

Counting citations in texts rather than reference lists to improve the accuracy of assessing scientific contribution

Citation frequency of individual articles in other papers more fairly measures their scientific contribution than mere presence in reference lists

Wen-Ru Hou¹⁾, Ming Li²⁾ and Deng-Ke Niu^{2)*}

Introduction

Being cited is a popular measure of the scientific contribution of a scientific paper and consequently a well-used measure of the academic reputation of the authors, their institutions, and the journal that published it [1–6]. In common citation index systems, like ISI Web of Science, Scopus, and Google Scholar, all citations are treated equally. However, all authors would agree that references listed in the bibliography of a paper often differ greatly in their contribution to that paper. Some references are indispensable; they directly stimulate hypotheses or provide essential methods. By contrast, some other refer-

ences are cited just for background information or are incidentally mentioned. An early analysis of 575 references in 30 articles published in *Physical Review* has shown that about 40% of the references are perfunctory, which raises doubts about the use of citations as a measure of scientific contribution [7].

To solve this problem, researchers have advocated a different strategy to that of simply counting the references in a bibliography, i.e. content analysis of references [7, 8]. With this approach, the functions of citations are classified by analyzing the contexts in which the references appear. Close reading and expert judgment limit the application of this strategy in large-scale analysis.

In recent years, the number of approaches for automatic classification of citations by key words or phrases has grown [9–11].

Here, we present a simple alternative approach to improve the accuracy of citations as a measure of scientific contribution: counting citations in texts. The underlying hypothesis is very simple. Those important references that make a major contribution to a given study appear in the text more frequently, while references providing only background information are mentioned just once in the text. By counting the appearance of each reference in the text, we can obtain a new citation frequency that reflects the scientific contribution of each reference more accurately.

We tested our hypothesis by examining first whether closely related references appear more frequently in texts. Then, we tested whether our approach could significantly reduce the typical faults in using citations as a measure of scientific contribution.

Keywords:

■ citation; credit; impact factor; reference list; scientific contribution

DOI 10.1002/bies.201100067

¹⁾ National Institute of Biological Sciences, Zhongguancun Life Science Park, Beijing, China

²⁾ Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, China

Abbreviations:

CRR, closely related reference; **LRR**, less related reference

Supporting information online

References that contribute more to a paper are cited more frequently in the text

Except for incorrect citations, all the references in a scientific paper should be assumed necessary to properly sup-

*Corresponding author:

Deng-Ke Niu

E-mail: dknui@bnu.edu.cn, dengkeniu@hotmail.com

port the science presented. To avoid the problems of subjectivity associated with manual classification of references, we instead organized references by relatedness. Suppose that a paper A published in 2008 has two references, papers B and C, both published in 2006. If paper B is very similar in content to paper A, while paper C is dissimilar, it is reasonable to assume that reference B has contributed more to paper A than reference C. According to our hypothesis, reference B should appear more frequently in the text of paper A than reference C.

However, it is still difficult to manually distinguish between closely related references (CRRs) and less related references (LRRs) on a large scale. In the ISI Web of Science, the relatedness between scientific papers is defined by the number of common references they cite. In the present study, we adopted this simple strategy to distinguish between CRRs and LRRs. For the reference list of a paper published in 2008, we focused only on references published in 2006. The two-year lag was selected based on the common research-publication cycle in the biological sciences: A research group might read a stimulating paper in 2006, carry out a new study based on that paper, and publish the results in 2008. Among the 2006 references of each 2008 paper,

we defined CRRs as references having 10 or more references in common with the 2008 paper and LRRs as having fewer than 10 common references. In total, we analyzed 651 papers published in 2008 under the categories of “Biochemistry & Molecular Biology” and “Genetics & Heredity” in the Web of Science. As shown in Fig. 1, CRRs were cited more frequently in texts than LRRs (Wilcoxon signed-ranks test, $p = 4 \times 10^{-28}$). On average, in the texts of each 2008 paper, each CRR appeared 3.35 times and each LRR appeared 1.88 times. The same pattern was observed when only five common references were used as the threshold between CRRs and LRRs and again when most and least related 2006 references were compared (data not shown).

Sometimes different aspects of the same reference are cited in different parts of a paper. In these cases, different appearances of the same reference contribute to the paper independently. This shows that counting citations in texts is a more accurate means of assessing the scientific contribution of such references than counting citations in the reference list. As we can see, counting citations in the text rather than in the reference list should be considered the natural null hypothesis reflecting the scientific contributions of cited refer-

ences. Rejection of the former rather than the later requires convincing reasons. The reason for the widespread usage of counting citations in reference lists is its convenience. Actually, accession of the full texts of numerous scientific papers with multifarious formats has limited the scale of our study. However, we believe that in an increasingly computer-automated future, accessibility and multifarious formats will not be an obstacle to the application of counting citations in texts.

Counting citations in texts returns credit to discoverers

If counting citations in texts is a more accurate measure of scientific contribution, we would expect that it can avoid or significantly reduce the negative features of using citations as a measure of scientific contribution.

The rapid growth in scientific literature has made it increasingly difficult for scientists to read every paper relevant to their research. This has made review papers, with their systematic examination of scholarly advances within given scientific disciplines over brief periods of time, particularly useful. As a result, review papers are often cited in contexts that properly should cite the original papers. In addition, the space constraints of many journals often force authors to cite one review paper instead of several original ones [12, 13]. This trend has led to the complaint that the practice of citing review papers “diverts academic credit from the discoverer” [12]. It is certainly true that review journals climb to the top of the ranking lists, and this diverts attention from the journals that published the primary research as well as away from the discoverers themselves [14]. Although the journal impact factor has been criticized, it remains the most popular measure of the overall quality of academic journals [15]. It has been widely misused as an indicator of the quality of the papers that a journal publishes. Although counting citations within the text does not get around limits to the number of citations, we expect that replacing raw counts of citations in reference lists with citations in text

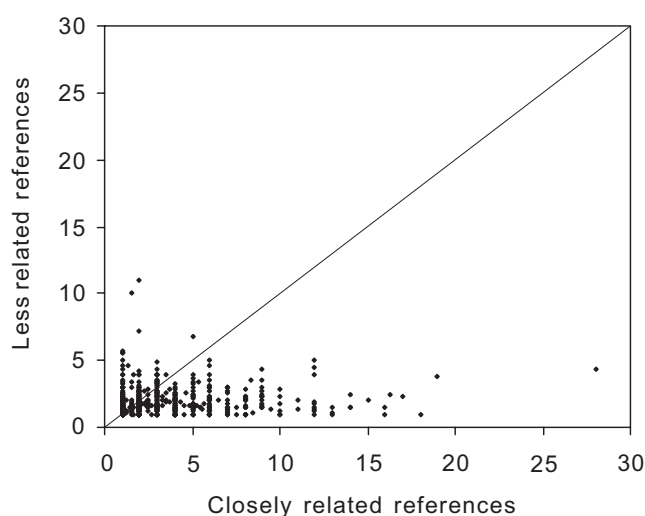


Figure 1. Frequencies of the appearance of closely related and less related references in the texts. We surveyed the references of 651 scientific papers published in 2008. The references published in 2006 were divided into closely related and less related references. If a 2008 paper had multiple closely related references, we used the mean value of the frequencies of the appearance of these references. The same strategy was applied to less related references. Closely related references appear more frequently in the texts of the 2008 papers than less related references (Wilcoxon signed-ranks test, $p = 4 \times 10^{-28}$).

would revert some credit to the discoverers and the journals in which the original research was published.

From two subject categories, “Biochemistry & Molecular Biology” and “Genetics & Heredity” in ISI Journal Citation Reports, we obtained a list of 404 journals. To avoid data bias, we retained only those journals that published 50–500 papers a year. Review journals were treated separately. A review journal, such as *Annual Review of Genetics*, would be retained if it published more than 10 papers a year. To analyze the frequency of citations in the texts, we had to access the full texts. So journals where we could not access the full texts were excluded from the study. For the convenience of parsing the citations in the texts, we included only those journals whose articles are available online with HTML format. In total, 75 journals were used. As a comparative study with the traditional citation analysis, we examined only the articles (meaning, overwhelmingly, those that report primary research results) and reviews as recognized by the Web of Science. In total, the 75 journals published 8,658 article and review papers in 2008, and 14,537 article papers and 2,580 review papers from 2006 to 2007. We counted the appearance of these 2006–2007 papers both in the texts and in the reference lists of the 2008 papers. As one might expect, in the reference lists of the 2008 papers, the citation frequency of the 2006–2007 review papers was 1.64 times that of the 2006–2007 article papers. In the texts of the 2008 papers, the citation frequency of the 2006–2007 review papers was decreased to 1.26 times that of the 2006–2007 article papers. The citation frequency of article papers increased more than that of review papers (2.18 vs. 1.68 times) when replacing counts of citations in reference lists by citations in the texts. Although review papers still have a higher citation frequency with our method, the disadvantage of citation counting is decreased, and greater credit is reverted to the discoverers.

Furthermore, we compared the 2-year journal impact factors calculated by counting citations in reference lists and counting citations in texts. The impact factor of a journal in a year is calculated following the method in ISI Journal Citation Reports: the quotient of

the number of current citations to articles published in the 2 previous years over the total number of articles the journal published in the 2 previous years. When counting citations in reference lists is replaced by counting citations in texts, the impact factors of almost all the journals increased, and the ranking of the impact factors of 67 among the 75 journals changed (see Supporting Information online). As shown in Fig. 2, the change in the ranking of the impact factor depends significantly on the percentage of review papers a journal published (Spearman's $\rho = -0.30$, $p = 0.008$). Among the 14 review journals we surveyed, in terms of the ranking of the impact factor, 11 journals decreased and the other 3 journals showed an increase of only 1. The top three journals that decreased in the ranking were all review journals: *Trends in Molecular Medicine*, *Trends in Microbiology*, and *Current Opinion in Chemical Biology*. By counting citations in the texts, the disadvantage of journal impact factors is reduced.

It is to be expected that other measures of scientific contribution based on citation counting, such as the h -index [2], could also be improved by counting citations in texts. Notably, some interdisciplinary works may produce smaller citation counts than their

counterparts within the same field, but otherwise also present critical values to the articles. These interdisciplinary references may therefore receive some compensation when setting the standard for CRRs.

Inaccurate or incorrect citations is a common problem in the present academic world [16]. We believe inaccurate or incorrect citations are for the most part incidentally mentioned and so are less likely to appear many times in a text. Counting citations in a text will reduce the undesirable impact resulting from inaccurate or incorrect citations.

Conclusions

Citations are widely used to measure the degree or level of scientific contribution. However, not all references contributed equally to the papers in which they are cited. By comparing the appearances of CRRs and LRRs in texts rather than in reference lists, we showed that counting citations in the text more accurately reflects the scientific contribution. In addition, we found that counting citations in text helped to avoid over-allocation of credit to review authors and to journals that publish review papers over the authors of original studies and journals that publish original studies. For

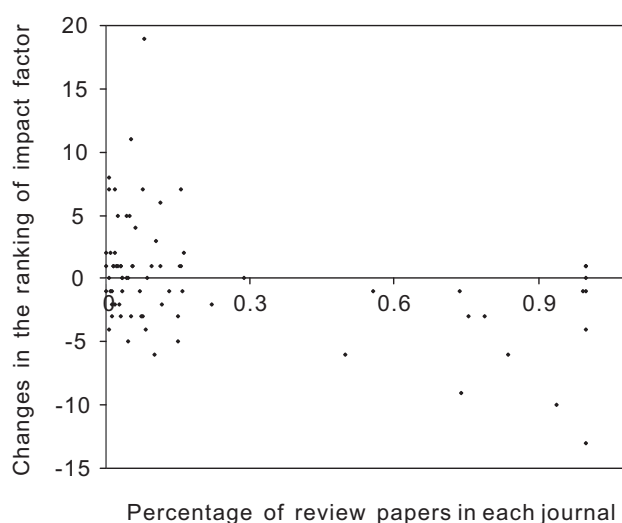


Figure 2. Journals with a high proportion of review papers decrease in the rank position according to impact factor when, instead, we used the method of counting the numbers of citations in the text of articles. We surveyed 75 biomedical journals. Their impact factors were calculated both by counting citations in the texts and counting citations in reference lists. When we replaced counting citations in reference lists by counting citations in the texts, the ranking of the impact factor of most journals changed. The change is significantly related to the percentage of review papers in each journal (Spearman's $\rho = -0.31$, $p = 0.006$).

these reasons, we advocate this new strategy for the measurement of scientific contribution: counting citations in text.

Acknowledgments

We thank Wan-Jiao Wang, Dr. Andrew Moore and an anonymous referee for helpful suggestions on the paper. This work was supported by the National Natural Science Foundation of China (grant no. 31071112).

References

1. **Garfield E.** 1972. Citation analysis as a tool in journal evaluation. *Science* **178**: 471–9.
2. **Hirsch JE.** 2005. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* **102**: 16569–72.
3. **Kelly CD, Jennions MD.** 2006. The h index and career assessment by numbers. *Trends Ecol Evol* **21**: 167–70.
4. **Aguillo IF, Bar-Ilan J, Levene M, Ortega JL.** 2010. Comparing university rankings. *Scientometrics* **85**: 243–56.
5. **Zhang CT.** 2010. Relationship of the h-index, g-index, and e-index. *J Am Soc Inf Sci Technol* **61**: 625–8.
6. **Winker K.** 2011. In scientific publishing at the article level, effort matters more than journal impact factors. *BioEssays* **33**: 400–2.
7. **Moravcsik MJ, Murugesan P.** 1975. Some results on the function and quality of citations. *Soc Stud Sci* **5**: 86–92.
8. **Chubin DE, Moitra SD.** 1975. Content analysis of references: adjunct or alternative to citation counting? *Soc Stud Sci* **5**: 423–41.
9. **Teufel S, Siddharthan A, Tidhar D.** 2006. Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Sydney, Australia: Association for Computational Linguistics.
10. **Jörg B.** 2008. Towards the nature of citations. In Eschenbach C, Grüninger M, ed; *Formal Ontology in Information Systems: Fifth International Conference (FOIS 2008)*. Germany: DFKI GmbH.
11. **Radoulov R.** 2008. *Exploring Automatic Citation: Classification*. Waterloo: University of Waterloo.
12. Anonymous. 2009. Credit where credit is due. *Nat Cell Biol* **11**: 1.
13. **Wiley S.** 2010. Down with reviews: review articles simply don't deserve all the citations they receive. *Scientist* **24**: 31.
14. Anonymous. 2003. Making impact. *Nat Cell Biol* **5**: 681.
15. Anonymous. 2005. Not-so-deep impact. *Nature* **435**: 1003–4.
16. **Pearce WJ.** 2010. Citations: too many, or not enough? *Scientist* **24**: 29.