To what degree are uni-citations perfunctory?
A case study

Dangzhi Zhao (dzhao@ualberta.ca), **Lucinda Johnston** (Lucinda.johnston@ualberta.ca)
School of Library and Information Studies, University of Alberta, Edmonton, Canada

INTRODUCTION

Central amongst critiques of the current practices of citation analysis has long been that it treats all citations equally, be they crucial to the citing paper or perfunctory. This problem is especially serious when tracing or assessing research impact. Weighting citations by how they are used in the citing paper has therefore long been proposed as a theoretically promising solution to this problem (Herlach, 1978; Narin, 1976). Recently, studies have experimented with weighing citations by the frequency with which they are referred to in the text (e.g., Ding, Liu, Guo, & Cronin, 2013; Hou, Li, & Niu, 2011; Zhu, Turney, Lemire, & Vellino, 2014), by the citation impact of citing papers (Ding & Cronin, 2011), or by the location and context in which they are cited (Boyack, Small, & Klavans, 2013; Jeong, Song, & Ding, 2014). Among many features of in-text citations, in-text frequency weighted citation analysis schemes have been found to be the most effective at characterizing essential citations (Zhu, Turney, Lemire, & Vellino, 2014).

If the signal to be detected in citation analysis is the direct and substantial flow of knowledge from the cited to the citing papers, perfunctory citations can be considered a serious source of noise because a high incidence of perfunctory citations (40% or more) has been repeatedly observed (Tabatabaei, 2013; Teufel, Siddharthan, & Tidhar, 2006; Moravcsik & Murugesan, 1975). There are two obvious types of approaches to dealing with noise: to filter out the noise, or to amplify the signal. The ultimately best approach is likely some combination of the two.

Compared to the signal amplification approach which is used by almost all frequency-weighted citation counting schemes found in the literature, the noise filtration approach attempts to make the fundamental qualitative distinction between those citations that represent real use by, or core impact on or connection with, the citing paper (which it aims to retain for analysis) and those that are merely mentioned in passing as related work or background information (which it aims to remove). By focusing on core connections in knowledge networks, this approach can help research evaluation become more sensitive to the distinction between essential and perfunctory impact of research, and can better capture "aboutness" of documents, the essence of subject indexing in knowledge representation and retrieval. Knowledge representation and retrieval systems that make use of citation links can therefore benefit from improved precision in computer-aided subject indexing and in their "More like this" features (Zhao & Strotmann, 2016). In addition, the signal amplification required to counter the very strong noise created by perfunctory citations (40% or more) tends to be so strong (N^2 is the minimal power of N required) that it can cause serious distortions (Zhao & Strotmann, 2016). Filtering out this strong noise before applying necessary signal amplification can avoid this potential technical problem.

The key and difficult question is now how to identify and then filter out perfunctory citations. We introduced recently a simple method for this: re-citation analysis, which focuses on re-citations – references that appear more than once in the text of a citing paper, by filtering out uni-citations – documents referenced only once in a work (Zhao & Strotmann, 2016). The basic assumption of re-citation analysis is that papers are very likely to be cited again and again in a publication that relies heavily on them, while perfunctory citations should appear once only in a citing paper almost by definition.

The present study is a preliminary test of this basic assumption underlying re-citation analysis. Its results will be important for assessing if the great potential that the noise filtration approach promises in improving citation analysis for research evaluation, knowledge network analysis, knowledge representation, and information retrieval (Zhao & Strotmann, 2015) can be realized by re-citation analysis. Results from this study may also shed light on other directions to realize this potential.

METHODOLOGY

We chose to examine articles in the field of library and information science, a field that we understand well and can therefore assess more accurately the functions that references have in the articles. This is a case study and like all case studies the generalizability of its results may be limited.

Specifically, we chose all Research Articles in a single issue of the Journal of the Association for Information Science and Technology – 2016 volume 67 issue 1, and coded all in-text citations in these articles as to their importance and function to the context in which they appear. There are 14 articles and 1473 in-text citations in total.

The coding scheme we used is the one developed in Tabatabaei (2013) for assessing the contribution of information science to other disciplines as reflected in citation contexts of highly cited JASIST papers because the coding purposes of these two

studies align very well. This coding scheme has five categories: Applied, Contrastive, Supportive, Reviewed, and Perfunctory, which is very similar to the five typical distinctions in citation classification schemes identified by Small (1982): (1) negative or refuted, (2) perfunctory or noted only, (3) compared or reviewed, (4) used or applied, and (5) substantiated or supported by the citing work. The detailed explanation and coding examples for each of the categories in the scheme provided in Tabatabaei (2013) helped ensure the accuracy and consistency of coding in the present study.

The 14 source articles were processed in a random order. Each of the in-text citations was classified into one of the five function categories, and its location in the text (introduction, lit review, methodology, findings, discussions, conclusions, other) was also recorded.

RESULTS AND DISCUSSION

About 50% of the in-text citations that we examined are perfunctory, confirming findings from previous studies. About 1/3 of the in-text citations are uni-citations (33% to be precise) and 67% are re-citations.

Figure 1 was a total surprise to us as we didn't see high percentage of essential citations (Applied, Contrastive, Supportive) in re-citations, or close to absence of essential citations in uni-citations. Instead, both uni- and re-citations show very similar distribution by function and are dominated by inessential citations (Reviewed and Perfunctory) – 74% of uni- and 78% of re-citations.
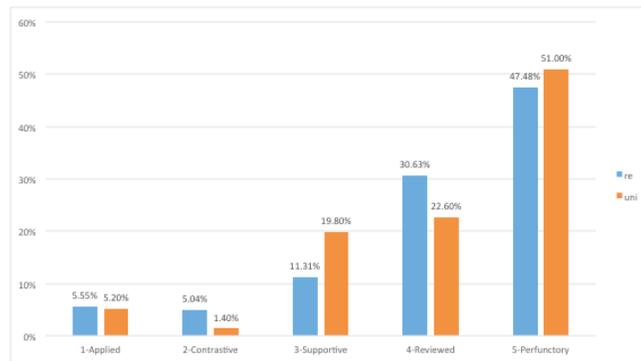

Figure 1. Distribution of uni- and re-citations by function.

This result means that re-citation analysis would exclude more than a quarter of uni-citations unfairly and close to half of the citations that it retains would be perfunctory. In other words, 8.6% of total citations and more than 1/3 of all essential citations would be removed but the percentage of perfunctory citations in the citations to be analyzed would remain more or less unchanged.

If removing uni-citations doesn't seem to be effective in filtering out perfunctory citations, what is a more effective way then? Figures 2 and 3 suggest that removing citations by location may be more effective. More than 90% of both uni- and re-citations in the Background and Related Studies/Literature Review sections are either perfunctory or reviewed. So are more than 90% of uni-citations and more than 80% of re-citations in the Introduction section.
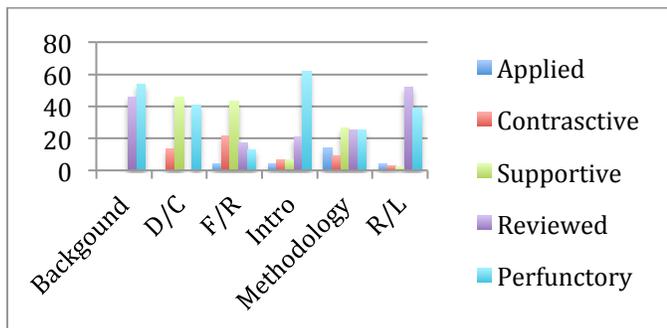

Figure 2. Percentage of re-citations in section total.
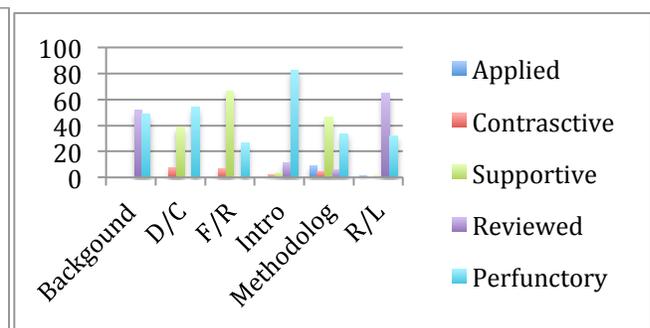
Figure 3. Percentage of uni-citations in section total.

CONCLUSION

The basic assumption underlying re-citation analysis that perfunctory citations correspond largely to uni-citations doesn't seem to hold. Removing citations by location seems more effective in filtering out inessential citations than removing uni-citations. The promises of noise-filtration approach to weighted citation counting mentioned above and explained in detail in

Zhao & Strotmann (2015) invite future studies to explore further whether removing citations by location is indeed an effective way to filter out inessential citations with a reasonable error rate.

REFERENCES

Boyack, K. W., Small, H., & Klavans, R. (2013). Improving the accuracy of co-citation clustering using full text. *Journal of the American Society for Information Science and Technology, 64*(9), 1759-1767.

Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. *Information Processing and Management, 47*, 80–96.

Ding, Y., Liu, X., Guo, C., & Cronin, B. (2013). The distribution of references across texts: Some implications for citation analysis. *Journal of Informetrics, 7*, 583-592.

Herlach, G. (1978). Can retrieval of information from citation indexes be simplified? Multiple mention of a reference as a characteristic of the link between cited and citing article. *Journal of the American Society for Information Science, 29*(6), 308-310.

Hou, W., Li, M., & Niu, D. (2011). Counting citations in texts rather than reference lists to improve the accuracy of assessing scientific contribution. *BioEssays, 33*, 724-727.

Jeong, Y. K., Song, M., & Ding, Y. (2014). Content-based author co-citation analysis. *Journal of Informetrics, 8*(1), 197-211.

Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science, 5*(1), 86-92.

Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Washington, D. C.: Computer Horizons.

Peritz, B. C. (1992). On the objectives of citation analysis: Problems of theory and method. *Journal of the American Society for Information Science, 43*(6), 448-451.

Shadish, W. R., Tolliver, D., Gray, M., & Gupta, S. K. S. (1995). Author judgements about works they cite: three studies from psychology journals. *Social Studies of Science, 25*(3), 477-498.

Small, H. (1982). Citation context analysis. *In B. J. Dervin & M. J. Voigt (eds.), Progress in communication sciences, 3* (pp. 287-310). Norwood, NJ: Ablex.

Teufel, S., Siddharthan, A., & Tidhar, D. (2006). Automatic classification of citation function. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (pp. 103-110)*. Stroudsburg, PA, USA.

Tabatabaei, N. (2013). *Contribution of information science to other disciplines as reflected in citation contexts of highly cited JASIST papers.* McGill University, Montreal, PQ.

Zhao, D., & Strotmann, A. (2016). Dimensions and uncertainties of author citation rankings: Lessons learned from frequency-weighted in-text citation counting. *Journal of the Association for Information Science and Technology*, 671-628.

Zhao, D., & Strotmann, A. (2015). Re-citation analysis: Promising for research evaluation, knowledge network analysis, knowledge representation and information retrieval? *Proceedings of the 15th International Society for Scientometrics and Informetrics Conference*, June 30 - July 3, 2015, Istanbul, Turkey.

Zhu, X., Turney, P., Lemire, D., & Vellino, A. (2014). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology*. Early view (DOI: 10.1002/asi.23179).