

Can conference papers have information value through Wikipedia?

An investigation of four engineering fields

Kuku Joseph Aduku*, Mike Thelwall**, and Kayvan Kousha***.

K.J.Aduku@wlv.ac.uk; M.Thelwall@wlv.ac.uk; K.Kousha@wlv.ac.uk

Statistical Cybermetrics Research Group, School of Mathematics and Computer Science,
University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK

Abstract

Wikipedia provides a widely-used overview of many academic fields, often referencing journal articles and books to justify its content. Previous studies have shown that these citations can, in turn, be used to help assess the knowledge transfer impact of the cited articles and books. Nevertheless, it is not known whether the same is true for conference papers. To fill this gap, citations in Wikipedia and Scopus were compared for conference papers (and journal articles) published in 2011 in four engineering fields that value conferences. Wikipedia citations had correlations that were statistically significantly positive only in Computer Science Applications, whereas the correlations were not statistically significantly different from zero in Building & Construction Engineering, Industrial & Manufacturing Engineering and Software Engineering. Conference papers were less likely to be cited in Wikipedia than were journal articles in all fields, although the difference was minor in Software Engineering. Hence, there is little evidence that Wikipedia citations are valuable as impact indicators in engineering fields, especially in the case of conference papers.

Keywords

Wikipedia citations; Citation analysis; Conference papers; Alternative indicators

Introduction

Wikipedia provides a widely-used overview of many academic fields, often referencing journal articles and books to justify its content. Previous studies have shown that these citations can, in turn, be used to help assess the knowledge transfer impact of the cited articles and books. Wikipedia has become a standard resource in education and perhaps also for researchers. In the USA, for example, over a third of college students used Wikipedia by 2013, despite concerns about its quality and reliability (McKerlich, Ives, & McGreal, 2013; Aibar et al., 2015., Knight & Pryke, 2012; Soules, 2015).

Although citation counts are widely used to support research evaluation, they can only reflect academic impacts, whereas research can also be useful outside of academia. Many alternative indicators derived from patents (Trajtenberg, 1990), the general web (Thelwall & Kousha, 2015) or the social web (Priem, Taraborelli, Groth, & Neylon, 2010) have therefore been proposed to reflect the different types of impacts that documents can have. The value of these alternative indicators has been assessed primarily through correlation tests on the basis that robust new indicators would be likely to have a positive correlation with citation counts even

if they reflected different types of value (Sud & Thelwall, 2014). On this basis, Mendeley reader counts are particularly promising impact indicators (Thelwall, M., Haustein, Larivière, & Sugimoto, 2013). Previous studies have mainly focused on journal articles because they are the primary scholarly outputs in most field but Mendeley readership counts have also been investigated for conference papers in engineering fields (Aduku, Thelwall & Kousha, 2016) finding statistically significant positive correlations with citation counts in computing fields. Similarly, Google Patents citations have been assessed in four engineering fields (Computer Science Applications, Software Engineering, Industrial & Construction Engineering and Building & Manufacturing Engineering), finding few Google Patents citations but positive significant correlations between Scopus citation counts and Google Patents citations for conference papers in one field: Computer Science Applications (Aduku, Thelwall & Kousha in press).

Wikipedia is integrated into scholarship through being cited in a minority of academic articles. A study of English Medical Science journals cited in PubMed and Medline found 1,433 articles from 1,008 journals citing 2,048 Wikipedia articles. These main cited definitions (31.6%) and processes descriptions (23.5%) (Bould et al., 2014). Similarly, between 2005 to 2009 articles in chemistry journals cited Wikipedia at least 370 times (Brazzeal, 2011).

Citations from Wikipedia to journal articles are relatively rare, with only 5% of the Scopus articles in a set of fields having at least one Wikipedia citation, whereas 33% of a set of academic monographs attracting one or more Wikipedia citations (Kousha & Thelwall, 2016). No studies have investigated Wikipedia citations to conference papers, however. The current study fills this gap by investigating Wikipedia citations and Scopus citation counts conference papers in four engineering-related Scopus subject categories: Building & Construction Engineering; Industrial & Manufacturing Engineering; Computer Science Applications and Computer Software Engineering. These fields were chosen because conferences are known to be important in computer science and other engineering-related fields. The results are compared with journal articles from the same fields.

Background

Wikipedia is a free online collaborative encyclopedia that has experienced exponential growth since 2002 (Kittur, Chi, Pendleton, Suh, & Mytkowicz, 2007). Because of its popularity, its coverage, currency, accuracy, and readability, have all been investigated (Mesgari, Okoli, Mehdi, Nielsen, & Lanamäki, 2015). Previous studies have analysed Wikipedia contributors (Jullien, 2012; Yasseri & Kertész, 2013) and used it as raw data for text-mining (Medelyan, Milne, Legg, & Witten, 2009).

In 2008, study by (Laurent & Vickers, 2009) shows that the English Wikipedia is rank as one of the highest search engine for health related information in comparison to MedlinePlus, NHS Direct, and the National Organisation of Rare Diseases. Medical keywords were searched as queries through special software to measure Wikipedia ranking and number of pages viewed among other internet search engines (Google, Google UK, Yahoo and MNS).

The result shows that Wikipedia is rank 71-85% above the MedlinePlus, NHS Direct online and the National Organisation of Rare Diseases also, Wikipedia articles was found to have more viewed pages than the topic in the MedlinePlus. The study concluded that English Wikipedia was highly ranked among the first 10 Google search hits for medical keywords with about 70% greater than other sources of medical search engine examined.

Wikipedia has been found to be a prominent resource that contains huge amount of health information(Heilman et al., 2011) which can easily be accessed by the general public and the health care professionals. Study shows that practicing physicians uses resources from Wikipedia to provide medical care, while junior physician use Wikipedia almost every week more often than all other medical websites. Statistics shows that medical articles on Wikipedia receive about 150 million page views every month as compared to other source for medical articles that receive only 60,000 views per month.

There is a substantial evidence that Wikipedia cites astronomy research (Thelwall, 2016). Older astronomy articles are less cited, as are newer articles with the citations perhaps aligning with the years in which the core astronomy content was added to Wikipedia. However, the proportion of the citations might depend on their publication year and might vary by language version of Wikipedia. In the health sector, studies have demonstrated the reliability of Wikipedia as a source for scientific journals, where 42 scientific articles compared with articles in the Encyclopedia Britannica found similar accuracy levels. Medical articles in Wikipedia are highly viewed and an important source of information for the public (Thomas, Eng, de Wolff, & Grover, 2013).

There have been exponential increase of Scopus publications citing Wikipedia over time. Study shows that in August 2015, about 61, 135 Scopus publications cited at least one Wikipedia article higher than Encyclopaedia Britannica articles with only 7,849 Scopus publications in the same year, similar study shows that in 2001, Scopus publications have 0 citation to Wikipedia articles and 12 citations to Wikipedia in 2003. However, in 2014 there was a dramatic increase of 8,579 Scopus citations to Wikipedia articles (Kousha & Thelwall, 2016).

Several investigations have shown Wikipedia citing Scientific publications. The study of (Halfaker & Taraborelli, 2015 in Kousha & Thelwall, 2016) shows that ISBN, PubMed, DOI and ArXiv identifiers in English Wikipedia matches to Books and Monographs. Hence, 35% cite books, 2% cite academic journals. Similar study by (Luyt & Tan, 2011 in Kousha & Thelwall, 2016) shows that 62% cited in the Wikipedia articles were internet sources for social media, the study suggests that citation to Wikipedia is more in books and monographs than in academic articles.

Research Questions

The primary goal of this paper is to assess the scholarly importance of Wikipedia citations in a conference-based engineering fields. The following research questions guide the study.

- How common are there Wikipedia citations to conference papers in engineering fields and, if so, do they correlate with Scopus citations?
- Is the situation different for engineering conference papers?

Methods

Bibliographic information and citation data for journal articles and conference papers in the four fields from 2011 was extracted from Scopus in March 2015. The year 2011 gives six years for papers to attract citations from Wikipedia.

Narrow subject categories were used to ensure comprehensive coverage. From the Scopus computer science category, a field in which conferences are arguably more important than journals, the two categories of Computer Science Applications and Computer Software were chosen. Conferences appear to be important throughout engineering subjects, and so from the broad Scopus Engineering category, Building & Construction and Industrial and Manufacturing Engineering were selected.

Spearman correlations were used to compare Scopus citations and Wikipedia citations because the data are skewed. The Spearman rank correlation formula was used to calculate 95% confidence intervals for the correlation coefficients. The formula used was $\tanh(\operatorname{arctanh}(r) \pm \frac{1.96}{\sqrt{n-3}})$. Here, r is the sample correlation and n is the sample size. A 95% confidence interval was calculated by taking the transformed estimate and adding and subtracting 1.96 times its standard error (Dowdy, Wearden Chilko, 2011. p. 245-246: in Aduku, Thelwall & Kousha, 2016).

Bing searches were used to extract and filter Wikipedia citations in December 2016 since Wikipedia does not allow large scale direct automatic citation searching. The free Webometric Analyst software (<http://lexiurl.wlv.ac.uk>) was used to perform automatic searches with the Bing API by searching for the key bibliographic information of a large set of articles and papers. From the “Make Searches” menu the option “Make Wikipedia Searches for a set of Scopus/WoS/ other journal articles or books” was used. This generates queries with the last names of the authors (up to a maximum of eight) and the conference paper or journal article title as a phrase search along with the publication year and the command `site:wikipedia.org/wiki/` to restrict the results to the Wikipedia site, as the following examples illustrate.

Gálvez-López "Real-time loop detection with bags of binary words" 2011 `site:wikipedia.org/wiki/`

Watanabe Kanou Kobayashi "Development of a steerable drill for ACL reconstruction" 2011 `site:wikipedia.org/wiki/`

The above queries were searched in Bing using Webometric Analyst and all the matches were recorded for each one. Similar queries have been previously used for journal articles and found to be accurate, but not for conference papers. Hence, a manual check was carried out by the first author on each citation to ascertain whether it contained a citation to the correct conference paper as shown in Table 1 and no errors were found. Therefore, it seems the search process is safe for conference papers.

Table 1. The results of manual checks of the Bing search results for Wikipedia citations to conference papers.

<i>Subject category</i>	<i>Citations</i>	<i>Correct Citations</i>
Computer Science Applications	35	35 (100%)
Software Engineering	168	168 (100%)
Industrial & Manufacturing Engineering	9	9 (100%)
Building & Construction Engineering	7	7 (100%)

Results

There are low proportions of journal articles and conference papers cited by Wikipedia (Table 2). This could be due to low coverage of the subject areas chosen in Wikipedia, although it seems more likely that most articles and papers in these fields are not useful for Wikipedia.

Table 2. Average numbers of Wikipedia and Scopus citations (geometric mean) and percentage cited for both journal articles and conference papers.

<i>Scopus Subject category</i>	<i>Journal articles</i>		<i>Conference papers</i>	
	<i>Wikipedia citations Geometric mean (% cited)</i>	<i>Scopus citations Geometric mean (% cited)</i>	<i>Wikipedia citations Geometric mean (% cited)</i>	<i>Scopus citations Geometric mean (% cited)</i>
Computer Science Applications	0.19459 (6.68%)	3.46 (80.9%)	0.00326 (0.42%)	0.53 (34.4%)
Software Engineering	0.01197 (1.36%)	3.43 (80.6%)	0.00119 (1.35%)	1.29 (54.7%)
Industrial & Manufacturing Engineering.	0.00429 (0.52%)	2.54 (71.3%)	0.00093 (0.12%)	0.18 (17.5%)
Building & Construction Engineering.	0.00706 (0.83%)	2.43 (71.7%)	0.00205 (0.23%)	0.20 (18.3%)

There are statistically significant correlations between Wikipedia citations and Scopus citation counts only in Computer Science Applications (Table 3). In the other three subject categories, the correlations are not statistically significantly different from zero. In Computer Science Applications, the correlation is higher for conference papers than for journal articles but in both cases the correlations are low.

Table 3. Spearman correlations between Wikipedia citations and Scopus citations for articles and conference papers in Scopus from 2011 in four engineering subjects.

<i>Scopus Subject Category</i>	<i>Articles</i>	<i>Conf. papers</i>	<i>Spearman correlation for conference papers (95% CI)</i>	<i>Spearman correlation for journal articles (95% CI)</i>
Computer Science Applications	5912	6700	0.274** (0.2517, 0.2960)	0.056** (0.0306, 0.0814)
Software Engineering	8231	8085	-0.001 (-0.0228, 0.0208)	0.020 (-0.0016, 0.0416)
Building & Construction engineering	8406	1753	0.040 (-0.0068, 0.0867)	0.002 (-0.0193, 0.0234)
Industrial & Manufacturing engineering	7354	5650	-0.017 (-0.4306, 0.0091)	0.012 (-0.0109, 0.0348)

**Statistically significant at p=0.01

Discussion

There are several limitations in this study. The magnitude of the Scopus citation counts depends solely on the coverage of Scopus and so they are likely to be underestimates. Similarly, the citations from Wikipedia were obtained from Bing searches and some may have been overlooked by the Bing search algorithm. The results may also vary by year and could be different for other types of engineering. Another limitation is that conferences vary in quality, importance and purpose (e.g., profession-oriented or academic-oriented) and so the scope of the current study (analysing all Scopus-indexed conference papers) may hide differences between conference types.

The four selected engineering fields have very low Wikipedia citation counts for both conference papers and journal articles. This study therefore shows that Wikipedia has very few citations to engineering fields.

Wikipedia counts and Scopus citation counts have statistically significant positive significant correlations for journal articles and conference papers for Computer Science Applications, but not in the other fields. Thus, overall, there is little evidence that Wikipedia citations reflect scholarly impact. This could be due to low numbers of Wikipedia citations, which makes it difficult to obtain a high correlation.

For conference papers, 0.42% of the papers in Computer Science Applications have at least one Wikipedia citation, 1.35% of the papers in Software Engineering have at least one Wikipedia citation, 0.12% of the papers in Industrial & Manufacturing Engineering have at least one Wikipedia citation and 0.23% of the paper in Building & Construction Engineering have at least one citation from Wikipedia. These proportions are too low to allow Wikipedia citations to be used to assess the impact of individual articles for most purposes but they could still be used to compare the impact of groups of articles using proportion cited indicators (Thelwall, 2017). For such a calculation to be credible, however, additional

evidence would be needed of the value of Wikipedia citations in engineering due to the absence of correlation evidence in three of the four fields analysed here.

Conclusion

The low number of Wikipedia citations for the four engineering categories (Computer Science Applications, Software Engineering, Industrial & Construction Engineering and Building & Manufacturing Engineering) show that Wikipedia citations are unlikely to form powerful indicators for these fields. Nevertheless, they can still be useful for comparing the impact of large sets of documents using the proportion cited indicator that is designed for datasets that are dominated by zeros (Thelwall, 2017).

References

- Aduku, K.J., Thelwall, M., Kousha, K. (2016), Do Mendeley reader counts reflect the scholarly impact of conference papers? An investigation of Computer Science and Engineering fields. In: Ismael Rafols, Jordi Molas-Gallart, Elena Castro-Martínez, Richard Woolley, Proceedings of the 21th International conference on science and technology indicators – Peripheries, frontiers and beyond, Valencia (Spain).
- Aduku, K.J., Thelwall, M., Kousha, K. (in press), Do Patents cite conference papers as often as journal articles in engineering? An investigation of four fields. *Scientometrics*.
- Aibar, E., Lladós-Masllorens, J., Meseguer-Artola, A., Minguillón, J., & Lerga, M. (2015). Wikipedia at university:
- Bould, M. D., Hladkowicz, E. S., Pigford, A. E., Ufholz, L., Postonogova, T., Shin, E., & Boet, S. (2014). References that anyone can edit: Review of Wikipedia citations in peer reviewed health science literature. *BMJ (Online)*, 348, g1585. doi: <http://dx.doi.org/10.1136/bmj.g1585>
- Brazzeal, B. (2011). Citations to Wikipedia in chemistry journals: A preliminary study. *Issues in Science and Technology Librarianship*, 67. doi: 10.5062/F4057CV7
- Kittur, A., Chi, E., Pendleton, B. a, Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *Algorithmica*, 1(2), 1–9.
- Knight, C., & Pryke, S. (2012). Wikipedia and the university, a case study. *Teaching in Higher Education*, 17(6), 649-659.
- Kousha, K., & Thelwall, M. (2016). Are wikipedia citations important evidence of the impact of scholarly articles and books? *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23694

- Heilman, J. M., Kemmann, E., Bonert, M., Chatterjee, A., Ragar, B., Beards, G. M., ... Laurent, M. R. (2011). Wikipedia: A key tool for global public health promotion. *Journal of Medical Internet Research*, 13(1). <http://doi.org/10.2196/jmir.1589>
- Laurent, M. R., & Vickers, T. J. (2009). Seeking Health Information Online: Does Wikipedia Matter? *Journal of the American Medical Informatics Association*, 16(4), 471–479. <http://doi.org/10.1197/jamia.M3059>
- McKerlich, R., Ives, C., & McGreal, R. (2013). Measuring use and creation of open educational resources in higher education. *International Review of Research in Open and Distance Learning*, 14(4), 90–103. <http://doi.org/10.1002/asi>
- Medelyan, O., Milne, D., Legg, C., & Witten, I. H. (2009). Mining meaning from Wikipedia. *International Journal of Human Computer Studies*, 67(9), 716–754.
- Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. Å., & Lanamäki, A. (2015). “The Sum of All Human Knowledge”: A Systematic Review of Scholarly Research on the Content of Wikipedia. *Journal of the Association for Information Science and Technology*, 66(2),
- Okoli, C., Mehdi, M., Mesgari, M., Nielsen, F. Å., & Lanamäki, A. (2014). Wikipedia in the eyes of its beholders: A systematic review of scholarly research on Wikipedia readers and readership. *Journal of the Association for Information Science and Technology*, 65(12), 2381–2403. <http://doi.org/10.1002/asi.23162>
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. <http://altmetrics.org/manifesto/>
- Soules, A. (2015). Faculty perception of Wikipedia in the California state university system. *New Library World*, 116(3-4), 213-226. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2053597_code728676.pdf?abstractid=2053597&mirid=2
- Sud, P., & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, 98(2), 1131-1143.
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PloS one*, 8(5), e64841.
- Thelwall, M., & Kousha, K. (2015). Web indicators for research evaluation. Part 1: Citations and links to academic articles from the Web. *El profesional de la información*, 24(5), 587-606.
- Thelwall, M. (2016). Does astronomy research become too dated for the public? Wikipedia citations to astronomy and astrophysics journal articles 1996-2014. *El Profesional de La Información*, 25(6), 893–900. [doi:10.3145/epi.2016.nov.06](https://doi.org/10.3145/epi.2016.nov.06)
- Thelwall, M. (2017). Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics*, 11(1), 128–151. [10.1016/j.joi.2016.12.002](https://doi.org/10.1016/j.joi.2016.12.002)
- Thomas, G. R., Eng, L., de Wolff, J. F., & Grover, S. C. (2013). An Evaluation of Wikipedia as a Resource for Patient Education in Nephrology. *Seminars in Dialysis*, 26(2), 159–163. <http://doi.org/10.1111/sdi.12059>

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics*, 21(1), 172-187.

Yasseri, T., & Kertész, J. (2013). Value production in a collaborative environment. *Journal of Statistical Physics*, 151(3-4), 414-439.