

Implications of references instead of citations for the bibliometric laws and science evaluation

Teresa Sillio^a and Isidro Aguillo^b

^a **INFO I+D** C/ Einstein, 13, Pabellón C, 1ª planta Campus Universitario de Cantoblanco, 28049 Madrid, SPAIN, tsillio@infoimasd.es
^b **CINDOC-CSIC**, Joaquín Costa, 22, 28002 Madrid, SPAIN

ABSTRACT

INTRODUCTION Citation indexes are produced by counting de-contextualised citations, which result from the transformation of references into citations. This causes an ambiguity of both signs which is relevant for citation indexes and for science evaluation

METHODOLOGY On one hand, from a qualitative point of view, it is performed a semiotic analysis in order to make a difference between the bibliographic references and the citations in terms of electronic reference linking. On the other hand, from a quantitative point of view, citations and references were counted within a relevant sample of papers from different scientific disciplines

RESULTS Systematically, the number of bibliographic references is higher than the number of citations in almost the 100 % of the articles analysed. In many cases, the number of references is twice as the number of citations. The 30,1 % of the citations were referenced several times, for, were counted just once

DISCUSSION By considering references instead of citations, the following bibliometric laws and indicators are affected,: Bradford / De Solla's law, direct academic journal citations, Bibliographic coupling, co-citation, impact factor and *h*-index

CONCLUSION A semiotic ambiguity between references and citations causes a relevant error in science evaluation that can be avoided by the use of electronic reference linking

INTRODUCTION

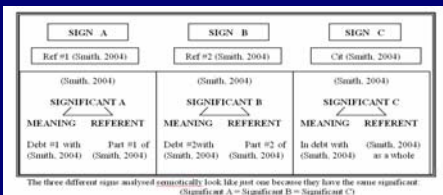
Scientific communication can be represented as a network of scientific papers (De Solla, 1965). Every scientific paper is in debt with all the precedent papers related. The value of a paper is expressed in its citation frequency (Wouters, 1999). The scientific value of each paper has been determined by quantitative criteria based on de-contextualised citation counts (Garfield, 1979, 1972). The production of citation indexes results from the transformation of references into citations (Wouters, 1999). Citation indexes count at most only one citation, no matter how many times it was mentioned. The meaning of a reference is not considered, treating the references as "de-contextualised citations". This has caused a confusion between the bibliographic references of a text and the final list of citations of the same text. References and citations are two different signs which should be analysed in terms of linking analysis –quantitative analysis-. In practice, scholarly journal publishing makes a distinction between articles, references and citations in terms of *URLs* (Caplan & Arms, 1999, Hitchcock, 2002, Barrueco, 2002) and citation databases of research literature like *Scopus*, *CiteSeer* or *Google Scholar* are context-sensitive because they are based on the Web

MATERIALS AND METHODS

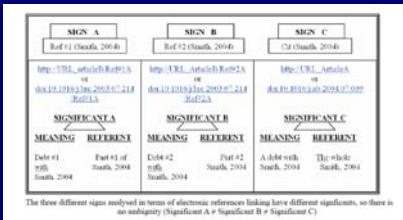
Methodology: Qualitative analysis The qualitative analysis is carried out in terms of semiotic analysis of signs (Cronin, 2000, Wouters, 1999), and in terms of linking analysis: An article B has two references to paper A (Smith, 2004), and both references link with the final single citation (Smith, 2004). The three different signs are i.e. Ref #1 (Smith, 2004), Ref #2 (Smith, 2004), and Cit (Smith, 2004). There is an ambiguity between the three of them due to a semiotic phenomenon called "homonymy". When the three signs are analysed in terms of semiotic analysis (significant, meaning, and referent) it is seen that what the three signs have in common is only the same **significant** but nothing else. The ambiguity disappears completely with the electronic reference linking (Caplan & Arms, 1999, Hitchcock, 2002) decomposition of these signs. The three signs in terms of semiotic analysis have the same significant, and that's the source of the error: Only the significant is considered within citation analysis, which performs a meaningless de-contextualised analysis. The source of the error of citation indexes is to extend the homonymy between the three significant [significant A = significant B = significant C] into a homonymy between the three signs, [sign A = sign B = sign C]: Article B (Howards, 2003) refers twice to article A (Smith, 2004), and article A (Smith, 2004) will be cited only once because of ambiguity

RESULTS

Before the qualitative approach, there was an ambiguity between references and citations



In terms of linking analysis the ambiguity is vanished



The 30,1 % of the citations were referenced several times, for, were counted just once

Journal	Article #1	Article #2	Article #3
Journal #1 (Social Sciences)	37,50 %	30,43 %	36,36 %
Journal #2 (Social Sciences)	10,0 %	17,64 %	11,76 %
Journal #3 (Humanities)	25,53 %	20,33 %	26,78 %
Journal #4 (Medicine)	43,42 %	48,57 %	58,16 %
Journal #5 (Medicine)	15,0 %	27,27 %	29,41 %
Journal #6 (Medicine)	45,90 %	36,97 %	41,30 %
Journal #7 (Biology)	10,52 %	26,19 %	48,88 %
Journal #8 (Mathematics)	52,94 %	26,31 %	22,58 %
Journal #9 (Chemistry)	26,53 %	53,57 %	29,16 %
Journal #10 (Engineering)	37,50 %	0,0 %	9,09 %
Average			30,1 %

Table 2. Percentage of citations referenced several times, but were counted just once

Methodology: Quantitative analysis These differences in *URLs* have offered the possibility to perform a quantitative analysis of references and citations. Once it has been stated out that references and citations are two different signs, a representative sample of 30 scientific papers has been quantitatively analyzed in order to determine whether or not the number of references is higher than the number of citations. These 30 articles have been extracted from ten journals of the Directory Open Access of Journals (*DOAJ*), all of them written in English and with an open access bases licence. The articles chosen were the articles number four, five and six of the first issue of year 2004. The ten electronic journals belong to the following scientific areas: Social Sciences (two journals), Humanities (one journal), Medicine (three journals), Biology (one journal), Mathematics (one journal), Chemistry (one journal) and Engineering (one journal)

DISCUSSION

By considering references instead of citations, other bibliometric laws and indicators are affected:

	CHANGES IN BIBLIOMETRIC LAWS AND INDICATORS
BIBLIOMETRIC LAWS: LOTKA'S LAW	There are no changes within the distribution of scientific productivity
BIBLIOMETRIC LAWS: ZIPP'S LAW	There are no changes within the frequency in the distribution of the words of a text
BIBLIOMETRIC LAWS: BRADFORD'S LAW / PRICE	There is a change: The 49 % of the existing documents are cited once, and it can be said as an hypothesis extracted from the results that those papers were referenced three times
INDICATORS: DIRECT ACADEMIC JOURNAL CITATIONS	There is a change: affecting the direct relation between documents: It used to be 1/1 with the citation system. With the reference system it is not 1/1 anymore, and now influence between articles can be weighted. Knowing which cited papers were more relevant for a paper. This information is more important than citations in order to determine which are the seminal papers of a discipline
INDICATORS: BIBLIOGRAPHIC COUPLING	There is a change: Two articles cited once by a citing paper, are similar, in a proportion of 1/1. But two articles referenced by a citing paper more than one time have a higher degree of similarity, a higher bibliographic coupling
INDICATORS: CO-CITATION / JACCARD COEFFICIENT	There is a change: Two articles that cite once the same paper are partially similar. But two articles that make several references to the same paper have a higher degree of similarity. The co-reference frequency, that is, the intersection of two sets of references is higher
INDICATORS: IMPACT INDEXES	There is a change: The average number of times that published papers are referenced to two years after publications (calculated based on a three-year period). The probability of being referenced if publishing in a journal
INDICATORS: H-INDEX	There is a change: Index h: the number of papers with citation number "h. A scientist has index h: if % of his/her $\frac{h}{2}$ papers have at least h references each, and the other $(2h - h)$ papers have no more than h references each
CITATIONS PAGE RANK	There is a change: In order to identify the "authoritative" authors, just like the PageRank algorithm does with WebPages, this "citation rank" should be a "reference rank"

Table 3. Implications of the consideration of references instead of citations for the bibliometric laws and science evaluation

CONCLUSION

As the value of an article is expressed in its citation frequency, the semiotic ambiguity between references and citations, causes the "reference /citation problem" what causes a relevant error with consequences in science evaluation, as text references are identified with the final citations. This error is avoided by the use of electronic reference linking. The "reference /citation problem" should be considered as an electronic reference linking problem because the future of citation indexes is the future of the Web