

Toward a more precise definition of self-citation

Stephen Carley · Alan L. Porter · Jan Youtie

Received: 22 February 2012 / Published online: 25 April 2012
© Akadémiai Kiadó, Budapest, Hungary 2012

Abstract The definition assigned to self-citations is nontrivial. This decision can affect research outputs in a number of ways. The current paper considers the self-citation definition used by the Web of Science, and compares this with an alternative definition, advanced in the present study, within the context of the work of an individual researcher. A discussion follows.

Keywords Self-citation · Citation indicators · Bibliometrics

The role that self-citations play in measuring research outputs is significant. The practice of citing one's work can meaningfully influence a number of metrics, including total citation counts, citation speed, the ratio of external to internal cites, Diffusion scores and Hirsch's h-indices. Although higher levels of aggregation do not see significant effects from self-citation, these effects become more prominent at meso- and micro-levels (Glanzel et al. 2006). Self-citations observed at these more disaggregated levels have been found to be especially influential in the h-indices of average and early career researchers rather than for very prominent researchers with high h-scores (Schreiber 2007). The effect of self-citation has been found to decrease over time, indicating that these types of citations are likely to be more of an issue for studies with briefer than longer time frames. In addition, self-citation is more likely to be prevalent the greater the number of authors per article (Schubert et al. 2006; Aksnes 2003; Rousseau 1999). This relationship between author numbers and self-citation suggests that research collaboration, including international collaboration, will be more apt to include self-citations in these papers' reference lists.

S. Carley (✉) · J. Youtie
School of Public Policy, Georgia Tech, Atlanta, GA, USA
e-mail: stephen.carley@gmail.com

A. L. Porter
Technology Policy and Assessment Center, Georgia Tech, Search Technology, Inc., Atlanta, GA, USA

J. Youtie
Georgia Tech Enterprise Innovation Institute, Atlanta, GA, USA

Table 1 WOS citation statistics for Newman's individual publications on WOS

Newman ISI unique ID #	Citing articles	Citing articles without self-citations	
A1996TT20800002	20	20	
000172575700001*	10	10	
A1996VH96100001*	8	8	
000174838500002*	6	6	
000224846300007	5	5	
000233752700005*	4	4	
000262416800002*	4	4	
000240863700001	1	1	
000275746700001*	1	1	
000288932700005	1	1	
000168824702172	0	0	
* Counted as a self-cite when all of Newman's ISIs are searched for simultaneously in WOS	000189300600002	0	0
	000283110100034*	0	0

Given their potential to significantly influence outcomes at the meso and micro levels, the definition one assigns to the practice of self-citation is nontrivial. The self-citation algorithm employed by the WOS¹ provides the point of departure in the following analysis. This algorithm was selected as the frame of reference because the data and methods employed by WOS are used on a mass scale. According to WOS, "Citing Articles without Self-citations" are defined in the following way: "This field displays the total num[b]er of citing articles minus any article that appears in the set of search results on the citation report."² To come to better terms with this definition, the authors of this study analyze specific examples within the WOS citation report to more clearly identify what citing articles do and do not count as a self-citation for WOS purposes.

The following analysis considers the work of Nils Newman, Director of New Business Development at Search Technology in Atlanta, Georgia, and former analyst with SRI and the Technology Policy and Assessment Center (Georgia Tech). Newman has 13 publications indexed on WOS, which is a convenient number for purposes of the present study. Table 1 reveals that when each of Newman's publications is individually searched for in WOS, the citation report indicates that the total number of citations is equal to total citations minus self-citations for each and every publication.³ In other words, Newman's publications on WOS, considered individually, do not contain self-citations according to the citation report.

We see from Table 2, however, that when Newman's publications are entered in the aggregate into the advanced search tool of WOS, the WOS citation report reveals that the total number of citations is greater than total citations minus self-citations. In other words, Newman's research, taken as a whole, does contain self-cites (but all of his publications had to be entered into the WOS search input simultaneously to produce this result). Self-cites appear for Newman's research in the aggregate because seven of the articles in the citing articles list also appear in the list of Newman's 13 publications (we note, however,

¹ see: www.isiknowledge.com/wos.

² see: http://images.webofknowledge.com/WOKRS55B6/help/WOS/hp_citation_report_citing_articles_wo_selfcites.html.

³ The statistics used in this paper were collected on February 17, 2012.

Table 2 WOS citation report for Newman's aggregate publications on WOS

Sum of the times cited	60
Sum of times cited without self-citations	47
Citing articles	47
Citing articles without self-citations	40
Average citations per Item	4.69
h-index	5

Table 3 Citation report for Newman's publications using the definition of self-citation advanced in the present study

Newman ISI	Citing articles	Citing articles without self-citations (new definition)	Citing articles without self-citations (WOS definition)
A1996TT20800002	20	17	17
000172575700001	10	7	7
A1996VH96100001	8	5	5
000174838500002	6	3	3
000224846300007	5	1	4
000233752700005	4	4	4
000262416800002	4	3	4
000240863700001	1	1	1
000275746700001	1	1	1
000288932700005	1	1	1
000168824702172	0	0	0
000189300600002	0	0	0
000283110100034	0	0	0

that not every citing paper that contains an author name which matches an author name on a Newman paper appears in Newman's list of self-cites on WOS). Comparing the 47 citing articles to the 40 citing articles without self-citations reveals that WOS treats the Newman ISI unique article identifiers (hereafter: "ISIs") with an asterisk (*) beside them in Table 1 as self-citations.

In light of the above findings, the definition of self-citation that is most appealing and intuitive to the authors of the present study is: a citing paper is termed a self-citation if any of its authors' names match any of the author names on the cited paper (i.e. if paper A cites paper B and any of the author names on the former also appear on the latter, this would constitute a self-citation) (Egghe and Rousseau 1990). WOS does not employ this definition, and the difference in citation outcomes is worth highlighting. The complications of second-order self-citations, which go beyond the scope of the present analysis, are not considered here. Table 3, which was built using a macro written by the first author, applies the self-citation definition advanced in this study to the work of Nils Newman.

We note from the above table that Newman's citation statistics change considerably when applying this definition. The definition advanced in this paper classifies more citing articles as self-citations than does the definition advanced by WOS. In particular, the above outcome results when a paper that does not contain Newman's name cites a paper that does contain Newman's name along with the name of the citing author. At least four instances of this occur in the above analysis. Hence, even when all of Newman's ISIs are entered into

WOS simultaneously, not all citing articles containing an author's name that is an exact match to an author's name on a Newman ISI will be identified as a self-citation in the WOS citation report.

This study advances the present definition of self-citation over the algorithm used by WOS for the following reasons: (1) the argument can be made that if a given author's name appears on both a citing and cited publication, this constitutes a self-citation (Egghe and Rousseau 1990), and (2) to identify self-citations using the definition employed by WOS one must simultaneously search for all of a given author's publications—not only can this be problematic for authors with common names and/or a large volume of publications, but it also fails to identify all matches for author names that appear on both citing and cited papers. By separating all self-citations to a given body of work, citation indicators will produce more meaningful results.

References

- Aksnes, Dag. (2003). A macro study of self-citation. *Scientometrics*, 56(2), 235–246.
- Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*. The Hague, The Netherlands: Elsevier.
- Glanzel, 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.
- Rousseau, R. (1999). Temporal differences in self-citation rates of scientific journals. *Scientometrics*, 44(3), 521–531.
- Schreiber, M. (2007). Self-citation corrections for the Hirsch index. *Europhysics Letters*, 78(3), 1–6.
- Schubert, A., Glanzel, W., & Thijs, B. (2006). The weight of author self-citations. A fractional approach to self-citation counting. *Scientometrics*, 67(3), 503–514.