTER&filetype=.pdf (accessed 23 April 2007).

Corresponding author
Mohammad Reza Davarpanah can be contacted at: mrdavarpanah@yahoo.com